Part III
Structural Foundations of Virus Properties and Functions

Chapter 10
Assembly of Simple Icosahedral Viruses

José M. Almendral

Abstract Icosahedral viruses exhibit elegant pathways of capsid assembly and maturation regulated by symmetry principles. Assembly is a dynamic process driven by consecutive and genetically programmed morphogenetic interactions between protein subunits. The non-symmetric capsid subunits are gathered by hydrophobic contacts and non-covalent interactions in assembly intermediates, which serve as blocks to build a symmetric capsid. In some cases, non-symmetric interactions among intermediates are involved in assembly, highlighting the remarkable capacity of capsid proteins to fold into demanding conformations compatible with a closed protein shell. In this chapter, the morphogenesis of structurally simple icosahedral viruses, including representative members of the parvoviruses, picornaviruses or polyomaviruses as paradigms, is described in some detail. Icosahedral virus assembly may occur in different subcellular compartments and involve a panoplia of cellular and viral factors, chaperones, and protein modifications that, in general, are still poorly characterized. Mechanisms of viral genome encapsidation may imply direct interactions between the genome and the assembly intermediates, or active packaging into a preformed empty capsid. High stability of intermediates and proteolytic cleavages during viral maturation usually contribute to the overall irreversible character of the assembly process. These and other simple icosahedral viruses were pioneer models to understand basic principles of virus assembly, continue to be leading subjects of morphogenetic analyses, and have inspired ongoing studies on the assembly of larger viruses and cellular and synthetic macromolecular complexes.

J.M. Almendral (✉)
Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM) and Department of Molecular Biology of the Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain
e-mail: jmalmendral@cbm.uam.es

M.G. Mateu (ed.), *Structure and Physics of Viruses; An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_10,

Keywords Icosahedral capsid • Triangulation number • Assembly intermediate • Protein contacts • Nuclear translocation • Protein folding • Nucleation • Packaging • Encapsidation • Virus factory • Maturation cleavage

Abbreviations

AAP	assembly-activating protein
AAV	adeno-associated virus
CBB	capsid building block
CP	capsid protein
CPV	canine parvovirus
DBD	DNA-binding domains
H1-PV	parvovirus H1
hr-t	host range-transforming
FPV	feline parvovirus
MEV	mink enteritis virus
MVM	minute virus of mice
NLM	nuclear localization motif
NLS	nuclear localization sequence
NPC	nuclear pore complex
PPV	porcine parvovirus
SV40	simian virus 40
VLP	virus-like particle
VP	viral protein
5x	five-fold axis
3x	three-fold axis
2x	two-fold axis.

10.1 Introduction

All viral entities have a capsid, which in structurally simple viruses is built up from one or a few types of protein subunits. At late stages of the intracellular phase of their life cycle (see Chap. 1) viruses perform capsid assembly, a process by which the structural capsid protein (CP) subunits are joined by maximal hydrophobic contacts and/or non-covalent interactions (and occassionally covalent bonds) to construct the viral particle. This process is essential for viruses to mature (become infectious) and release progeny, and in the simple icosahedral viruses proceeds by strict principles of genetic economy and symmetry (see Chap. 2). This chapter reviews the main stages in the assembly and genome encapsidation of small or medium sized viruses with a relatively simple icosahedral architecture, exemplified by three distinct virus models: (i) Adeno-associated virus (AAV) and minute virus

of mice (MVM), as respective representative members of the *Dependovirus* and *Parvovirus* genera of the *Parvoviridae*, a family of single-stranded (ss) DNA viruses with a T = 1 capsid (25 nm in diameter) assembling in the nucleus; (ii) Poliovirus, genus *Enterovirus* of the *Picornaviridae*, RNA(+) viruses with a pseudoT = 3 capsid (30 nm in diameter) assembling in the cytoplasm; and (iii) Polyomavirus and Simian Virus (SV40), members of the *Polyomavirus* genus of the *Polyomaviridae*, double-stranded (ds) DNA viruses with an all-pentamers T = 7d capsid (45 nm in diameter) assembling in the nucleus. In these so-called "simple" icosahedral viruses, assembly occurs through an orchestrated pattern of interactions of the capsid subunits to form complexes or assembly intermediates, whose composition and conformation usually change along the process. The structural dynamics undergone by the assembly intermediates must also fulfill another important function, which is to traffic within the infected cell towards the compartment where the genome is being replicated. It is the accumulation of assembly intermediates at a specific compartment what triggers genome encapsidation and maturation, allowing the virus to finally propagate in nature.

10.2 Icosahedral Capsids: Symmetry and Genetic Regulation

10.2.1 Structural Principles in Icosahedral Capsid Assembly

In simple icosahedral viruses, the capsid is formed by many copies of one or a few protein subunits that assemble by making multiple contacts to build a hollow shell of proper size and symmetry. The regular icosahedron is formed by a defined number of copies of a capsid building block (CBB), which can be built up by a single CP subunit, by several identical CPs, or by non-identical CPs. The CBBs are related by two-fold (2×), three-fold (3×), and five-fold (5×) symmetry axes, and CPs within the CBBs establish regular interactions with their neighbors depending on their position in relation to these icosahedral axes. The Caspar and Klug (1962) theory explained how some multiples of 60 identical subunits could be arranged with similar (quasi-equivalent) interactions, according to the rule $T = h^2 + hk + k^2$, where h and k are integers and T is called the triangulation number (see Chap. 2 for a detailed explanation of capsid icosahedral symmetry and quasi-equivalence).

In the members of the *Parvoviridae*, the capsid is a T = 1 perfect icosahedron, formed by a total of 60 CPs (termed viral proteins, VPs, in this and some other virus families), which include two to three variant types (VP1, VP2 and VP3) with identical amino acid sequence and fold except for short streches of sequences at the C- or N-termini, which are intrinsically disordered and are not observed in the X-ray structures of the virus particles (Fig. 10.1a). The topology of the capsid surface differs among the parvoviruses due to the characteristic prominence of peptide loops and spikes at the 3x symmetry axes, and the depth and contour of the depression surrounding the 5x axes [1, 2, 7–10]. In Poliovirus and related picornaviruses the protein shell is built up of 60 copies of a fundamental subunit (the protomer) composed of three different proteins (termed VP1, VP2 and VP3)

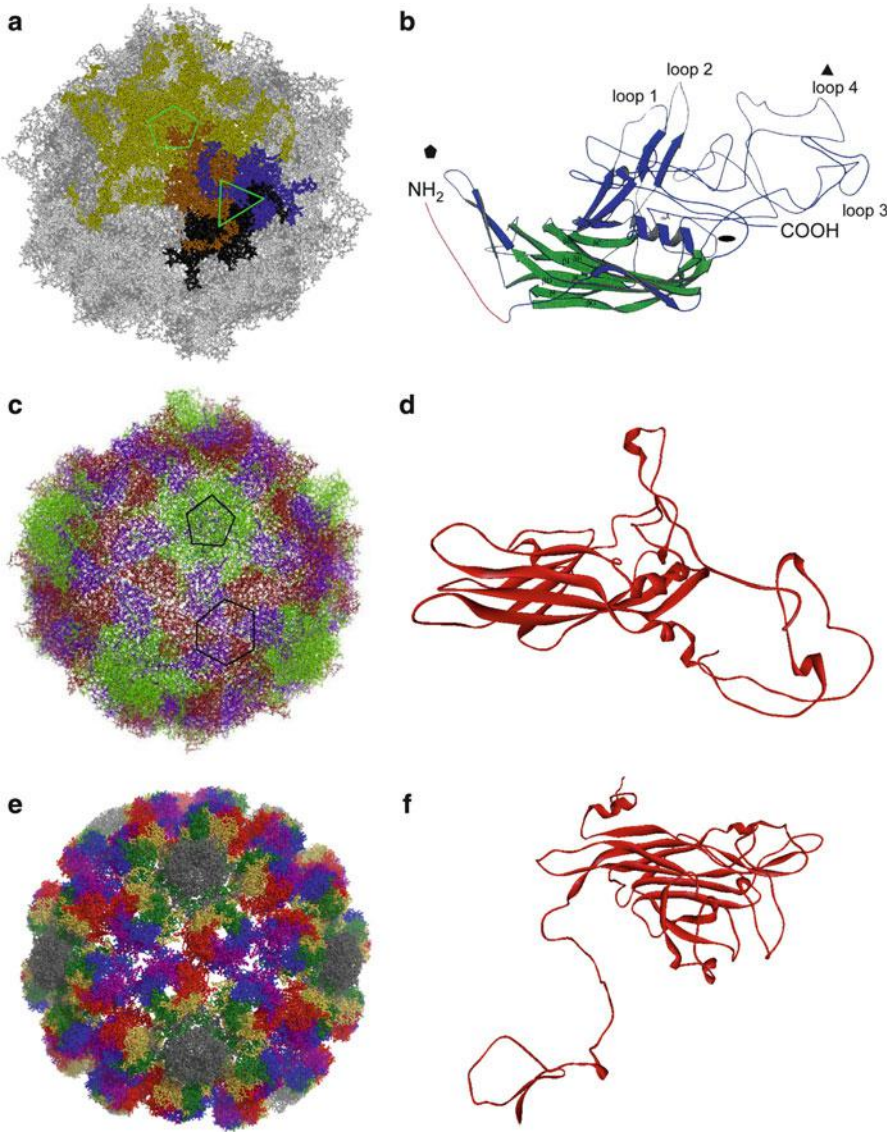

Fig. 10.1 Atomic structure of icosahedral virus particles and capsid subunits whose assembly is described in this chapter. (**a**) Structure of Parvovirus MVM (p and i strains; [1, 2]). Different *colours* distinguish the subunits surrounding the 5x axes (*pentagon*) from those interdigitated at the 3x axes (*triangle*). (**b**) Folding of the VP1 and VP2 subunits in the MVM structure. Note the prominent loops projecting away from the capsid surface. (**c**) Structure of Poliovirus, a picornavirus [3, 4]. The VP1 subunits around the 5x axes (*pentagon*), and the alternating VP2 and VP3 subunits around the 3x axes (*hexagon*) are respectively shown in *green, red* and *violet*. (**d**) Folding of the VP1 subunit in the capsid of Poliovirus. The β-strands (*arrows*) forms two antiparallel sheets juxtaposed in a wedgelike structure. (**e**) Structure of the SV40 virus showing the organization of the VP1 pentamers [5]. The 5x-coordinated pentamers are represented in gray and the VP1

which are not related in amino acid sequence but have a similar fold, and a small extended polypeptide (VP4). VP4 is located at the inner surface of the protein shell and remains covalently linked to the N-terminus of VP2 until the final stages of virus assembly and maturation. The interactions between the VP1 subunits around the 5x axes are not equivalent to those engaging the alternating VP2 and VP3 subunits in the 3x axes, accounting for the features of the surface of the capsid (Fig. 10.1c), including the conspicuous protrusion of the VP1 subunits at the 5x axes [3, 4]. In polyomaviruses such as SV40, the icosahedral capsid is formed by 360 identical CP subunits of a single protein (termed VP1), whose arrangement does not follow the quasi-equivalence rules of Caspar and Klug, as the basic structural elements (capsomers) are 72 pentamers displayed in a T = 7d surface lattice ([5]; Fig. 10.1e). Each pentamer contains in addition one copy of either of two other proteins (termed VP2 and VP3), which share most of their amino acid sequences. The capsomers located around each 5x axis are referred to as the pentavalent pentamers, and those capsomers arranged around each of the 3x axis are referred to as the hexavalent pentamers.

Although the parvovirus, poliovirus and polyomavirus capsid proteins are not related in amino acid sequences, the core of all these proteins are folded in their capsids as a β-sheet structure termed β-barrel jelly roll (or Swiss-roll β-barrel, or an eight-stranded antiparallel β-barrel), a wedge-shaped structure comprising two antiparallel β-sheets. The topology of all these capsid subunits is, thus, similar (Fig. 10.1b, d and f). The major structural differences between them are in the loops that connect the strands, which are particularly prominent in parvoviruses (Fig. 10.1b), and in the N- and C- terminal segments that extend from the central β-barrel domains. The remarkable similarity of these β-barrel jelly rolls among viral proteins that do not share primary amino acid sequences, and belong to different families of viruses with unrelated biological properties, suggests that it may represent one of the few structural solutions allowing proteins to be packaged in icosahedral capsids, or be a testimony of a common ancestral evolutionary history (see also Chaps. 2 and 7).

10.2.2 Synthesis of the Capsid Subunits: Setting the Assembly Scenario

For successful assembly, the synthesis of the capsid proteins must be tightly regulated during the virus life cycle, in order to satisfy at least three key requirements: (i) quantity, viruses must induce the accumulation of high amounts of capsid

Fig. 10.1 (continued) subunits of the pentamers in hexameric arrays are represented in different *colors*. (**f**) folding of the VP1 subunit as observed in the SV40 capsid. The β-barrel (jellyroll) is radial to the capsid surface. The figure was prepared using the Pymol program (http://www.pymol.org/) and the VIPER resources [6], using the atomic coordinates of MVM (1MVM, 1Z1C), Poliovirus (2PLV), and SV40 (1SVA), deposited in the Protein Data Bank (PDB)

proteins in the host cells, in order to compete with the vast amount and diversity of preexisting cellular proteins; (ii) timing, capsid proteins synthesis must reach maximum levels by the time the viral genome is being replicated, so it can be efficiently packaged; and (iii) stoichiometry, capsid subunits must be synthesized at the proper ratio to ensure assembly of an infectious viral particle. Genetic regulation of viral gene expression is therefore crucial for a successful assembly process. Mechanisms controlling viral gene expression involve multiple networks which are out of the scope of this book, so only a brief outline of those involved in assembly are mentioned below.

In Poliovirus (see [11] for a review of its life cycle), the genomic RNA is released in the cytoplasm by the incoming virion, and translated as a single open reading frame to produce a very large polyprotein, which undergoes cotranslational *cis*-cleavage by the viral protease 2A. This cleavage releases from the N-terminus the precursor polyprotein myristoyl-P1, which contains the CP sequences. In this and related picornaviruses, CPs are synthesized to high levels by the combination of the translational competence of the genomic RNA with an effective host protein shut-off induced by viral proteases, whereas the proper protein stoichiometry results from the cleavage of a common precursor.

In parvoviruses, polyomaviruses and other DNA viruses, the stoichiometry of capsid proteins is regulated mainly at the level of splicing of their messenger RNAs. In the parvovirus MVM for instance, site-directed mutagenesis at the minor splicing sites, or independent cDNA cloning, allows to obtain genomic clones expressing either VP protein [12]. VP2 alone can form a capsid which can encapsidate the viral genome, but VP1 is necessary for the infectivity of the particles due to specific domains residing at its N-terminal segment [13, 14]; and (Fig. 10.3c). A 1:5 ratio of VP1:VP2 is found for the soluble synthesized proteins, as well as in the assembled capsid, and preserving this ratio is critically important for an ordered assembly avoiding protein aggregates [16]. In polyomaviruses VP1, VP2, and VP3 levels in the infected cells are regulated by alternative splicing from a common trascript, occurring soon after viral DNA replicative intermediates accumulate. In these small DNA viruses protein shut-off is not a major mechanisn to counteract host protein synthesis, but the nuclear accumulation of protein products prior assembly (see below) facilitates their interactions at certain nuclearly confined enviroments.

10.3 Capsid Building Blocks and Assembly Intermediates

The capacity of the viral structural proteins to form capsids ultimately result from their folding and self-assembly properties, which are conferred by the encoded amino acid sequences. Virus capsids are assembled from CBBs which may be CP monomers or, in many cases, CP oligomers. These stable oligomeric CBBs may be considered as the first stable intermediates of the capsid assembly process. The use of *in vitro* assembly systems in combination with diverse theoretical approaches has allowed the investigation of fundamental principles of virus capsid self-assembly

starting from CBBs. The theory of capsid self-assembly is outside the aims of this chapter, which is dedicated to the description of assembly processes in different viruses from *in vitro* and *in vivo* experimental evidences supported by structural and functional studies. The reader is referred to Chap. 1 for a brief overview, and to Chap. 19 for a detailed description on the thermodynamic and kinetic aspects of the assembly of simple virus capsids. These studies are generally consistent with experimental observations on the assembly of very simple virus capsids, frequently carried out in controlled *in vitro* experiments. The models support assembly as a nucleated cooperative process in which CP or CBB concentration is critical, with a lag phase reflecting the time required to build up an assembly line of intermediate structures. The intermediates are expected to be at very low concentration, but this steady state of intermediates is required for efficient assembly in any stepwise reaction.

10.3.1 Structure of CBBs and Assembly Intermediates

The stepwise assembly pathway has been well characterized in Poliovirus (and other picornaviruses including human rhinovirus and foot-and-mouth disease virus), as discrete intermediates are stable enough to be isolated. A common strategy that many viruses adopt to build blocks for capsid assembly is to initiate assembly while the structural units are linked into a polyprotein precursor. In Poliovirus, the first intermediate of the assembly pathway is an inmature structural unit (unprocessed protomer) formed by a folded polyprotein (termed P1). P1 contains three structural domains. These domains are split apart from each other upon cleavage at specific sites in the linker sequences by the viral $3CD^{pro}$ protease. The result is a processed protomer (5S protomer) which sediments as a 5S particle and is formed by one copy each of VP0, VP3, and VP1, which correspond to the cleaved P1 domains (Fig. 10.2a). It is unclear when the β-barrel of these proteins core fold, but unprocessed P1 of picornaviruses is recognised by panels of virus-neutralising antibodies elicited against discontinuous epitopes of mature virions, strongly suggesting that the domains in unprocessed P1 are already folded like the mature CPs in the assembled capsid. The structures of the isolated unprocessed or processed P1 protomers have not been solved for any picornavirus yet; however, in the virus capsid structure VP1, VP2 and VP3 form an intricate network of intermolecular interactions among the surfaces of their β-barrel domains, which must greatly stabilize the protomer and contribute to the early stages of the assembly. VP4 remains covalently linked to VP2 in the VP0 protein until virus assembly is completed and maturation occurs (see below).

The 5S precursor is followed in the assembly line by the 14S pentamer (Fig. 10.4a), which is formed by oligomerization of five 5S protomers, and is stabilized by extensive protein-protein interactions and by others mediated by myristate chains incorporated in the five VP0 N-termini. These multiple interactions determine a molecular interlocking of the five protomers in the pentamer, conferring high stability and directionality to the whole assembly pathway.

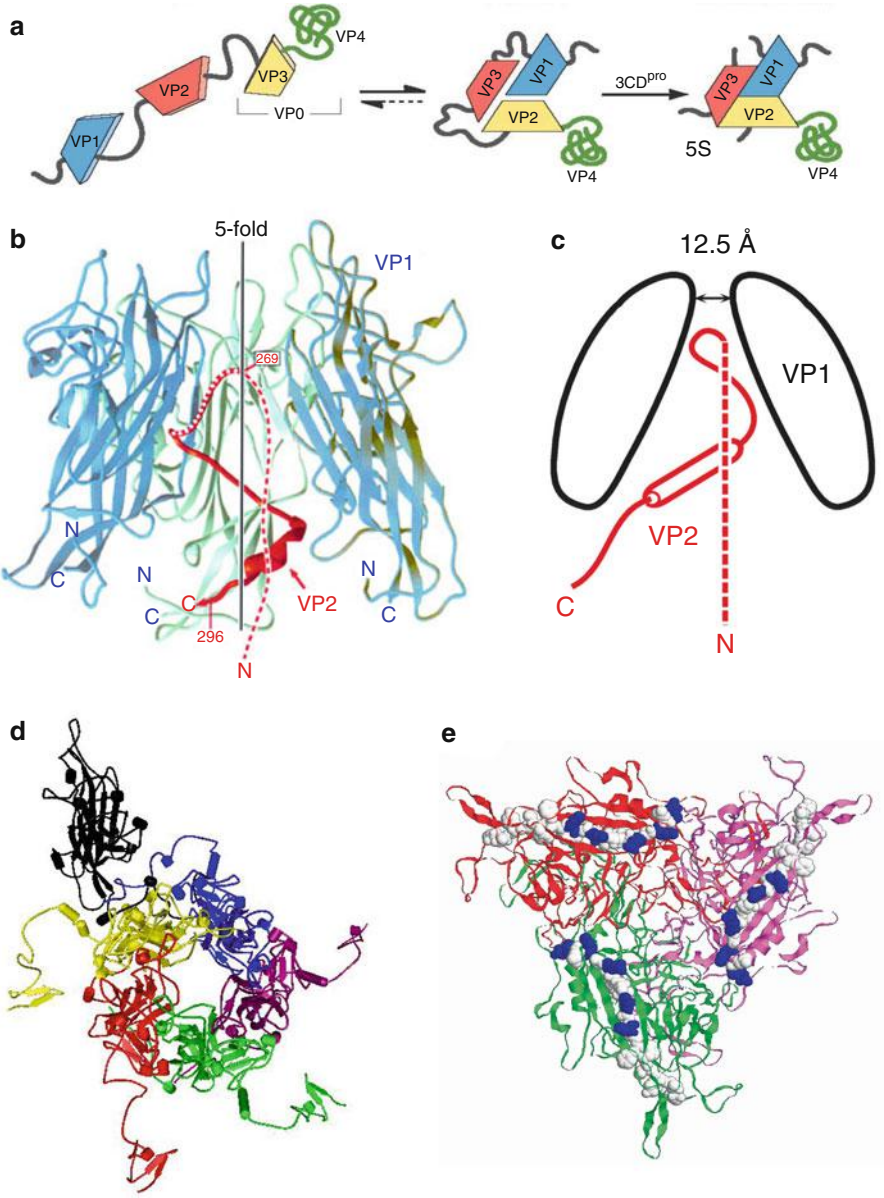

Fig. 10.2 Structural models of assembly intermediates in icosahedral viruses. (**a**) Poliovirus P1 polyprotein precursor showing the four proteins forming the heterometric structural unit (protomer). Cleavage by $3CD^{pro}$ protease yields the 5S assembly intermediate. (Adapted from Flint et al. (2009) Principles of Virology, ASM Press, with permission). (**b**) Structure of the VP1-VP2 assembly intermediate of polyomavirus. VP2 is shown in red and the three VP1 monomers that form contacts with VP2 in *green* (*middle*) and *blue* (*left* and *right*). (**c**) Schematic representation of the VP1/VP2 interaction in this complex. VP2 (*red*) enters in the VP1 pentamer (*black*) from the base, continuing to the *upper* part of the conical depression. It then loops back to interact

In polyomaviruses as well as in parvoviruses, the CP subunits interact in the cytoplasm before nuclear capsid formation. The composition of the assembly intermediates is characteristic for each virus system and plays a role of paramount importance in the correct timing and spatial ordering during assembly. A relatively stable cytoplasmic assembly intermediate seems to be a common need for these viruses completing capsid assembly and maturing within the nucleus. In Polyomavirus and SV40, stable pentamers of VP1 are formed in the cytoplasm. Each VP1 pentamer binds either one VP2 or one VP3 protein that becomes allocated in the axial cavity of the pentamer (Fig. 10.2b), and the VP1-VP2/3 complex is stabilized by strong hydrophobic interactions [17]. The contacts between subunits significantly alter the configuration of the VP1 pentamer, as demonstrated by changes in epitope accesibility and by direct structural insights obtained from the crystal structure (Fig. 10.2b, c). In SV40, transient disulphide bridges are established intramolecularly, and subsequently intermolecularly, as the monomers assemble into pentamers, which facilitate the folding and interdigitation of structural elements [19]. The noncovalent interactions and covalent bonds collectively conform a stable cytoplasmic CP pentameric complex (Fig. 10.2d), which is the major assembly intermediate in these viruses.

In parvoviruses, taking the murine MVM as a reference model, trimers of VP subunits assemble in the cytoplasm. The VP1 (82 kDa) and VP2 (63 kDa) proteins synthesized at a VP1:VP2 1:5 M ratio assemble into two types of trimers (Fig. 10.2e), which are produced in stoichiometric amounts. The larger trimer (200 kDa) is a heterotrimer formed by one VP1 and two VP2 subunits, whereas the smaller (180 kDa) is a VP2-only homotrimer [16]. In the formation of the cytoplasmic trimer, the VP2 protein may act as a scaffolding factor assisting VP1 to acquire a proper folding, as deletion mutants of the VP1-specific region undergo extensive ubiquitination degradative reaction that can be significantly prevented by co-expression of VP2 [14]. The assembly of VP cytoplasmic trimers for these viruses was structurally supported first by the higher stability of the trimer (measured by the buried surface area on oligomer formation) as compared to putative dimers or pentamers of CP subunits, due to the multiple contacts in the intertwined loops of the subunits around the 3x symmetry axes (Fig. 10.1a). Furthermore, crosslinking experiments and mutations disrupting the intertrimer interfaces in the MVM capsid allowed the isolation of stable trimers (see below).

Fig. 10.2 (continued) specifically with the inner face of VP1, forming a hairpin-like structure. (**b**, **c**: Reprinted from [17], with permission). (**d**) Structure of the VP1 pentamer in the SV40 capsid. (**e**) Structure of the VP trimer in the MVM capsid. Residues involved in NLM function (see *below*) are highlighted. (**d**, **e**: prepared using the Pymol program (http://www.pymol.org/) and the 1SVA and 1MVM respective atomic coordinates deposited in the PDB)

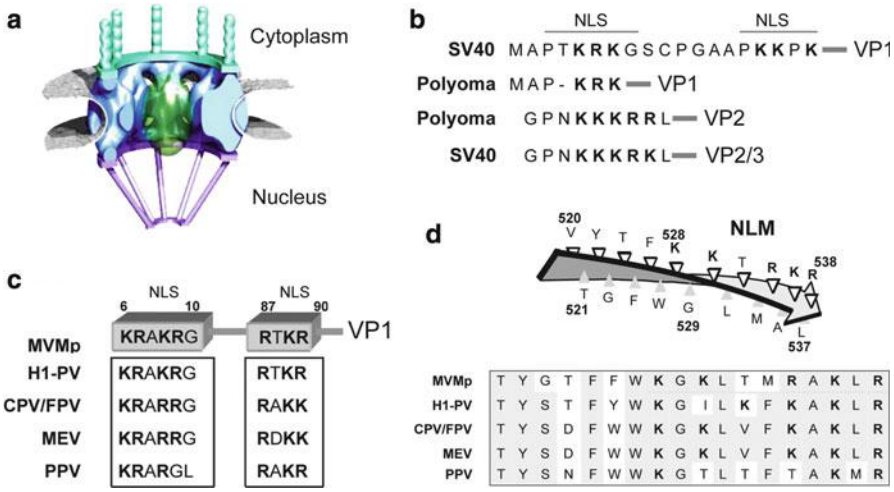

Fig. 10.3 Nuclear localization sequences in polyomavirus and parvovirus capsid proteins. (**a**) Basic architecture of the nuclear pore complex (NPC). The central ring is illustrated with the filament facing the cytoplasm and the basket protruding into the nucleus. (**b**) Examples of identified NLS in the VP proteins of Polyomavirus and SV40. (**c**) Two conserved domains with NLS activity identified in the VP1 N-terminal sequence of several parvoviruses: MVM, parvovirus H1 (H1-PV), CPV, feline parvovirus (FPV), mink enteritis virus (MEV), porcine parvovirus (PPV). (**d**) Configuration of the NLM in the capsid subunits of MVM and conservation of the NLM sequence in related parvoviruses. Basic residues contributing to nuclear targeting of assembly intermediates are shown in *bold*. (Adapted from [15] with permission)

10.3.2 Intracellular Traffic of CBBs and Assembly Intermediates

For capsid formation, CBBs or other assembly intermediates must accumulate in the subcellular compartment where the viral genome replicates, and at the right time. In some viral systems, the capsid subunits must traffic from the site of synthesis in the cytoplasm to the assembly compartment. This traffic is directed by protein signals in CBBs or intermediates that are accessible to the transport machineries of the cell. The nature of the signals and the configuration of the intermediate exposing them, which will determine the transport route accessed by the intermediate and its fate in the cell, are thus key elements in the viral assembly process.

In polyomaviruses and parvoviruses, the viral genome is replicated in the nucleus, and the CPs of these viruses are therefore karyophilic polypeptides that traverse the nuclear membrane. Translocation of proteins into the nucleus imposes two restrictions, firstly by the need of nuclear localization sequences (NLS) to access the cellular transport machinery, and secondly by the size of the complex to be transported, which cannot exceed 25–30 nm, the functional diameter of the aperture in the nuclear pore complex (NPC) [20], a supramolecular structure embedded in the nuclear envelope (Fig. 10.3a). The exchange of macromolecules

across the NPC is mostly mediated by proteins of the importin β (karyopherin β) superfamily, which comprises importins and exportins. The NLS are, in the so-called conventional configuration [21], single or bipartite stretches of basic amino acids that access, generally upon direct binding to a protein adaptor, the importin α/β transport pathway [22]. However, many karyophilic proteins harbor non-classical NLS that may bind karyopherin β1 directly, or access the alternative karyopherin β2/transportin import pathway. A current active area of research is the identification of the transport routes accessed by the viral proteins in specific cell hosts. Functional conventional NLS are found in the major VP1 as well as in the minor VP2/3 subunits of Polyomavirus [23, 24] and SV40 [25] (Fig. 10.3b). It is remarkable the sequence conservation among the structural proteins of these viruses at the NLS domains, compared to the high sequence divergence in the rest of these proteins. In polyomaviruses the VP1-VP2/3 cytoplasmic complex is translocated through the NPC by the functional cooperation of the multiple NLS displayed by the VP subunits [26].

In MVM, used as a representative molecular model of the *Parvoviridae*, the two types of VP trimers are major CBBs, or stable assembly intermediates, translocating across the nuclear membrane. The protein subunits within the trimer cooperate for nuclear transport, as both VP1 and VP2 proteins genetically depleted of functional nuclear transport sequences can be co-transported into the nucleus by expressed intact subunits [14]. Indeed VP1 and VP2 carry independent NLS and efficiently target the nucleus of transfected cells when singly expressed [12]. VP1 harbours two conventional NLS at its N-terminal specific region (Fig. 10.3c), which function independently. These NLS are required for nuclear translocation of the expressed VP1 subunits in MVM [14] and in canine parvovirus, (CPV) [27], and also for the MVM virion to initiate infection [14], suggesting that they are exposed out of the virus shell during the cell entry process. Similar separate regions with basic amino acids essential for assembly and infectivity were identified in the capsid proteins of AAV [28]. In addition, both VP1 and VP2 (the major capsid-forming polypeptide) contain, in their common folded sequence, a structured domain with nuclear targeting capacity (named NLM) [15]. In the assembled capsid the NLM is localized in the amphipatic ß-strand I at the inner capsid surface (Fig. 10.3d). The NLM is the only functional nuclear targeting sequence identified in VP2. The structured NLM, highly conserved in many members of the *Parvoviridae*, shows, when displayed in the capsid structure (Fig. 10.2e), all its charged basic amino acids placed within the side of the strand facing the solvent at the interior surface of the capsid, and the hydrophobic amino acids oriented towards the protein core. This structuration of the NLM probably occurs only upon folding and trimerization of the VPs, which may occur concomitantly. Folding and oligomerization of the capsid subunits probably occur in the cytoplasm, leading to trimeric assembly intermediates acquiring nuclear transport competence. The structuration of the two types of trimers translocating across the NPC may exert a quality control role in the virus morphogenetic flow, as misfolded subunits (not exposing the NLM) or oligomers with aberrant structures and protein composition would not access the nucleus [15, 16].

10.4 Forming the Capsid

10.4.1 Contacts and Structural Changes in CBBs and Assembly Intermediates

The CBBs or initial stable intermediates of capsid assembly may be found in an assembly-incompetent state, incapable of establishing productive interactions between them to accommodate the final configuration of the capsid [29]. Thus, the assembling oligomers must frequently undergo conformational changes during the late stages of the assembly pathway. In Poliovirus and other picornaviruses, the similar β-barrel folding of the VP1, VP2 and VP3 proteins that facilitate their interactions to form the 60 structural subunits (5S protomers) and the 14S pentameric CBBs (Fig. 10.4a), also favours the subsequent assembly of these latter intermediates into a complete viral particle [30]. For this, the extensive interactions among the β-barrels of adjacent proteins help to form a rather rigid protein shell with a dense network of intersubunit interactions. Inside the capsid, a network of additional protein contacts stabilizes the mature particle. These contacts are largerly contributed by the long (40–80 residues) N-terminal arms of the three VP subunits, which have a similar path inside the capsid in different picornaviruses, although their primary amino acid sequence is drastically different. These interactions are most extensive at the 5x axes, where the N-termini of five VP3 molecules are arranged in a tubelike parrallel β-sheet. The conformation of these arms is ordered only after capsid assembly, so it is difficult to evaluate the contribution of the interactions involving these arms in the final particle stability based on structural data only.

In members of the *Parvoviridae*, capsid formation in the natural infection starts in the nucleus. Indeed the nucleolus was identified as the subcellular compartment where AAV assembly inititates, colocalizing with the non-structural replicative viral Rep proteins; subsequently, capsid accumulation spreads across the entire nucleus [31]. In MVM, as the assembly intermediates (mainly trimers) accumulate in the nucleus, the nucleation reaction is triggered and multiple non-covalent interactions are established between amino acids localized at the edges of the binding trimers. A few evolutionary conserved residues involved in presumably strong intertrimer contacts were found to be necessary for capsid assembly (Fig. 10.4c). These residues buried a large hydrophobic surface upon trimer association, or formed buried intertrimer hydrogen bonds or salt bridges [18]. Assembly intermediates other than the CBBs are difficult to isolate in parvoviruses, presumably because of the high efficiency of the assembly reaction in which intermediates are highly transient and accumulate at very low levels, in agreement with theoretical studies on the assembly of simple virus capsids (see Chap. 19). A genetic approach to trap unstable assembly intermediates based on the introduction of disruptive mutations at some of the residues that in the capsid are involved in contacts between MVM trimers yielded high amounts of trimers, without evidence of any other larger intermediate [16]. These trimers could be isolated by

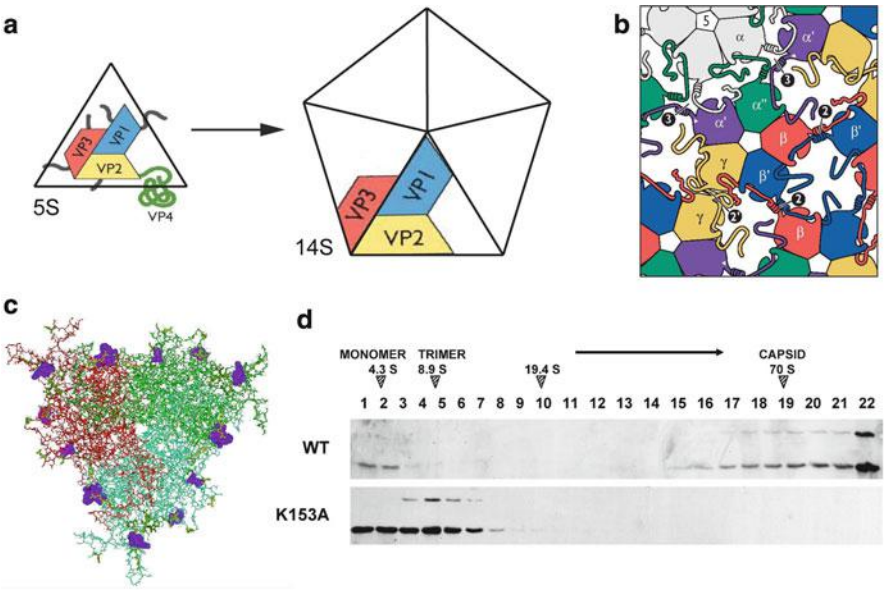

Fig. 10.4 Interactions among assembly intermediates to form the capsid. (**a**) Assembly of the 5S protomer into a 14S pentamer in Poliovirus. The protomer is not identical to the icosahedral asymmetric unit in the capsid. (**b**) Contacts between capsomers in SV40. Four pentamers assembled in the SV40 capsid are shown. Each subunit of the pentamers projects an arm that makes different contacts with the jellyroll of the subunit in another pentamer. (Adapted from [5], with permission). (**c**) Capsid formation in Parvovirus: a VP trimer is represented; some residues involved in major inter-trimer contacts of MVM capsid and required for capsid assembly are shown as spacefill models and coloured *violet*. (Adapted from [18], with permission). (**d**) Sedimentation analysis in sucrose gradients of the K153A mutant of the MVM capsid proteins. This mutant lacks a side chain that is important for establishing critical contacts between MVM trimers during capsid assembly. The sedimentation position of complete capsids, trimers, and VP monomers are indicated. (Adapted from [16], with permission)

sedimentation in sucrose gradients (Fig. 10.4d), and were competent in a nuclear transport assay performed in permeabilized cells [32], indicating that their configuration properly expose the nuclear transport sequences, and thus mutations precluded just the final capsid formation stage.

The acquisition of competence for assembly by MVM trimers involves conformational changes that have been indirectly detected. An MVM-induced monoclonal antibody recognizing a discontinuous epitope located at each 3x axis of the capsid (which corresponds to the center of each trimer) failed to react with non-assembled trimers accumulated in the cell nucleus or isolated *in vitro* [16, 32]. Trimers must therefore change their conformation during nuclear capsid assembly, and the process may be triggered by some external factor and/or phosphorylation of the subunits (see below). These induced conformational rearrangements would reorganize some residues located at the vertex of the thee-fold axes, creating a new epitope.

In the viral particle of Polyomavirus and SV40, all the capsomers are pentamers, even though the capsid shell is built by subunits arranged in an icosahedral T = 7 lattice. Therefore the architecture of this virus does not fullfil the arrangement of pentameric and hexameric capsomers predicted by the quasiequivalent theory for capsids with T >1 values ([33]; see Sect. 10.2.1). However the SV40 capsid is stable enough, and this is accomplished in part by virtue of unique intercapsomers bonds established during assembly by the C-terminal domain of the VP1 protein (Fig. 10.4b). This domain acts as a connecting arm which, under six different configurations, makes contacts between the pentavalent pentamers (around each 5× axis) and the hexavalent pentamers (around each 3× axis). The different types of contacts in which the arms are involved act as multiple clamps holding the subunits together, and thus ensuring the stability of the capsid shell [5].

The VP1 protein of Polyomavirus can self-assemble into pentamers in heterologous cell expression systems [34], and the purified pentamers can associate to form capsid-like assemblies (VLPs) that become stabilized at low ionic strength by calcium. Interestingly, these VLPs assemble without the minor virion protein components, VP2 and VP3, suggesting that the non-equivalently related subunits of the penta- and hexavalent capsomeres may spontaneously switch their bonding specificity during assembly. VP1-only capsids however do not represent the physiological assembly pathway observed in the natural virus infections, in which the assembly of the VP1 pentamers is driven upon regular interactions involving VP2 and VP3 with the viral genome DNA (see below), and empty capsids are thought to be minor abortive assembly by-products.

10.4.2 Compartments, Factors, and Protein Modifications Influencing Capsid Formation

Icosahedral capsid assembly is influenced by environmental conditions in the cell and also by multiple molecular factors (in addition to the VP protein themselves) that may be encoded by the cells or by the virus genomes. These factors may act at specific subcellular compartments, and their contribution may critically determine in many cases the efficiency of the assembly process *in vivo*. Some of these factors have been identified in several viral systems, although multiple experimental evidences suggest that many others remain to be identified. For instance in Poliovirus, the rate of assembly of the structural proteins *in vitro* is reduced by at least two orders of magnitude compared to the rate observed in infected cells, and the empty capsids formed showed altered conformation, unless the reaction is primed by 14S pentamers isolated from infected cells. This experiment implies that proper folding, interactions, and/or modifications of the proteins forming the pentameric 14S intermediate, are essential for assembly to proceed successfully. A candidate for this function is the cellular Hsp70 chaperone, which associates with the P1 polyprotein precursor during its folding to form the 5S structural unit.

The cellular modulations of virus assembly may involve complex signalling pathways. For parvoviruses, capsid assembly during natural infections occurs with high efficiency in the nuclear compartment of infected host cells. However virus-like particles (VLPs) devoid of nucleic acid may be formed in the cytoplasm of heterologous expression systems (*e.g.* recombinant baculovirus expressing VP2 of MVM in insect cells) at low efficiency [35, 36]. This distinct assembly efficiency may be accounted, at least in part, by post-translational modifications of the capsid subunits. In MVM-infected cells, VPs and native capsid become extensively post-translationally modified by phosphorylation [37], whereas VLPs purified from heterologous insect cells were not phosphorylated [32]. In spite of the absence of modifications, these VLPs showed a 3-D structure identical to the native capsid and virus [2, 35], indicating that phosphorylation is not important for the icosahedral T = 1 ordering. The 2D-tryptic phosphopeptides analysis of native MVM capsid subunits resulted in a complex pattern of phosphoserine and phosphothreonine residues which was different for VP1 compared to VP2 [37]. In the host cell systems studied, the phosphorylation of the VP subunits of MVM was mainly catalyzed by the cytoplasmic activity of the Raf-1 kinase of the MAPK signalling pathway [32], and this modification was crucial for the acquisition of nuclear transport competence by the trimers. Phosphorylation by cellular kinases may be a general strategy to connect capsid assembly with host cell physiology, ensuring a spatially and timely regulated process maximizing virus yield.

As in parvoviruses, in Polyomavirus the efficiency of nuclear assembly is regulated by phosphorylation of the capsid subunits. Polyomavirus is highly tumorigenic in mouse, but host range transforming (hr-t) mutants of this virus defective in tumour induction are blocked in virion assembly when infecting non-permissive cells [38], although viral DNA and capsid proteins are synthesized to wild type levels. In purified Polyomavirus particles the several VP1 species identified by 2-D electrophoresis are generated by acetylation and phosphorylation of threonine and serine residues of the initial translation product. The hr-t mutants failed to assemble the complete (240S) viral particle, correlating with the lack of acidic forms of VP1 [39]. These acidic forms resulted from phosphorylation of threonine residues, and at least one of the phosphotreonines was shown to be essential for the encapsidation of the viral minichromosome [45]. In SV40, phosphate groups are added in natural infections on serine and threonine residues flanking the NLS of the VP proteins. Phosphorylation was shown to act on the activity of these NLS indirectly regulating the nuclear accumulation of VPs and virus assembly.

An example of a factor encoded by the viral genome favouring assembly was identified in studies with the parvovirus AAV. Capsid assembly in AAV begins in the nucleolus, and spreads throughout the nucleus at later stages of infection. In addition to the three capsid proteins, VP1, VP2 and VP3, the *cap* gene also encodes, by using an alternative open reading frame, the assembly-activating protein (AAP) that is essential for capsid assembly [31]. The AAP factor targets newly synthesized capsid proteins to the nucleolus, and becomes stabilized upon co-expression of the capsid protein VP3 of the same virus serotype, or from an AAV serotype of the same assembly group. The assembly-promoting activity of AAP is mediated by

interaction between two hydrophobic domains in the N-terminal region of the molecule with the C-terminus of the VP proteins, which forms the capsid protein interface at the 2× symmetry axes [40]. AAP seems to act as a scaffolding factor able to change the conformation of non-assembled VP molecules.

10.5 Genome Encapsidation and Virus Maturation

10.5.1 Poliovirus Cytoplasmic Maturation

Pathways of genome packaging adopted by some icosahedral viruses have been, and continue to be a matter of debate, as it is quite difficult in many cases to distinguish between "concerted assembly", in which the capsid is formed as a result of the ordered association of the protein subunits with the genome, from the "sequential assembly" that occurs when the genome is encapsidated into a preformed capsid, which may require an active stage (see Chap. 12 for detailed descriptions of mechanisms of encapsidation referred mostly to structurally complex viruses). Figure 10.5a illustrates this dilema in Poliovirus. In this system, most evidences support a "concerted" assembly of the 14S pentamer condensing around the RNA genome. Alternatively, an empty procapsid containing 60 copies of the VP0-VP3-VP1 structural unit could be transiently required at low concentrations for packaging. The apparent equilibrium between pentamers and an empty capsid of low stability makes it difficult to demonstrate whether the RNA is encapsidated by association with the pentamers or becomes inserted into the capsid [11].

Whichever encapsidation pathway is dominant, the maturing Poliovirus virion must undergo a number of modifications prior to becoming infectious. Maturation involves lipid modification and proteolytic cleavage at a specific site, making assembly an irreversible process. Along the final stages of assembly, a molecule of the fatty acid myristate is added post-translationally to the N-terminus of each VP4 subunit. This lipid mediates the interaction of the β-sheet formed by VP3 N-termini with a second β-sheet structure containing strands contributed by both VP4 and VP1 molecules. This feature of the capsid does not form until final stages of the maturation, when proteolytic processing liberates VP2 and VP4 from their precursor VP0, and this reaction is associated with a significant increase in the stability of the viral particle, priming it for entry into a new host cell.

10.5.2 Polymorphic Nuclear Maturation of Polyomaviruses

To form the SV40 and Polyomavirus virions, the VP1-pentamers become associated with single copies of a minor capsid protein (either VP2 or VP3) and, once imported into the nucleus, interact with the replicating DNA. Although the

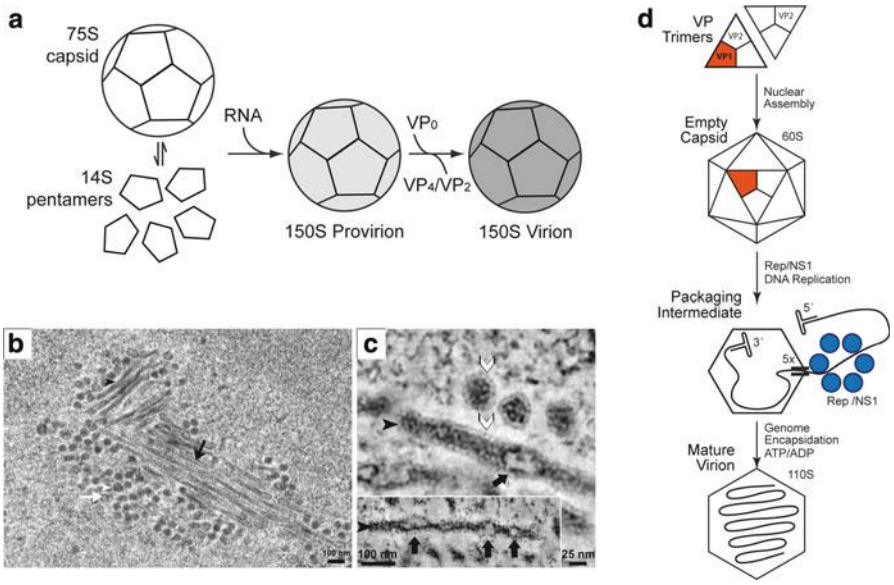

Fig. 10.5 Genome pakaging and virus maturation. (**a**) Late stages of Poliovirus assembly and virion maturation cleavage [30]. (**b**) Organization of the Polyomavirus factory. Electron microscopy of plastic-embedded thin sections of the nucleus of mouse fibroblasts showing tubular structures adjacent to a virus cluster (*black arrowhead*, "full" tubular structure; *white arrowhead*, full virion; *black arrow*, "empty" tubular structure; *white arrow*, empty virion). (**c**) Spherical virions and tubular structures seen in the nucleus of Polyomavirus-infected cells showing a lighter and well-arranged density at their periphery (*white arrowhead*) corresponding to capsid protein density, and a dense core (*black arrowhead*) suggesting DNA. (**b**, **c**: Adapted from [41]). (**d**) Integrative model of major stages in parvovirus assembly and maturation: I. VP capsid subunits assemble into trimers in the cytoplasm (*top image*). II. Inmature empty capsids (60S) form in the nucleus (*top center*). III. Genome packaging reaction (*bottom center*). IV. Mature viral particle (110S) (*bottom image*)

virus structure is known to atomic resolution, how the virions mature in the nucleus during productive infections is only partly understood. In a stepwise model, the capsid proteins would be sequentially added to and arranged on the viral DNA, resulting in its condensation and packaging to form the virion. The process implies multiple molecular recognition events of the viral genome by the CP subunits, in which DNA sequences near the viral origin of replication and the T-antigen (major replicating non-structural Polyomavirus protein) play important functions. Several DNA-binding domains (DBD) localized in all three VP proteins contribute to the packaging process. Major DBD of VP2 and VP3 were found in their C-terminal segments, whereas the DBD of VP1 was localized overlapping with the N-terminal bipartite NLS [42]. This VP1 N-terminal sequence of 15 amino acids is not visible in the virion crystal due to structural disorder, but it likely extends into the virion core to interact with the viral DNA.

The dynamics of maturation of the Polyomavirus and SV40 virions is not precisely known yet, but most evidences suggest a polymerization of the CP subunits onto the viral genome, acting as a scaffold. Although the intermediates are not well defined, the final nuclear assembly process seems to proceed through large, polymorphic structures that may serve as virus factories (Fig. 10.5b, c) (see Chap. 14 for a description of virus factories). Electron microscopy reveals tubular structures in the nucleus adjacent to clusters of virions. They share an organization consisting of a protein shell surrounding an electron dense DNA core, suggesting that they are the main factories able to yield mature virions budding from their ends, although the resolution mechanism is unknown.

10.5.3 Viral Genome Encapsidation in a Pre-formed Parvovirus Capsid

The final stages of parvovirus virions assembly and maturation are a subject of current active research, but in the best known cases it proceeds through the formation of large amounts of empty capsids in the nucleus, which serve as preformed substrate for active genome encapsidation. In MVM for example, empty capsids are first detected accumulated in the nucleus of synchronized cells preceding the accumulation of DNA-filled virions. A consensus model for parvovirus maturation, based on the available current data obtained in the AAV and MVM systems [43, 44] is outlined in Fig. 10.5d, although many details of the ssDNA packaging mechanism are still poorly understood. The reaction proceeds through a "packaging stage" in which the major non-structural proteins, NS1 in the *Parvovirus* genus and Rep 78/52 in AAV, mediate the association of the empty capsid with DNA replicative intermediates. These multifunctional proteins harbor helicase, DNA binding, and endonuclease activities that are critical for genome replication, and form oligomers *in vitro* in the presence of ATP. A Rep/NS1 hexameric packaging motor is probably formed at the five-fold symmetry axes, injecting the ssDNA genome through the channel of the capsid, which acts as the portal for genome encapsidation. For the packaging reaction the helicase activity of the non-structural proteins is essential. Genome encapsidation operates in the 3′ to 5′direction, mediated by specific contacts with the resolving hairpins and ATP hydrolysis. Encapsidation is coupled to DNA replication, ending as the capsid is filled with the complete ssDNA genome, which establishes regular contacts with some internal capsid residues and become icosahedrally ordered along certain sequences. The Rep/NS1 subunits may remain bound to the genome outside of the viral particle. The reader is referred to Chap. 12 for comparisons with equivalent mechanisms of dsDNA and dsRNA packaging into preformed bacteriophage capsids.

10.6 Perspectives and Conclusions

The assembly of simple icosahedral viruses proceeds through elegant genetically encoded and virus-specific pathways, which direct the self-association of the synthesized asymmetric protein subunits into a symmetric viral capsid shell. The general mechanism of capsid assembly involves first the oligomerization of CP subunits into early stable intermediates, usually of one major type per virus, which constitute the building blocks (CBBs). Capsid assembly then proceeds along a pathway regulated by ordered interactions between activated CBBs, most likely through a cascade of second-order reactions (see Chap. 19). The general architecture and the organization of the capsid subunits in the intermediates resembles those in the final capsid. However, the assembly process implies further structural rearrangements within and between the capsid subunits, including the establishment of multiple inter-protein non-covalent interactions, protein cleavages and covalent bonds (in some cases), and intracellular traffic of the assembly intermediates, prior to their final condensation into a closed container. Assembly is a dynamic process, implying that many of these interactions among subunits may have a transient morphogenetic role, and will not be preserved in the assembled virion. Functional capsid assembly cannot be understood without concomitant or later packaging of the genome, a complex process that viruses solve through molecular recognition patterns between the replicating genomes and either the assembly intermediates or a preformed empty capsid (see Chap. 12). During the final stages of morphogenesis of many viruses, cleavage of the viral particle leads to a mature infectious virion (see Chap. 13 for a description of maturation strategies in different viruses, including structurally complex bacteriophages). Virion morphogenesis in the cell, even for simple viruses, is frequently mediated by additional viral and or cellular factors, and may involve specific cellular factories (see Chap. 14). The efficiency of assembly critically determines some important features of viral fitness, such as the virus yield per host cell, or the acquisition of enough stability to prevail in natural enviroments.

The intensive research on small icosahedral viruses over the past decades have provided detailed information on the atomic structure of viral particles, and on the overall organization and mechanisms of assembly *in vitro* and in the host cells. Despite these advances, our knowledge on some crucial steps of the assembly of even structurally simple viruses is still poor or unclear in many respects, as exemplified by the following unsolved problems: the detailed structure of assembly intermediates in picornaviruses and parvoviruses; the cellular factors and transporters regulating VP traffic in the cell; the enzimology of genome encapsidation in parvoviruses; the processes of protein-protein and protein-nucleic acid recognition leading to the maturation of Poliovirus and Polyomavirus; or the resolution of the mature virions from the Polyomavirus factories. Ongoing research on these and other subjects related with morphogenesis promises exciting insights into essential aspects of the structural biology of simple icosahedral viruses in the near future.

Acknowledgements I gratefully acknowledge the collaborative work of Mavis Agbandje-Mckenna (University of Florida) on solving parvovirus MVMp capsid structure to atomic resolution, of Mauricio G. Mateu (CBMSO, Universidad Autónoma de Madrid) on MVM assembly, and of Thomas L. Benjamin (Harvard Medical School, Massachussets) on aspects related to Polyomavirus assembly. The many inspiring conversations along the years with regular participants to the "Virus Assembly" FASEB meeting (Vermont Academy, USA), in particular with Peter Tattersall (Yale University, Connecticut), are also deeply acknowledged. This review stands from the experience gained by the author on Parvovirus assembly thanks to the enthusiastic work of Eva Hernando, Eleuterio Lombardo, Beatriz Maroto, Juan C. Ramírez, Laura Riolobos, and Noelia Valle. Support on figures design by Jon Gil-Ranedo and Jorge Sánchez is also acknowledged. Research in JM Almendral's laboratory is currently supported by grant SAF 2011-29403 from the Spanish Ministerio de Ciencia e Innovación.

References and Further Reading

1. Agbandje-McKenna M, Llamas-Saiz AL, Wang F, Tattersall P, Rossmann MG (1998) Functional implications of the structure of the murine parvovirus, minute virus of mice. Structure 6:1369–1381
2. Kontou M, Govindasamy L, Nam HJ, Bryant N, Llamas-Saiz AL, Foces-Foces C, Hernando E, Rubio MP, McKenna R, Almendral JM, Agbandje-McKenna M (2005) Structural determinants of tissue tropism and *in vivo* pathogenicity for the parvovirus minute virus of mice. J Virol 79:10931–10943
3. Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht HJ, Johnson JE, Kamer G, Luo M, Mosser AG, Ruecker RR (1985) Structure of a human common cold virus and functional relationship to other picornaviruses. Nature 307:145–153
4. Hogle JM, Chow M, Filman DJ (1985) Three-dimensional structure of poliovirus at 2.9 A resolution. Science 229:1358–1365
5. Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC (1991) Structure of simian virus 40 at 3.8-A resolution. Nature 28:278–284
6. Natarajan P, Lander GC, Shepherd CM, Reddy VS, Brooks CL III, Johnson JE (2005) Exploring icosahedral virus structures with VIPER. Nat Rev Microbiol 3:809–817
7. Tsao J, Chapman MS, Agbandje M, Keller W, Smith K, Wu H, Luo M, Smith TJ, Rossmann MG, Compans RW, Parrish CR (1991) The three-dimensional structure of canine parvovirus and its functional implications. Science 251:1456–1464
8. Xie Q, Bu W, Bhatia S, Hare J, Somasundaram T, Azzi A, Chapman MS (2002) The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. Proc Natl Acad Sci U S A 99:10405–10410
9. Gurda BL, Parent KN, Bladek H, Sinkovits RS, DiMattia MA, Rence C, Castro A, McKenna R, Olson N, Brown K, Baker TS, Agbandje-McKenna M (2010) Human bocavirus capsid structure: insights into the structural repertoire of the parvoviridae. J Virol 84:5880–5889
10. Kaufmann B, Simpson AA, Rossmann MG (2004) The structure of human parvovirus B19. Proc Natl Acad Sci U S A 101:11628–11633
11. Racaniello VR (2001) Picornaviridae: the viruses and their replication. In: Knipe D, Howley P (eds) Fields virology. Lippincot Williams & Sons, New York, pp 685–722
12. Tullis GE, Lisa RB, Pintel DJ (1993) The minor capsid protein VP1 of the autonomous parvovirus minute virus of mice is dispensable for encapsidation of progeny single-stranded DNA but is required for infectivity. J Virol 67:131–141
13. Zadori Z, Szelei J, Lacoste MC, Li Y, Gariepy S, Raymond P, Allaire M, Nabi IR, Tijssen P (2001) A viral phospholipase A2 is required for parvovirus infectivity. Dev Cell 1:291–302

14. Lombardo E, Ramírez JC, García J, Almendral JM (2002) Complementary roles of multiple nuclear targeting signals in the capsid proteins of the parvovirus minute virus of mice during assembly and onset of infection. J Virol 76:7049–7059
15. Lombardo E, Ramirez JC, Agbandje-McKenna M, Almendral JM (2000) A beta-stranded motif drives capsid protein oligomers of the parvovirus minute virus of mice into the nucleus for viral assembly. J Virol 74:3804–3814
16. Riolobos L, Reguera J, Mateu MG, Almendral JM (2006) Nuclear transport of trimeric assembly intermediates exerts a morphogenetic control on the icosahedral parvovirus capsid. J Mol Biol 357:1026–1038
17. Chen XS, Stehle T, Harrison SC (1998) Interaction of polyomavirus internal protein VP2 with the major capsid protein VP1 and implications for participation of VP2 in viral entry. EMBO J 17:3233–3240
18. Reguera J, Carreira A, Riolobos L, Almendral JM, Mateu MG (2004) Role of interfacial amino acid residues in assembly, stability, and conformation of a spherical virus capsid. Proc Natl Acad Sci USA 101:2724–2729
19. Li PP, Nakanishi A, Clark SW, Kasamatsu H (2002) Formation of transitory intrachain and interchain disulfide bonds accompanies the folding and oligomerization of simian virus 40 Vp1 in the cytoplasm. Proc Natl Acad Sci USA 99:1353–1358
20. Raices M, D'Angelo MA (2012) Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions. Nat Rev Mol Cell Biol 13:687–699
21. Robbins J, Dilworth SM, Laskey RA, Dingwall C (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell 64:615–623
22. Mattaj IW, Englmeier L (1998) Nucleocytoplasmic transport: the soluble phase. Annu Rev Biochem 67:265–306
23. Chang D, Haynes JI, Brady JN, Cosigli RA (1992) Identification of a nuclear localization sequence in the polyomavirus capsid protein VP2. Virology 191:978–983
24. Moreland RB, Garcea RL (1991) Characterization of a nuclear localization sequence in the polyomavirus capsid protein VP1. Virology 185:513–518
25. Ishii N, Minami N, Chen EY, Medina AL, Chico MM, Kasamatsu H (1996) Analysis of a nuclear localization signal of simian virus 40 major capsid protein Vp1. J Virol 70:1317–1322
26. Ishii N, Nakanishi A, Yamada M, Macalalad MH, Kasamatsu H (1994) Functional complementation of nuclear targeting-detective mutants of simian virus 40 structural proteins. J Virol 68:8209–8216
27. Vihinen-Ranta M, Wang D, Weichert WS, Parrish CR (2002) The VP1 N-terminal sequence of canine parvovirus affects nuclear transport of capsids and efficient cell infection. J Virol 76:1884–1891
28. Grieger JC, Snowdy S, Samulski RJ (2006) Separate basic region motifs within the adeno-associated virus capsid proteins are essential for infectivity and assembly. J Virol 80:5199–5210
29. Johnson JE (1996) Functional implications of protein-protein interactions in incosahedral viruses. Proc Natl Acad Sci USA 93:27–33
30. Hogle JM (2002) Poliovirus cell entry: common structural themes in viral cell entry. Annu Rev Microbiol 56:677–702
31. Sonntag F, Schmidt K, Kleinschmidt JA (2010) A viral assembly factor promotes AAV2 capsid formation in the nucleolus. Proc Natl Acad Sci U S A 107:10220–10225
32. Riolobos L, Valle N, Hernando E, Maroto B, Kann M, Almendral JM (2010) Viral oncolysis that targets Raf-1 signaling control of nuclear transport. J Virol 84:2090–2099
33. Caspar DLD, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol XXVII:1–24
34. Salunke DM, Caspar DL, Garcea RL (1986) Self-assembly of purified polyomavirus capsid protein VP1. Cell 46:895–904

35. Hernando E, Llamas-Saiz AL, Foces-Foces C, McKenna R, Portman L, Agbandje-McKenna M, Almendral JM (2000) Biochemical and physical characterization of parvovirus minute virus of mice virus-like particles. Virology 267:299–309
36. Yuan W, Parrish CR (2001) Canine parvovirus capsid assembly and differences in mammalian and insect cells. Virology 279:546–557
37. Maroto B, Ramírez JC, Almendral JM (2000) Phosphorylation status of the parvovirus minute virus of mice particle: mapping and biological relevance of the major phosphorylation sites. J Virol 74:10892–10902
38. Garcea R, Benjamin T-L (1983) Host range of trasnforming gene of polyoma virus plays a role in virus assembly. Proc Natl Acad Sci USA 80:3613–3617
39. Garcea RL, Ballmer-Hofer K, Benjamin TL (1985) Virion assembly defect of polyomavirus *hr-t* mutants: underphosphorylation of major capsid protein VP1 before viral DNA encapsidation. J Virol 54:311–316
40. Naumer M, Sonntag F, Schmidt K, Nieto K, Panke C, Davey NE, Popa-Wagner R, Kleinschmid JA (2012) Properties of the AAV assembly activating protein AAP. J Virol 86:13038–13048
41. Erickson KD, Bouchet-Marquis C, Heiser K, Szomolanyi-Tsuda E, Mishra R, Lamothe B, Hoenger A, Garcea RL (2012) Virion assembly factories in the nucleus of polyomavirus-infected cells. PLoS Pathog 8:1–15
42. Li PP, Nakanishi A, Shum D, Sun PC, Salazar AM, Fernandez CF, Chan S-W, Kasamatsu H (2001) Simian virus 40 VP1 DNA-binding domain is functionally separable form the overlapping nuclear localization signal and is required for effective virion formation and full viability. J Virol 75:7321–7329
43. King JA, Dubielzig R, Grimm SW, Kleinschmidt JA (2001) DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. EMBO J 20:3282–3291
44. Plevka P, Hafenstein S, Li L, D'Abramo A, Cotmore SF, Rossmann MG, Tattersall P (2011) Structure of a packaging-defective mutant of minute virus of mice indicates that the genome is packaged *via* a pore at a 5-fold axis. J Virol 85:4822–4827
45. Li M, Garcea RL (1994) Identification of threonine phosphorylation sites on the polymavirus major capsid protein VP-1: relationship to the activity of middle T antigen. J Virol 68:320–327

Further Reading

Carrasco L, Almendral JM (eds) (2006) Virus patógenos. Hélice and Fundación BBVA, Madrid
Flint SJ, Enquist LW, Racaniello VR, Skalka AM (2009) Principles of virology, 3rd edn. ASM Press, Washington, DC
Kerr JK, Cotmore SF, Bloom ME, Linden RM, Parrish CR (eds) (2006) Parvoviruses. Hodder Arnold, London
Harrison SC (2006) Principles of virus structure. In: Knipe DM, Howley PM (eds in chief) Fields virology, 5th edn. Lipincott Williams and Wilkins, Philadelphia, p 59–98
Reddy VS, Johnson JE (2005) Structure-derived insights into virus assembly. Adv Virus Res 64:45–68
Zlotnick A, Fane BA (1997) Mechanisms of icosahedral virus assembly. In: Agbandje-McKenna M, McKenna R (eds) Structural biology. RSC Publishing, Cambridge, pp 180–202

Chapter 11
Structure and Assembly of Complex Viruses

Carmen San Martín

Abstract Viral particles consist essentially of a proteinaceous capsid protecting a genome and involved also in many functions during the virus life cycle. In simple viruses, the capsid consists of a number of copies of the same, or a few different proteins organized into a symmetric oligomer. Structurally complex viruses present a larger variety of components in their capsids than simple viruses. They may contain accessory proteins with specific architectural or functional roles; or incorporate non-proteic elements such as lipids. They present a range of geometrical variability, from slight deviations from the icosahedral symmetry to complete asymmetry or even pleomorphism. Putting together the many different elements in the virion requires an extra effort to achieve correct assembly, and thus complex viruses require sophisticated mechanisms to regulate morphogenesis. This chapter provides a general view of the structure and assembly of complex viruses.

Keywords Virus structure • Virus assembly • Symmetry • Capsid • Cementing proteins • Envelope • Symmetry mismatch • Scaffold • Maturation • Virus evolution

Abbreviations

3D	Three-dimensional
ABV	*Acidianus* bottle-shaped virus
ATV	*Acidianus* two-tailed virus
cryo-EM	Cryo-electron microscopy
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA

C. San Martín (✉)
Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: carmen@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_11,

EMDB	Electron Microscopy Data Bank
GON	Group of nine
GOS	Group of six
PBCV-1	*Paramecium bursaria Chlorella* virus-1
PDB	Protein Data Bank
Sid	Size determination protein
SNDV	*Sulfolobus neozelandicus* droplet-shaped virus
ssDNA	Single-stranded DNA
SSIP-1	*Salisaeta* icosahedral phage 1
ssRNA	Single-stranded RNA
STIV	*Sulfolobus* turreted icosahedral virus

11.1 Introduction

A viral particle consists essentially of a proteinaceous capsid with multiple roles in protection of the viral genome, cell recognition and entry, intracellular trafficking and controlled uncoating. Evolutionary forces have caused viruses to adopt different strategies to achieve these goals. Simple viruses (Chap. 10) generally build their capsids from a number of copies of the same, or a few different proteins, organized into a symmetric oligomer. In the case of complex viruses, capsid assembly requires further elaborations. What are the main characteristics that define a structurally complex virus?

Structural complexity on a virus often, but not necessarily, derives from the need to house a large genome, in which case a larger capsid is required. However, capsid or genome sizes by themselves are not determinants of complexity. For example, flexible filamentous viruses can reach lengths in the order of microns, but most of their capsid mass is built by a single capsid protein arranged in a helical pattern [1]. On the other hand, architecturally complex viruses such as HIV have moderate sized genomes (7–10 kb of single-stranded (ss) RNA) [2]. Structurally complex viruses incorporate a larger variety of components into their capsids than simple viruses. They may contain accessory proteins with specific architectural or functional roles; or incorporate non-proteic elements such as lipids.

The elaborated composition of complex virus particles often involves a rupture of the basic symmetry rules (Chap. 2), from a range of symmetry mismatches in icosahedral shells, to completely asymmetric or pleomorphic capsids. The more subtle departure from symmetry is the case when identical subunits occupy similar but slightly different environments, as in the case of quasi-equivalence in icosahedral shells (see Chap. 2). In other cases, virion components with different symmetry may interact with each other, forming a *symmetry mismatch* at the interface. In the extreme case, identical components may form morphological units with no

symmetry, or even assemble in a completely different manner for each realization of the virion – this property is called *pleomorphism*. Recent advances in cryo-electron microscopy (cryo-EM) (Chap. 3) and X-ray crystallography (Chap. 4) are helping to unveil the organization of complex viruses in great detail, including features that depart from strict icosahedral symmetry. Understanding the architectural details of asymmetric capsids is the most challenging problem, since structural biology techniques heavily rely on the use of symmetries to reach high resolution detail. Cryo-electron tomography (Chap. 3) is helping to advance our understanding of these viruses, although the resolution currently attained is still in the 3–5 nm range.

The presence of many different elements in the virion entails an extra effort to achieve correct assembly. Accordingly, complex virus morphogenesis requires sophisticated mechanisms, tightly regulated in space and time. Here we provide a general view of all these variations in complexity, finishing with a consideration on the evolutionary insights provided by structural studies on complex viruses.

11.2 Molecular Composition of Complex Viruses

11.2.1 Different Proteins with Specific Roles

A characteristic feature of complex viruses is the presence of multiple proteins in the virion, playing specific architectural or functional roles during the viral cycle. For example, different proteins may occupy the sixfold and fivefold coordinated positions in the icosahedral net (see Chap. 2). The specific architectural role of proteins at the fivefold vertices is often combined with a specific functional role, as will be described in Sect. 11.3. In icosahedral viruses with triangulation numbers $T > 1$, mobile terminal regions of the capsid proteins may adopt different conformations depending on their position in the capsid. In this way they act as molecular switches, enabling the same protein to occupy the different quasi-equivalent environments (see Chap. 2). In complex viruses, these mobile arms may still exist, but they often appear combined with a variety of minor capsid proteins, required for correct assembly of the virion. These *cementing proteins* can be considered as detached molecular switches, required to modulate the variety of interactions needed for assembly and stability of a complex capsid. One case where the intricate capsid organization includes: (i) biochemically different hexameric and pentameric capsomers; (ii) a network of mobile arms; and (iii) cementing polypeptides, has recently been described in great detail: adenovirus [3, 4].

Adenoviruses infect vertebrates. They enclose their dsDNA genome within a *pseudo* $T = 25$ icosahedral shell with a diameter of 950 Å between vertices. Trimers of the major capsid protein, hexon, constitute all the sixfold capsomers, while the vertices are occupied by pentamers of a protein named penton base. Although there is no sequence similarity between hexon and penton base, both have

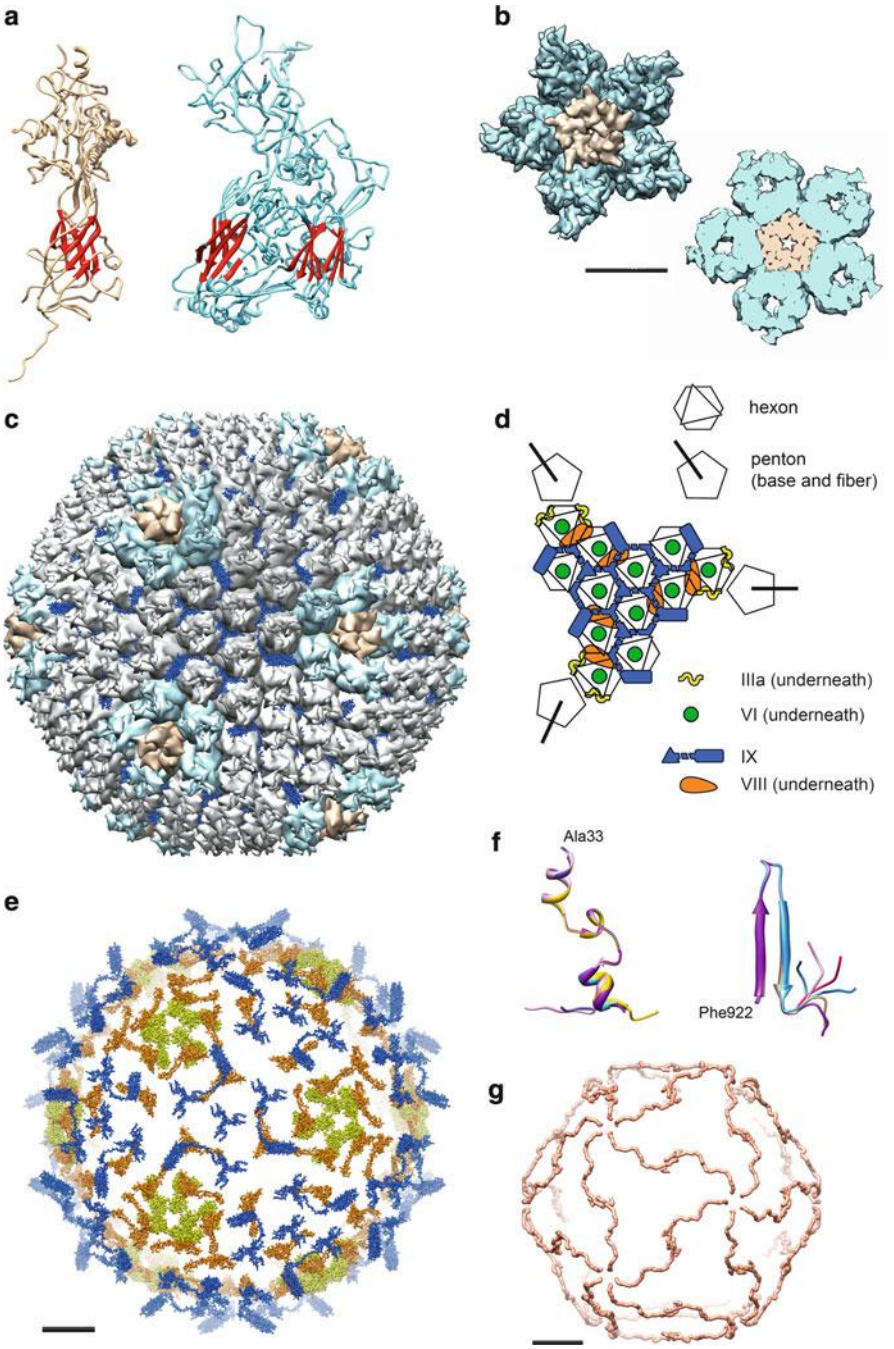

Fig. 11.1 Molecular composition of complex viruses: capsid proteins, cementing proteins, and mobile arms. (**a**) Structure of adenovirus penton base (*beige*) and hexon (*cyan*) monomers, with

a very similar fold based on the eight-stranded β-barrel motif, or "jelly roll", which allows appropriate interdigitation of the different capsomers to form the closed shell [5, 6]. Penton base folds as a single jelly roll. The hexon monomer folds as a double β-barrel jelly-roll perpendicular to the capsid surface (Fig. 11.1a). Thanks to this fold, trimers have a pseudo-hexagonal shape and can occupy sixfold symmetry positions in the icosahedron. Each capsid facet is formed by 12 trimers of hexon. The general icosahedral architecture of adenovirus can be described as two different systems of tiles. Nine hexon trimers form the central plate of each facet, known as GON (Group of Nine). The five peripentonal hexon trimers, together with the penton base, form the second tile system, known as GOS (Group of Six) (Fig. 11.1b). Numerous double-stranded (ds) DNA viruses share the basic jelly-roll fold in their capsid proteins and assemble capsids organized as combinations of tiles based on pentagonal and pseudo-hexagonal building blocks. In larger viruses, the tiles composed by the pentameric vertex and surrounding capsomers are known as *pentasymmetrons*, while the triangular tiles centered at the threefold icosahedral axis are called *trisymmetrons*. This efficient architectural solution facilitates construction of very large icosahedral capsids; triangulation numbers as large as $T = 169$ have been described, and larger ones are likely to exist [8, 9].

In adenovirus, apart from hexon and penton base, there are at least four other proteins making up the icosahedral shell. Minor capsid proteins IIIa, VI, VIII and IX are required for correct capsid assembly and occupy specific positions in the capsid (Figs. 11.1c, d, e), forming specialized networks that stabilize the two systems of tiles [3]. Polypeptide IX is the only cementing protein located on the outer part of the adenovirus capsid. It has an extended structure and forms a sort of hairnet keeping together the hexon trimers in each GON, and binding GONs to

Fig. 11.1 (continued) the β-barrel jelly roll motifs highlighted in *red*. The molecules are oriented so that the external capsid surface is up. Notice the extended N-terminal arm in penton base directed towards the interior of the virion (Unless otherwise indicated, all ribbon and surface structure representations in this chapter, as well as fullerene models, were prepared using USCF Chimera software [7]). (**b**) One penton base pentamer and the five peripentonal hexon trimers form the GOS, the adenovirus pentasymmetron. *Left*, view of the GOS from outside the capsid. *Right*, a slab showing the tight interdigitation at the base of the capsomers to close the shell. Color scheme as in (**a**). (**c**) Adenovirus capsid, seen along a threefold icosahedral axis. The GOS is colored as in (**b**), and the rest of the hexon trimers are in *gray*. These form the central plate of the icosahedral facet, termed GON, or trisymmetron in the more general terminology. In *blue*, the cementing protein polypeptide IX. (**d**) Schematics showing the location of cementing proteins in an adenovirus capsid facet. Reproduced from [4]. (**e**) Cementing proteins in the adenovirus capsid. The view is as in (**c**), with the pentons and hexons removed to reveal both the external and internal networks of accessory proteins. These are colored as in (**d**). Notice that polypeptide VI has not been traced in the high resolution structure. (**f**) Mobile arms in adenovirus hexon. Different conformations of the hexon N-terminal (*left*) and C-terminal (*right*) regions in the capsid are shown. Panels (a), (b), (c), (e) and (f) made from atomic coordinates deposited at the Protein Data Bank (PDB; Protein Data Bank is at http://www.pdb.org) with entry ID 3IYN. (**g**) Minor capsid proteins as size determinants. The bacteriophage PRD1 "tape measure" protein P30 forms a cage beneath the capsid surface (PDB ID 1W8X). Scale bars correspond to 100 Å

GONs across the icosahedral edges. The N-terminal domains of three IX monomers join *via* hydrophobic interactions at the icosahedral and local threefold axes in the GONs forming triskelion structures. A long, unstructured domain of each monomer runs in a different direction towards the facet edges, where the C-terminal α-helix joins with the C-terminal helices of another three copies of IX, different from those forming the N-terminal triskelion, to create a leucine 4-helix bundle. On the interior of the shell, each GON is further stabilized by copies of polypeptide VIII located around the icosahedral threefold symmetry axis. Also on the inner capsid surface, polypeptide IIIa mediates the interaction between penton base and the peripentonal hexons, to keep each GOS together. Finally, IIIa and VIII cooperate to bind each GOS to its five surrounding GONs. The remaining minor capsid protein, polypeptide VI, has not been unequivocally traced so far, but has been assigned to density within an internal cavity present in each hexon trimer.

Mobile regions of hexon and penton base also play a role in the extensive interaction networks in the adenovirus capsid. Due to their flexibility, these regions could not be traced in the crystal structures of the isolated proteins, but they adopt ordered conformations when they are within the capsid context. The N- and C-termini of the hexon monomer, located at the innermost part of the capsomer, adopt a total of 5 (N-) and 6 (C-) different conformations to establish interactions between neighbouring hexons, or between hexons and minor capsid proteins (Fig. 11.1f). Similarly, for each penton base monomer an N-terminal arm extends away from the β-barrels that form the main body of the protein towards the viral core, interacting with two IIIa monomers along the way, and therefore contributing to anchor the penton within the GOS. Interestingly, some of the interactions between cementing proteins and hexons, and among cementing proteins, occur by β-sheet augmentation. That is, the interaction is mediated by a β-strand from one of the proteins binding to the edge of a β-sheet in the other. This observation tells about the intricate organization of the capsid and makes us wonder about the difficulty of assembling all elements together. The fact that no high resolution structure is available for any of the minor capsid proteins in isolation suggests that they may require the virion context to fold properly.

Although in general it is understood that cementing proteins are required for correct viral assembly, it is difficult to pinpoint their exact role in morphogenesis. Some of them are dispensable for assembly, but required to reach structural stability; this is the case of adenovirus polypeptide IX [10]. Others are thought to play the role of "molecular rulers", determining capsid size. This role was proposed when the structure of bacteriophage PRD1 was solved by protein crystallography (see also Chap. 17). It was then found that minor capsid protein P30, required for assembly, runs beneath the icosahedral edges, from the vertex to the twofold symmetry axis. Thanks to its extended conformation, two copies of the 83-residue polypeptide can cover the 300 Å length of the capsid edge, and act as a tape measure during morphogenesis [11] (Fig. 11.1g). A further complication for determining the role of minor virion components in assembly comes from the fact that, in keeping with the genetic economy principle, they often play other roles different from the purely architectural one during the viral cycle. A remarkable example of this phenomenon is illustrated by adenovirus polypeptide VI [4]. This protein is

involved in disrupting the endosomal membrane, so that the virus can escape into the cytosol after internalization. It also has a role in facilitating virion traffic to the nucleus along the microtubular network; acts as an activator of the adenoviral gene expression; and promotes transport of newly synthesized hexon to the nucleus. Finally, a C-terminal peptide of polypeptide VI activates the viral protease for maturation.

Other additional proteins may be incorporated to the viral particles and play fundamental roles for viability. Elucidating the organization of these additional components within the virion is not straightforward, since they usually do not follow defined symmetry rules, and their disposition may even change between particles of the same virus. Notable examples are viral proteases, such as the maturation protease VP24 in herpesvirus [12], or the adenoviral protease AVP [4]; and molecular motors involved in nucleic acid translocation, such as dsDNA packaging ATPases in bacteriophage and herpesviruses [13, 14], or the dsRNA packaging ATPase in cystoviruses (bacteriophage Φ6) [15]. Viruses with RNA genomes must carry their own replication and transcription enzymes, to supply RNA metabolism functions absent in the cell [16]. Viruses that carry out their replication in the cytosol (*e.g.*, vaccinia) must also supply DNA and RNA processing enzymes whose cellular counterparts are only present in the nucleus [17]. Some dsDNA viruses encapsidate basic proteins that help screen the nucleic acid charge repulsion, to facilitate compaction of the genome within the reduced capsid space. These basic proteins can be of cellular origin, such as in Simian Virus 40, which uses histones to pack its minichromosome [18]. Baculovirus [19], adenovirus [20], mimivirus [21] and poxviruses [17] encode their own DNA compacting proteins. The genomes of negative strand ssRNA viruses usually appear in the form of ribnucleoproteic structures [22]. More information on the packaging motors and on the organization of nucleic acids within viral capsids can be found in Chap. 12.

11.2.2 Membranes

Apart from the genome and structural proteins, a large number of viruses incorporate lipidic layers into their architecture. Lipid bilayers (membranes) are widely extended in biological entities such as cells and organelles, and are ideally suited for enclosing a defined volume and separating it from neighboring compartments or the surrounding environment. This is the same function they play in viral capsids. For viruses, membranes are particularly advantageous, since they can readily be taken from the cell, are highly scalable in size, and do not consume coding space in the genome. More detail on how viruses sequester cell membranes for their own use can be found in Chap. 14.

Viruses with external membranes are called *enveloped viruses*. Envelopes form a protective layer, blocking entry of aggressive chemicals or enzymes into the viral particle. They are composed not only by lipids but also by protein and sugars

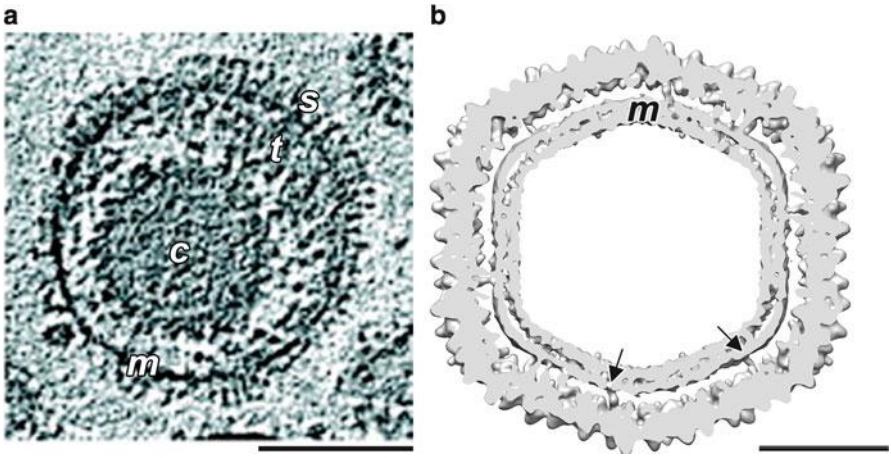

Fig. 11.2 Viruses with lipid bilayers. (**a**) Slice of a cryo-electron tomography reconstruction of herpes simplex virus-1 showing the icosahedral capsid (*c*), tegument (*t*), membrane (*m*) and glycoprotein spikes (*s*). Scale bar, 1,000 Å (Modified from [25]). (**b**) Bacteriophage PRD1 has an internal membrane (*m*) that adopts an icosahedral profile due to interactions with capsid proteins (*arrows*). Shown is a slab of the cryo-EM map deposited at the Electron Microscopy Data Bank (EMDB; Electron Microscopy Data Bank is at http://www.ebi.ac.uk/pdbe/emdb/) with ID EMD_1013, viewed along a twofold icosahedral axis. The scale bar corresponds to 100 Å

(glycoproteins). An implication of this arrangement is that in enveloped viruses infectivity is linked to membrane integrity, because the viral attachment proteins which recognize the host cell receptors and trigger internalization are in the viral envelope. Such is the case of influenza virus hemaglutinins [23], or Env proteins in retroviruses [2]. In the simpler enveloped viruses, such as Alphaviruses, the glycoproteins completely cover the lipid surface and follow the same icosahedral organization as the inner shell (T = 4) [24]. In other cases, such as retroviruses or herpesviruses (Fig. 11.2a), the structure directly in contact with the genome (nucleocapsid or capsid) is not in contact with the membrane, and the organization of glycoprotein spikes is irregular and does not reflect the organization of the virion inner contents [25, 26].

The membrane can also be located beneath the icosahedral shell, such as in Tectiviruses (PRD1) and structurally related viruses [11, 24]. In these cases, the membrane itself adopts an icosahedral layout forced by interactions with the capsid proteins (Fig. 11.2b). In PRD1, some of the mobile arms in its major capsid protein are embedded in the outer leaflet of the membrane, contributing to enforce the icosahedral shape [11]. Icosahedral viruses with internal membranes have a large complement of virion-encoded membrane proteins: approximately half of the 18 proteins present in the PRD1 virion are membrane proteins [27]. The membrane in PRD1 can undergo a large conformational change and protrude forming a tube from one of the vertices. It has been proposed that this tube has a function in injecting the

viral genome into the host cell, similar to that played by tail structures in other bacteriophages [28] (see Chap. 17).

Some of the most architecturally complex viruses are enveloped. For example, herpesviruses have a T = 16 icosahedral capsid (1,500 Å in diameter), formed by one major capsid protein and several accessory proteins. This capsid is surrounded by a thick tegument layer, containing at least 13 different viral proteins and also some cellular components. Capsid and tegument are enveloped by a membrane with more than 12 different types of viral glycoproteins [12, 25]. Large dsDNA viruses infecting aquatic eukaryotic microorganisms, including the giant *Acanthamoeba polyphaga* Mimivirus (diameter ~750 nm), have internal membranes like bacteriophage PRD1 [21, 29]. Asfarviruses (African swine fever virus, diameter ~200 nm) have both an internal membrane surrounded by an icosahedral shell and a loose external envelope [30]. Other examples of complex, lipid-containing viruses will be described in Sect. 11.4.

11.3 Departures from Symmetry in Quasi-Icosahedral Capsids

11.3.1 Layers with Different T Numbers

In Sect. 11.2.2, it was pointed out that viruses could be organized in multiple layers, intercalating protein (ordered or not) and lipids. In other cases such as adenovirus, multiple cementing proteins combine to form a single icosahedral capsid, while additional components (*e.g.*, dsDNA condensing proteins) do not show any symmetrical organization. In yet another instance, concentric icosahedral protein shells are formed. Remarkably, these shells may have different triangulation numbers, including some not predicted by the theory of quasi-equivalence. This type of organization is most prominently present in the dsRNA Reoviruses.

Rotaviruses and mammalian orthoreoviruses are the best characterized members of the *Reoviridae* family [16]. The mature rotavirus virion has a diameter of approximately 100 nm, and is organized in three concentric layers composed by four different proteins (Fig. 11.3). The innermost layer, or core shell, is formed by 120 molecules of protein VP2. Because of the number of protein monomers in this shell, it is described as having a T = 2 triangulation number, a conformation not allowed in the Caspar and Klug formalism [31] (Chap. 2). In fact, the VP2 monomers adopt two different conformations to yield a T = 1 icosahedron of asymmetric dimers. VP2 works as a platform for assembly of the next layer, formed by 260 trimers of protein VP6 in a T = 13 lattice. VP6 must be able to adopt not only the different conformations required for the T = 13 quasi-equivalence requirements, but also those extra ones required to compensate for the symmetry mismatch with the VP2 layer. The third, and outermost layer, is also a T = 13 icosahedron composed by 260 trimers of glycoprotein VP7, plus 60 spikes formed by trimers of VP4 that protrude from peripentonal positions. The VP4 spikes must

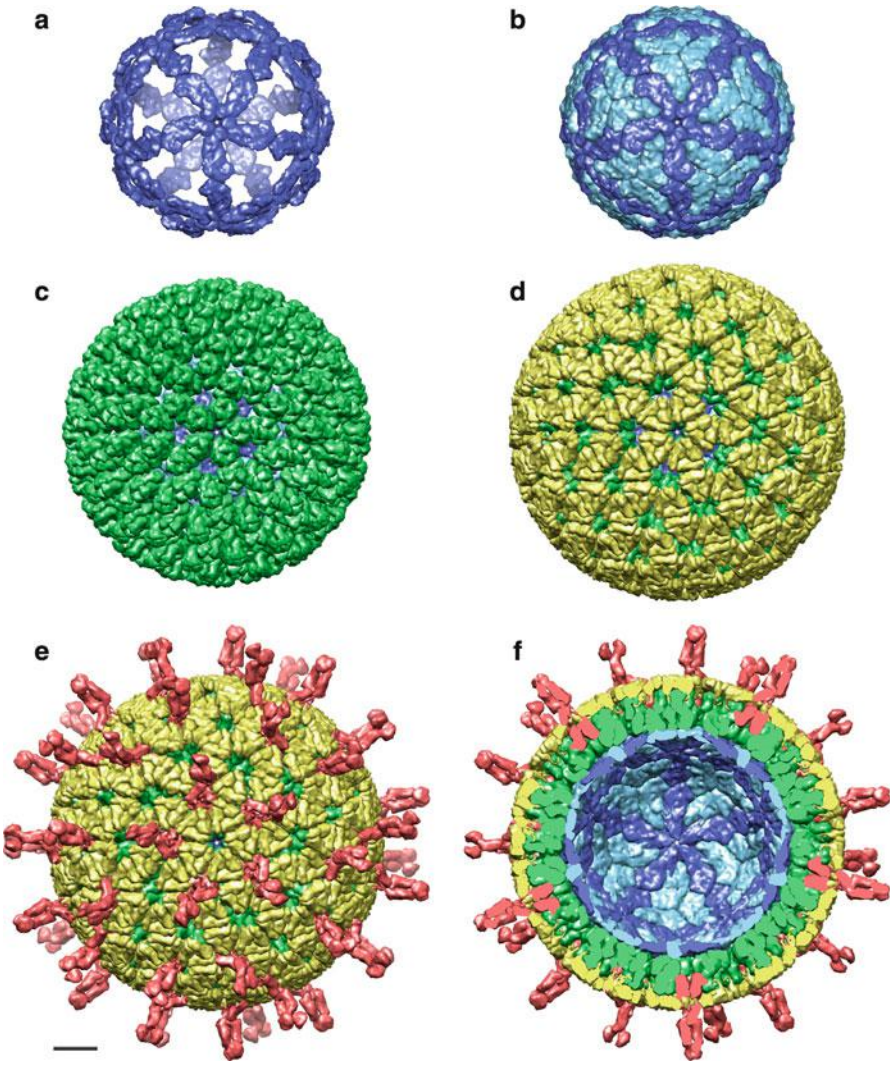

Fig. 11.3 Multiple layered viruses: rotavirus capsid architecture. Panels (**a**) to (**e**) show the consecutive building layers of the mature rotavirus virion. (**a**) Only one type of VP2 monomer in the core shell is represented. (**b**) The VP2 core shell, a T = 1 particle formed by asymmetric dimers. (**c**) The intermediate T = 13 VP6 layer. (**d**) VP7 glycoprotein layer. (**e**) The complete virion with VP4 spikes. (**f**) Cut out representation of the virion depicting the layered organization. The view is along a fivefold icosahedral symmetry axis. The scale bar represents 100 Å. Figure prepared using atomic coordinates from PDB IDs 3N09 and 3IYU

be proteolytically cleaved for the virus to be infectious. Remarkably, after cleavage the fragments remain non-covalently associated on the virion surface, but undergo an intriguing conformational modification, changing from a trimeric arrangement to

a mixture of trimeric, dimeric and monomeric associations. The VP1 and VP3 transcription enzymes are also part of the virion, and are located beneath the core shell surface.

The orthoreovirus capsid is also triple layered, and shares architectural similarities with rotavirus, namely the mismatch between "T = 2" and T = 13 symmetries. However, the composition is more complex, with six different proteins instead of four. The major differences appear in the outermost layers, probably reflecting differences in the viral entry mechanism. Instead of having 60 short spikes distributed in the icosahedral facet, orthoreovirus displays large turreted structures combined with a long, flexible fiber in each of the 12 vertices.

Why do viruses have different layers? As in the case of membranes, protein layers help to separate different compartments, and most likely also different functions along the infectious cycle. dsRNA viruses need to keep their genome confined within the core shell at all times during infection, to protect it from aggressions by cellular nucleases, and to prevent antiviral reactions triggered by accumulation of dsRNA. In rotavirus, the double layered particle formed by VP2 and VP6 is the transcriptionally competent form of the virus. The external layers carry the viral components in charge of initial interaction with the host: recognition, attachment and entry. These are shed once entry into a new host cell has been accomplished [32]. In reovirus however, the turrets are not lost upon entry like the other external layers, but form part of the double layered, transcriptionally active form of the virus [33].

11.3.2 Symmetry Mismatches

In the previous section, we have seen that in reoviruses there is a symmetry mismatch between two concentric shells with different triangulation numbers. Nevertheless, the two layers still follow icosahedral symmetry, and therefore it has been possible to study their organization at a very detailed level. Symmetry mismatches (two elements with different symmetries in direct interaction) are frequent in icosahedral viral structures, particularly at the vertices, where proteins involved in genome translocation or host attachment reside. Solving the organization of mismatched features represents a remarkable challenge for structural biology techniques, due to the predominance of icosahedral symmetry in the complete virion that obscures them. In the *Cystoviridae* representative bacteriophage Φ6, a hexameric ssRNA packaging ATPase occupies multiple fivefold vertices of the empty procapsid [15]. In dsDNA bacteriophages, 12-fold portal structures occupy one of the vertices (Sect. 11.3.4 and Chaps. 12 and 17). Host recognition elements often take the shape of elongated fibers protruding from the fivefold capsomers. The oligomerization state and number of fibers per vertex varies, and is usually at odds with the pentameric architecture of the capsomers.

In human adenoviruses, a trimer of the fiber forms a non-covalent complex with a pentamer of penton base (Fig. 11.4a). The C-terminal domain of fiber is the main

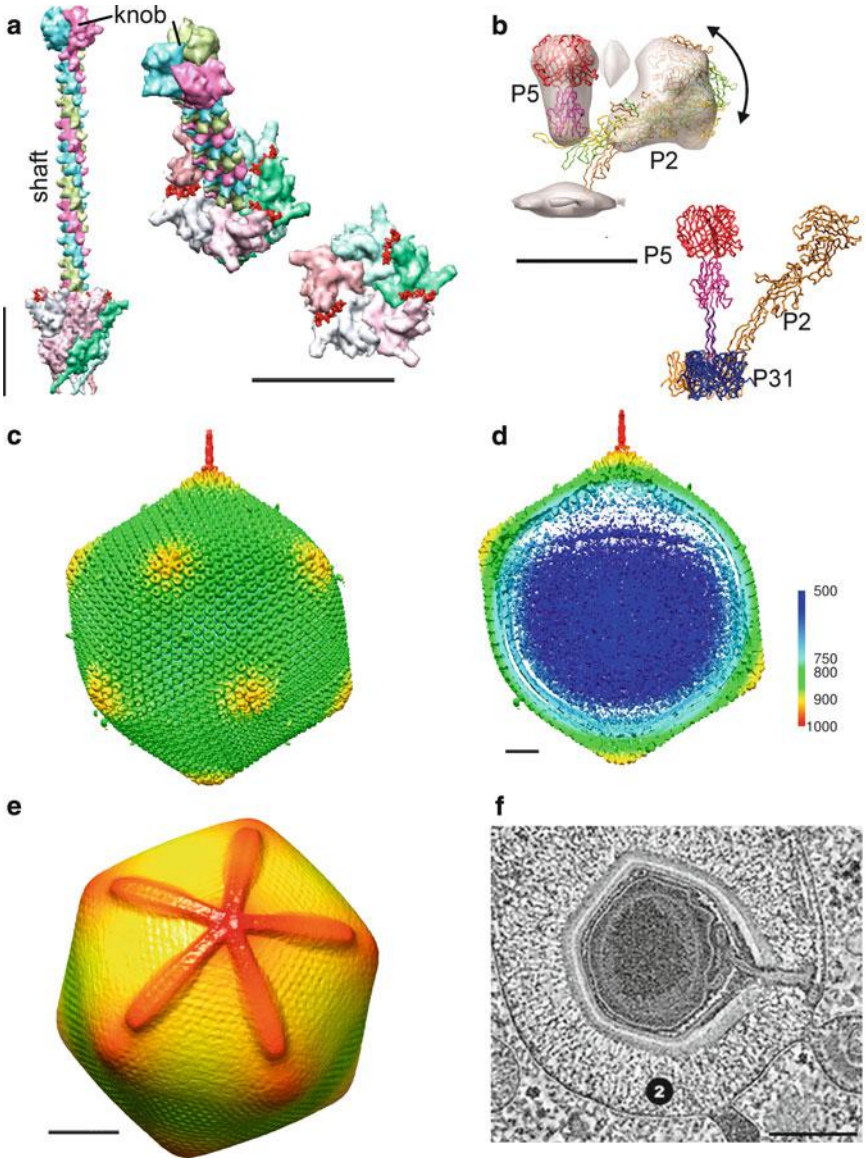

Fig. 11.4 Symmetry mismatches and singular vertices. (**a**) Adenovirus vertex complex, composed by pentameric penton base (compare with Fig. 1a and b) and the trimeric fiber (knob and distal shaft from PDB ID 1QIU, rest of the shaft modeled). In *red*, three N-terminal fiber peptides bound to penton base (PDB ID 3IZO). *Left*: *side view*; *center*, *oblique view*; *right*, *top view* of the penton base with the fiber shaft and knob removed to highlight the position of the N-terminal fiber peptides. (**b**) A model for the double fiber complex in bacteriophage PRD1 vertex. *Top*: high resolution structures of trimeric P5 and monomeric P2 fitted to a cryo-EM reconstruction of the vertex region (*side view*). The *arrow* indicates variability in the position of P2. Bottom: a model of the complex interactions between the three vertex components. P31 forms the fivefold capsomers

player in trimerization and forms the distal knob, responsible for binding to the adenovirus primary receptor. The knob is followed by a shaft of variable length and flexibility, depending on the serotype. The knob domain has an eight-stranded β-sandwich fold similar to that of the major capsid protein, and the shaft forms a triple β-spiral [36]. Finally, the N-terminal region is responsible for binding to penton base. At the proximal part the shaft becomes frayed and three flexible N-terminal peptides spread out to attach to the penton base pentamer. The crystal structure of penton base in complex with an N-terminal fiber peptide [5] showed the peptide bound to a groove on the outer surface of the pentamer formed by the interface between two penton base monomers, reaching radially from the penton center to the outer rim. Five peptides were observed with equivalent density, implying that all fiber binding sites are equivalent. Therefore, in the virus the three N-terminal tails may adopt two different arrangements: either they occupy three consecutive grooves, or two of them are in neighboring binding sites and the third one is flanked by two empty grooves. In the recently solved cryo-EM atomic structure of the complete virion, density for the proximal part of the fiber shaft was observed protruding from the center of the penton base pentamer [3, 37]. Density for the fiber shaft is blurred by the enforced fivefold symmetry, but it could be observed that its base interacts with a hydrophobic ring at the rim of a narrow channel in the center of the penton. This hydrophobic interaction may allow relative rotation of fiber on penton to accommodate the symmetry mismatch, while the N-terminal tails secure the binding to the penton groves. Fiber binding to the adenovirus receptor in the cell results in its release from the capsid, which in turn induces a cooperative conformational change in the penton base pentamer. This change is thought to play a role in preparing pentons for release at a later stage of adenovirus entry and uncoating [4].

In orthoreovirus, the receptor binding protein σ1 also forms a flexible trimeric spike attached to the fivefold turret [16, 38]. Intriguingly, some viruses can hold two fibers attached to the same vertex: this is the case of fowl adenovirus type-1, where two fibers of different length can be observed bound to the same penton base [39]. Bacteriophage PRD1 also has two different spikes (proteins P2 and P5) attached to a single pentamer of the vertex protein P31 (Fig. 11.4b). Moreover, in PRD1 each spike has a different oligomerization state. P5 is a trimer resembling the structure of the adenovirus fiber; while the other, P2, is a monomer with a pseudo-β propeller

Fig. 11.4 (continued) (pentons) in PRD1. This study showed that the two fibers interact with each other at the icosahedral capsid level, and that P2 can move relatively to P5 (Reproduced from [34]. With permission). (**c**) The PBCV-1 capsid (EMDB ID EMD_1597) showing the special vertex with a spike, and (**d**), a central slab where the asymmetry in internal contents can be appreciated. Color key indicates color changes with map radius. (**e**) 3D map of the Mimivirus capsid (EMDB ID EMD_5039) showing the starfish feature (*red*). (**f**) Slice of a tomographic reconstruction showing a Mimivirus particle within a phagosome. The viral membrane is extruded through the open stargate (Modified from [35]. Scale bars represent 100 Å in panels (**a**) and (**b**); 200 Å in (**c**) and (**d**); and 1,000 Å in (**e**) and (**f**)

head. The role of each spike in PRD1 host recognition and attachment is not fully clarified [27, 34] (see also Chap. 17).

The biological significance of these symmetry mismatches has long intrigued virologists. For the dodecameric nucleic acid packaging motors, the mismatch may allow conformational changes required for the translocation function (see Chap. 12). In the case of host recognition fibers, it is possible that the mismatch facilitates flexibility to scan for and attach to the viral receptor, as well as fiber removal upon binding, a step required to initiate the cascade of signals in both cell and virion for appropriate entry and/or genome delivery.

11.3.3 Special Vertices

Special (or singular) vertices have been found in many icosahedral dsDNA viruses. Singular vertices play key roles in genome packaging and ejection; they may also represent initial or final points in the assembly pathway of the shell. They represent a rupture of icosahedral symmetry (one vertex different from the other 11), and often include a symmetry mismatch (protein with non-fivefold symmetry occupying a fivefold coordinated position in the icosahedral net). Dodecameric proteins involved in genome packaging are found in a single vertex in tailed bacteriophages (see Sect. 11.3.4, and Chap. 12), as well as in herpesviruses [13, 14]. The best characterized case of special vertex is the portal in tailed bacteriophages, which connects the icosahedral head with the conspicuous tail that is characteristic of this viral family.

Giant dsDNA viruses infecting eukaryotic microorganisms also have singular vertices. *Paramecium bursaria Chlorella* virus-1 (PBCV-1) has a 190 nm diameter icosahedral capsid surrounding a lipid bilayer and dsDNA genome. A 250 Å long spike protrudes from one of the capsid vertices [8] (Figs. 11.4c, d). The peripentonal capsomers around the singular vertex seem to be structurally different from the rest. A ring-shaped density is observed near the singular vertex inside the capsid, which may correspond to a portal structure involved in genome packaging; however, there is no indication of symmetry mismatch between the fivefold vertex and this ring [29]. The spike is too thin to be used as a DNA ejection tube; besides PBCV-1 is thought to deliver its genome into the host by fusion of the internal membrane with the host one. It has been proposed that the function of the PBCV-1 spike is to puncture the cell wall to initiate the fusion process. The capsid side holding the spike is disassembled upon attachment to the host [29].

The giant Mimivirus has a 500 nm large icosahedral capsid structurally related to those of adenovirus, bacteriophage PRD1, and PBCV-1, covered by 125 nm long fibers. Early images of Mimivirus showed a starfish-shape feature with five arms reaching from one of the vertices to the five neighbouring ones (Fig. 11.4e). The arms of the starfish are inserted between adjacent facets, opening a gap between them. The starfish is an independent macromolecular assembly that remains together when detached from the virion [21], and is the only part of the capsid

not covered by fibers. When Mimivirus enters the cell by phagocytosis, a remarkable structural change occurs, whereby the five icosahedral facets in contact with the starfish feature open, leading to the structure called "stargate" [35]. The internal viral membrane is extruded through the stargate, to fuse with the phagosome membrane and release the viral DNA into the cytosol (Fig. 11.4f). Tailed phages use their special vertex both for genome packaging and delivery (Sect. 11.3.4; Chaps. 12, 17). In Mimivirus however, the stargate vertex is used for genome delivery, but not for packaging, which occurs instead *via* an aperture located in the icosahedral facet.

Remarkably, the asymmetry originated by the singular vertex in both PBCV-1 and Mimivirus reflects in an asymmetry of the internal virion contents [8, 21]. The viral genome and surrounding membrane do not occupy the full internal volume of the capsid. Rather, a gap exists between the DNA core and the side of the capsid containing the special vertex. This gap forms a pocket where viral enzymes required for cell membrane penetration may be contained. It may also contain structural elements required to precisely determine the asymmetric location and shape of the genome within the virion. However, these elements have not been identified or imaged so far.

Other viruses, such as PRD1 or adenovirus, have been reported to have singular vertices, based on genetic, biochemical and immunolabeling assays [40, 41]. However, for these viruses no structural information on the singular vertex is available yet, possibly due to the lack of large conspicuous features (such as tails) that would help calculation of three-dimensional (3D) maps without imposing full icosahedral symmetry.

11.3.4 The Extreme Case: Heads, Tails and Baseplates (Tailed Phages)

Tailed bacteriophages (order *Caudovirales*) are among the best described and more complex of the non-enveloped viruses. Their virions are composed by several functionally specialized morphological units, arranged according to different symmetries and connected *via* multiple symmetry mismatches (see also Chap. 17).

Bacteriophage capsids (heads) contain the dsDNA viral genome. They are icosahedrally ordered, but details vary among the different viruses. For example, HK97 or T7 phages have a strictly icosahedral head with a $T = 7$ net and a single major capsid protein occupying both the sixfold and fivefold coordinated positions. Others, like T4, have elongated, prolate icosahedral heads; two different proteins form the hexameric and pentameric capsomers, and several minor proteins are located at specific positions on the head surface [42] (Fig. 11.5a). One of the 12 vertices in the head is different from the other 11. Instead of a regular penton, it contains the portal structure or connector. This specialized structure is critical during assembly, since it contains the machinery needed to select the viral genome

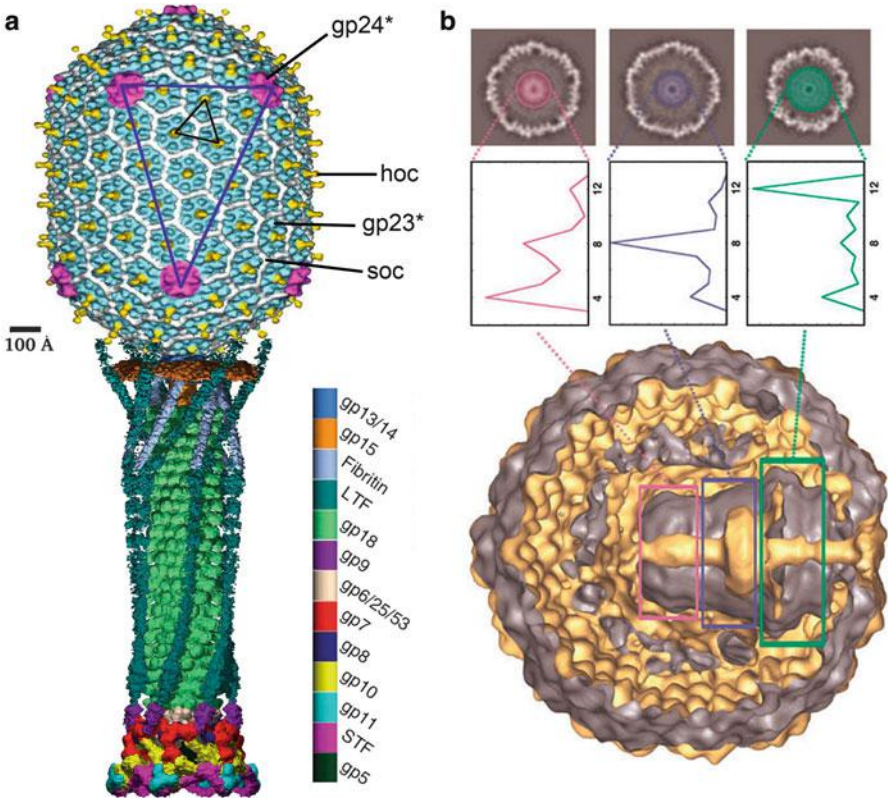

Fig. 11.5 Structure of tailed bacteriophages. (**a**) Bacteriophage T4 virion. The structures of the head and tail/baseplate have been solved separately, and merged to compose a representation of the complete virion. The tail is shown in its extended conformation (Modified from [43, 44]. With permission). (**b**) 3D map of the bacteriophage T7 procapsid, showing the internal core complex. The plots show the dominant symmetry for each core region (Reproduced from [45]. With permission). Diameter of the T7 procapsid = 510 Å

and pump it into the head (see Chap. 12). In addition, it links the head to the feature that differentiates these viral families: the tail.

Bacteriophages in the *Caudovirales* order are further classified in three groups, depending on the tail morphology: long, contractile (*Myoviridae*); long, non-contractile (*Siphoviridae*); and short, non-contractile (*Podoviridae*). In *Myoviridae* (representative: T4), the tail is composed by two layers of protein, one of them (the outer one) contractile [46] (Fig. 11.5a). The tail connects the head to a distal structure called the baseplate, formed by at least 16 different proteins in T4. Fibers with different lengths protrude from the baseplate; fibers may also be present at the portal region and the icosahedral head. In siphoviruses (representative: phage λ), the tail lacks the outer contractile sheath. Research on the structure of non-contractile tails has unveiled a crucial structural element: the tape measure protein,

whose length determines that of the tail by limiting the stacking of tail protein rings to a defined number. The baseplate composition is variable among the different viruses in this family, with some having only a simpler element called the tail tip complex. Fibers project both laterally from the periphery or longitudinally from the very the tip of the tail tip complex [47]. Podoviruses (representatives: T7, P22) have short tails, with a fiber complement that may include long, thin fibers (T7) or thick spikes (P22) [48]. Fibers, tails and baseplates or tail tips form the complex machinery required to initiate infection by recognizing and attaching to the host, and delivering the viral genome through the many layers protecting the bacterial cell (see Chap. 17).

Tailed bacteriophages deviate from the icosahedral symmetry due to their conspicuous genome delivery apparatus; additionally, they are a compendium of symmetry mismatches. Icosahedral (prolate or not) heads have a singular vertex where a fivefold symmetric capsomer is replaced by a 12-fold ring of the portal protein [13]. The portal complex is connected to the tail, which in general follows sixfold symmetry along the tube and baseplate. In the case of myoviruses however, a further mismatch may exist, since the contractile sheath presents helical symmetry, and it is not yet clear if the inner tube follows the arrangement of the sheath or the sixfold symmetry observed in non-contractile tails [46]. Additionally, some podoviruses such as T7, incorporate an internal proteic structure referred to as the core. This structure grows from the portal vertex towards the capsid center, and is thought to serve as a spindle for wrapping the DNA. In T7, the core presents eightfold and fourfold symmetries [45] (Fig. 11.5b). Finally, another symmetry mismatch may appear when the packaging motor binds to the portal vertex in the prohead during encapsidation. Reported oligomeric states for components of packaging motors include pentamers (T4 gp17, Φ29 pRNA), octamers (SF6 small terminase) and tetramers (λ terminase). However, for some of these motors it is not clear if the oligomerization states found in recombinant proteins are the same than in the prohead (immature capsid) context [13].

11.4 Asymmetric Virus Particles

11.4.1 Brick-Shaped Viruses

Poxviruses are large, enveloped dsDNA viruses apparently lacking any kind of high order symmetry in their capsids. The representative of the poxvirus family is vaccinia virus (Fig. 11.6a). The complex composition, large size, and asymmetric organization; plus the sensitivity of these viruses to the different preservation methodologies for EM, have caused the architecture of vaccinia to be a subject of debate for a long time. To complicate things even more, vaccinia virions exist in three different infectious forms carrying a different number of envelopes: mature virions, wrapped virions, and extracellular virions. The mature virion of vaccinia is

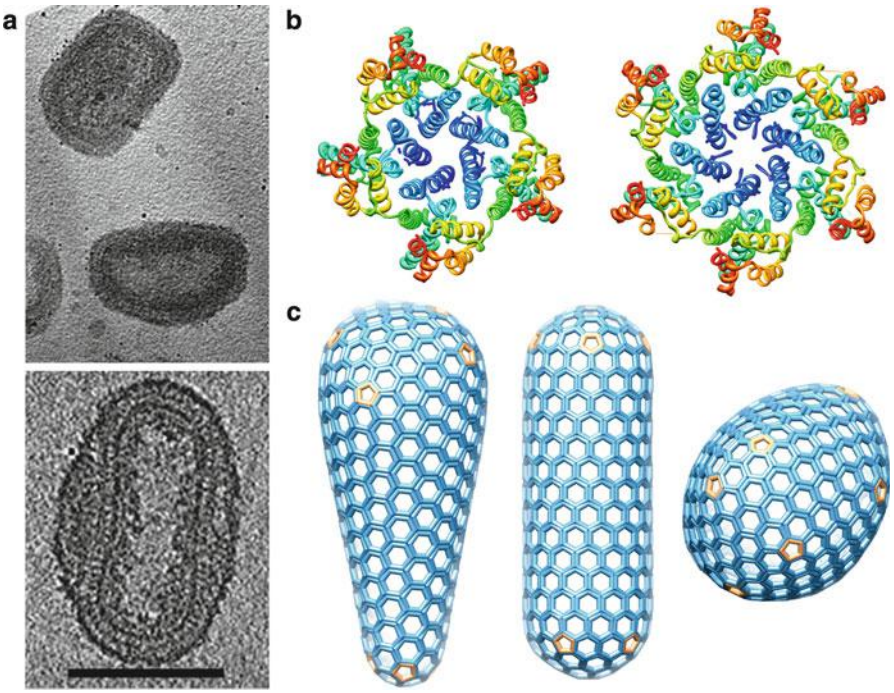

Fig. 11.6 Asymmetric and pleomorphic viruses. (**a**) Cryo-electron tomography reconstruction of mature vaccinia virions. *Top*: central section showing two virions in different orientations. *Bottom*: Central section of a virion where the dumbbell shape of the core can be appreciated. The bar represents 2,000 Å (Reproduced from [49]. With permission). (**b**) Structure of HIV CA protein assembled as a pentamer (*left*; PDB ID 3P05) or a hexamer (*right*; PDB ID 3H4E). (**c**) Example of fullerene-like objects generated from 12 pentamers and a variable number of hexamers

a brick-shaped enveloped particle with approximate dimensions 350 × 250 × 150 nm (varying depending on the imaging technique used), composed by at least 75 different proteins [17]. Mature virions are intracellular forms of the viral particle. Wrapped virions are also found in the cells, and consist of mature virions surrounded by two additional membranes derived from the Golgi cisternae. Wrapped virions leave the cell by fusing with the cell membrane, leaving one of their envelopes behind, to produce the extracellular virion.

The 200 kbp genome of vaccinia is contained in a core with an elongated dumbbell shape, surrounded by a protein capsule (core wall). The core also contains a variety of viral enzymes involved in RNA metabolism, required for the virus to replicate in the cytosol. The outer part of the core wall has striated appearance (palisade layer) while the inner part is smooth. It is not known if these different appearances are due to the existence of two chemically different layers or if there is only one asymmetrically organized layer. The extremes of the dumbbell rest adjacent to the envelope, while the central part is surrounded by electron-dense material (lateral bodies) of unknown function. *In vitro* disruption studies suggest

that the dsDNA in the core is in complex with condensing proteins [50]. However, the condensing proteins have not been identified yet.

11.4.2 Pleomorphic Viruses

Pleomorphic viruses not only do not follow high symmetry rules when forming the infectious particle, but may even adopt a wide range of sizes, shapes and composition from particle to particle, making each virion unique. Because of their intrinsic variability, the structural organization of pleomorphic viruses cannot be deduced from structural biology techniques based on averaging data from many identical virions, such as X-ray crystallography or cryo-EM analyses. The advent of electron tomography to visualize single virus particles (see Chap. 3) has started to reveal the architectural details of this kind of macromolecular machines, which includes many important pathogens for humans. Examples of pleomorphic viruses include retroviruses (HIV); orthomyxoviruses (influenza); coronaviruses (SARS-coronavirus); and paramyxoviruses (measles) [2, 51–53]. In addition, atomic force microscopy can be used for surface visualization of any kind of single virus particles (see Chap. 8) and holds great potential for imaging pleomorphic viruses in liquid, in close to physiological conditions. Possible deformations by adhesion to a solid base should be, however, considered in this case.

Pleomorphism is most pronounced among enveloped viruses, since the lipid envelope readily adapts different shapes and sizes. But also proteins with a tendency to form symmetric aggregations can give rise to pleomorphic capsids. The capsid protein of retroviruses (CA) can assemble into either hexamers or pentamers, in much the same way as capsid proteins of icosahedral viruses (Fig. 11.6b). Recombinant CA forms only hexamers in certain conditions, giving rise to tubular oligomers or flat, ordered sheets; while when pentameric oligomerization is enforced, $T = 1$ icosahedral particles are formed [54, 55]. However, when CA hexamers and pentamers associate to form the closed mature capsid that contains the nucleocapsid complex including the ssRNA genome, they do it in such manner that the pentamers are not distributed regularly within the hexamer lattice. Even if a fixed number of 12 pentamers is incorporated into each capsid, the asymmetry of their distribution results in asymmetrical structures that can adopt shapes ranging from roughly spherical to roughly conical, and can be modeled using the geometrical principles governing fullerene cones (Fig. 11.6c).

11.4.3 A Glimpse of the Weird Shapes of Archaeal Viruses

In the last years numerous new microorganisms living in extreme environments have been described, and with them their corresponding infecting viruses [56]. The most abundant repertoire of archaeal viruses reported so far is that of dsDNA

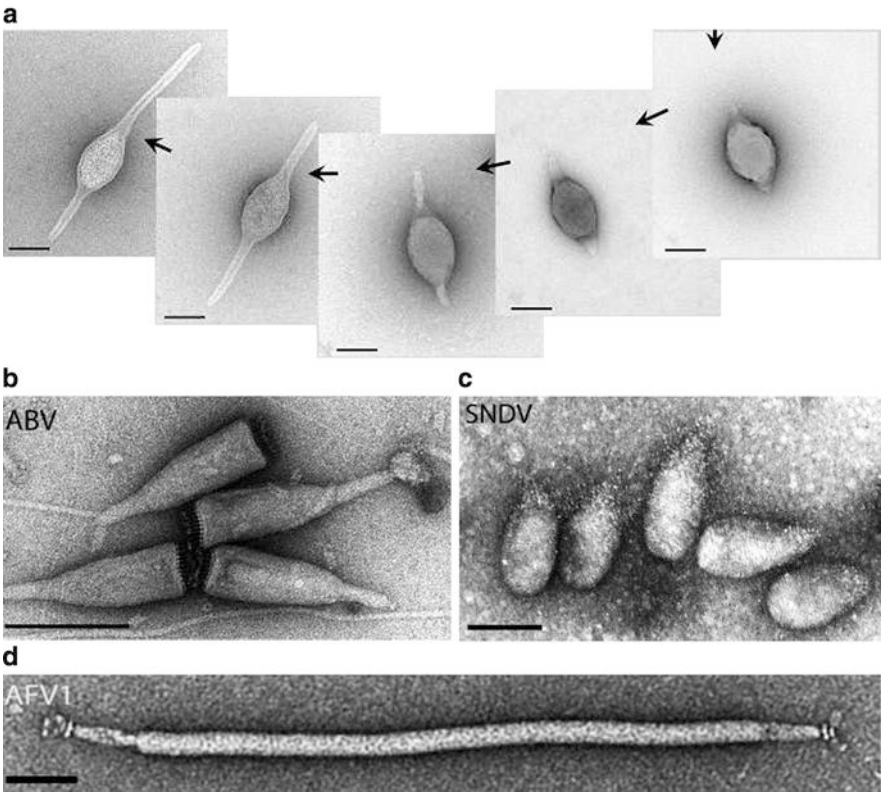

Fig. 11.7 Examples of archeoviruses with unusual shapes. (**a**) From *right* to *left*: extracellular elongation of the two tails in the fusiform ATV virion. (**b**) Bottle-shaped and (**c**) droplet-shaped viruses. (**d**) A Lipothrixvirus representative (AFV-1) with claw-shaped structures at both ends (Panel (**a**) reproduced from [60]; (**b**), (**c**) and (**d**) reproduced from [56]. With permission). Scale bars represent 100 nm

viruses. Only a few ssDNA archaeal viruses have been described, enclosing their genome in pleomorphic enveloped particles [57]. It is still not clear if RNA archaeal viruses exist [58]. Some dsDNA viruses infecting archaea follow the general architectural types previously known for bacterial and eukaryotic viruses. For example, *Sulfolobus* turreted icosahedral virus (STIV), *Haloharcula hispanica* SH1, or *Salisaeta* icosahedral phage 1 (SSIP-1) [59] are icosahedral, tailless viruses with an internal membrane, structurally similar to bacteriophage PRD1 [11] or the algae virus PBCV-1 [8]. Some archaeal viruses with tailed bacteriophage morphologies have also been reported [56].

Intriguingly, other viral families discovered in archaea are unique to this branch of life, with morphologies never observed before (Fig. 11.7). According to their overall organization, they are classified into fusiform, droplet or bottle-shaped, and linear viruses. Within each class, viruses with different genome types, sizes, and

bearing no sequence similarity can be found. Fusiform viruses are very abundant in habitats dominated by archaeal microorganisms. They have spindle-shaped virions with tails of variable length protruding from the spindle poles. One fusiform virus, *Acidianus* two-tailed virus (ATV) not only has an unusual shape, but is able to assemble new structural features after leaving the host cell. When propagated at temperatures slightly suboptimal for its host (75 °C), isolated virions look like ~0.2 μm long lemons. However, when temperature is raised in the absence of the host cell, these viruses grow two filamentous tails of variable lengths, one from each pole. This is the only known example of a virus with extracellular assembly, but it is likely that others exist [60]. The tails end in an anchor-like structure, thought to be involved in attachment to the host.

Acidianus bottle-shaped virus (ABV) and *Sulfolobus neozelandicus* droplet-shaped virus (SNDV) are the only known members of the two viral families termed *Ampullaviridae* and *Guttaviridae* [56]. The enveloped ABV virion contains a conical core formed by a supercoiled nucleoprotein filament. A brush of short filaments protrudes from the bottom of the bottle, but host attachment seems to occur at the opposite side of the virion. Little is known about the architecture of the SNDV virion, except for its droplet shape and the presence of a tuft of long fibers at its narrower pole. Finally, linear archaeal viruses can form stiff rods (*Rudiviridae*) or flexible filaments (*Lipothrixviridae*). Rudiviruses are relatively simple in composition, with no envelope and only a few proteins arranged in particles of variable length, usually related to that of the genome. Lipothrixviruses are enveloped, and the ends of their filamentous capsids are capped with structures of varied shapes (spider legs, pincers, bottle brushes), probably involved in attachment to the host.

11.5 Sophisticated Regulation of Assembly and Maturation

In simple viruses, assembly can occur either in a single step where the newly replicated nucleic acid associates with capsid protein subunits during co-assembly, or in a two-step process where an empty capsid is assembled first and the viral nucleic acid is packaged afterwards (see Chaps. 10, 12). For complex viruses, putting together the many different pieces in their proper places at the appropriate time requires elaborated regulation of the morphogenesis process. In the following sections we discuss some of the strategies used by viruses to achieve assembly of complex capsids. The additional steps required by enveloped viruses to coordinate assembly of proteic elements with recruitment of membranes from the cell will be described in Chap. 14.

11.5.1 Separate Assembly Lines

In viruses with complex chemical composition, the different morphological components are often built separately, forming subassemblies that will be later

put together along carefully regulated pathways. Some of these subassembly reactions may also require chaperones, either of cellular or viral origin, as is also the case for some simpler viruses (Chap. 10). For example, in adenovirus, capsid protein oligomers are formed in the cytosol before being transported to the nucleus, where viral assembly takes place. This preassembly step includes hexon trimerization, which requires a viral chaperone (L4 100K) [61]; and piecing together the vertex complex, composed by a pentamer of penton base bound to a trimer of the fiber protein [4]. In adenovirus, however, the precise temporal order of incorporation of major and minor capsid proteins is not yet understood. Similarly, in herpesvirus hexamers and pentamers of the major capsid protein VP25 are formed previous to particle assembly.

The best described examples of subassembly formation and integration into a virion come from the order *Caudovirales*, more specifically from the *Myoviridae* prototype bacteriophage T4 [42]. In these long tailed bacteriophages, the head, fibers, and tail form separately (Fig. 11.8). Bacteriophage T4 head assembly starts from an initiation complex containing the portal protein gp20 bound to a cellular membrane. This complex recruits the components of a scaffolding core composed by eight different types of viral proteins, among them the main scaffolding protein gp22 and the viral protease (gp21). This core will later be coated by the hexameric and pentameric capsomers (gp23 and gp24) to form the procapsid, or prohead. Once coating is complete, gp21 is activated and cleaves all components of the procapsid except the portal protein. The prohead is detached from the membrane and most of the cleaved peptides exit the particle, opening the way for entry of the viral genome. Genome packaging leads to the large structural rearrangements involved in capsid maturation (Sect. 11.5.3 and Chap. 13), to produce the head in its final form. Finally, the head completion (or neck) proteins gp13, gp14, gp2 and gp4 attach to the portal to form the interface between the head and the tail.

The tail is in turn formed from several preassembled pieces. For contractile tails, the baseplate is assembled first, and used as a seeding point for assembly of the inner tube and contractile sheath. In bacteriophage T4, tail assembly involves 19 different proteins and seven viral chaperones. To form the T4 baseplate, proteins gp6, gp7, gp8, gp10, gp11, gp25 and gp53 assemble in the form of hetero-oligomeric wedges. Six wedges bind around a central hub containing gp5 and gp27. Proteins gp9 and gp12 (the short tail fiber) are then inserted at the gaps between wedges, and the interface between wedges and hub is sealed by proteins gp48 and gp54. This seal is the starting point from which the gp19 inner tail tube will grow. The length of the tube is controlled by a tape measure protein gp29, which extends from the hub to the tube end where the tail capping protein gp3 will bind. The tail sheath gp18 assembles around the inner tube, and finally the tail terminator protein gp15 binds to gp3 and the last row of gp18 subunits, making the tail ready to bind to the neck proteins in the head.

The final stage of tail assembly is incorporation of the long tail fibers to the base plate. The fibers also assemble independently, even starting from separate

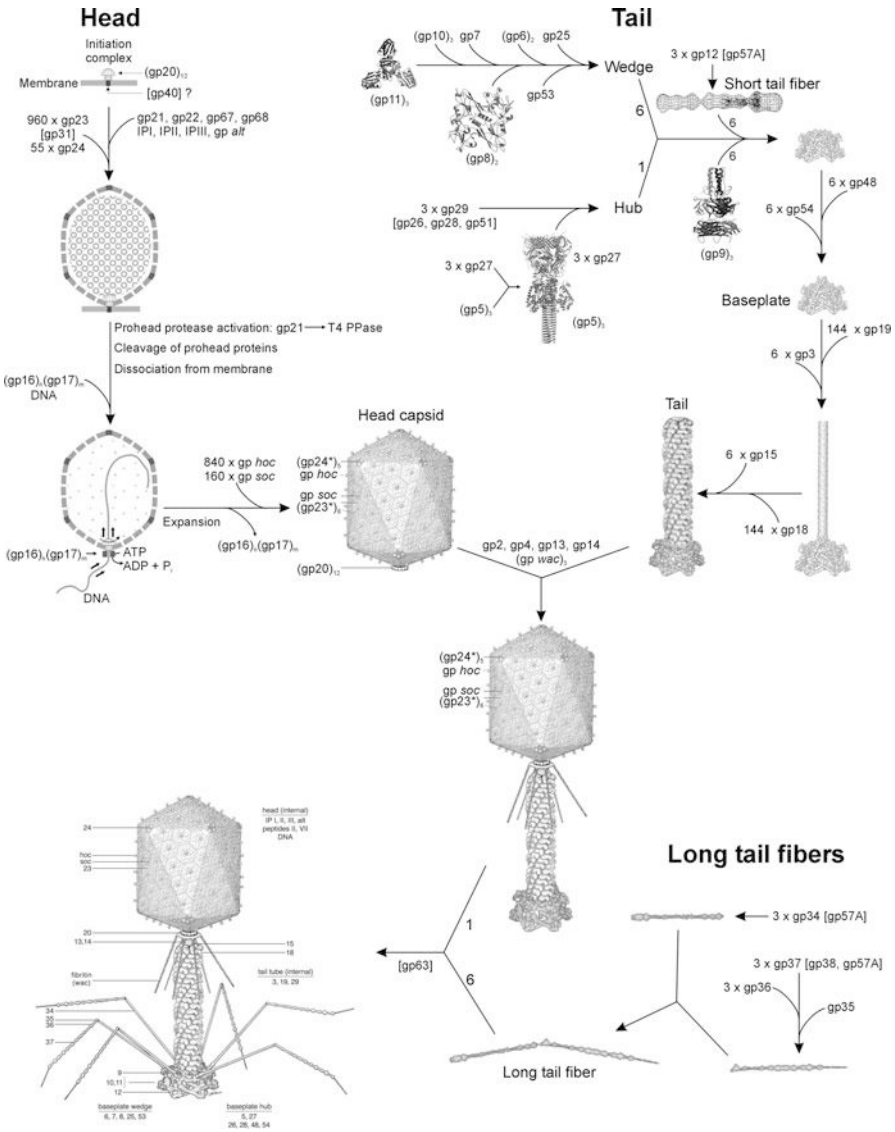

Fig. 11.8 Separate assembly lines. Schematic depiction of the complex bacteriophage T4 morphogenesis pathway (Reproduced from [42]. With permission)

subassemblies. The long tail fiber of bacteriophage T4 is kinked; the proximal part of the fiber before the kink is formed by a single protein (gp34), while the distal part contains three different proteins: gp35, gp36, and gp37. The proximal and distal parts assemble separately, and then join before attaching to the baseplate. All long

tail fiber proteins form trimers, except monomeric gp35 (notice the symmetry mismatch) that sits at the interface between the distal and proximal half fibers. Interestingly, the long fibers cannot join the tail until it is bound to the DNA-filled head. A similar assembly pathway has been described for non-contractile long tails. In podoviruses however, the short tail is not assembled as a separate entity, but it grows outward from the portal vertex on the virion capsid. The careful temporal regulation of these assembly processes is proved by the fact that, when any of the structural proteins is absent, viral morphogenesis is interrupted and the assembly intermediates previous to the disrupted step accumulate in the cell.

11.5.2 Scaffolding

Scaffolding elements are crucial for accurate assembly of large viral capsids. They are present in assembly intermediates (*e.g.* procapsids), but absent in the final, infective product. Their role is to facilitate interactions between capsid elements at the early stages of assembly, by promoting nucleation – that is, putting together the viral proteins that may be highly diluted within the crowded cellular context. Scaffolds are also thought to stabilize weak interactions at the initial stages of assembly [62], while simultaneously allowing flexibility for mistake corrections. This last function is most important in large capsids, where the number of interactions to be checked for errors is correspondingly large. Finally, scaffolds have a role in determining the size and shape of viral capsids.

The most studied scaffold proteins are those present in tailed bacteriophages. For example, bacteriophage P22 (*Podoviridae*) scaffold is a 33 kDa protein. In the early stages of P22 morphogenesis, a procapsid is formed by 415 copies of the capsid protein, with approximately 300 molecules of scaffold inside. Unlike the capsid protein, the scaffold does not follow icosahedral symmetry; therefore, little is known about its organization in the assembly intermediate [45]. Scaffold proteins have been quite refractory to structural studies. Nuclear magnetic resonance and crystallographic studies on the scaffold of P22 and Φ29 indicate that they have a helical fold. Biophysical analyses indicate that many of them share an extended, rod-like shape and a tendency to dimerize in solution [63]. However, an equilibrium between different oligomeric forms seems to be required to achieve correct capsid assembly. Kinetic studies have revealed that in phage P22, scaffold is predominantly a dimer during assembly, but the presence of free monomers is absolutely required to complete the head. Kinetically trapped intermediates are observed when monomers are depleted by decreasing the ionic strength, while restoring it eliminates the trap and allows elongation to proceed. Phage scaffolds are usually ejected from the procapsid immediately before genome packaging. In P22 and Φ29, the intact protein exits the shell, and can be recycled in a new round of assembly. In other cases, the scaffold is removed *via* cleavage by a viral-encoded protease.

In spite of their apparently simple organization, some small bacteriophages, such as the Microvirus representative ΦX174 (T = 1), encode both internal and external scaffold proteins [63]. The ΦX174 internal scaffold protein (protein B) helps in the early stages of assembly by preventing aggregation of the capsid protein F into aberrant oligomers, and ensuring the recruitment of the vertex spike protein G. The C-terminal region of protein B (24 aminoacids) interacts with the capsid and can be observed in the crystal structure of the procapsid, while the rest is disordered and appears to be largely tolerant to mutations. On the contrary, the external scaffold protein D is highly ordered and sensitive to mutations. Protein D is absolutely required for elongation (to assemble capsid pentamers into a spherical particle), while protein B helps to make assembly efficient but is not strictly required: in the absence of B, viral particles can be formed, but the process requires overexpression of protein D and takes as much as ten times longer than in the presence of both scaffolds. It is believed that scaffold redundancy confers an evolutionary advantage to ΦX174 by facilitating extremely rapid replication cycles.

In the absence of scaffold, many phage capsid proteins have been observed to self-assemble into aberrant oligomers (tubes, elongated shells, T = 1 icosahedrons). It follows from these observations that scaffolding proteins are involved in determining the correct curvature in the interactions between capsomers, so that they can form a closed shell of the appropriate size to hold the viral genome. A remarkable proof of this size determination role comes from the P2/P4 phage system [64]. Phage P2 is a member of the *Myoviridae* family, with an assembly pathway similar to P22 that includes a T = 7 procapsid formed with the help of an internal scaffolding protein, gpO. Remarkably, P2 gpO combines the function of scaffold and protease, able to cleave itself and other shell components at later maturation stages. In the presence of P4, however, the whole assembly pathway is altered. P4 is a replicon that can exist either as a prophage or a free plasmid, and does not code for any major structural proteins. However, P4 is able to sequester players of the P2 assembly line to form small T = 4 capsids where only its smaller genome can fit, excluding that of P2 (Fig. 11.9). This parasitic process is achieved by synthesis of a P4-encoded scaffold protein, Sid (SIze Determination protein). Sid forms an external scaffold on P2 assembly intermediates, forcing a narrower curvature and therefore a smaller icosahedral net. Unlike the internal scaffolds, P4 Sid forms a dodecahedral ordered cage that can be observed in cryo-electron microscopy reconstructions.

Although one can generally speak about scaffolding proteins, and indeed many viruses have such proteins, scaffolding functions can also be performed by flexible regions of the capsid proteins, which establish interactions during assembly that are later removed *via* conformational changes or cleavage by viral proteases. For example, bacteriophage HK97 (*Siphoviridae*) does not encode a scaffold protein. Instead, a 103 residue stretch at the N-terminus of the capsid protein, known as the delta-domain, performs the scaffold function [65]. The delta domain is located towards the interior of the capsid and mediates interactions between capsomers during assembly. Once the procapsid is completed, and before the DNA is packaged, the delta domain is cleaved out by the viral protease, allowing the transition to

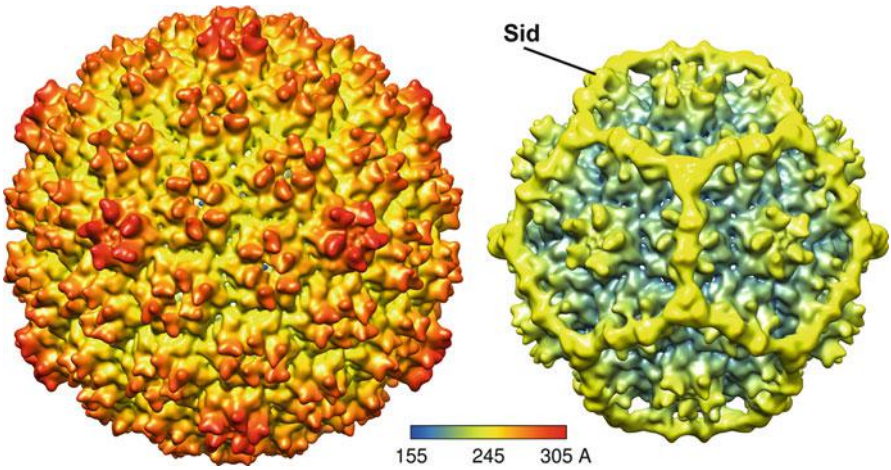

Fig. 11.9 Capsid size determination by scaffolding proteins: the P2/P4 phage system. *Left*: bacteriophage P2 T = 7 procapsid (EMDB ID EMD_5406). *Right*: the parasitic P4 T = 4 procapsid (EMDB ID EMD_5405). Note the Sid external scaffold restricting the shell size. The view is along a twofold icosahedral axis. Original maps have been filtered to a lower resolution for clarity. Color key indicates color changes with map radius

the next assembly intermediate [62]. Adenovirus may use a dual system, including both a separate scaffolding protein [66] and flexible regions of minor capsid proteins removed by the viral protease during maturation [67].

11.5.3 Maturation

In complex viruses, piecing together a number of proteins into a capsule to host the genome is far from producing the final, infectious form. Instead, the newly formed particles (procapsids) need to undergo a series of morphological and/or stability changes to acquire their full infectious potential. This process is known as maturation. There is a double goal for maturation on the viral cycle: first, to produce virions stable enough to protect the genome from aggressive conditions in the extracellular milieu; and second, to prepare the viral particle for correct delivery of the genome into the new host cell.

In dsDNA bacteriophages, maturation encompasses large structural changes and protein rearrangements in the capsid, concomitant with scaffold removal and genome packaging. The capsid changes from a weak, labile object to a highly stable shell, ready to withstand the high internal pressure imposed by the tightly packed DNA inside [62] (see Chaps. 9, 12, 18, and 19). In other viruses, such as polio [68] or adenovirus [67], maturation does not end with an extremely stable shell, but with a metastable one. This difference with respect to the bacteriophage case is likely related to the mode of infection of these eukaryotic viruses. Instead of

ejecting the genome across the plasma membrane leaving the capsid behind, polio and adenovirus are internalized in the cell, and must be disassembled within in a concerted fashion to ensure exposure of the genome at the appropriate place and time for successful replication. Maturation prepares these viruses to start the programmed uncoating sequence upon reception of the appropriate signal, for example attachment to the receptor, or pH changes along the endocytosis pathway. Interestingly, in adenovirus maturation is related to genome packaging in quite a unique way. The adenoviral protease, which is the main maturation agent, is packaged together with the viral genome thanks to its dsDNA binding ability, and uses the dsDNA itself as a cofactor to increase its catalytic activity several orders of magnitude [4].

Maturation processes are not restricted to icosahedral capsids: pleomorphic enveloped viruses such as retroviruses also undergo extensive structural rearrangements to become fully infectious [2]. A more extensive discussion on maturation for several different viruses can be found in Chap. 13.

11.6 Perspectives and Conclusions

In the past decade, structural studies on complex viruses have greatly benefited from technical improvements in structural biology techniques such as those described in Chaps. 3, 4, 5, 6, 7, 8, and 9. As more details are known, it is becoming clear that many complex viruses that infect hosts far apart in evolution share common structural solutions.

For example, adenovirus, which infects vertebrates, has a striking structural similarity to PRD1, a bacteriophage with an internal membrane. The parallels between adenovirus and PRD1 extend from their DNA replication mechanisms, to their capsid architecture and the folding of their major capsid proteins [3, 11]. In the last years, more members of the PRD1-adenovirus family have been described or predicted, and the lineage now extends from viruses infecting bacteria or archaea, to the large nucleo-cytoplasmic DNA viruses such as Asfarvirus, Iridovirus and the giant Mimivirus [9]. All these viruses are built from the same kind of double 8-stranded β-barrel, pseudo-hexagonal capsomers arranged in different tiling systems, with triangulation numbers ranging between $T = 21$ and $T = 169$, and reaching up to $972 < T < 1{,}200$ for the giant Mimivirus [21]. Intriguingly, even a scaffold protein of the non-icosahedral vaccinia virus folds as a double barrel pseudo-hexamer, indicating a possible common ancestor with icosahedral dsDNA viruses [69].

Adenovirus and PRD1 are not the only cases indicating an evolutionary relationship between animal and bacterial viruses. Herpesviruses, which infect all sorts of animal organisms, share many structural characteristics with tailed bacteriophage [14]. They follow a similar assembly pathway, starting from an empty procapsid formed with the help of scaffold, and maturing to a more angular shell *via* large structural rearrangements upon DNA packaging. Like tailed bacteriophage, one of

the vertices is different from the rest and contains a 12-fold symmetric portal structure involved in genome encapsidation. Herpesvirus capsid protein folds with a topology very similar to that of the HK97 phage family; and like tailed phages, the packed genome forms concentric shells when observed in icosahedrally averaged EM reconstructions. Finally, structural parallels also exist between Cystoviruses (dsRNA bacteriophage, representative Φ6) and Reoviruses [70].

The fact that many complex viruses with different hosts share a common structural solution has evolutionary implications. First, since the design has been conserved throughout time, even after all traces of sequence similarity have disappeared, it must be a highly efficient arrangement. Second, it suggests that the architecture was established in the early stages of evolution, before the branches of the evolutionary tree diverged into the three kingdoms known today (archaea, bacteria and eukarya). One could wonder, then, if all complex viruses existing today fall within a limited number of structural solutions selected by their success early in evolution [71]. However, discovery of the unique architectures of archaeal viruses indicates that other structural solutions exist. Advances in high throughput virus isolation and structural characterization techniques will contribute to clarify this question.

In summary, complex viruses incorporate a wide range of molecules into their capsids, including specialized host interaction, genome packaging and cementing proteins; and in some cases membranes, either internal or external. Accommodation of the different components often requires deviations from high order symmetry, from mismatches to pleomorphism; and involves complex regulation of the assembly dynamics. Key elements in this regulation are: separate assembly lines, scaffold elements, and maturation processes. Host and virus evolution probably act hand in hand to optimize viral particle structure and morphogenesis.

Acknowledgements Work in the San Martín lab is funded by grants BFU2010-16382 and FIS2010-10552-E/FIS2011-16090-E from the Ministerio de Ciencia e Innovación of Spain. José R. Castón (CNB-CSIC) is acknowledged for careful reading of the manuscript.

References and Further Reading

1. Stubbs G, Kendall A (2012) Helical viruses. Adv Exp Med Biol 726:631–658
2. Goff SP (2007) Retroviridae: the retroviruses and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA (eds) Fields virology, vol II. Lippincott Williams & Wilkins, Philadelphia, pp 1999–2069
3. Liu H, Jin L, Koh SB, Atanasov I, Schein S, Wu L, Zhou ZH (2010) Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. Science 329:1038–1043
4. San Martín C (2012) Latest insights on adenovirus structure and assembly. Viruses 4:847–877
5. Zubieta C, Schoehn G, Chroboczek J, Cusack S (2005) The structure of the human adenovirus 2 penton. Mol Cell 17:121–135
6. Rux JJ, Kuser PR, Burnett RM (2003) Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution x-ray crystallographic, molecular modeling, and sequence-based methods. J Virol 77:9553–9566

7. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera–a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612
8. Cherrier MV, Kostyuchenko VA, Xiao C, Bowman VD, Battisti AJ, Yan X, Chipman PR, Baker TS, Van Etten JL, Rossmann MG (2009) An icosahedral algal virus has a complex unique vertex decorated by a spike. Proc Natl Acad Sci U S A 106:11085–11089
9. Krupovic M, Bamford DH (2008) Virus evolution: how far does the double beta-barrel viral lineage extend? Nat Rev Microbiol 6:941–948
10. Colby WW, Shenk T (1981) Adenovirus type 5 virions can be assembled *in vivo* in the absence of detectable polypeptide IX. J Virol 39:977–980
11. Abrescia NG, Cockburn JJ, Grimes JM, Sutton GC, Diprose JM, Butcher SJ, Fuller SD, San Martín C, Burnett RM, Stuart DI, Bamford DH, Bamford JK (2004) Insights into assembly from structural analysis of bacteriophage PRD1. Nature 432:68–74
12. Roizman B, Knipe DM, Whitley RJ (2007) Herpes simplex viruses. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA (eds) Fields virology, vol II. Lippincott Williams & Wilkins, Philadelphia, pp 2501–2601
13. Feiss M, Rao VB (2012) The bacteriophage DNA packaging machine. Adv Exp Med Biol 726:489–509
14. Cardone G, Heymann JB, Cheng N, Trus BL, Steven AC (2012) Procapsid assembly, maturation, nuclear exit: dynamic steps in the production of infectious herpesvirions. Adv Exp Med Biol 726:423–439
15. Huiskonen JT, Jaalinoja HT, Briggs JA, Fuller SD, Butcher SJ (2007) Structure of a hexameric RNA packaging motor in a viral polymerase complex. J Struct Biol 158:156–164
16. Trask SD, McDonald SM, Patton JT (2012) Structural insights into the coupling of virion assembly and rotavirus replication. Nat Rev Microbiol 10:165–177
17. Condit RC, Moussatche N, Traktman P (2006) In a nutshell: structure and assembly of the vaccinia virion. Adv Virus Res 66:31–124
18. Germond JE, Hirt B, Oudet P, Gross-Bellark M, Chambon P (1975) Folding of the DNA double helix in chromatin-like structures from simian virus 40. Proc Natl Acad Sci U S A 72:1843–1847
19. Tweeten KA, Bulla LA, Consigli RA (1980) Characterization of an extremely basic protein derived from granulosis virus nucleocapsids. J Virol 33:866–876
20. Giberson AN, Davidson AR, Parks RJ (2012) Chromatin structure of adenovirus DNA throughout infection. Nucleic Acids Res 40:2369–2376
21. Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kostyuchenko VA, Chipman PR, Suzan-Monti M, Raoult D, McPherson A, Rossmann MG (2009) Structural studies of the giant mimivirus. PLoS Biol 7:e92
22. Ruigrok RW, Crepin T, Kolakofsky D (2011) Nucleoproteins and nucleocapsids of negative-strand RNA viruses. Curr Opin Microbiol 14:504–510
23. Palese P, Shaw ML (2007) Orthomyxoviridae: the viruses and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA (eds) Fields virology, vol II. Lippincott Williams & Wilkins, Philadelphia, pp 1647–1689
24. Huiskonen JT, Butcher SJ (2007) Membrane-containing viruses with icosahedrally symmetric capsids. Curr Opin Struct Biol 17:229–236
25. Grunewald K, Desai P, Winkler DC, Heymann JB, Belnap DM, Baumeister W, Steven AC (2003) Three-dimensional structure of herpes simplex virus from cryo-electron tomography. Science 302:1396–1398
26. Forster F, Medalia O, Zauberman N, Baumeister W, Fass D (2005) Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. Proc Natl Acad Sci U S A 102:4729–4734
27. Butcher SJ, Manole V, Karhu NJ (2012) Lipid-containing viruses: bacteriophage PRD1 assembly. Adv Exp Med Biol 726:365–377

28. Grahn AM, Daugelavicius R, Bamford DH (2002) Sequential model of phage PRD1 DNA delivery: active involvement of the viral membrane. Mol Microbiol 46:1199–1209
29. Zhang X, Xiang Y, Dunigan DD, Klose T, Chipman PR, Van Etten JL, Rossmann MG (2011) Three-dimensional structure and function of the paramecium bursaria chlorella virus capsid. Proc Natl Acad Sci U S A 108:14837–14842
30. Tulman ER, Delhon GA, Ku BK, Rock DL (2009) African swine fever virus. Curr Top Microbiol Immunol 328:43–87
31. Caspar DLD, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol 27:1–24
32. Baker M, Prasad BV (2010) Rotavirus cell entry. Curr Top Microbiol Immunol 343:121–148
33. Danthi P, Guglielmi KM, Kirchner E, Mainou B, Stehle T, Dermody TS (2010) From touchdown to transcription: the reovirus cell entry pathway. Curr Top Microbiol Immunol 343:91–119
34. Huiskonen J, Manole V, Butcher S (2007) Tale of two spikes in bacteriophage PRD1. Proc Natl Acad Sci U S A 104:6666–6671
35. Zauberman N, Mutsafi Y, Halevy D, Shimoni E, Klein E, Xiao C, Sun S, Minsky A (2008) Distinct DNA exit and packaging portals in the virus acanthamoeba polyphaga mimivirus. PLoS Biol 13:e114
36. van Raaij MJ, Mitraki A, Lavigne G, Cusack S (1999) A triple β-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. Nature 401:935–938
37. Liu H, Wu L, Zhou ZH (2011) Model of the trimeric fiber and its interactions with the pentameric penton base of human adenovirus by cryo-electron microscopy. J Mol Biol 406:764–774
38. Chappell JD, Prota AE, Dermody TS, Stehle T (2002) Crystal structure of reovirus attachment protein σ1 reveals evolutionary relationship to adenovirus fiber. EMBO J 21:1–11
39. Hess M, Cuzange A, Ruigrok RWH, Chroboczek J, Jacrot B (1995) The avian adenovirus penton: two fibres and one base. J Mol Biol 252:379–385
40. Christensen JB, Byrd SA, Walker AK, Strahler JR, Andrews PC, Imperiale MJ (2008) Presence of the adenovirus IVa2 protein at a single vertex of the mature virion. J Virol 82:9086–9093
41. Gowen B, Bamford JK, Bamford DH, Fuller SD (2003) The tailless icosahedral membrane virus PRD1 localizes the proteins involved in genome packaging and injection at a unique vertex. J Virol 77:7863–7871
42. Leiman PG, Kanamaru S, Mesyanzhinov VV, Arisaka F, Rossmann MG (2003) Structure and morphogenesis of bacteriophage T4. Cell Mol Life Sci 60:2356–2370
43. Kostyuchenko VA, Chipman PR, Leiman PG, Arisaka F, Mesyanzhinov VV, Rossmann MG (2005) The tail structure of bacteriophage T4 and its mechanism of contraction. Nat Struct Mol Biol 12:810–813
44. Fokine A, Chipman PR, Leiman PG, Mesyanzhinov VV, Rao VB, Rossmann MG (2004) Molecular architecture of the prolate head of bacteriophage T4. Proc Natl Acad Sci U S A 101:6003–6008
45. Agirrezabala X, Martin-Benito J, Caston JR, Miranda R, Valpuesta JM, Carrascosa JL (2005) Maturation of phage T7 involves structural modification of both shell and inner core components. EMBO J 24:3820–3829
46. Leiman PG, Shneider MM (2012) Contractile tail machines of bacteriophages. Adv Exp Med Biol 726:93–114
47. Davidson AR, Cardarelli L, Pell LG, Radford DR, Maxwell KL (2012) Long noncontractile tail machines of bacteriophages. Adv Exp Med Biol 726:115–142
48. Casjens SR, Molineux IJ (2012) Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. Adv Exp Med Biol 726:143–179
49. Cyrklaff M, Risco C, Fernández JJ, Jiménez MV, Estéban M, Baumeister W, Carrascosa JL (2005) Cryo-electron tomography of vaccinia virus. Proc Natl Acad Sci U S A 102:2772–2777

50. Kuznetsov Y, Gershon PD, McPherson A (2008) Atomic force microscopy investigation of vaccinia virus structure. J Virol 82:7551–7566
51. Harris A, Cardone G, Winkler DC, Heymann JB, Brecher M, White JM, Steven AC (2006) Influenza virus pleiomorphy characterized by cryoelectron tomography. Proc Natl Acad Sci U S A 103:19123–19127
52. Barcena M, Oostergetel GT, Bartelink W, Faas FG, Verkleij A, Rottier PJ, Koster AJ, Bosch BJ (2009) Cryo-electron tomography of mouse hepatitis virus: insights into the structure of the coronavirion. Proc Natl Acad Sci U S A 106:582–587
53. Liljeroos L, Huiskonen JT, Ora A, Susi P, Butcher SJ (2011) Electron cryotomography of measles virus reveals how matrix protein coats the ribonucleocapsid within intact virions. Proc Natl Acad Sci U S A 108:18085–18090
54. Pornillos O, Ganser-Pornillos BK, Yeager M (2011) Atomic-level modelling of the HIV capsid. Nature 469:424–427
55. Cardone G, Purdy JG, Cheng N, Craven RC, Steven AC (2009) Visualization of a missing link in retrovirus capsid assembly. Nature 457:694–698
56. Pina M, Bize A, Forterre P, Prangishvili D (2011) The archeoviruses. FEMS Microbiol Rev 35:1035–1054
57. Pietila MK, Atanasova NS, Manole V, Liljeroos L, Butcher SJ, Oksanen HM, Bamford DH (2012) Virion architecture unifies globally distributed pleolipoviruses infecting halophilic archaea. J Virol 86:5067–5079
58. Bolduc B, Shaughnessy DP, Wolf YI, Koonin EV, Roberto FF, Young M (2012) Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated yellowstone hot springs. J Virol 86:5562–5573
59. Aalto AP, Bitto D, Ravantti JJ, Bamford DH, Huiskonen JT, Oksanen HM (2012) Snapshot of virus evolution in hypersaline environments from the characterization of a membrane-containing salisaeta icosahedral phage 1. Proc Natl Acad Sci U S A 109:7079–7084
60. Haring M, Vestergaard G, Rachel R, Chen L, Garrett RA, Prangishvili D (2005) Virology: independent virus development outside a host. Nature 436:1101–1102
61. Hong SS, Szolajska E, Schoehn G, Franqueville L, Myhre S, Lindholm L, Ruigrok RW, Boulanger P, Chroboczek J (2005) The 100K-chaperone protein from adenovirus serotype 2 (subgroup C) assists in trimerization and nuclear localization of hexons from subgroups C and B adenoviruses. J Mol Biol 352:125–138
62. Johnson JE (2010) Virus particle maturation: insights into elegantly programmed nanomachines. Curr Opin Struct Biol 20:210–216
63. Prevelige PE, Fane BA (2012) Building the machines: scaffolding protein functions during bacteriophage morphogenesis. Adv Exp Med Biol 726:325–350
64. Marvik OJ, Sharma P, Dokland T, Lindqvist BH (1994) Bacteriophage P2 and P4 assembly: alternative scaffolding proteins regulate capsid size. Virology 200:702–714
65. Huang RK, Khayat R, Lee KK, Gertsman I, Duda RL, Hendrix RW, Johnson JE (2011) The prohead-I structure of bacteriophage HK97: implications for scaffold-mediated control of particle assembly and maturation. J Mol Biol 408:541–554
66. Hasson TB, Ornelles DA, Shenk T (1992) Adenovirus L1 52- and 55-kDa proteins are present within assembling virions and colocalize with nuclear structures distinct from replication centers. J Virol 66:6133–6142
67. Pérez-Berná AJ, Ortega-Esteban A, Menéndez-Conejero R, Winkler DC, Menéndez M, Steven AC, Flint SJ, de Pablo PJ, San Martín C (2012) The role of capsid maturation on adenovirus priming for sequential uncoating. J Biol Chem 287:31582–31595
68. Hogle JM (2002) Poliovirus cell entry: common structural themes in viral cell entry pathways. Annu Rev Microbiol 56:677–702
69. Bahar MW, Graham SC, Stuart DI, Grimes JM (2011) Insights into the evolution of a complex virus from the crystal structure of vaccinia virus D13. Structure 19:1011–1020
70. Bamford DH, Burnett RM, Stuart DI (2002) Evolution of viral structure. Theor Popul Biol 61:461–470
71. Abrescia NG, Bamford DH, Grimes JM, Stuart DI (2012) Structure unifies the viral universe. Annu Rev Biochem 81:795–822

Further Reading

Agbandje-McKenna M, McKenna R (2011) Structural virology. RSC Publishing, Cambridge
Flint SJ, Enquist LW, Racaniello VR, Skalka AM (2009) Principles of virology. ASM Press, Washington, DC
Rixon FJ, Chiu W (2003) Studying large viruses. Adv Protein Chem 64:379–408
Rossmann MG, Rao VB (2012) Viral molecular machines. Adv Exp Med Biol, vol 726, Springer, New York

Also especially recommended for further reading are references [4, 9, 24, 32, 42, 54, 56, 62] *listed above*.

Chapter 12
Nucleic Acid Packaging in Viruses

Ana Cuervo, María I. Daudén, and José L. Carrascosa

Abstract Viruses protect their genetic information by enclosing the viral nucleic acid inside a protein shell (capsid), in a process known as genome packaging. Viruses follow essentially two main strategies to package their genome: Either they co-assemble their genetic material together with the capsid protein, or they assemble first an empty shell (procapsid) and then pump the genome inside the capsid with a molecular motor that uses the energy released by ATP hydrolysis. During packaging the viral nucleic acid is condensed to very high concentration by its careful arrangement in concentric layers inside the capsid. In this chapter we will first give an overview of the different strategies used for genome packaging to discuss later some specific virus models where the structures of the main proteins involved, and the biophysics underlying the packaging mechanism, have been well documented.

Keywords Bacteriophage • Capsid • Connector • DNA • Electron microscopy • Encapsidation • Ejection • Helical symmetry • Icosahedral symmetry • Maturation • Molecular motor • Nucleocapsid • Packaging • Portal • RNA • Shell • Terminase • X-ray diffraction • Virus

Abbreviations

ATP	Adenosine triphosphate
bp	Base pair
BPMV	Bean pod mottle virus
BTV	Bluetongue virus

A. Cuervo • M.I. Daudén • J.L. Carrascosa (✉)
Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: jlcarras@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_12,

Cdom	Carboxy domain
ds	Double-stranded
FHV	Flock house virus
HCMV	Human cytomegalovirus
mRNA	messenger RNA
Ndom	Amino domain
NPC	Nucleoprotein complex
nt	Nucleotides
NTP	Nucleotide triphosphate
PaV	Pariacoto virus
pRNA	prohead RNA
ss	Single stranded
STMV	Satellite tobacco mosaic virus
TMV	Tobacco mosaic virus

12.1 Introduction

Viruses are mainly composed by a carrier of genetic information (either DNA or RNA), and a protective and multifunctional container, usually made of proteins and, eventually, lipidic components [1] (see Chap. 2). Several main aspects have to be considered in virus construction. One is that it must consume the minimum possible genetic information, making use of geometric principles to build a large container using a limited set of proteins. In addition, the virus has to be easy to assemble to facilitate the highest possible progeny production but, at the same time, this assembly process has to incorporate mechanisms to select the viral genome and reject cellular components. Then, the virus particle must be able to actively participate in its release from the infected cell, the transfer to other possible hosts, their proper and accurate recognition and, finally, the delivery of their genetic information to start a new infection cycle.

It is clear that to carry out all these processes in so many different environments the virus particle must be a quite flexible vehicle able to trigger and undergo structural changes to provide at the appropriate moment the required functionality: the resistance needed to survive the harsh extracellular environment may be superfluous and even counter-productive in the well-controlled intracellular conditions. The propagation strategy is also reflected in the virus structure. In certain plants, the viral propagation allows each particle to carry partial genomic information, as very many viruses will co-infect the cells (*i.e.*, the genome is found multipartite in different viral particles). On the other hand, viruses infecting most animal cells present the whole genome either on a single molecule or in several segments within the same viral particle to ensure efficient infection. Also, the requirements to

penetrate the eukaryotic cell membrane (see Chaps. 15 and 16) are very different to those required to traverse the thick-walled bacterial envelope (see Chap. 17). Due to these and many other variables, different virus particles have evolved different survival strategies. Thus, viruses provide excellent model systems for the study of biotechnological solutions at the nanoscale.

Viruses offer a unique gradient of structural solutions to encapsidate and protect their genetic material, from the simplest ones to those revealing unexpected levels of sophistication. The co-assembly of the viral RNA and one (or a few) proteins to build a helical particle is one of the simplest arrangements, widely used by single stranded (ss) RNA and ssDNA viruses. While easy to build, these viral assemblies may be too rigid to adapt to different environments, and they have to be fully disassembled to allow viral genome replication and expression.

A more complex structural solution is offered by the enveloped RNA viruses, where a flexible nucleocapsid is formed by the interaction of the RNA and multiple copies of a protein (in a way resembling the simple solution mentioned above); but in this case, the nucleoprotein assembly is based on very different types of RNA-protein contacts, and may adopt different structural conformations to carry out the transcription and replication of the genetic message. Proper protection to transfer this complex from cell to cell is conferred by the inclusion of the nucleocapsid inside a lipidic envelope derived from the cellular membranes after its modification with viral proteins. The extreme in this type of strategy is offered by the more complex viruses (human immunodeficiency virus, adenoviruses, herpesviruses, vaccinia virus), where the nucleocapsid is enclosed by several layers or envelopes of different composition. Examples of these viruses are given in Chap. 11 and several other chapters of this book.

An alternative strategy, used by many viruses, is to incorporate the nucleic acid while the capsid is being formed in a co-assembly process. This solution is adopted by a large number of non-enveloped ssRNA viruses with icosahedral capsids. Icosahedral symmetry is a preferred geometrical solution to build a closed viral container (see Chap. 2). Although the assembly details vary from one virus type to another (see Chap. 10), it seems that the interaction of small segments of the RNA molecule and certain domains of the shell protein facilitates their mutual recognition, and their action as chaperones that assist each other to drive the co-assembly process.

The extreme rigidity and other structural properties of double-stranded (ds) nucleic acids impose severe restrictions to the possible ways to enclose these molecules within a protein container. Nevertheless, the evident evolutionary advantages of dsDNA as a genomic substrate have driven viruses to explore efficient solutions to encapsidate this type of DNA molecule, the best known example being many bacteriophages. These viruses assemble first a protein container (called prohead), in which dsDNA is later packaged using a complex machinery made of different viral proteins. This process requires an exquisite selection procedure to encapsidate the right (viral) DNA, followed by an energy-consuming process that translocates the DNA up to quasi-crystalline densities inside the virus shell. These viruses can be considered as a paradigm of efficient

DNA aggregation: there is no other biological example for such an ordered and condensed DNA conformation, an order that plays a critical role in the release of the DNA during infection by these phages. Nevertheless, this packaging solution is expensive in terms of information (several viral proteins are involved), as well as in energy consumption (one ATP molecule is hydrolyzed by each two base pairs packaged), but the reward is worth the expense: These viruses exhibit the largest infection efficiency by far, as each viral particle is potentially able to infect a host bacteria.

In the following sections we will briefly review examples of viruses using different viral strategies for nucleic acid packaging, making emphasis in common underlying mechanisms and solutions selected by viral evolution.

12.2 Structural Features of the Packaged Nucleic Acids

The extended use of X-ray fiber and crystal diffraction and electron microscopy has provided a great amount of information on the structure of viral particles (see Chaps. 3, 4, and 7). Due to their geometrical shapes and the existence of symmetries (see Chap. 2), these and other structural methods have provided a comprehensive insight into the way the protein components of viral capsids are organized at the molecular or even atomic level. Nevertheless, the same cannot be said for the nucleic acid component of virions, as it is usually far less ordered than the protein counterpart, and in most cases the way the nucleic acid is organized in each viral particle is not identical to those in others. This section reviews the main features of the way the nucleic acids interact with the protein capsids for protection, mobility and functionality.

12.2.1 RNA Viruses

Tobacco mosaic virus (TMV), a representative positive ssRNA virus with helical symmetry has provided a most successful model for biophysical studies for more than 100 years. The structure of this rigid rod-shaped virion was solved at near-atomic resolution using X-ray fiber diffraction and later by electron cryo-microscopy [2, 3]. The virus is organized as a stack of identical subunits of a single capsid protein that follows a helical path with 17 subunits per ring, leaving an internal cavity (Fig. 12.1a). The ssRNA molecule is located near this inner region, with three nucleotides binding to each of the capsid subunits.

The type of protection for the RNA provided by TMV-like viral particles is limited in functional terms due to their rigid nature. A more flexible solution is adopted by negative ssRNA viruses. The Mononegavirales [8] are characterized by the assembly of a helical-symmetry based nucleoprotein complex (NPC) essentially made by the ssRNA and a specific protein (the nucleocapsid protein). Unlike the

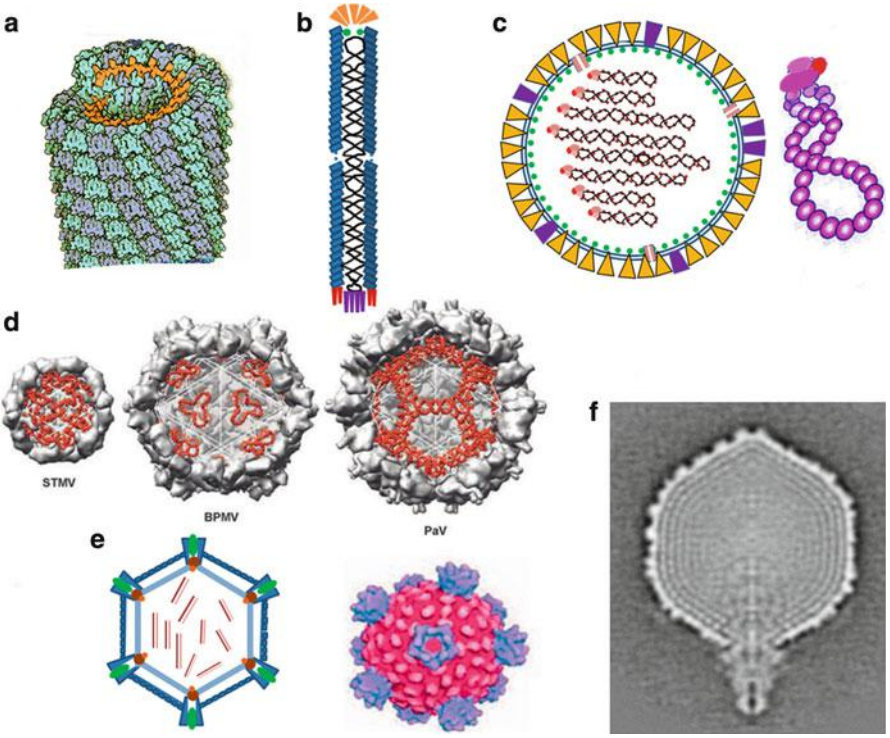

Fig. 12.1 Structural alternatives for nucleic acid encapsidation in viruses. (**a**) Structure of TMV. The viral RNA is colored *orange* (Adapted from the RCSB PDB. With permission). (**b**) Bacteriophage M13. The viral DNA is depicted as a superhelical molecule inside the virus shell. (**c**) Influenza virus. The different structural components are indicated (Adapted from [4]. With permission). The viral ribonucleoprotein complex is shown as a beaded string with the polymerase at one end. (**d**) Structures of STMV, BPMV and PaV particles sectioned to show the partial organization of the viral RNA molecule (*red*) (Reproduced from [5]. With permission). (**e**) Reovirus and the structure of the internal capsid solved by electron cryo-microscopy (Reproduced from [6]. With permission). (**f**) Section of the mature phage T7 structure obtained by electron cryo-microscopy. The DNA is shown as concentric punctuated layers inside the capsid (Reproduced from [7]. With permission)

helical TMV particle, NPCs have a remarkable intrinsic flexibility [9]. Instead of resembling a rigid rod, NPCs have a herring-bone aspect by electron microscopy, revealing different conformations with helical symmetry. These nucleoproteins are enclosed into a near-spherical or elongated lipidic envelope (integrating viral specific proteins) to yield the infectious viral particle (Fig. 12.1c). Examples of these viruses include the *Rhabdoviridae* (like the bullet-shaped vesicular stomatitis virus, or the rabies virus), the *Paramyxoviridae* (respiratory syncytial virus, Sendai virus or measles virus), and the *Orthomyxoviridae* (influenza virus). A general characteristic of all these viruses is that the NPCs (either one or, in the case of viruses with segmented genomes, several NPCs containing a ssRNA molecule each,

as in influenza virus) are released inside the cell and used as substrates for RNA replication and transcription. This means that the RNA-protein assembly must be flexible enough to allow the RNA polymerase (which also forms part of the NPC) to act on the viral RNA substrate without complete disassembly of the nucleoprotein.

Although the viral nucleocapsid proteins of different viruses (N-proteins) show a rather low sequence identity, electron microscopy and X-ray diffraction studies have revealed that NPCs share a basic structural organization that is relatively conserved. The NPC helix has a pitch of around 6–7 nm and 10–13 subunits per turn. Each protein subunit is associated with nine nucleotides, yielding a threading of the proteins into a super-helical complex. The RNA moiety is located in a conserved RNA-binding cavity formed between the carboxyl and amino terminal domains of the protein. This region presents a number of positively charged residues that can interact with the RNA phosphate groups, resulting in the exposure of certain bases of the RNA in the outer face of the NPC, fully available for functional recognition of the RNA polymerase. Three dimensional reconstructions of different NPCs based on electron microscopy studies showed differences in the overall superhelicity and different parameters of the basic helical assembly, and revealed that these complexes have an intrinsic flexibility that accommodates the conformations required in each of the different stages of the viral life cycle, either inside the cell or when packaged inside the viral envelope (reviewed in [10]).

A different solution for the successful storage of the viral genetic information based on ssRNA molecules is found in many icosahedral viruses [5]. Icosahedral symmetry-based capsids are found in about half the virus families. Among them, there is an important group containing positive ssRNA that includes human pathogens such as poliovirus, rhinovirus and hepatitis A, as well as many insect and plant viruses (*e.g.*, flock house virus (FHV) and satellite tobacco mosaic virus (STMV)). In other ssRNA viruses the icosahedral capsid is enclosed by an outer lipidic envelope, as in the *Togaviridae* and *Flaviviridae*. In many of these viruses, the RNA, besides carrying the genetic information, plays an important role in the assembly of the viral particle and in the definition of the shape and size of the capsid. Most of the knowledge we have on the function of the viral RNA in the assembly and life cycle of these viruses comes from biochemical and genetic studies. Knowledge of the RNA structure within viral particles has been largely prevented by the fact that all or a high percentage of the viral RNA is not sufficiently ordered inside the virion. Both X-ray crystallography and three dimensional reconstructions from electron cryo-microscopy images are based on the presence of order and symmetry in identical particles and, thus, only those parts of the RNA sufficiently ordered can be solved with a resolution comparable to that obtained for the capsid. In general, these viruses present ordered RNA density in contact with the capsid inner wall corresponding to around 13 % of the total RNA in the case of FHV, 20 % in bean pod mottle virus (BPMV), or 44 % in STMV. In all cases, the packaged RNA shows a dramatic increase of secondary structure compared to the cytoplasmic free RNA, including the formation of intrachain duplex RNA stems (up to 80 % of the total RNA is found in stem-loop structures within the

STMV capsid), as well as non-covalent base-pairing between RNA segments in those cases of multipartite RNA genomes.

The combination of X-ray diffraction results and electron microscopy reconstructions (see Chap. 7) has provided some insights on the organization of the RNA inside the icosahedral capsids of these viruses. As mentioned above, 30 dsRNA ordered segments are visible in STMV; they are 9 bp long each and follow the twofold axes of the icosahedral shell. By contrast, in BPMV the ordered RNA segments are located at the threefold axes (Fig. 12.1d), and in FHV (with a bipartite RNA genome) ordered double stranded RNA stretches, 10 bp long, are also observed along the capsid twofold axes. Interestingly enough, the location of these RNA moieties close to the interfaces between neighboring capsid subunits allow them to act as part of molecular switches to generate two types of quasi-equivalent intersubunit contacts required for the formation of either planar or more wedge-shaped interfaces along the twofold icosahedral axes (see Chap. 2). In certain cases it has been possible to see the overall topology of most of the RNA within the virion, as in the case of FHV and in the related Pariacoto virus (PaV), where the encapsidated RNA is organized as a dodecahedral cage of double-stranded dsRNA intrachain stems closely following the capsid twofold axes of the icosahedral shell (Fig. 12.1d) [5, 11].

There is evidence that, in some ssRNA viruses, the RNA plays a specific role in the promotion and direction of viral assembly [11, 12]. The fact that the RNA is folded in precise structures near the capsid and interacts with the capsid proteins supports a condensing role of the RNA in the nucleation and first assembly steps. There is also evidence that the capsid proteins could work as chaperones to facilitate the folding of the RNA in a subset of structures compatible with capsid geometry. Most probably, the mutual interactions between capsid proteins and viral RNA interplay to yield the virion definitive structure [5]. In addition, the existence of specific secondary structures in the RNA seems to play a fundamental role as packaging signals for the specific encapsidation of the viral RNA.

Another type of packaging solution is found in dsRNA viruses, which includes mammalian orthoreoviruses and certain fungal viruses and bacteriophages. All of them share many functional and structural characteristics, like the presence of concentric icosahedral capsids [13]. The study of the structure of icosahedral dsRNA viruses is an excellent example of the advantages of the use of hybrid methods: Combining X-ray diffraction of structural components and electron microscopy reconstructions of complete virions allows to obtain quasi-atomic maps of the different virus assemblies (see Chap. 7). The best studied representatives of this type of viruses are the *Reoviridae* family [14] and the *Cystoviridae* bacteriophage Φ6, which present a multipartite genome (three segments in Φ6 and 10–13 segments in reoviruses), together with several concentric icosahedral capsids (normally 2), and an outer lipidic envelope in the case of the *Cystoviridae*.

A common trend in all these viruses, except for the *Birnaviridae*, is the presence of an inner icosahedral capsid built by 120 protein subunits [15]; this capsid might be considered as a functional counterpart of the helical NPCs described above, but

in this case the capsid plays an active role in the viral RNA metabolism, namely in RNA replication, transcription and release from the capsid. The inner capsid in these viruses presents specific turret-like structures at the fivefold vertices (Fig. 12.1e). These assemblies, which contain the viral RNA polymerase, are involved in the generation and extrusion of the viral messenger RNA to the cell. In fact, this capsid maintains the dsRNA enclosed inside during the whole viral cycle, thus preventing a cellular response to the presence of the dsRNA molecules. Also, there is solid evidence that the RNA segments of the multipartite genome of these viruses are actually packaged into preformed inner core capsids, although in this case only a single vertex would be involved (see below).

All these fascinating functions carried out by the inner core have to be performed under quite extreme conditions: The dsRNA segments are densely packaged inside the capsid building concentric layers with spacings between 2.5 and 3.0 nm [16]. This high RNA density inside the capsid, together with the proposed RNA packaging mechanism, resembles the case of the dsDNA bacteriophages, which is also discussed below.

12.2.2 DNA Viruses

The *Inoviridae* are ssDNA viruses which present some of the simplest helical capsids. They infect bacteria and some of them, including M13, Fd and related phages and PF1, have been the subject of detailed studies. M13 has been widely used as a substrate for biotechnological approaches [17] (see also Chap. 22). The capsid protein has a cylindrical shape with the negatively charged amino terminal segment on the outside, and the positively charged carboxyl terminal segment inside, lining a channel where the DNA is enclosed (Fig. 12.1b). X-ray fiber diffraction studies of phage PF1 have shown that the DNA follows the helical parameters of the protein helix assembly (1.6 nm pitch), with lysine and arginine residues inserting between the DNA bases, to stabilize the phosphate charges [18]. Although the details of the protein-nucleic acid are specific in each case, the overall assembly strategy of these viruses closely resembles that of the ssRNA helical viruses mentioned above.

As in the case of the RNA viruses, the most widely preferred geometry of ssDNA viruses is the spherical shape using icosahedral symmetry. Most of these viruses infect mammals and other vertebrates (parvoviruses), while there are several examples infecting plants (geminiviruses) and bacteria (phage ΦX174). The resolution of the structure of canine parvovirus showed that 13 % of the total viral DNA is organized in defined structural stretches. These stretches are 11 nucleotides long and interact with capsid protein pockets at 60 icosahedral symmetry-related positions [19]. There is a low level of sequence specificity in these DNA-protein contacts, and there are no hints of DNA-binding motifs in the capsid subunits. In fact, DNA-binding domains are generated after capsid assembly, thus representing a clear mean to stabilize correct assemblies. A singular aspect in parvovirus is the

fact that the interaction of the DNA with the capsid protein is based on the existence of a series of well conserved polar residues in the inner face of the capsid, instead of the more common interaction of the nucleic acid with basic residues found in many other viruses.

The dsDNA viruses provide some of the best known examples of sophisticated strategies to package the nucleic acid in a most efficient way. Most of these viruses have a single linear dsDNA molecule enclosed within an icosahedral capsid, although there are examples of filamentous dsDNA viruses (*Lipothrixviridae*), and others show more complex architectures that include lipidic envelopes (herpesviruses, poxviruses). dsDNA viruses as a group can infect every cell type, from bacteria (some bacteriophage families), invertebrates (*Baculoviridae*, *Iridoviridae*) and vertebrates (*Papovaviridae*, *Adenoviridae*). The best known examples of this type of viruses include the caudovirales, a widely extended bacteriophage order characterized by their icosahedral capsid attached to a tail which is instrumental for the phage-host interaction [20].

Caudovirales present a common maturation strategy, building first a proteinaceous prohead, which is later filled up with the dsDNA (for a review see Ref. [7]). The physical properties of the dsDNA molecule clearly impose severe constrains on the way it can be packaged inside a spherical (icosahedral) container. The persistence length of the dsDNA (around 50 nm) is related to a minimum diameter in the capsid of these viruses: most of them have diameters from 50 to 400 nm, and they present a complex DNA packaging machinery (see Sect. 12.4) which sits in a unique icosahedral vertex of the viral capsid (the portal). This machinery has to deal with the formidable task to fill up the viral capsid with dsDNA up to quasi-crystalline densities (more than 500 mg/ml). Electron cryo-microscopy has revealed that the DNA is organized in layers following a traverse spool [21] (Fig. 12.1f). The distance between adjacent layers is 2.5–2.7 nm, suggesting that the DNA is tightly arranged in a close hexagonal packaging. Depending on the capsid geometry, the DNA seems to adopt a different topology: fully icosahedral capsids (T7, ε15, P22) shows the coaxial spool geometry, while in elongated icosahedral variants (prolate shells in T4, Φ29) the DNA strands adopt an orientation more parallel to the longitudinal axis of the viral capsid [22]. In all these topologies the structure of the outer layers is always better defined than the inner layers: it seems clear that the inner DNA region must present sharp geometrical discontinuities (bends) to allow its accommodation in the very restricted available space. In this context, the presence of inner capsid proteins must play an important role. These components can be individual small proteins (probably instrumental in shielding the electrostatic repulsion between phosphates of neighboring DNA segments); alternatively, they can form a complex inner core connected to the portal vertex (as in T7, ε15 and myovirus) (Fig. 12.1f). The presence of this structure along the DNA translocating path appears to be related with both the way the DNA is accommodated in the inner shell surface and the ejection process [7].

An important factor to be considered in these viruses is that this highly condensed DNA inside the viral head has to be also efficiently injected into the

host bacterium. This imposes important constraints to the possible topological arrangements of the DNA, especially taking into account that although the packaging and release of DNA are performed along the same capsid vertex (the portal), the respective molecular mechanisms in which they are based are different. While the packaging reaction consumes cellular ATP to develop up to 50 pN forces (see Sect. 12.5.1 and Chap. 9), the ejection of DNA initially depends on structural changes in the tail apparatus that triggers the release of the potential energy stored in the packaged viral head.

Different studies by electron cryo-microscopy have shown that the DNA packaging pathway in these viruses follows an initial stage where the DNA is pushed inside the capsid in a poorly ordered way. Only when the percentage of DNA is around 60–70 % of the total to be packaged in phages as T3 or Φ29, the layers of DNA start to organize, probably from the layer in contact with the inner face of the capsid shell [23, 24]. At higher DNA concentrations the multiple layer topology is generated until the acquisition of the final coiled spool structure.

12.3 Reorganization of the Viral Capsid During Nucleic Acid Packaging

The viruses that actively package their genetic material (with energy consumption) start their morphogenetic pathway by forming an empty shell (prohead) that will be subsequently filled up with the nucleic acid by a molecular motor. This packaging machinery translocates the viral genome to a very high concentration inside the head (see Sect. 12.2). Nucleic acid encapsulation correlates with a change in capsid morphology in a process known as virus maturation (see Chap. 13). These conformational changes have been best characterized for dsRNA and dsDNA bacteriophages. Although associated in many cases to nucleic acid packaging, maturation may occur before, during and/or after packaging, and is described in detail in Chap. 13.

12.3.1 Conformational Changes During RNA Packaging

Active packaging in RNA viruses has only been described for Φ6 *Pseudomonas* bacteriophage [25]. Φ6 is composed by three concentric layers that enclose a segmented dsRNA genome (Fig. 12.2a) [27]. As it was previously mentioned (Sect. 12.2.1) this bacteriophage shares functional and structural characteristics (common capsid fold and common head symmetries) with other members of the *Reoviridae* family (like bluetongue or Rotavirus). Nevertheless it remains unclear whether these viruses share as well a common packaging mechanism [28].

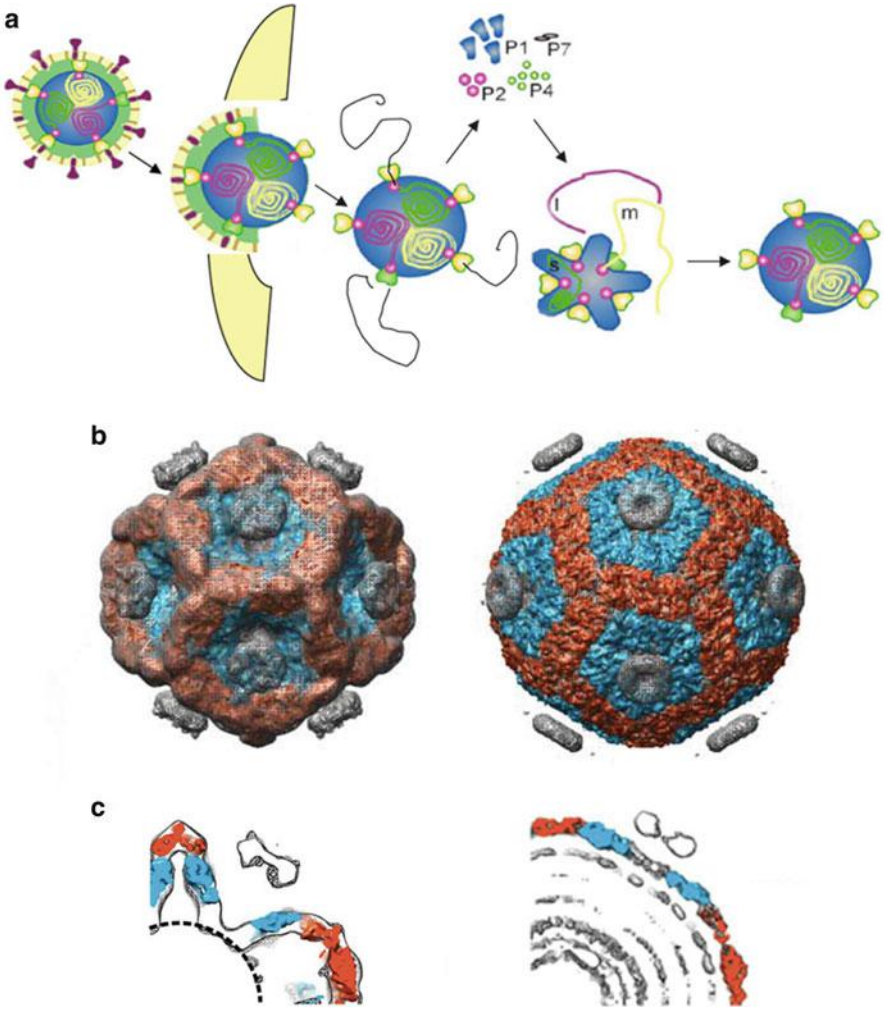

Fig. 12.2 Conformational changes during dsRNA packaging. (**a**) Morphogenetic pathway of Φ6 bacteriophage showing (from *left* to *right*): viral entry inside bacteria, extrusion of mRNA templates, synthesis of viral proteins, procapsid assembly and RNA packaging. (**b**) *Left*, model of Φ6 prohead before RNA packaging showing cup-like structures at the fivefold symmetry axes. *Right*, model of the capsid after RNA packaging. P1 (capsid protein) class A monomers are shown in *blue* and P1 class B monomers are shown in *red*. (**c**) Sections of capsid structural models before (*left*) and after (*right*) RNA packaging, showing the conformational change that takes place in the P1 monomers; color coding for the P1 monomers is the same used in (b) (Figures in panels (b) and (c) are reproduced from [26]. With permission)

In the case of the *Cystoviridae* family, the multi-layered capsid is also enveloped by a lipid bilayer. The two first layers are lost during viral entry and the innermost shell is delivered inside the cytoplasm activating RNA synthesis (Fig. 12.2b) [28]. During assembly cystoviruses form an empty procapsid, whose main component is protein P1 [25]. The procapsid shows an icosahedral shape composed by cup-like structures assembled on the fivefold vertex [28]. The fivefold vertex of the capsid is occupied by the RNA-dependent polymerase (P2), the packaging NTPase (P4) and the packaging factor (P7). The structure of the procapsid differs from the RNA-filled head. During RNA packaging the prohead structure expands sequentially to a spherical shaped structure with turrets projecting at the fivefold vertex. This conformational change has been suggested to play a role orchestrating RNA segment packaging order by sequentially exposing different genome interaction regions [28]. During this capsid rearrangement the internal volume of the capsid is increased by 2.4-fold [26].

The inner capsid is organized in a T2 lattice and the asymmetric unit is composed by a dimer of two P1 monomers, A and B. The class A monomers surround the fivefold axes, while the class B monomers assemble the pentamers together (Fig. 12.2c). Capsid expansion can be explained by local conformational changes at the P1 monomer interfaces. The most dramatic changes are produced at the A-P1 monomers interface that changes from an angular conformation to a nearly flat position (Fig. 12.2c) [26].

12.3.2 Expansion and Reinforcement of Caudoviral Procapsids During DNA Packaging

As we have mentioned above, complex dsDNA bacteriophages (caudovirales) are among the best known viruses that package their dsDNA into preformed viral proheads. These proheads (Fig. 12.3a) are composed by an outer shell, an inner scaffold, and in some cases a core-like structure, as discussed in Sect. 12.2 [20]. There is also a complex machinery located at the portal vertex of the prohead (see Sect. 12.4), which is directly involved in the translocation of the DNA inside the preformed protein container [30]. The onset of the DNA packaging is correlated with a number of changes in these proheads: there is a massive rearrangement of the shell subunits, leading to the expansion of the head diameter by about 10 %, the scaffold is disassembled and their subunits are either proteolyzed or recycled in new prohead assemblies, and the core components undergo also structural transformations [31]. In several viruses there is also incorporation of new proteins to the viral shell and/or chemical modification of shell proteins (cleavage, crosslinking). It is worth emphasizing that all these rearrangements have to be carried out in situ, maintaining the structural integrity of the viral particle while the

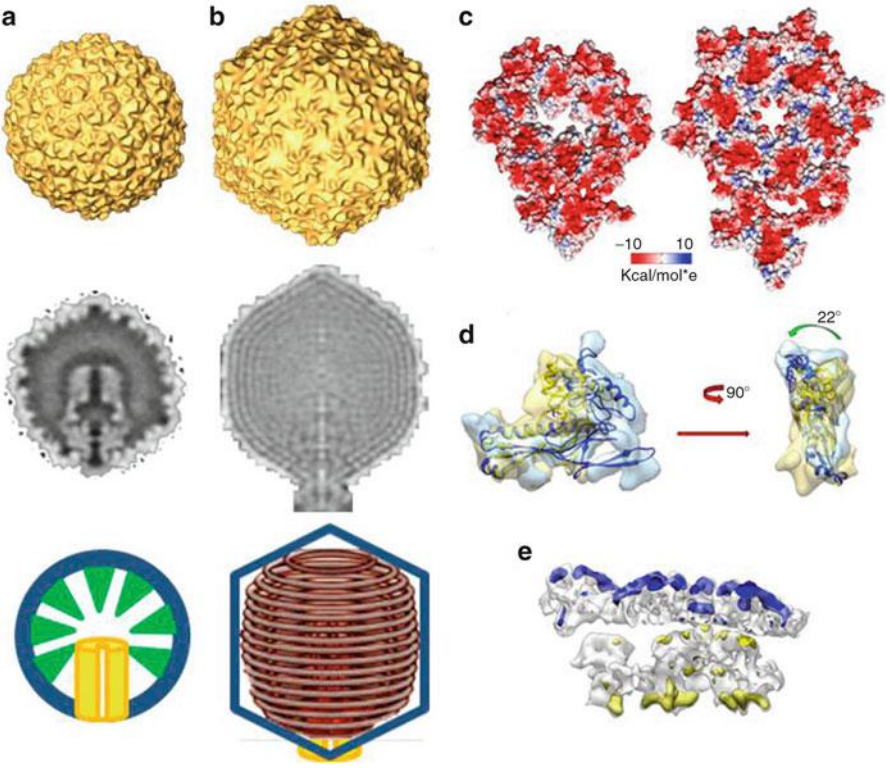

Fig. 12.3 Reorganization of *Caudovirales* capsids during DNA packaging and maturation. (**a**) Structure of the prohead of phage T7: electron cryo-microscopy reconstruction (*top*), section revealing the core and scaffolding components (*middle*) and schematic outline of the basic structural components (*bottom*). (**b**) Structure of the mature viral capsid of phage T7. The capsid is more angular than the prohead (*top*), the DNA and the core form a tight complex inside the capsid (*middle*), with the ordered DNA tightly coiled around the core (*cartoon at bottom*). (**c**) Electrostatic potential in the inner surfaces of the prohead (*left*) and mature head (*right*) of phage T7. (**d**) Two views of the relative rearrangements of the domains of the T7 main shell protein subunit when comparing the prohead (*yellow*) and mature head (*blue*). (**e**) Movement of the N-terminal domain of the main shell subunit of phage T7 from the inner side of the prohead shell (*yellow*) to the outer side of the mature head shell (*blue*) (Figures (c), (d) and (e) reproduced from [29]. With permission)

DNA molecule is actively being incorporated inside the shell interior. This reflects the exquisite architectural design of these proteins containers. The consequences of the structural rearrangements are fundamental for the packaging reaction: The viral capsid volume becomes almost 50 % bigger, the scaffolding protein is released and leaves the capsid, and a new inter-subunit interaction pattern is created (Fig. 12.3b). This reorganization makes more stable the viral shell and, at the same time, creates

a new inner shell surface (Fig. 12.3c): while the inner face of the prohead shell is quite uniformly electropositive, the mature head inner surface presents negatively charged domains that might be important for the interaction of the first layers of DNA to start the organization of the packaged DNA inside the capsid [29].

All these changes are closely related to the incorporation of the DNA, as the release of the scaffolding and the increment of the capsid volume facilitate the entrance of larger DNA molecules up to high concentrations. At the same time, the shell subunit rearrangement, based on local tilting, rotation and radial translations of the subunit domains (Fig. 12.3d), generates an increment in the contact areas between subunits, thus increasing the stability of the particle [29]. This increment might be further reinforced in different viral systems by either the incorporation of new, accessory proteins (as in T4, λ and ε15), or by covalent inter-subunit cross-linking (HK97) (reviewed in [32]). An additional aspect to be considered is the fact that during the shell reorganization, the shell protein domain that is involved in the putative contacts with the scaffolding protein changes its position from the inner side of the shell to the outside (Fig. 12.3e). This rearrangement not only facilitates (and probably induces) the release of the scaffolding protein, but also prevents the eventual competition between the DNA molecule and the scaffolding protein subunits for the interaction with the inner side of the shell. The final result is a more stable capsid, securely enclosing the DNA, and ready to incorporate the tail complex to produce the final infective viral particle.

12.4 Components of the Packaging Machinery

The packaging machinery is a molecular motor that transforms nucleoside triphosphate (NTP) hydrolysis energy into mechanical work, leading to nucleic acid translocation inside the viral head. These motors have evolved from the simplest ones present in dsRNA viruses to very complicated transient multi-complexes present in dsDNA bacteriophages.

12.4.1 dsRNA Packaging

Viruses containing a dsRNA genome present a unique packaging machinery. The existence of RNA as genomic material forces them to use a specialized viral RNA polymerase to replicate and generate its mRNA transcripts [27, 33]. To avoid degradation by host RNAses the viral polymerase operates inside the intact viral apparatus. The strategy of these viruses consists in replicating their RNA as it is translocated inside the viral head, and it requires the coordination of two viral complexes, the polymerase and the RNA translocase, both sitting at the fivefold vertex (Figs. 12.2 and 12.4a and Sect. 12.3.1). RNA packaging has been only characterized for dsRNA phages belonging to the *Cystoviridae* family [25, 27],

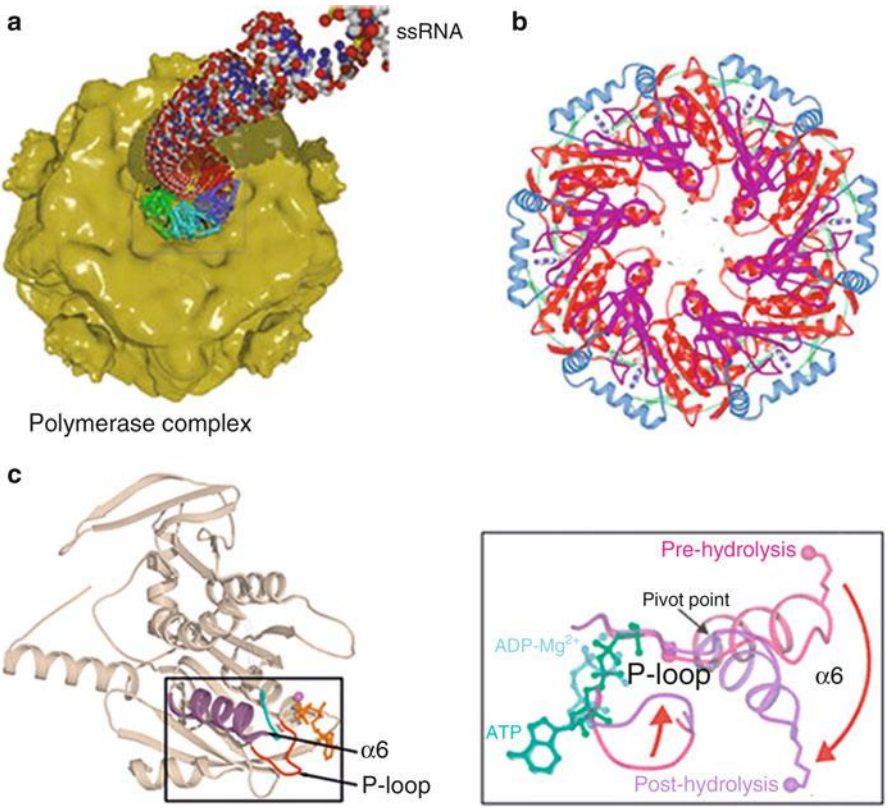

Fig. 12.4 P4 the packaging motor of Φ12 bacteriophage. (**a**) Model of the polymerase complex of Φ12 bacteriophage showing the P4 position at the fivefold vertex of the capsid translocating a ssRNA molecule (Reproduced from [27]. With permission). (**b**) Model of the atomic structure of the P4 hexamer (Reproduced from [34]. With permission). (**c**) *Left*, cartoon representation of the atomic model of one P4 monomer before ATP hydrolysis. *Right*, close-up of the region highlighted in the *left panel* showing the conformational change that takes place after ATP hydrolysis (Adapted from [27]. With permission)

and it is not clear at the moment whether this mechanism is also shared with other segmented dsRNA viruses belonging to the *Reoviridae* family [35].

Viral RNA translocases are structural proteins that are also present in the mature virus. They build the channel for ssRNA entry during packaging and exit of the messenger RNA (mRNA) transcript during the first steps of viral infection (Fig. 12.4a). RNA translocating systems are simple packaging machines as they are only constituted by one protein. P4, the packaging motor of bacteriophage Φ12, is an NTPase (it does not have ATP specificity), and it is the only one whose structure is known up to date [27]. Nevertheless this bacteriophage presents homology with other virus belonging to the *Cystoviridae* family (Φ6–Φ13). P4 shares important similarities with other helicases: the sixfold symmetry and the

characteristic Rossmann fold, consisting in a twisted eight-stranded β-sheet surrounded by five helices responsible for nucleotide binding (Fig. 12.4b) [27]. The crystal structure of P4 shows that this protein presents a conserved central core that, together with the C-terminal domain, constitutes the Rossmann fold. Despite lacking sequence homology, this nucleotide binding domain is structurally similar to the one of RecA, the T7 helicase and the bacterial conjugation protein TrW. The central channel is only wide enough to accommodate ssRNA. This channel is flanked by helix α6 and the L1 and L2 loops (Fig. 12.4b). Hydrogen-deuterium exchange and mutation of the lysines present in these loops prevents RNA packaging, indicating that these amino acids are implicated in RNA interaction and translocation [27]. Structural determination of two different nucleotide interacting protein intermediates suggested a mechanism for RNA translocation: The phosphate binding region (P-loop) changes from a down to up conformation after NTP hydrolysis, indicating that this modification could act as a molecular lever pumping the RNA inside the procapsid (Fig. 12.4c) [36].

The packaging mechanism can be subdivided into three steps: RNA recognition, RNA loading into the hexamer, and RNA translocation. The current model suggests that RNA recognition is mediated by conformational changes into the P1 capsid protein (see Sect. 12.3.1). RNA loading has been postulated to happen by ring opening, as is the case of the Rho protein. The interaction of the RNA molecule triggers NTPase activity leading to the translocation of the RNA molecule [36]. Once the molecule is inside the capsid, the viral polymerase generates the dsRNA molecules present in the mature viral particle.

12.4.2 dsDNA Packaging

As mentioned above, the DNA packaging mechanism is well conserved in dsDNA bacteriophages and in some animal viruses such as herpesviruses, showing common essential components. In this section we present a brief review of the most important of them to gather additional functional knowledge about this process.

The Portal Complex

The portal complex is located at a unique vertex of the prohead, and it comprises the connector (that builds a channel to accommodate the DNA) and the terminase, that is involved not only in the selection and processing of the DNA to be packaged but also in the ATP-driven DNA translocation. Although the icosahedral head has 12 pentameric vertices, the position of the connector and the transient interaction of the terminase is confined to a specific one. Moreover, in some viruses, such as T7 and ε15, there is a proteinaceous internal structure, called the core, also interacting with

Table 12.1 Viral components involved in dsDNA packaging

	Connector		Large terminase		Small terminase		dsDNA
Virus	Gene product	Mass (Kda)	Gene product	Mass (Kda)	Gene Product	Mass (Kda)	Length (Kb)
λ	gpB	59	gpA	73	gpNu	20	48.5
SPP1	gp6	57	gp2	49	gp1	21	45.9
T4	gp20	61	gp17	70	gp16	18	166
T7	gp8	59	gp19	67	gp18	20	40
P22	gp1	83	gp2	58	gp3	19	43.4
Φ29	gp10	35	gp16/pRNA	39/58	gp3	31	19.3
HSV1	UL6	74	UL15	81	UL28	85	152

this vertex. This core is not required for the prohead assembly, but it seems to be essential for infectivity and, as it has been revised in Sect. 12.2.2, it may facilitate the topological ordering of the dsDNA genome during packaging and/or release. Consequently, through a special vertex the DNA is both pushed by the portal complex and, once the virus is formed, ejected during the infection process [30, 37].

The function of the portal complex is defined not only by the communication among its own components but also by the interaction with the DNA substrate. The linear dsDNA molecule varies in length for each viral system: from 19.3 Kb in the case of Φ29 to 166 Kb in T4 (Table 12.1). Most of them (except Φ29 and some related phages) produce head-to-tail multimers, or concatemers, of DNA as a substrate for packaging. Concatemeric DNA is formed by genomic units linked by tandem repeats generated during the replication process. The recognition of the terminal sequences by the viral proteins ensures the specific packaging of the viral genome from that of the host. Moreover, the terminal repeats indicate the cutting region between genomic units, avoiding the loss of base pairs in the 5′ end, and ensuring the packaging of a single molecule. The specific recognition sequences, named cos or pac sequences, differ depending on the viral system. They can be identical in both extremes as in T7, T3 and λ; non-unique and repeated at the ends as in P22 and T4; or miscellaneous, with a protein covalently attached, as in Φ29. Besides this sequence recognition, two cleavages are required for the unit length packaging: the initiation cut (that generates the free end from which packaging starts) and the terminal cut (that delimitates the genome unit). Those cleavages can be either sequence specific, as in λ and T7, or sequence independent as in T4, P22 and SPP1. In the latter case, the connector acts as a termination sensor that produces the "headful-signal" when the genome is already packaged [38, 39].

The packaging proteins and their size, together with the genome length of several of the most abundant of those viruses (tailed bacteriophages) are presented in Table 12.1. Connector proteins and large terminases present wide variations in size, although the size of small terminases is more similar in different phages (except for HSV1, significantly higher). Even though the dsDNA molecule is always linear, its length differs considerably from one virus to another [38, 40].

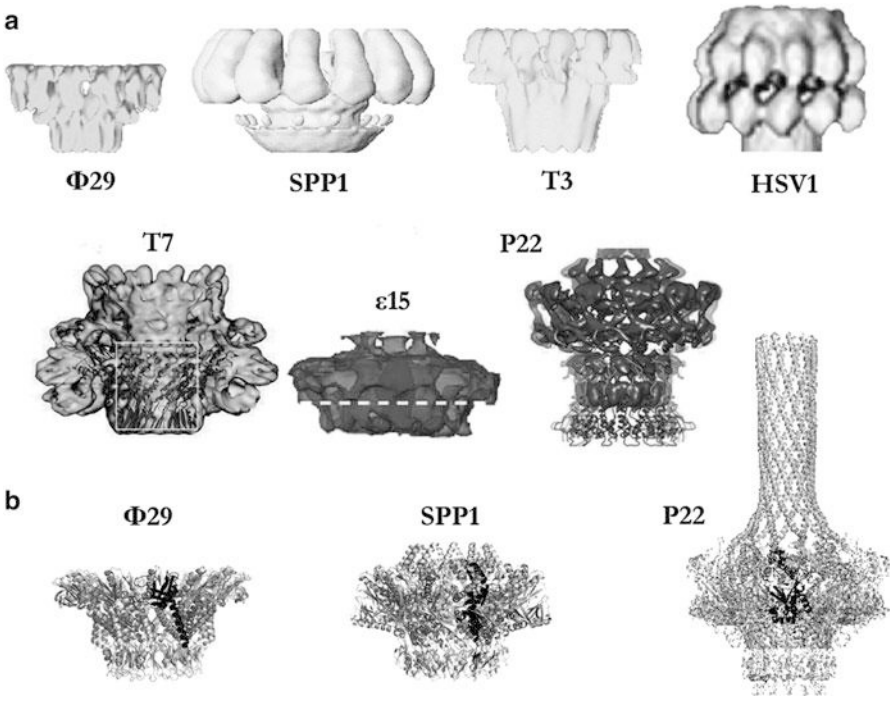

Fig. 12.5 Oligomeric connectors structures in dsDNA viruses. (**a**) Electron microscopy three-dimensional reconstructions of oligomeric connectors in different viruses. The structures share a toroidal morphology with wings protruding, a central channel (visible in the phage T7 connector) and a stem region in the basal part. The models present 12-fold rotational symmetry, except for the phage SPP1 connector model that was obtained for in its 13-mer form. (**b**) Atomic structures of the oligomeric connectors solved by X-ray crystallography. The conserved central domain is highlighted. PDB codes: phage Φ29 (1H5W), phage SPP1 (2JES) and phage P22 (3LJ5) (Reproduced from [41]. With permission)

The Connector

Although connector proteins do not share sequence homology, and they present large variations in size, they show a common architecture. Connectors are dodecameric assemblies with a toroidal morphology showing a conspicuous channel in the center. This ring shape is common to protein complexes involved in DNA translocation such as helicases, sliding clamps and certain exonucleases and topoisomerases, suggesting the implication of the channel in the DNA interaction. As shown in Fig. 12.5, several oligomeric connectors have been solved using electron microscopy and X-ray crystallography. These propeller-like structures can be divided in several domains (see Fig. 12.5a, b): the crown (apical part in HSV1, T7, ε15 and P22), the wing (central part, with maximum diameter) and the stem (the tight region in the base of the structures). The central domain is the most conserved, and it comprises two helices and an additional extended α-β domain

present in all bacteriophage connectors (see the highlighted ribbons in Fig. 12.5b). The other two domains, the crown and the stem, present variations in length and complexity (note the huge α-helical tube protruding from the crown of the P22 connector, in Fig. 12.5b). Considering the structural similarities among all connector proteins it is reasonable to assume that the smaller one would represent the minimum core to perform the main common functions while the supplemental domains, present in larger connectors, would correlate with the acquisition of additional features [41, 42].

All the connectors inserted in the viral head are built by 12 subunits, although the overproduction of this protein may induce the formation of different oligomeric forms. The symmetry mismatch between the dodecameric connector and the pentameric vertex plays an essential role during DNA translocation and it is also involved in the tail attachment to the portal. Moreover, the tenfold symmetry of the dsDNA might act as a mediator among the rotational symmetries of the portal complex and the vertex. The toroid oligomeric symmetry builds a channel with a diameter wide enough to fit the dsDNA molecule. The overall surface charge of the channel is mainly electronegative (especially at the entrance and exit) with rings of positive charges scattered all over the walls that may interact with the electronegative DNA molecule. Furthermore, SPP1 and Φ29 connectors present loops protruding inside the channel that may be involved in retaining the DNA inside the head [41].

The interactions between the connector and other viral components are fundamental for the portal functioning, as it has been shown by mutational analysis. The connector has been proposed to nucleate the prohead assembly by the interaction with the scaffolding proteins. The connector also interacts with the viral ATPase, serving as a docking point and modulating the packaging motor activity. In Φ29, an unusual case, the connector directly interacts with a specific virus-encoded small RNA molecule (pRNA), which encircles its narrow end acting as a bridge between the connector and the ATPase (see below in this Section). The P22 connector presents a helical barrel that would mediate the orderly DNA filling of the head and it also regulates the delivery pressure during DNA ejection. Additionally, the connector interacts with either the tail proteins or other proteins involved in the closure of the channel after DNA packaging. In fact, the connector has been proposed to be the sensor of the headful mechanism, defining the quantity of DNA to be packaged. Finally, to develop all its functions, during and after DNA packaging the connector probably undergoes conformational changes, which would not be necessarily irreversible as it was previously suggested [41, 43].

The Terminase Complex

The second component of the portal is the terminase complex, proposed to be the macromolecular motor that converts chemical energy from ATP hydrolysis into mechanical movement of DNA during phage morphogenesis. Beside the ATPase activity, most terminases also contain the endonuclease that cuts concatemeric

DNA into genome lengths. This nomenclature derives from the phage λ proteins that were first shown to be required for the formation of the termini of the packaged DNA. Terminases are generally hetero-oligomers built by a small protein involved in DNA recognition, and a large protein containing the ATPase and nuclease activities and a motif for docking at the portal vertex. Phage Φ29 is an unusual case, as its large terminase protein (gp16) does not have nuclease activity (the Φ29 DNA replicates as a unit length genome), and it functions in coordination with a small packaging RNA (pRNA) required to dock the gp16 onto the portal. Although this pRNA may be considered as part of the large terminase, it would be described separately below in this section. Even though there is no significant overall sequence similarity, terminase proteins from different phages contain well-conserved patches of amino acid sequences, or structural motifs that are required for packaging [38]. The structural information available about terminases, mainly obtained by X-ray crystallography, facilitates correlating their topology and interactions to their function along DNA packaging.

The large terminase subunit. Only a few atomic structures of large terminases have been obtained due to their flexible, conformationally heterogeneous nature (Fig. 12.6a). The first crystallographic structure presented was the gp17 amino terminal domain (Ndom) from RB49 phage, which shares 72 % sequence identity with its counterpart in phage T4. Afterwards, the carboxyl terminal domain (Cdom) of the same protein and the full-length crystallographic structure of gp17 from phage T4 (Fig. 12.6a, right panel) were solved. Recently, the atomic structures of the Cdom of both G2P and UL89, from phage SPP1 and human cytomegalovirus (HCMV) respectively, were obtained (Fig. 12.6a, left and central panel).

Sequence alignments show that the functional signatures of the ATPase domain of viral terminases are conserved and they are similar to those of the translocating monomeric SF2 helicases, restriction endonucleases, and protein translocases. Large terminases consist of two domains, an N-terminal ATPase domain that powers DNA translocation and a C-terminal nuclease domain that generates the termini of the viral genome. The Ndom of T4 consists of the classic nucleotide binding Rossmann fold (see Sect. 12.4.1). It contains the Walker A and B and the catalytic carboxylate, often found in proteins that bind and cleave ATP. Biochemical analysis of large terminases homologues from phages Φ29, λ, SPP1, P22, T3 and T4 confirm the ATPase activity of this domain. The nuclease Cdom presents an also conserved RNase H-like fold formed by seven β-sheets sandwiched between two clusters of α-helices. This basic fold (pointed by an asterisk in Fig. 12.6a) displays variations in length and shows almost no amino acid sequence identity [38].

The over expressed and purified large terminases from many phages exhibit different oligomeric states. T4 gp17, SPP1 G2P, P22 gp3 and HCMV UL89 exist essentially as monomers, while λ gpA oligomerizes in solution. Nevertheless, during the packaging machinery assembly, the terminase stoichiometry may be remodelled. The analysis of Φ29 and T4 large terminases bound to the portal vertex has shown electron densities that are consistent with five subunits being present. In the case of pRNA, six-prohead bound fluorescent molecules were reported, whereas

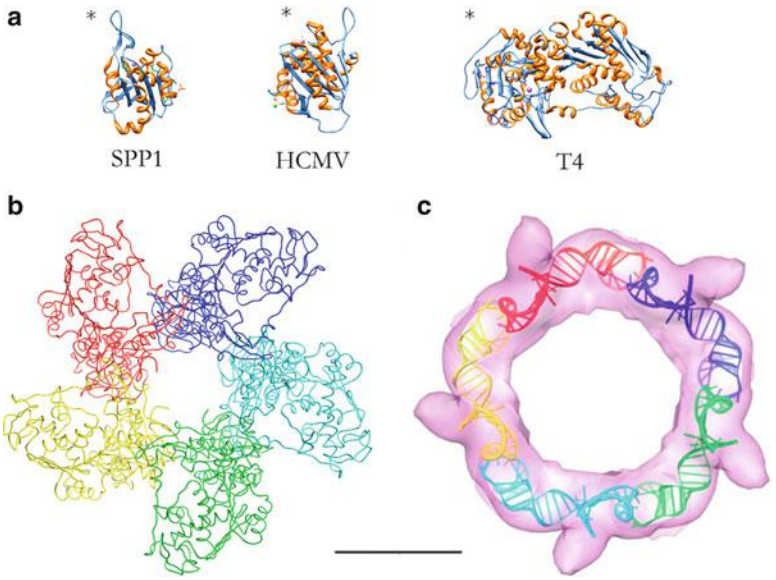

Fig. 12.6 Atomic structures of the large terminases and the pRNA solved by X-ray crystallography. (**a**) Large terminase monomers: nuclease domains of G2P and UL38 from phage SPP1 and HCMV, respectively; and complete structure of gp17, from phage T4. The nuclease domain (*asterisk*) shows an RNase H/integrase-like fold constituted by a bunch of parallel and anti-parallel β-sheets surrounded by several α-helix. (**b**) Frontal view of the atomic model of the pentameric gp17 terminase from phage T4. Each monomer is depicted in a different colour. (**c**) Flexible docking of the pRNA crystal structure into its electron cryo-microscopy pentameric envelope (Reproduced from [44]. With permission). The scale bar corresponds to 50 Å and refers to images in panels (b) and (c). Images in (a) and (b) were generated from atomic coordinates deposited in the PDB. Access numbers: phage SPP1 (2WC9), HCMV (3N4Q), phage T4 (monomer 3CPE, pentamer 3EZK) and phage Φ29 pRNA (3R4F)

recent studies support its interaction with the prohead forming a pentameric ring. Figure 12.6b represents the pentameric atomic model of T4 gp17, and the flexible fitting of the pRNA crystal structure into its electron cryo-microscopy pentameric envelope is presented in Fig. 12.6c. The oligomeric ring-like form, together with the diameter of the channel (40 Å in T4 gp17 and 82 Å in pRNA), allow the passage of the dsDNA molecule [45]. The terminase pentameric stoichiometry fits well with the tenfold DNA symmetry, but it mismatches with the 12-fold connector symmetry. The latter can be largely reconciled if there were interactions between the terminase and the pentameric vertex of the prohead.

During packaging the large terminase interacts with the connector (inserted in the prohead), with the DNA (first to translocate it and finally to cleave it) and with the DNA attached-small terminase. The binding to the connector-prohead is localized in the C-terminal region of most of the terminases, and it is based on charge-charge interactions. As suggested above, DNA is translocated through the channel of the large terminase, and the cleavage is carried out at a catalytic groove

of its Cdom. Finally, genetic and biochemical studies show that the interaction with the small subunit is localized on the Ndom of the large subunit [45, 46].

As the site for ATP binding and DNA cleavage resides in the large terminase, it has been proposed to undergo a conformational change in response to ATP hydrolysis to physically move the DNA. However, it has also been hypothesized that the energy from ATP cleavage is transmitted from the terminase to the connector, which in turn moves the DNA. The packaging models, together with the physical properties of the packaging process, will be revised in Sect. 12.5.

The small terminase subunit. Besides the DNA recognition function, small terminases perform a regulation of the large terminase enzymatic activities during DNA translocation. The small subunit is more variable in amino acid sequence that the other two motor proteins. Even though the Φ29 phage genome does not encode any small terminase, the gp3 protein primes its DNA developing an analogous function. In the last decade several structures of these proteins have been obtained (Fig. 12.7) using structural techniques such as nuclear magnetic resonance spectroscopy, electron cryo-microscopy and X-ray crystallography. Biochemical and structural data revealed how the domain organization is highly conserved among small terminases. The Ndom is the DNA binding domain, the central domain acts as the oligomerization domain and finally the C-terminal is in charge of the large terminase interaction. The N-terminal fragment of λ gpNu1 and the Shigella phage Sf6 g1p (Fig. 12.7a and c, respectively) share a similar winged helix-turn-helix fold. As shown in Figs. 12.7c, d, and e, the central domain is formed by two conserved α-helices and the Cdom share a characteristic β-barrel. The oligomerization of the small terminase produces multimer rings that vary in size from octamers to decamers (or larger). These data suggest that although oligomerization is important, the stoichiometry or a defined inner diameter does not appear to be strictly essential for its function. Alternatively, interactions with other components of the packaging machinery may select for a precise stoichiometry *in vivo*. Hence, the rings formed by the small terminases present variable channel diameters, from 10 to 24 Å in their narrow region. As a consequence, the wrap around model of DNA interaction has been increasingly favoured against the channel mediated one. Finally, despite the conserved overall structure of the small terminase, no conserved residues were located on the exterior surface indicating that the interaction with the large terminase or connector protein is mediated by the overall architecture of the portal complex and by the distinctive shape of the small terminase [45, 48].

Other Components: The pRNA

Bacteriophage Φ29 is an especial case because of the presence of a unique 174 nucleotides RNA molecule that forms part of its packaging motor together with the connector and the large terminase. Since this RNA was found to bind to proheads, it was named pRNA, and it is not needed to assemble proheads suggesting that pRNA attaches to proheads after capsid assembly. Its role is transitory and likely limited to

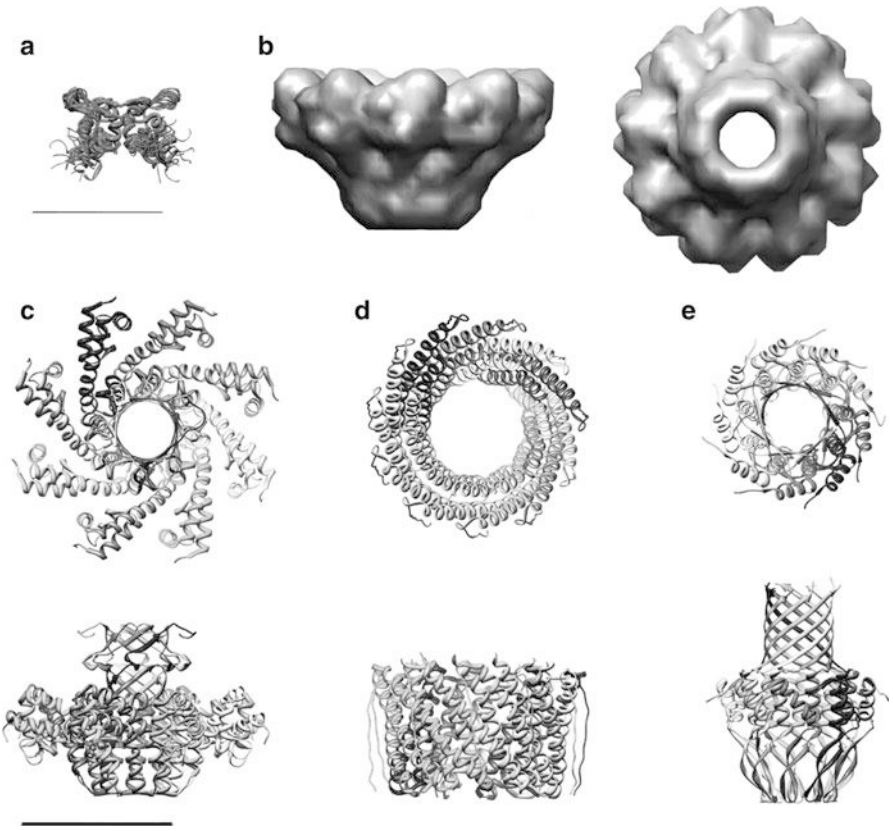

Fig. 12.7 Structures of the small terminases solved by different techniques. (**a**) DNA binding domain of the phage λ gpNu1 dimer solved by RMN. The scale bar corresponds to 25 Å. (**b**) Electron microscopy three dimensional reconstruction of the phage P22 gp3 nonamer at 18 Å-resolution (Reproduced from [47]. With permission). (**c**) Atomic structure of the octamer of g1p from the *Shigella flexneri* phage Sf6. (**d**) Crystal structure of the central region of the undecamer of gp16 from the 44RR phage (T4-like family). (**e**) Atomic structure of the nonamer of G1P from the *Bacillus subtilis* SF6 phage (SPP1-like family). In panels (c) to (e) frontal views (*upper panels*) and lateral views (*lower panels*) are shown. The scale bar corresponds to 50 Å and refers to sections (b), (c), (d) and (e). Images were generated from atomic coordinates deposited in the PDB. Access numbers: phage λ (1J9I), phage Sf6 (3HEF), phage 44RR (3TXQ) and phage SF6 (3ZQP)

DNA packaging, as pRNA is not present in mature Φ29. Thus, the connector, the pRNA and the ATPase form three concentric rings, and the DNA-gp3 is believed to be threaded through a continuous channel along their shared central axis and into the phage capsid [49].

The pRNA secondary structure consists of two domains: domain I comprises the first 117 bases and is separated from domain II by a 13 base single-stranded region. Although the full-length wild type is a 174-base transcript, a 120-base construct

encompassing domain I and lacking 54 nucleotides 3'-domain II is fully competent for packaging assembly *in vitro*. Despite the domain II function is not currently understood, its conservation in all Φ29 relatives suggests it may have an essential function *in vivo*. The domain I shows two functional modules: one necessary for prohead binding (helices C, D and E, and the three loops) and the other required for DNA packaging activity (the majority of the A helix), which provides the attachment site for the large terminase [50].

Formation of intermolecular pseudo-knots (also conserved in phage Φ29 relatives) between adjacent pRNAs would promote circular interactions resulting in an oligomeric ring-like structure. Nevertheless, the pRNA stoichiometry when bound to the prohead has been controversial. The hexameric state was supported by the analytical ultracentrifugation detection of both dimers and hexamers of pRNA [51]; and also by single particle fluorescence quenching experiments [52]. More recently, electron cryo-microscopy analysis definitely showed that bases 22–84 form a pentameric ring-like structure that binds to the prohead (see above and Fig. 12.6c), that five pRNA A-helices extend as spokes from this central ring, and that the terminase attaches to the distal end of those A-helix spokes [44]. The fitting on the electron cryo-microscopy envelope also revealed that although the pRNA was close to the connector, it made much more extensive contacts with the fivefold prohead vertex than with the dodecameric connector. This is in accordance with the effective binding of pRNA to connector-less proheads [50].

12.4.3 Other ATPase-Like Motors

Inside the cells there are proteins involved in the transport of nucleic acid between different compartments that have some resemblance to the viral DNA translocation motors. These proteins are ATPases sharing a RecA-like fold and assembled into propeller-like structures with a central channel that serves to transport the nucleic acid [53]. Some of these proteins play essential roles in chromosome segregation between two twin cells (as FtsK), or in bacterial conjugation to exchange genetic information between two different bacteria (as TrwB). FtsK translocates dsDNA and presents also a ring hexameric assembly. The structure is divided into two domains assembled through a central linker as in T4 ATPase (see Sect. 12.4). This structure suggests that during translocation the two domains would open like a jaw moving the DNA in an inchworm movement with a step size of 2 bp/ATP, similar to the one found in the viral DNA packaging motors in bulk experiments [54]. TrwB translocates ssDNA, and it is also an ATPase with a toroid structure assembled into a hexamer. Its atomic structure showed that it presents a RecA like fold. It has been suggested that ATP hydrolysis will trigger a conformational change allowing opening the central channel, which is too narrow in the solved atomic structure to translocate the dsDNA molecule [55].

12.5 Models for Nucleic Acid Packaging

The structural characterization of different components involved in viral nucleic acid packaging allowed proposing different models to understand the genome translocation process. Recent studies on single molecule experiments permit defining the mechanical characteristics of these motors (force generation, velocity, processivity and motor steps) using an optical tweezers setting (see Chap. 9). These experiments have been mostly carried for dsDNA bacteriophages. In this section we will first describe the properties of different machineries and then we will describe the mechanical models that have been proposed during the years. These models try to understand the coupling between ATP hydrolysis energy production and the conformational changes that lead to nucleic acid pumping inside the viral head.

12.5.1 Biophysical Properties of the Packaging Motor

The accurate measurement of viral genome packaging kinetics is difficult without the synchronization of different populations. This process is even harder for dsDNA motors where the motor only assembles transitory but it is not a structural part of the mature virus (see Sect. 12.4.2). While biochemical bulk assays can only determine the packaging efficiency by means of the total time to package the nucleic acid, single-molecule experiments allow collecting statistics of single events, and they have allowed observing for the first time pauses where the motor eventually disassembles from the packaged substrate (see Chap. 9) [56].

The optical tweezers setting usually consists in a viral prohead assembled with the packaging ATPase in the presence of an ATP analogue, then a microsphere carrying DNA is introduced and the variations in the nucleic acid extension values are carefully followed while different forces are applied to the system [45] (see Chap. 9 for technical details). These experiments have revealed the speed of the packaging motor for three different dsDNA bacteriophages: T4, Φ29 and λ [45, 56]. The speed values turned out to vary from 180 bp/s for Φ29 to 1,800 bp/s for T4, on average. The speed of the motor seem to be closely related to the length of the phage genome (see Table 12.1 and Sect. 12.4.2), thus phages with larger genomes have adapted to faster velocities in order to complete DNA encapsulation within 2–3 min from the total viral cycle of 20–30 min. It has also been observed that the speed decreased at late stages of packaging, when the internal pressure of the capsid increases. For example, λ packaging speed drops threefold when 90 % of the genome is inside the capsid.

Another biophysical value determined is the step size, or the translocation length of the motor after one ATP molecule hydrolysis. In bulk assays this value was determined to be 2 bp/ATP. Single molecule experiments have shown however, that he step size cannot be taken as a fixed value, as it can vary according to DNA symmetry and filling of the capsid. High resolution optical tweezers recently

showed that the motor translocates in four 2.5 bp steps (see Chap. 9). This strongly suggests a high coordination of the subunits, as the terminase is composed of five subunits and only four of them bind ATP before the hydrolysis and translocation are triggered [56].

Optical tweezers have also revealed that packaging motors are able to generate forces up to 50-60 pN at low capsid filling (see Chap. 9). These forces are 20–25 times that of myosine II, implicated in muscle contraction. Such high values may be essential to overcome internal capsid pressure, electrostatic repulsive forces and DNA bending (see [45] and Chap. 19).

The DNA packaging machinery is considered as a motor able to transform the chemical energy of ATP hydrolysis in a mechanical work consisting in nucleic acid translocation. One of the most interesting features to understand the mechanism of nucleic acid packaging is the coupling of these two events. During nucleic acid translocation the motor must interact with DNA and translocate the step size, release the DNA and set the process to zero to restart a new cycle. The ATP reaction can be divided in several steps consisting in ATP docking, nucleophilic attack of the γ-phosphate, and release of reaction subproducts. Recent studies with optical tweezers suggest that Φ29 engages DNA after ATP binding and that DNA is translocated right before ADP release [56].

12.5.2 dsRNA Packaging Models

Considering the large differences described between DNA and RNA packaging motors (see Sect. 12.4) it would be reasonable to think that RNA packaging machineries would show different characteristics from the ones described for DNA packaging machines. Although no single-molecule experiments have been reported up to date, the atomic resolution of two different RNA translocase conformations (P4 protein, see Sect. 12.4.1), and accurate biochemical studies have given strong structural insights into the molecular mechanistics of the dsRNA motors [27]. The main outcome from these data shows several interesting similarities with the DNA packaging motors. Structural snapshots showed that the P-loop of P4 changes its conformation from up to down in the presence of two different ATP analogues suggesting that this movement could help to pump the RNA inside the prohead (see Sect. 12.4 and Fig. 12.4). It has been proposed that P4 would follow a sequential translocation mechanism. In this mechanism the coordination between different subunits would be essential. Communication between adjacent subunits could happen through “arginine fingers” commonly found in ATPase motors. In this model ATP hydrolysis will trigger a conformational change that place the arginine finger of the adjacent subunit into the ATPase active site starting the next ATP hydrolysis [36]. It is supposed that the ssRNA structure could be similar to the A-dsRNA form presenting 11 bases per turn. It has been suggested that this symmetry mismatch would be solved by the L1 loop that would act as a grommet correcting the RNA position [39].

12.5.3 dsDNA Packaging Models

As previously described in Sect. 12.4, dsDNA packaging motors are more complex than dsRNA packaging motors. Usually their packaging machinery is composed by the dodecameric connector that sits in only one vertex of the prohead, and the terminase complex composed by the viral ATPase, which transiently associates to the prohead during packaging. Despite large efforts invested in the characterization of these motors, their precise molecular mechanism of translocation remains unclear. The models proposed can be classified in two categories. (1) Connector-driven models, where conformational changes induced by ATP hydrolysis in the connector structure lead to DNA packaging. (2) ATPase-driven models, where the connector is considered as a passive actor in the translocation mechanism while conformational changes in the ATPase directly translocate DNA. It is important to point out that both models consider the terminase as the ATP-consuming element of the motor. The difference lies in the element that actively pushes the DNA inside the capsid. In the case of the connector-driven models the terminase regulates the connector activity so it can push the DNA inside the capsid [50].

The first proposed model was based on the symmetry mismatch between the connector and the prohead, 12-fold and 5-fold respectively. This lack of symmetry would lead to a weak interaction between both components allowing the rotation of the connector inside the capsid. In this model connector rotation would lead to nucleic acid packaging inside the viral head [38]. The atomic structure determination of the connector allowed proposing later on the precise conformational changes inside the structure that would lead to DNA translocation. Observation of the Φ29 atomic structure showed that the helices that built the channel wall are not in a straight position. It was thus suggested that the straightening of one single helix by 12° rotation at the bottom of the connector would lead the expansion of the structure. This conformational change would be followed by the top of the protein leading to the contraction of the complex. This conformational change would be induced by the sequential hydrolysis of the ATP by the terminase (Fig. 12.8a) [38]. Another model based on the Φ29 connector structure suggested that the positive lysine ring inside the structure acts as an electrostatic grip on the negatively charged DNA molecule (Fig. 12.8b). This electrostatic interaction, together with the rotation of the connector would let the DNA to switch between two lysine rings and to be translocated without any additional connector structural rearrangement [50]. SPP1 connector pseudoatomic structure also inspired a different model in which the central channel is closed by a loop belt, suggesting that a sequential conformational change in these loops could act as a molecular lever to translocate DNA inside the capsid. In this model ATP hydrolysis will lead to 12° connector rotation and loop sliding and returning to the initial conformation [38].

Connector-driven models have been challenged after it was proved that the connector does not rotate inside the capsid. The recently solved structures of the T4 terminase have reinforced the terminase-driven models. The carefully biochemical and structural analysis carried out in T4 suggested the localization of the

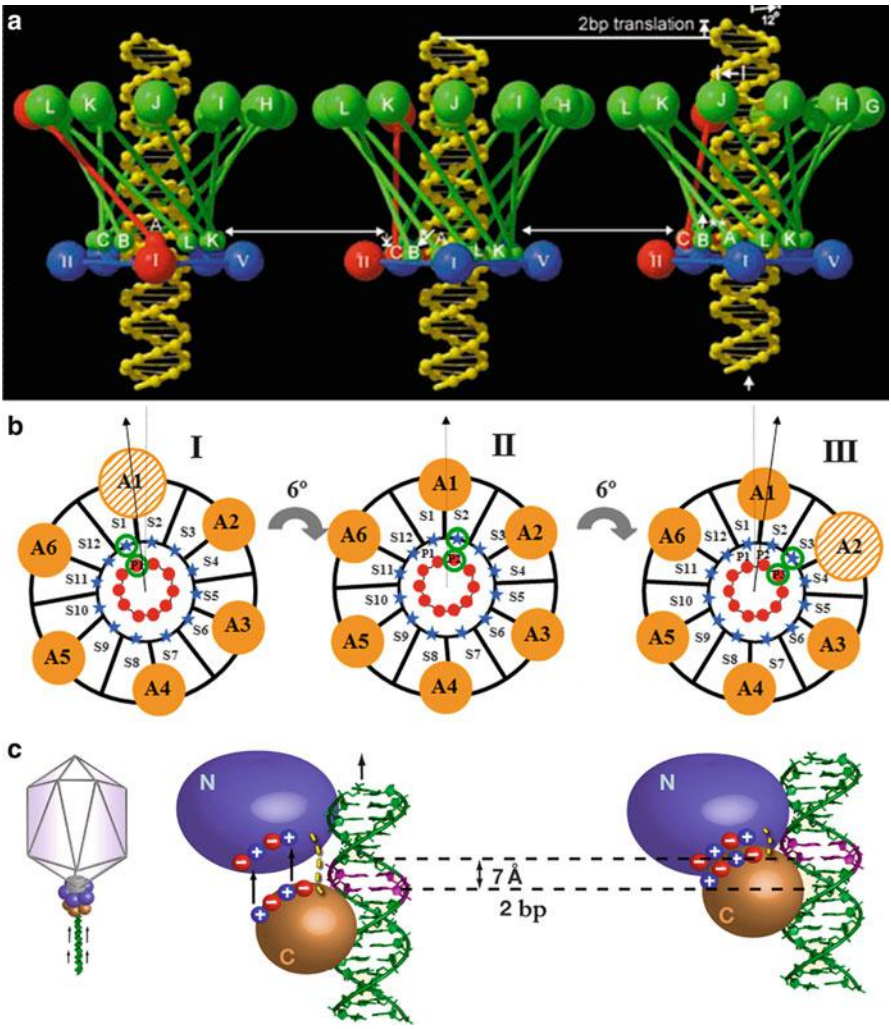

Fig. 12.8 Mechanical models of DNA packaging. (**a**) Compression-relaxation model showing the ATPase represented in *blue spheres* and the connector in *green*. The subunit of the connector suffering the conformational change is represented in *red*. First the subunit is straightened from the bottom (*centre*), then the movement is followed by the upper part consisting in translocation of 2 bp. (**b**) DNA gripping model showing the 12 subunits of the connector in *black* and the ATPase subunits as *yellow circles*. The ring of lysines located in the connector channel acts as an electrostatic docking point for the DNA (Figures in panels (a) and (b) are reproduced from [50]. With permission). (**c**) Terminase electrostatic model showing the two terminase domains Ndom in *blue* and Cdom in *orange*. During DNA packaging the two domains get closer and are pulled out by electrostatic repulsion allowing DNA translocation (Reproduced from [57]. With permission)

structural elements involved in ATP hydrolysis and their coupling to DNA translocation (see Sect. 12.4). It was shown that the terminase is composed by two domains: the Ndom which presents the ATPase activity and the nuclease Cdom (see Sect. 12.4). Both domains are joined by a flexible linker that allows their motility. It was proposed a model based on electrostatic interactions where ATP hydrolysis triggers a conformational change in the Ndom aligning both domains. This domain neighbouring will cause the pulling of the Cdom in an inch-worm like movement driven by the electrostatic forces conducing to DNA packaging (Fig. 12.8c) [58].

12.5.4 Differences Between Nucleic Acid Packaging and Ejection

Viral capsids are not impermeable containers and they present small pores that are necessary to exchange salts and ions with the media. Nevertheless these pores are too small to allow viral genome exit during infection. The infection process requires either capsid disassembly (as it occurs in most animal viruses) or the presence of a channel large enough to accommodate the viral genome. In viruses following an active packaging mechanism (dsDNA and dsRNA bacteriophages) the nucleic acid exits the capsid through the portal/translocase channel, at the fivefold vertex (see Chap. 17), the same passage that is used during genome packaging (see Sect. 12.4). These viruses package their genome in ordered layers (see Sect. 12.1) inside the capsid condensing its nucleic acid at a high density and accumulating an important potential energy (reflected in an internal pressure of up to 60 atmospheres; see Chaps. 9, 18, and 19). This high pressure requires the presence of protein complexes in the portal axis able to retain the nucleic acid inside the capsid. In dsDNA viruses this task is accomplished either by conformational changes in the portal channel or by protein plugs, like the tail proteins or the gatekeeper proteins [59] (see Chap. 17).

Even though nucleic acid packaging and ejection follow the same channel pathway they are not inverse processes. While packaging requires ATP consumption, genome delivery uses the energy stored in the nucleic acid during the condensation process [60]. Nevertheless this internal energy decreases as the genome is liberated and, thus, it is believed that other mechanisms may participate, as it is the case of proteins inside the host cell that pull the nucleic acid when the replication process is initiated (see [61] and Chap. 19). Differences are also evidenced by the fact that both processes have very different speeds, at least as measured in *in vitro* experiments: while DNA packaging is carried out at an average speed of 700 bp/s, DNA release can attain 60,000–70,000 bp/s.

In dsDNA bacteriophages there are strong evidences that the DNA entry and exit are carried out by different protein complexes. While DNA translocation is done by the packaging motor, DNA delivery uses the tail machine [62] (see Chap. 17). As some protein loops may as well be retaining the exit of the DNA, genome delivery requires conformational changes to modify the ejection channel in order to allow DNA passage [59]. Some viruses with dsRNA genomes present similar

characteristics, as in the first stages of the viral cycle they extrude their mRNAs trough the channels at the fivefold vertex (Fig. 12.2 and Sect. 12.3). During this process the positive strand from dsRNA is displaced to generate the mRNA substrate. Once more, this process is not energy-consuming and P4 protein serves as a passive channel for ssRNA.

12.6 Perspectives and Conclusions

Viruses protect their genetic material against the outside media or the cellular degrading enzymes by enclosing it inside a protein capsid which also serves other biological functions. Viral particles have evolved different packaging strategies according to their lifestyles, type of nucleic acid (RNA or DNA, single- or double-stranded), amount of genetic material to be packaged and other variables. A widespread strategy consists in the co-assembly of capsid and nucleic acid; in these cases the nucleic acid may act as a scaffold to promote capsid assembly, and the capsid proteins may help in the structuration of the nucleic acid inside the virion. In the simplest cases, this co-assembly strategy results in a helical structure that specifically encloses the viral nucleic acid. Packaging pathways involving condensation of capsid and nucleic acid are used by many bacterial, plant or animal ssRNA or ssDNA viruses.

Many other icosahedral viruses use dsRNA or dsDNA as genetic material. Packaging of long, stiff double-helical nucleic acid molecules pose additional problems; these viruses evolved sophisticated mechanisms for the packaging of these nucleic acids into preformed capsids. Viral particles actively packaging their double-helical nucleic acid genomes into spherical containers usually organize the nucleic acid in concentric layers inside the capsid. This arrangement facilitates genome uncoating by ejection without capsid disassembly.

The packaging machinery in those dsDNA and dsRNA viruses is a molecular motor that transforms the chemical energy of NTP hydrolysis into mechanical work to pump the nucleic acid inside the capsid. During the packaging process the initial immature capsid (prohead) suffers a series of conformational changes to become a mature capsid. The nucleic acid is transported inside the capsid through a channel formed by a protein complex present at a fivefold axes. These proteins present similarities with other nucleic acid translocases, showing ring-like structures with a central channel that serves for genome transport. In dsRNA bacteriophages this protein is named RNA-translocase and constitutes the simplest packaging motor described up to date. Structural snapshots obtained by X-ray crystallography suggest a mechanical mechanism consisting in the movement of a loop that would sequentially change conformation to transport the genome. dsDNA bacteriophages present a more complicated machinery, formed by a multi-protein complex that assembles transiently to the capsid. In these viruses the translocation pore is built by a protein, the connector, that sits at one single vertex of the capsid. The connector serves as a docking point for a viral complex, the terminase, that attaches transiently

to the capsid during packaging. The terminase complex is essentially composed by an ATPase and a secondary component, either a protein or RNA, which enhances the activity of the ATPase and specifically recognizes the viral genome. The properties of this multicomponent motor have been characterized using single molecule techniques (optical tweezers), which have revealed the biophysical characteristics of what constitutes one of the most powerful motors described to date. Several models have been proposed to couple the chemical energy of ATP hydrolysis with the mechanical action of DNA pumping; nevertheless the precise molecular mechanism remains unknown. The current hypothesis suggests that the connector will play a passive role during packaging helping with ATPase regulation, and in some cases acting as a sensor to indicate the degree of capsid filling. The active pumping of the DNA will be carried out by the ATPase structural reorganization moving the DNA inside the capsid.

Together, these results show that genome packaging is a very complex mechanism essential for virus survival. Its importance have forced some viruses to incorporate sophisticated machineries that share structural and mechanistic properties with other protein complexes playing essential tasks during the metabolism of genetic material in living cells.

Acknowledgements We acknowledge Jaime Martin-Benito for his contribution to Fig. 12.1c. This work was supported by Grant BFU2011-29038 from the Spanish Ministry of Science and Innovation.

References and Further Reading

1. Hunter E (2007) Virus assembly. In: Knipe DN, Howley PM (eds) Fields virology, vol 1. Lippincott Williams, Philadelphia, pp 141–168
2. Bhyravbhatla B, Watowich SJ, Caspar DL (1998) Refined atomic model of the four-layer aggregate of the tobacco mosaic virus coat protein at 2.4-Å resolution. Biophys J 74:604–615
3. Sachse C, Chen JZ, Coureux PD, Stroupe ME, Fandrich M, Grigorieff N (2007) High-resolution electron microscopy of helical specimens: a fresh look at tobacco mosaic virus. J Mol Biol 371:812–835
4. Palese P, Shaw ML (2007) Orthomyxoviridae: the viruses and their replication. In: Knipe DM, Howley PM (eds) Fields virology, vol 2. Lippincott Williams and Wilkins, Philadelphia, pp 1647–1689
5. Schneemann A (2006) The structural and functional role of RNA in icosahedral virus assembly. Annu Rev Microbiol 60:51–67
6. Spencer SM, Sgro JY, Dryden KA, Baker TS, Nibert ML (1997) IRIS explorer software for radial-depth cueing reovirus particles and other macromolecular structures determined by cryoelectron microscopy and image reconstruction. J Struct Biol 120:11–21
7. Agirrezabala X, Martin-Benito J, Caston JR, Miranda R, Valpuesta JM, Carrascosa JL (2005) Maturation of phage T7 involves structural modification of both shell and inner core components. EMBO J 24:3820–3829
8. Lamb RA (2007) Mononegavirales. In: Knipe DM, Howley P (eds) Fields virology, vol 1. Lipincott Williams and Wilkins, Philadelphia, pp 1357–1361
9. Green TJ, Zhang X, Wertz GW, Luo M (2006) Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. Science 313:357–360

10. Coloma R, Valpuesta JM, Arranz R, Carrascosa JL, Ortin J, Martin-Benito J (2009) The structure of a biologically active influenza virus ribonucleoprotein complex. PLoS Pathog 5: e1000491
11. Bunka DH, Lane SW, Lane CL, Dykeman EC, Ford RJ, Barker AM, Twarock R, Phillips SE, Stockley PG (2011) Degenerate RNA packaging signals in the genome of satellite tobacco necrosis virus: implications for the assembly of a T=1 capsid. J Mol Biol 413:51–65
12. Dykeman EC, Grayson NE, Toropova K, Ranson NA, Stockley PG, Twarock R (2011) Simple rules for efficient assembly predict the layout of a packaged viral RNA. J Mol Biol 408:399–407
13. Prasad BV, Prevelige PE Jr (2003) Viral genome organization. Adv Protein Chem 64:219–258
14. Coombs KM (2006) Reovirus structure and morphogenesis. Curr Top Microbiol Immunol 309:117–167
15. Reinisch KM, Nibert ML, Harrison SC (2000) Structure of the reovirus core at 3.6 a resolution. Nature 404:960–967
16. Caston JR, Ghabrial SA, Jiang D, Rivas G, Alfonso C, Roca R, Luque D, Carrascosa JL (2003) Three-dimensional structure of penicillium chrysogenum virus: a double-stranded RNA virus with a genuine T=1 capsid. J Mol Biol 331:417–431
17. Sidhu SS (2001) Engineering M13 for phage display. Biomol Eng 18:57–63
18. Liu DJ, Day LA (1994) Pf1 Virus structure: helical coat protein and DNA with paraxial phosphates. Science 265:671–674
19. Chapman MS, Rossmann MG (1995) Single-stranded DNA-protein interactions in canine parvovirus. Structure 3:151–162
20. Johnson JE, Chiu W (2007) DNA packaging and delivery machines in tailed bacteriophages. Curr Opin Struct Biol 17:237–243
21. Cerritelli ME, Cheng N, Rosenberg AH, McPherson CE, Booy FP, Steven AC (1997) Encapsidated conformation of bacteriophage T7 DNA. Cell 91:271–280
22. Petrov AS, Boz MB, Harvey SC (2007) The conformation of double-stranded DNA inside bacteriophages depends on capsid size and shape. J Struct Biol 160:241–248
23. Comolli LR, Spakowitz AJ, Siegerist CE, Jardine PJ, Grimes S, Anderson DL, Bustamante C, Downing KH (2008) Three-dimensional architecture of the bacteriophage phi29 packaged genome and elucidation of its packaging process. Virology 371:267–277
24. Fang PA, Wright ET, Weintraub ST, Hakala K, Wu W, Serwer P, Jiang W (2008) Visualization of bacteriophage T3 capsids with DNA incompletely packaged *in vivo*. J Mol Biol 384:1384–1399
25. Poranen MM, Bamford DH (2012) Assembly of large icosahedral double-stranded RNA viruses. Adv Exp Med Biol 726:379–402
26. Huiskonen JT, de Haas F, Bubeck D, Bamford DH, Fuller SD, Butcher SJ (2006) Structure of the bacteriophage phi6 nucleocapsid suggests a mechanism for sequential RNA packaging. Structure 14:1039–1048
27. Kainov DE, Tuma R, Mancini EJ (2006) Hexameric molecular motors: P4 packaging ATPase unravels the mechanism. Cell Mol Life Sci 63:1095–1105
28. Mindich L (2012) Packaging in dsRNA viruses. Adv Exp Med Biol 726:601–608
29. Ionel A, Velazquez-Muriel JA, Luque D, Cuervo A, Caston JR, Valpuesta JM, Martin-Benito J, Carrascosa JL (2011) Molecular rearrangements involved in the capsid shell maturation of bacteriophage T7. J Biol Chem 286:234–242
30. Valpuesta JM, Carrascosa JL (1994) Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. Quart Rev Biophys 27:107–155
31. Steven AC, Heymann JB, Cheng N, Trus BL, Conway JF (2005) Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. Curr Opin Struct Biol 15:227–236
32. Gertsman I, Gan L, Guttman M, Lee K, Speir JA, Duda RL, Hendrix RW, Komives EA, Johnson JE (2009) An unexpected twist in viral capsid maturation. Nature 458:646–650

33. Lawton JA, Estes MK, Prasad BV (2000) Mechanism of genome transcription in segmented dsRNA viruses. Adv Virus Res 55:185–229
34. Mancini EJ, Kainov DE, Grimes JM, Tuma R, Bamford DH, Stuart DI (2004) Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. Cell 118:743–755
35. McDonald SM, Patton JT (2011) Assortment and packaging of the segmented rotavirus genome. Trends Microbiol 19:136–144
36. Mancini EJ, Tuma R (2012) Mechanism of RNA packaging motor. Adv Exp Med Biol 726:609–629
37. Cuervo A, Carrascosa JL (2012) Bacteriophages: structure. In: eLS. Wiley, Chichester, pp 1–7
38. Rao VB, Feiss M (2008) The bacteriophage DNA packaging motor. Annu Rev Genet 42:647–681
39. Sun S, Rao VB, Rossmann MG (2010) Genome packaging in viruses. Curr Opin Struct Biol 20:114–120
40. Mettenleiter TC, Klupp BG, Granzow H (2006) Herpesvirus assembly: a tale of two membranes. Curr Opin Microbiol 9:423–429
41. Cuervo A, Carrascosa JL (2012) Viral connectors for DNA encapsulation. Curr Opin Biotechnol 23:529–536
42. Carrascosa JL, Valpuesta JM (1999) Bacteriophage connectors: structural features of a DNA translocating motors. In: recent research developments in virology. Transworld reseach network. Trivadrum 1:449–465
43. Feiss M, Rao VB (2012) The bacteriophage DNA packaging machine. Adv Exp Med Biol 726:489–509
44. Ding F, Lu C, Zhao W, Rajashankar KR, Anderson DL, Jardine PJ, Grimes S, Ke A (2011) Structure and assembly of the essential RNA ring component of a viral DNA packaging motor. Proc Natl Acad Sci U S A 108:7357–7362
45. Casjens SR (2011) The DNA-packaging nanomotor of tailed bacteriophages. Nat Rev Microbiol 9:647–657
46. Rao VB, Black LW (2010) Structure and assembly of bacteriophage T4 head. Virol J 7:356
47. Nemecek D, Lander GC, Johnson JE, Casjens SR, Thomas GJ Jr (2008) Assembly architecture and DNA binding of the bacteriophage P22 terminase small subunit. J Mol Biol 383:494–501
48. Buttner CR, Chechik M, Ortiz-Lombardia M, Smits C, Ebong IO, Chechik V, Jeschke G, Dykeman E, Benini S, Robinson CV, Alonso JC, Antson AA (2012) Structural basis for DNA recognition and loading into a viral packaging motor. Proc Natl Acad Sci U S A 109:811–816
49. Anderson D, Grimes S (2005) In: Catalano CE (ed) Viral Genome packaging machines: genetics, structure and mechanism. Kluwer Academic/Plenum Publishers, New York
50. Morais MC (2012) The dsDNA packaging motor in bacteriophage ϕ29. Adv Exp Med Biol 726:511–547
51. Guo P, Lee TJ (2007) Viral nanomotors for packaging of dsDNA and dsRNA. Mol Microbiol 64:886–903
52. Shu D, Zhang H, Jin J, Guo P (2007) Counting of six pRNAs of phi29 DNA-packaging motor with customized single-molecule dual-view system. EMBO J 26:527–537
53. Hingorani MM, O'Donnell M (1998) Toroidal proteins: running rings around DNA. Curr Biol 8:R83–R86
54. Massey TH, Mercogliano CP, Yates J, Sherratt DJ, Lowe J (2006) Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. Mol Cell 23:457–469
55. Gomis-Ruth FX, Moncalian G, Perez-Luque R, Gonzalez A, Cabezon E, de la Cruz F, Coll M (2001) The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. Nature 409:637–641
56. Chemla YR, Smith DE (2012) Single-molecule studies of viral DNA packaging. Adv Exp Med Biol 726:549–584
57. Williams RS, Williams GJ, Tainer JA (2008) A charged performance by gp17 in viral packaging. Cell 135:1169–1171

58. Sun S, Kondabagil K, Draper B, Alam TI, Bowman VD, Zhang Z, Hegde S, Fokine A, Rossmann MG, Rao VB (2008) The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. Cell 135:1251–1262
59. Tavares P, Zinn-Justin S, Orlova EV (2012) Genome gating in tailed bacteriophage capsids. Adv Exp Med Biol 726:585–600
60. Purohit PK, Inamdar MM, Grayson PD, Squires TM, Kondev J, Phillips R (2005) Forces during bacteriophage DNA packaging and ejection. Biophys J 88:851–866
61. Roos WH, Ivanovska IL, Evilevitch A, Wuite GJ (2007) Viral capsids: mechanical characteristics, genome packaging and delivery mechanisms. Cell Mol Life Sci 64:1484–1497
62. Vinga I, Sao-José C, Tavares P, Santos M (2006) Bacteriophage entry in the host cell. In: Wegrzyn G (ed) Modern bacteriophage biology and biotechnology. Research Signpost, Kerala, pp 165–205

Further Reading

Flint SJ, Enquist LW, Racaniello VR, Skalka AM (eds) (2009) Principles of virology, 3rd edn. ASM Press, Washington
Knipe DM, Howley PM (eds) (2007) Fields virology, 5th edn. Lippincott Williams & Wilkins, Philadelphia
Patton JT (ed) (2008) Segmented double-stranded RNA viruses. Structure and molecular biology. Caister Academic Press, Norfolk
Rossmann MG, Rao VB (eds) (2012) Viral Molecular Machines. Adv Exp Med Biol vol. 726, Springer, New York

Also especially recommended for further reading are references [5, 20, 27, 45] *listed above*.

Chapter 13
Virus Maturation

Laura R. Delgui and José F. Rodríguez

Abstract The formation of infectious virus particles is a highly complex process involving a series of sophisticated molecular events. In most cases, the assembly of virus structural elements results in the formation of immature virus particles unable to initiate a productive infection. Accordingly, for most viruses the final stage of the assembly pathway entails a set of structural transitions and/or biochemical modifications that transform inert precursor particles into fully infectious agents. In this chapter, we review the most relevant maturation mechanisms involved in the generation of infectious virions for a wide variety of viruses.

Keywords Capsid • Capsomer • Envelope • Glycoprotein • Occlusion body • Polyhedra • Polyprotein • Procapsid • Protease • Scaffold • Structural rearrangement • Virion • Virus egress • Virus entry • Virus maturation • Virus morphogenesis.

Abbreviations

3D	Three dimensional.
AcMNPV	*Autographa californica* nucleopolyhedrovirus
CA	Capsid protein of HIV
DNA	Deoxyribonucleic acid
ER	Endoplasmic reticulum

L.R. Delgui
Laboratorio de Biología Celular y Molecular, Instituto de Histología y Embriología Mendoza (IHEM), Facultad de Ciencias Médicas, Instituto de Ciencias Básicas, Universidad Nacional de Cuyo, Mendoza, Argentina

J.F. Rodríguez (✉)
Department of Molecular and Cell Biology, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: jfrodrig@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_13,

ESCRT	Endosomal sorting complexes required for transport
gp	Glycoprotein
HA	Hemagglutinin
HB-sAg	HBV surface antigen
HBV	Hepatitis B virus
HIV-1	Human immunodeficiency virus type 1
HSV-1	Herpes simplex virus type 1
IAV	Influenza A virus
kbp	Kilobase pairs
MA	Matrix protein of HIV
NC	Nucleocapsid protein of HIV
NωV	*Nudaurelia capensis* ω virus
ORF	Open reading frame
PR	Protease of HIV
RNA	Ribonucleic acid
T	Triangulation number
VLP	Virus-like particle
VP	Virus protein
WNPV	*Wiseana* nucleopolyhedrovirus

13.1 Introduction

The requirement of a maturation step during virus morphogenesis is directly related with the building strategy used by most viruses. As for many other large multi-subunit complexes, the assembly of virus particles is initiated by the establishment of weak interactions between structural subunits. This is critical for the correct positioning of the particle building blocks and the prevention of the formation of aberrant structures during the assembly process [39] (see Chaps. 10 and 11, 19). As a result of this initial step, labile immature particles, generally known as procapsids or provirions, are formed (Chaps. 10 and 11).

Due to their intrinsic fragile nature, procapsids are unable to withstand neither the internal pressure that may be exerted in those cases where the viral genome is tightly packaged (see Chap. 12) nor the external physical challenges inherent to the virus life-cycle that, in most cases, involve: (i) the egress of the particles from the infected cell to the extracellular milieu; (ii) survival in a highly hostile external environment; and (iii) interaction(s) with cellular receptors and the subsequent entry into a new host cell (Chaps. 15–17). The maturation process, generally triggered by one or more proteases present within the procapsid, involves a solid-state phase transition that, depending on the virus model, varies from subtle to major conformational rearrangements. This process takes place in a highly

ordered stepwise manner designed to consolidate the virion structure, and thus allowing it to achieve the structural properties required to be fully infectious [42].

The maturation of enveloped viruses generally requires a further step. To become infectious, glycoprotein envelope components involved in cell receptor interactions must undergo complex glycosylation pathways for their egress from the host cell. Additionally, these viruses depend upon proteolytic processing of their glycoprotein components for the activation of the membrane fusion events required for the internalization of the nucleocapsid to cell cytoplasm [43] (see Chap. 16).

Finally, some viruses secure their extracellular endurance by including a fraction of the newly formed virus progeny into discrete, highly resistant virus-derived proteinaceous superstructures. These superstructures, generally known as polyhedra or occlusion bodies, play a major role in the virus life-cycle providing a long-term stability that allows occluded virions to survive harsh environmental conditions for prolonged periods of time [8].

13.2 Immature Virus Particles and Maturation Strategies

Mechanisms directing the assembly of immature particles largely depend upon the structural complexity of the virus particle. Thus, whilst procapsids from simple viruses such as parvoviruses (see Chap. 10), nodaviruses or tetraviruses are capable of self-assembling without a direct contribution of auxiliary elements, the assembly of more complex viruses requires the participation of virus-derived structural components, *i.e*. the virus genome (see Chap. 12) and/or scaffolding polypeptides (see Chap. 11), different from those found as integral constituents in the mature capsid. As described above, maturation is a transition process that transforms inert immature particles into fully infectious virions. Indeed, regardless of their structural complexity, maturation processes have been documented in all virus models for which the assembly pathway has been characterized in depth.

The maturation of procapsids from simple viruses generally involves the autocatalytic processing of capsid polypeptides. This increases the stability of the particle and confers full virus infectivity. The presence of scaffolding proteins in procapsids from more complex viruses is essential to ensure the fidelity of the assembly process, and thus avoid the generation of aberrant dead-end structures. A comprehensive review about the role and structure of scaffolding polypeptides has been recently published [24]. Scaffolding polypeptides can be categorized in two main groups: (i) icosahedrally ordered, external scaffolds; and (ii) internal, core-like, scaffolds. The highly complex prolate cores found in T4-like bacteriophages also playing a critical scaffolding role and not fitting the categories described above have been described in detail in Chap. 11. The function of scaffolding polypeptides is intrinsically transient. Hence, during maturation they are either expelled out from the procapsid or displaced to a different topological location within the structure of the mature virion. Although the maturation of some viruses does not involve the

cleavage of scaffolding elements, in most cases it requires the activation of one or more well-coordinated sets of proteolytic processing events. This brings about the alteration of both particle size, either expansion or contraction, and of the shape of the original procapsid structure. Finally, maturation enveloped viruses involves an additional step consisting on the proteolytic cleavage of glycosylated envelope polypeptides responsible for the interaction with specific cell receptors, and the subsequent fusion of the virus envelope and the host cell membrane.

Following there is a description of well-characterized maturation processes corresponding to three virus groups with increasing levels of complexity.

13.3 Tetravirus Capsid Maturation

Tetraviruses are small positive-stranded RNA viruses that infect different insect species belonging to the *Lepidoptera* order [16]. Tetravirus virions are nonenveloped icosahedrons with a T = 4 symmetry enclosing a bipartite positive-stranded RNA genome. The virus particle is built by 240 copies of the capsid polypeptide (70-kDa) known as the α protein. During particle assembly, the α protein undergoes an autocatalytic processing event, directly associated to particle maturation, that generates the β (62-kDa) and the γ (8-kDa) polypeptides [1, 31].

Characterization of the tetravirus maturation process has mainly used the *Nudaurelia capensis* ω virus (NωV) as a model. The NωV capsid assembly and maturation process has been finely dissected using recombinant baculovirus-based expression systems. Expression of the NωV coat polypeptide in insect cells from recombinant baculoviruses leads to the assembly of virus-like particles (VLP) that, when exposed to an acidic environment, mature to form particles structurally identical to authentic virions [44].

The NωV procapsid to capsid transition involves a major, pH-induced, conformational change involving large-scale movements that drastically reduce the particle diameter (>15 %), and modify its external appearance. VLPs purified at near-neutral pH (pH = 7.6) have a diameter of 485 Å, a highly spherical appearance and a conspicuously porous surface (Fig. 13.1). These VLPs are thought to faithfully mimic the procapsid transient structure produced during NωV assembly in infected cells, a process that takes place in an acidic cellular compartment. The capsid protein subunits that form baculovirus-derived procapsids remains in its original uncleaved α state. Incubation of procapsids in acidic solutions (pH = 5) causes a major structural transition. The resulting particles have a diameter of 395 Å, a well-defined icosahedral aspect, and a smooth surface (Fig. 13.1). Although the structural transition from procapsid to capsid is very fast, taking place within the first 100 milliseconds of incubation at low pH, the self-processing of the coat protein is a much slower process that requires several hours to be completed. Interestingly, the procapsid to capsid transition is reversible, and capsids can be re-expanded to the procapsid stage by raising the pH. The transition becomes irreversible only when ca. 15 % of the

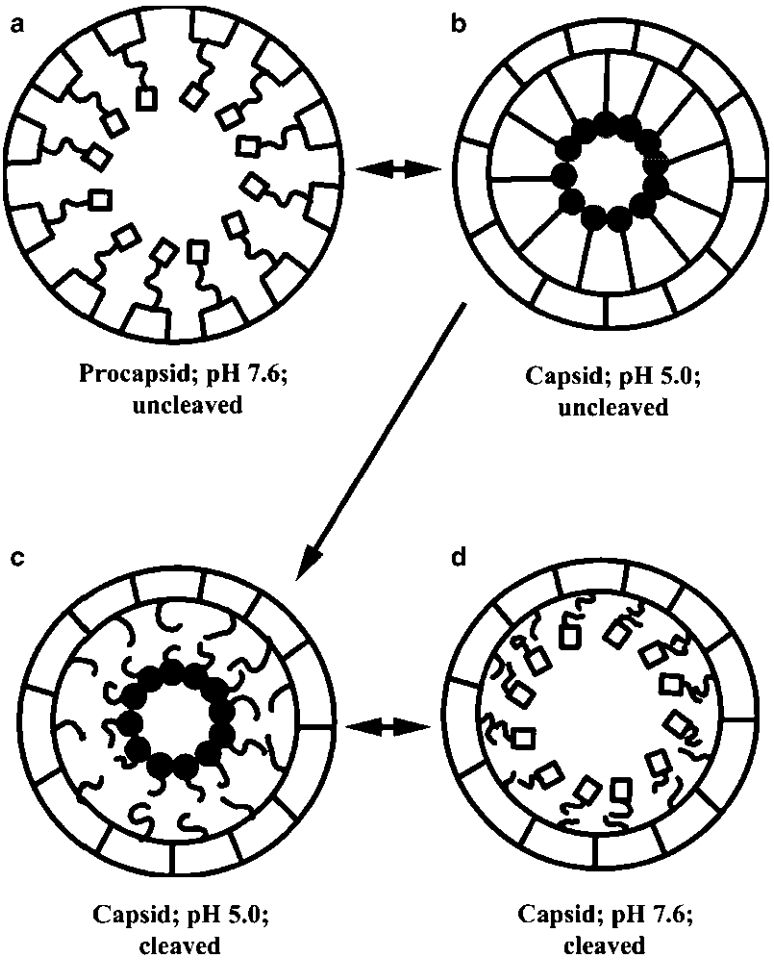

Fig. 13.1 Tetravirus capsid maturation. Schematic representation of the relationship of coat protein cleavage and the pH-induced conformational change in NωV. The region of the helical domain of each coat protein subunit believed to be responsible for the helix-coil transition represents residues 1–44 and 571–644 of the 644-residue protein. The diagram depicts this region as either α-helix (*opened squares*) or a random coil (*black circles*). (**a**) The particles were purified as procapsid at pH 7.6, with no cleavage occurring in the subunits. (**b**) Upon lowering the pH to 5.0, protonation of specific residues causes the helix to transform into a coil, initiating the quaternary structural rearrangement. If no cleavage occurs in the capsid at pH 5.0, the process is reversible. (**c**) Cleavage of wild-type particles locks the particles in the capsid conformation. (**d**) Altering the pH of the particles allows the helix-coil to operate reversibly, but cleavage uncouples this pH-driven engine from the rest of the cargo subunit, causing the surface of the particles to resemble capsids (Reproduced from [44]. With permission)

procapsid α coat protein subunits have been proteolytically cleaved and converted into the mature capsid β protein form.

Expression of mutant versions of the α coat protein lacking the self-cleavage site leads to the assembly of procapsids structurally identical to those obtained by expression of the wild-type α polypeptide. These mutant procapsids also undergo the procapsid to capsid transition when incubated at low pH. However, the lack of the scissile bond for the α protein autoproteolytic processing completely blocks the consolidation of the mature capsid stage [28].

Comparisons of procapsid and capsid secondary structures using Fourier transform infrared spectroscopy revealed that the α-helical content of the procapsid is higher than that of the capsid, thus indicating that local refolding of internal helical regions of the α capsid protein to coil is critical for the maturation process. Two alternative models have been proposed to explain the maturation process. The first one assumes that a helix to coil transition of γ-peptide region of the immature α capsid proteins found in the procapsid constitutes the driving force for the capsid rearrangement [44]. In this scenario, the autoproteolytic cleavage of the α protein would switch off the driving force for conformational change from the rest of the capsid, thus making the transition irreversible once cleavage has occurred. An alternative model, based on information gathered from comparisons of the crystal structure of both procapsids and capsids, suggests that charge repulsion amongst clusters of acidic amino acid residues found at subunit interfaces are responsible for particle re-expansion when incubated above their isoelectric point [19].

A very important aspect of the NωV capsid maturation process is the observation that although mature capsid particles are more stable to pH and ionic conditions as well as more resistant to proteolysis than procapsids, they are far more sensitive than the latter to thermal stress. This suggests that the end product of the maturation process is the formation of a metastable virion capable of safely transporting the virus genome to its target cell but designed to easily release its genetic cargo upon the correct physical stimuli, *i.e.* the interaction with its cognate receptor(s).

13.4 Herpesvirus Nucleocapsid Maturation

Members of the *Herpesviridae* family are highly complex double-stranded DNA viruses infecting a wide variety of animal species [11]. Viruses from the three herpesvirus subfamilies - *Alpha*-, *Beta*-, and *Gammaherpesvirinae*- exhibit significant genome conservation, and a common virion morphology. The innermost virion structure is the icosahedral capsid that contains the viral DNA and is surrounded by the tegument, a proteinaceous matrix lacking a defined structure, and by a lipidic envelope.

The herpesvirus nucleocapsid assembly process, that strongly resembles that of double-stranded DNA bacteriophages [35], has been characterized in detail using the human herpes simplex virus type 1 (HSV-1), prototype of the *Simplexvirus* genus of the *Alphaherpesvirinae* family, as virus model. The mature HSV-1 capsid

is an unusually large and complex structure of 125 nm in diameter, with a T = 16 icosahedral surface lattice. The capsid is built by two types of subunits: (i) hexamers and pentamers formed by VP5, the major capsid polypeptide; and (ii) triplexes, located between and connecting pentons and hexons, formed by a heterotrimer built by one VP19c molecule and a two copies of the VP23 polypeptide (Fig. 13.2a). One of the 12 particle vertices is occupied by the portal complex, a dodecameric structure that forms a channel that is essential for the packaging, and probably the release, of the viral DNA genome [4].

HSV-1 capsid assembly takes place in the nucleus of infected cells and is initiated by formation of a spherical procapsid containing two concentric protein layers: the procapsid shell and an underlying scaffold. In addition to proteins found in the mature capsid, procapsid assembly involves the participation of two additional virus-encoded polypeptides, pre-VP22a, the scaffolding polypeptide, and pre-VP21, the precursor of the protease, responsible for procapsid maturation. Although, as shown by experiments carried out using baculovirus-based procapsid artificial assembly systems [46], the presence of the portal complex is not essential for procapsid assembly. However, under physiological conditions it appears to play an important role for the initiation of the assembly process. Indeed, the portal complex is absolutely indispensable for the incorporation of the 152-kbp virus genome into the nascent procapsid, and for the subsequent cleavage and extrusion of the scaffolding polypeptide under native conditions.

The procapsid to capsid maturation process involves a massive, cooperative and irreversible rearrangement of the capsid shell. This process is promoted by the activation of the protease that releases the interaction between the scaffold and the shell layers, thus triggering a series of conformational changes, mainly relative rotations, affecting the major capsid protein VP5 forming the capsomers.

It has been shown that procasids produced by a protease-defective temperature-sensitive HSV-1 mutant can be purified from infected cells grown at non-permissive temperature. These procapsids are able to slowly mature when incubated at permissive temperature, thus allowing a close scrutiny of the maturation process using cryo-electron microscopy and three-dimensional (3D) image reconstruction [14, 21] (see Chap. 3). Comparison of the 3D images allowed assessing the existence of 17 clearly distinguishable intermediate maturation stages. In the procapsid, the neighboring capsomers (hexamers and pentamers) are held together mainly by interactions with the surrounding triplexes that exhibit a clear morphogenetic role. Upon maturation, triplexes are transformed into molecular clamps stabilizing the interaction of their three surrounding capsomers that come together at the inner base of the capsid layer. Indeed, the most conspicuous change associated to maturation is the formation of a floor, a continuous network of VP5 interactions, at the base of the capsid layer that reinforces particle stability. To form the floor connections, the VP5 base domains must rotate ca. 40° with respect to their external domains. A second major rotation affects the protrusion domains of the VP5 molecules found at hexons. In addition to their increased stability, at its

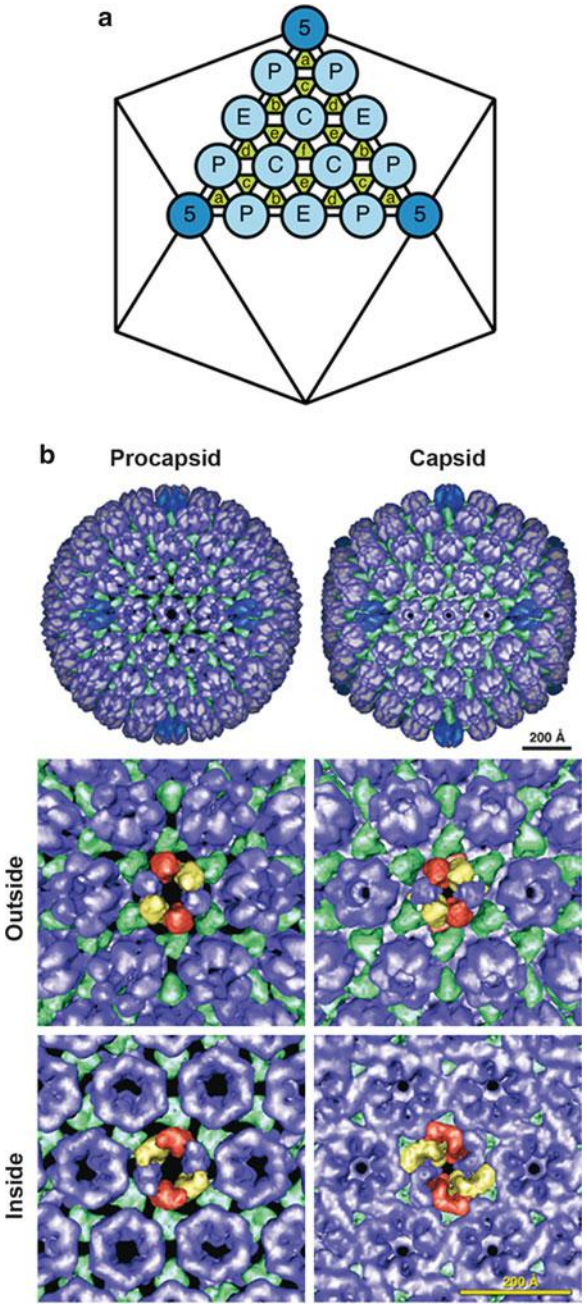

Fig. 13.2 Herpes simplex virus capsid maturation. (**a**) Herpes virus surface lattice. Diagram showing the positions of the quasi-equivalent capsomers and triplexes. The peripentagonal, edge and central hexons and pentons are denoted by P, E, C and 5, respectively. The six triplexes, Ta-Tf, are denoted by a–f (Reproduced with permission from [21]). (**b**) Molecular anatomy of the procapsid and the mature capsid. VP5 hexons are colored in *light blue*, pentons in *dark blue*, and triplexes in *green*. Particles are viewed along a two-fold symmetry axis. (**c**) Images show enlarged areas centered on an E hexon at a two-fold symmetry axis for the procapsid and the mature capsid and for the outer and inner surfaces, respectively. The three quasi-equivalent pair of VP5 subunits in the E hexon are *yellow*, *red* and *blue*, respectively. Other VP5 subunits are *blue* (Reproduced from [21]. With permission)

final maturation stage capsids have a more compact aspect, lacking intercapsomer gaps, and show an icosahedral contour that differentiates them from the original more globular procapsids (Fig. 13.2b).

13.5 Maturation of the Human Immunodeficiency Virus Capsid

The human immunodeficiency virus type 1 (HIV-1) is the best-characterized member of the Retrovirus family, a diverse group of enveloped viruses with a positive-stranded RNA genome. Retroviruses share a similar virion structure enclosing a dimeric genome, and some general replicative properties, *i.e.* the reverse transcription and the ability to integrate their genomes into the DNA of their host cells [45]. The highly complex HIV-1 assembly and maturation processes have been recently reviewed in depth [3].

The HIV-1 particle has a globular appearance with a mean diameter of about 120 nm. The envelope contains about 70 copies of the protruding Env protein complex which is built by a trimer of the glycoprotein gp120 bound to a trimer of the transmembrane gp41 polypeptide that forms a stem-like structure and anchors the complex to the viral envelope [5, 6]. The inner region of the viral particle is formed by proteins derived from the group-specific antigen (Gag) polypeptide, the structural polyprotein precursor. The geometric structure of the mature HIV-1 capsid is a fullerene cone, a conical hexagonal net closed at both ends through the introduction of 12 pentagonal defects [13]. This structure, built by ca. 1.500 molecules of the capsid protein (CA) in the form of hexameric and pentameric rings, encloses the molecular replicative machine, a ribonucleoprotein complex formed by nucleocapsid protein (NC) tightly bound to the virus genome and associated to the reverse transcriptase, the integrase, the protease [7] and the accessory protein Vpr. The core is surrounded by a discontinuous layer formed by molecules of matrix protein (MA).

The Gag protein precursor encompasses the above-mentioned proteins MA, CA and NC as three major structural components: MA, the membrane binding domain; CA, the capsid domain; and NC, the nucleocapsid domain, which interacts with and recruits the viral RNA. CA and NC are separated by a linker peptide, SP1, and downstream of NC are two further peptide domains, SP2 and p6 (Fig. 13.3a, b). p6 is responsible for the recruitment of the cellular endosomal sorting complexes required for transport (ESCRT) elements. The assembly of the spherical HIV-1 procapsid is exclusively driven by Gag precursor proteins, Gag (Pr55) and Gag–Pro-Pol (Pr160), targeted to the plasma membrane by myristoylation of their N-terminal domains. Accumulation of Gag precursors promotes the assembly of incomplete spheres underneath specialized host cell membrane microdomains containing virus-encoded glycoproteins. The assembly of these immature structures prompts the recruitment of components of the ESCRT machinery that drive

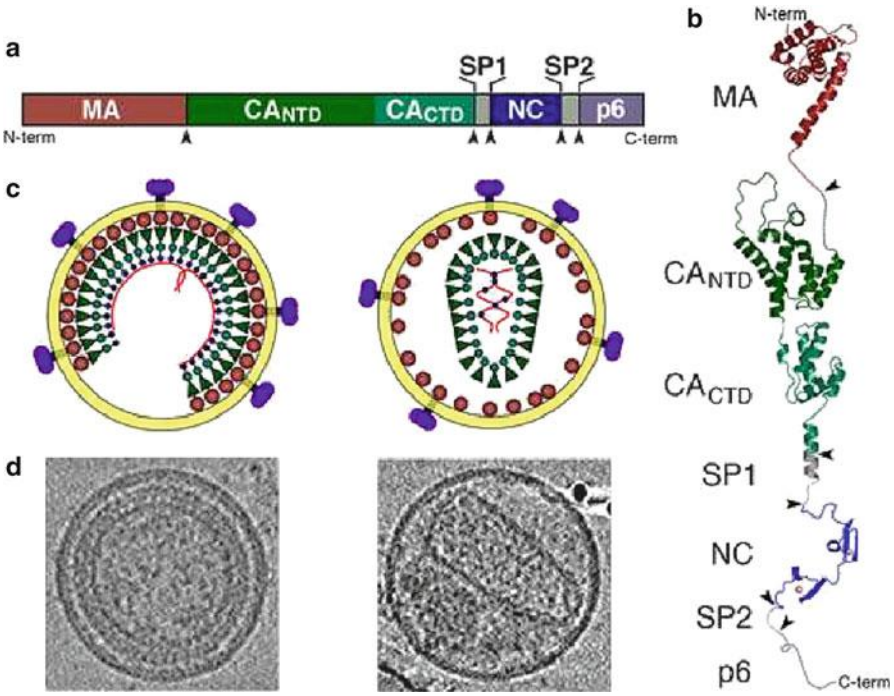

Fig. 13.3 HIV-1 particle maturation. (**a**) HIV-1 Gag polyprotein domain structure, showing the locations of MA, CA (N-terminal domain CANTD and C-terminal domain CACTD), SP1, NC, SP2, and p6. (**b**) Structural model of the extended Gag polypeptide, derived from high-resolution structures and models of isolated domains. *Dashed lines* represent unstructured and linker regions. PR cleavage sites are indicated by the *arrowheads* in (**a**) and (**b**). (**c**) Schematic models of the immature (*left*) and mature (*right*) HIV-1 virions. (**d**) Central slices through cryo-electron microscopy tomograms of immature (*left*) and mature (*right*) HIV-1 particles. The spherical virions are approximately 130 nm in diameter (Reproduced from [13]. With permission)

the scission of the membrane and the release of budding non-infectious immature virus particles from the cell [13] (see also Chap. 14). The Gag molecules in the immature particle are extended and oriented radially, with their N-terminal MA domains associated with the viral membrane and their C-terminal p6 domains facing the interior of the particle (Fig. 13.3c, d). The immature capsid lattice is stabilized primarily by lateral protein-protein interactions.

As it is the case for the maturation of procapsids of many icosahedral viruses, maturation of the HIV-1 procapsid involves a series of dramatic morphological changes and is directly associated to the activation of the protease (PR) activity that takes place during or immediately after budding. Although the PR polypeptide has been extensively characterized, its activation mechanism is not completely understood. In order to be fully active, PR has to dimerize and the activity in the Gag-Pro-Pol precursor is negligible. It is thought that Gag trafficking regulates PR activation by preventing premature PR dimerization until the Gag molecules coalesce at the

plasma membrane, and that an as yet unknown mechanism(s), other than the assembly-mediated PR dimerization, must play an important role in this process. The activated PR cleaves Gag in five positions, thus releasing the MA, CA, NC, and p6 proteins as well as the SP1 and SP2 peptides. Indeed proteolytic maturation is essential for infectivity, and PR inhibitors are of outmost importance in current antiretroviral therapies.

The five proteolytic cleavage sites in Gag are cleaved at very different rates *in vitro*. The fastest cleavage is that between SP1 and NC taking place 400-fold faster than that between CA and SP1 [34]. Additionally, the detection of reproducible processing intermediates in lysates from infected cells or upon partial inhibition of PR, suggests that cleavage is a stepwise process. The order of cleavage it is thought to involve the initial processing separating NC-p6 from the membrane-bound N-terminal part of Gag. The secondary cleavages would separate MA from CA-SP1 and p6 from NC-SP2, and final processing would release the two spacer peptides from the C-termini of CA and NC, respectively. After proteolytic cleavage, the MA layer is thought to remain associated with the viral membrane, whereas NC and the RNA are condensed into the ribonucleoprotein complex layer surrounded by the viral CA capsid core.

13.6 The Role of Glycosylation in Virus Maturation

One very important structural feature of many viruses is that their surface is covered with glycoproteins, and glycosylation is vital for virus replication and infectivity. Proper protein folding, attachment to host-cell receptors and evasion of host-immune responses are events where glycosylation is involved in some viral systems. In recent years, glycosylation has become an additional means of therapy by interfering with viral host cell entry or egress or by preventing the correct assembly of virions [30].

13.6.1 Hepadnavirus Glycosylation: Its Importance in Virus Assembly and Egress

Hepatitis B virus (HBV), a member of the *Hepadnaviridae* family, is a human pathogen causing acute and chronic liver disease that eventually leads to liver cirrhosis and hepatocellular carcinoma [12]. HBV particles are double-shelled spheres with a DNA-containing inner nucleocapsid and an outer envelope composed of cellular lipids and three structurally related virus-encoded proteins termed small (S), middle (M) and large (L) proteins, which are together referred as to as HBV surface antigen (HB-sAg). Virion assembly is initiated by insertion of the envelope proteins into the endoplasmic reticulum (ER) membrane proceeding at

pre-Golgi membranes, where cytosolic nucleocapsids are packaged by transmembrane envelope proteins. Virions then bud into intraluminal cisternae and leave the cell *via* the constitutive secretory pathway [23]. L, M and S proteins are derived from a single open reading frame (ORF) by employing three different translation start sites that divide the ORF into three domains: the N-terminal pre-S1 domain; the middle pre-S2 domain; and the C-terminal S domain common to S, M and L proteins (Fig. 13.4a) [29]. All three proteins possess a partially utilized *N*-glycosylation site at Asn-146 of the S domain and are also thought to have similar three-dimensional structures. The major difference between S and M proteins is the presence of 55 additional amino acid residues corresponding to the pre-S2 domain on M, and importantly the additional glycan site within this region at Asn-4. Although the L protein also contains the pre-S2 domain, this glycosylation site is not utilized (Fig. 13.4b) [18]. Under normal circumstances, after synthesis, HBsAg leaves the ER, passes through the Golgi stacks and is secreted within 3 h. However, when infected cells are treated with glucosidase inhibitors, enzymes that mediate the first steps in the trimming of terminal sugars pathway, the HBsAg molecules are detained and displaced to the Golgi, and even returned to the ER [25, 26]. On the other hand, detailed studies based on the use of site-directed mutagenesis revealed that the common glycan in L, M and S proteins does not play a role in virus secretion. Instead, the removal of the pre-S2 glycan site prevents the secretion of enveloped virus, rendering the M glycosylation at the ER as a crucial event in the formation of the HB viral particle [29].

For glycoproteins, the processing of the initial oligosaccharide precursor from the $Glc_3Man_9GlcNAc_2$ to the $Glc_1Man_9GlcNAc_2$ glycoform in the ER can lead to an interaction with chaperones such as calnexin. Calnexin binds only to glycoproteins containing $Glc_1Man_9GlcNAc_2$ structures to assist in their folding and anchors the polypeptides to the ER until they have achieved their correct folding conformation [17]. The HBV envelope M glycoprotein associates to calnexin. This interaction strictly depends on the glycan at Asparagine (Asn)-4, specific for M, while the common Asn-146-linked glycan is not involved [48]. It is hypothesized that the M protein acts as a dominant negative scaffolding glycoprotein, and that its misfolding destabilizes the viral envelope, thus hindering viral particle secretion.

A second glycosylation modification also restricted to the M protein has been described. It has been shown that the M proteins carry, at least in part, a single *O*-linked carbohydrate substituent, which could be identified as GalNAcα-, Gal(β1-3)GalNAcα- or Neu5Ac(α2-3)Gal(β1-3)GalNAcα-unit with the threonine in position 37 (Thr-37) of pre-S2 having the highest potential to be a O-glycosylated site [38]. Glycoproteins with O-linked glycans have been found in a number of enveloped viruses, although the distinct functions of viral *O*-glycoproteins remain obscure. For pre-S2 *O*-glycan, a masking of the respective peptide sequence has been suggested. Supporting this hypothesis mouse monoclonal antibodies directed against epitopes involving Thr-37 or neighboring amino acid residues have never been identified. In contrast, antibodies recognizing an epitope encompassing the *N*-glycan linked to Asn-4 are readily available [38].

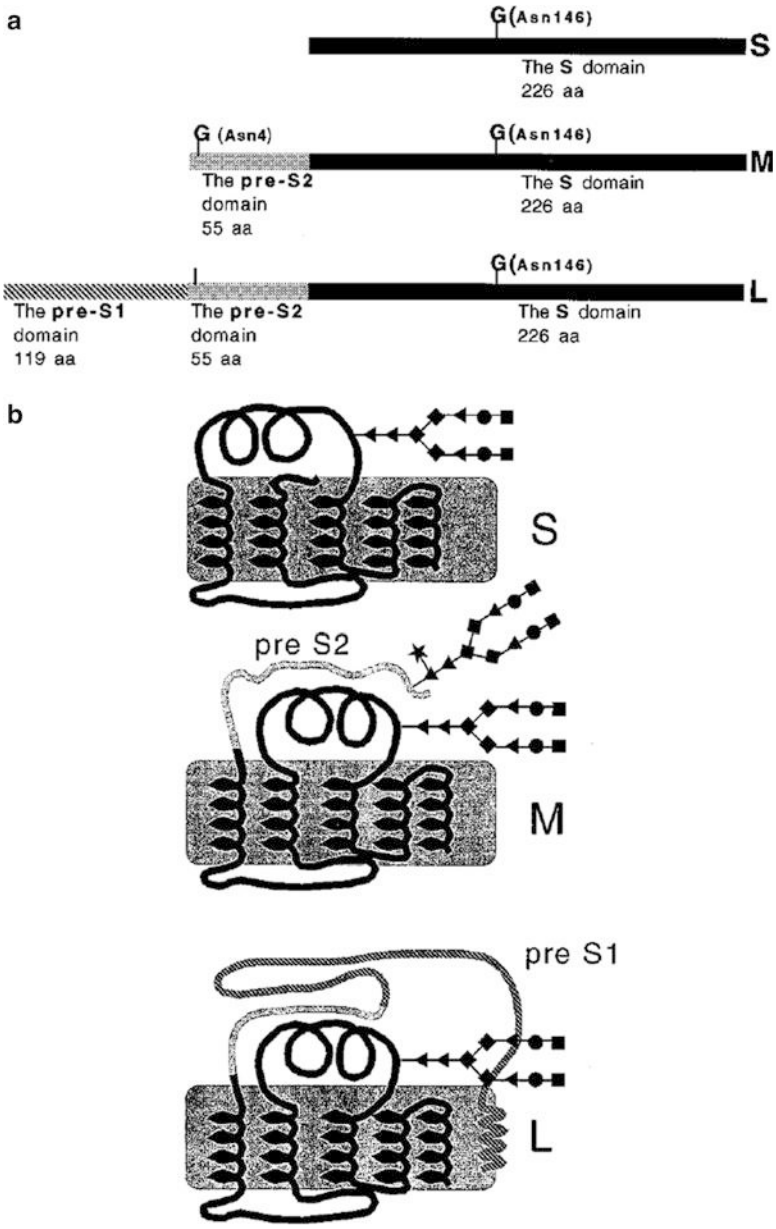

Fig. 13.4 Maturation of hepatitis B virus (HBV) envelope glycoproteins. (**a**) The HBV envelope proteins. All three proteins have a common N-linked glycosylation site at position 146 of the S domain (*marked with a G*). The M protein contains an additional glycan site at amino acid 4 of the pre-S2 domain (*marked with a G*). The L protein, while containing the pre-S2 glycosylation sequon, only utilizes the shared S glycan site (1). (**b**) Structural diagram of the three HBV envelope proteins with attached N-linked glycans. The pre-S1 and pre-S2 domains are indicated. The glycan structures are as follows: mannose (*diamonds*); N-acetylglucosamine (*triangles*); galactose (*circles*); sialic acid (*squares*); fucose (*stars*) (Reproduced from [29]. With permission)

13.6.2 Glycosylation in Orthomyxovirus Replication

Influenza A virus (IAV) is the only species of the Influenza virus A genus belonging to the *Orthomyxoviridae* family that causes annual epidemics and recurring pandemics with potentially severe consequences for public health and global economy [40]. The IAV hemagglutinin (HA) is the virion surface glycoprotein that attaches the virus to glycan receptors on host cell and mediates the fusion of the viral envelope with the membranes of endocytic vesicles to initiate the infectious process (Chap. 16). Moreover, HA is the virion component that stimulates the generation of protective antibodies. All of these important functions of HA are mediated by *N*-linked glycosylation involving attachment of complex glycans to an asparagine residue in a consensus sequence [25].

HA is a homotrimeric integral membrane protein where the monomers are synthesized on membrane-bound ribosomes from a precursor that is then glycosylated and cleaved into two smaller polypeptides: the HA_1 and HA_2 subunits. Each monomer has an ectodomain consisting of a globular head, which harbors the glycan receptor-binding site (HA_1) and a stem region that anchors the protein to the membrane (HA_2) [41]. The IAV HA contains 3–9 N-linked glycosylation sites per subunit, depending on the virus strain (Fig. 13.5). Amino acids sequence analysis has revealed that there is considerable variation in both the number and location of potential glycosylation sites among different HA subtypes and even among variants from a single subtype which is believed to be involved in the evolution of influenza viruses [51]. Variation in protein glycosylation is a more efficient mechanism than the direct mutation of amino acids for the virus to escape the surveillance of the host immune system. This is due to the fact that the glycans are host-derived and hence considered as "self" by the immune system [47]. However, highly conserved glycosylation sites at Asn-12 and Asn-478 (the numbering corresponds to H7) and a further semiconserved site at Asn-28 is [32], the three of them located within the stem region of the HA molecule (Fig. 13.5), have been described to be related to protein folding and maturation [15, 50].

Cleavage of precursor HA by endoproteases occurs late during the transport in the trans-Golgi network. Based on detailed studies employing site-specific mutagenesis on the influenza virus strain A/FPV/Rostock/34 conserved-glycosylation sites of HA and cleavage rates of HA mutants, Roberts and collaborators examined their role in protein transport [36]. It was found that Asn-28 oligosaccharide plays a dominant role in promoting trimerization and proper HA transport since the loss of the carbohydrate at this site interferes with the rapid formation of a transport-competent form of HA early after synthesis. However, only the loss of all three conserved glycosylation sites results in the accumulation of this HA protein at the ER. A highly detailed model for the folding of HA in its natural environment has been proposed indicating that *N*-linked glycans direct the molecular choreography for a ribosome-bound nascent chain as it emerges mediating interactions with chaperones and a foldase, ERp57 [10]. Studies reflecting the impact of each HA glycosylation site on production of the mature infective virus progeny are still

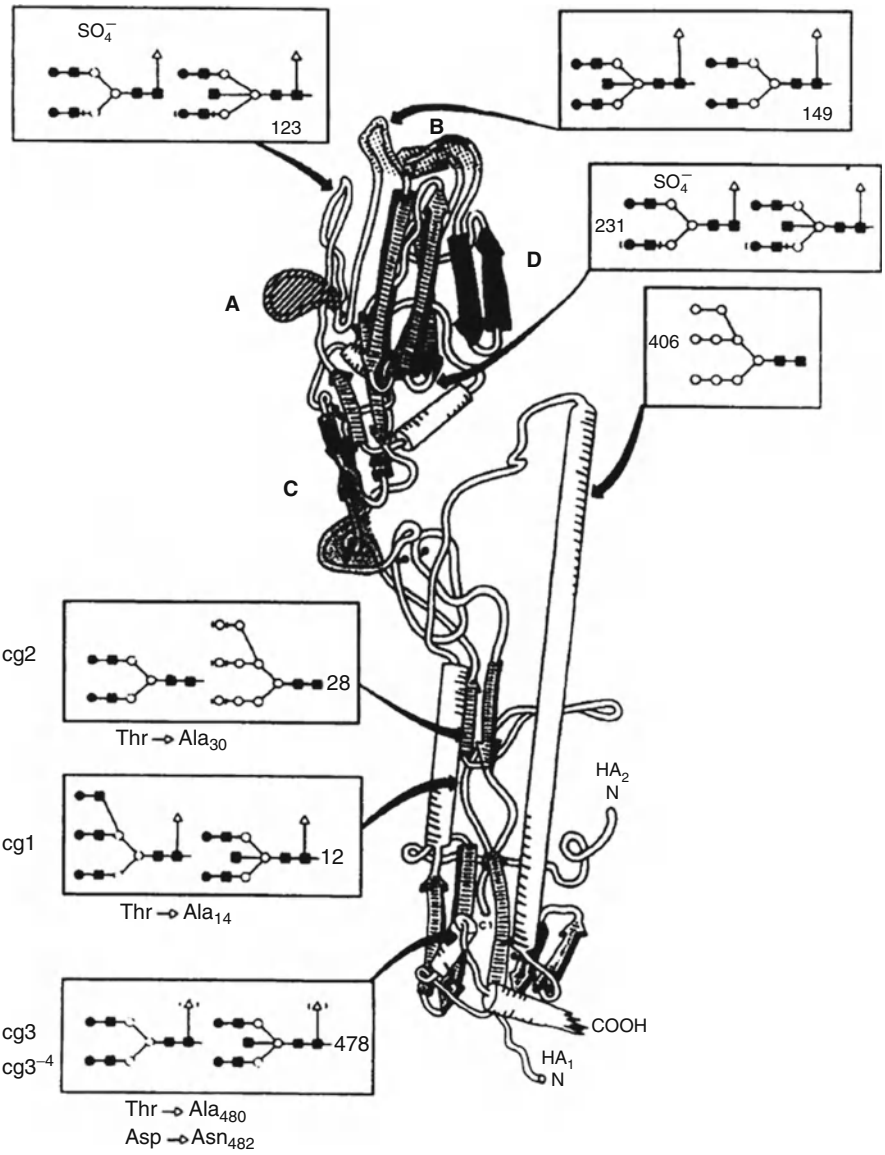

Fig. 13.5 Structural model of the HA monomer; the carbohydrate structures present at the seven glycosylation sites of the HA of influenza virus strain A/FPV/Rostock/34 (H7N1) [50] are shown. The positions of the asparagine residues are indicated in accordance with the H7 HA amino acid sequence. The conserved glycosylation sites are designated cgl (Asn-12) and cg2 (Asn-22) in the HA, subunit and cg3 (Asn-478) in the HA2 subunit and correspond to H3 numbering as residues 22, 38, and 483, respectively (Reproduced from [36]. With permission)

lacking, but it is tempting to hypothesize that correct folding, transport and maturation of HA is a central issue in this regard.

Finally, the amino acid sequence of the HA and, hence, the location of its *N*-linked oligosaccharides are determined by the viral genome, which is replicated by a virus-encoded RNA-dependent RNA polymerase. This enzyme lacks editing functions, thus mutations in all of the viral genes occur at a high frequency. On the other hand, the composition and structure of the oligosaccharides put onto the HA at the various sites is determined by biosynthetic and trimming enzymes provided by the host cell. Thus, the plasticity of the viral genome and the host specificity of the glycosylation machinery can, together, create virus populations that are more heterogeneous in structure and function than those potentially developed by either process alone. This diversity is considered to be responsible for the survival of these viruses in a variety of biological niches and for their ability to overcome the inhibitory effects of neutralizing antibodies and antiviral agents. The glycoprotein nature of the viral HA is therefore a key factor in enabling these viruses to retain their prevalent position amongst the re-emerging human infections.

13.7 Virus Polyhedra: Virus-Derived Assemblages for Long-Term Survival

A small number of viruses belonging to different families undergo a maturation step that is subsequent to the assembly of infectious particles. During the last phase of the replication cycle of these viruses, a fraction of the newly assembled particles are embedded into proteinaceous superstructures, known as polyhedra or occlusion bodies, formed by virus-encoded proteins. Virions trapped inside polyhedra are known as occluded viruses. Occlusion bodies are released from infected cells and play a critical role ensuring the survival of occluded viruses for very long periods (several years) of time after their discharge from the infected organism into the environment. Although the presence of occluded viruses has been mainly associated to three families of insect viruses namely baculoviruses [37], cypoviruses [33], and entomopoxviruses [2], they have also been found in viruses infecting other animal species, *e.g.*, cowpox virus, ectromelia virus, and raccoonpox virus belonging to the poxvirus family that infect different mammal species [22], as well as in different plant viruses [27].

13.7.1 Baculovirus Polyhedra

The *Baculoviridae* is a large family of viruses infecting arthropods mainly belonging to the insect orders *Lepidoptera*, *Diptera* and *Hymenoptera*. Baculovirus virions are rod-shaped enveloped particles enclosing a large double stranded DNA

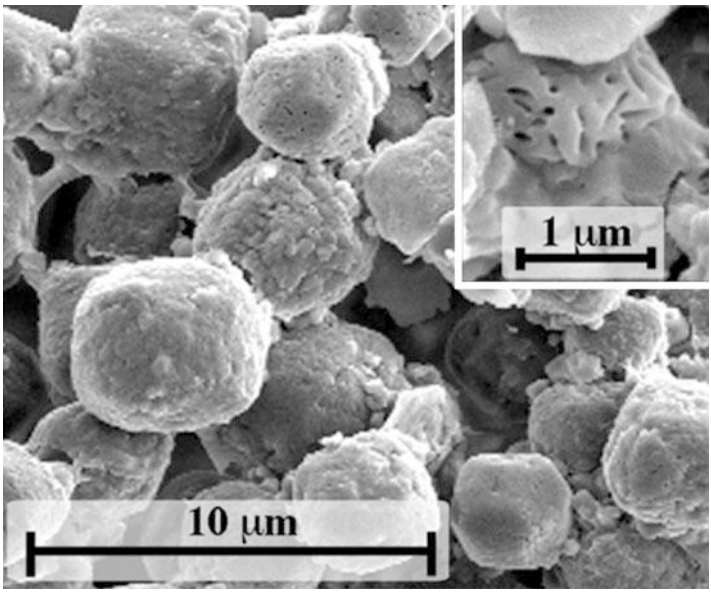

Fig. 13.6 Scanning electron microscopy of polyhedra purified from larvae of porina moths (*Wiseana* spp.) infected with *Wiseana* nucleopolyhedrovirus (WNPV). The *inset* shows an enlarged image area in which imprints of WNPV particles are apparent (Reproduced from [9]. With permission)

(88–160-kbp) genome [20]. Baculovirus polyhedra have been extensively characterized. Two baculovirus genera occlude their virions into protein crystals (Fig. 13.6). The Alphabaculoviruses form a single nuclear polyhedra embedding up to 200 virions surrounded by an electron-dense carbohydrate layer, known as calyx, containing a virus-encoded 34-kDa polypeptide [49]. In contrast, members of the Betabaculovirus genus, generally known as granuloviruses, occlude their virions into smaller cytoplasmic polyhedra usually containing a single virus particle [37]. The matrix of occlusion bodies is composed by a single protein, termed polyhedrin, with a molecular size of 29-kDa. This protein is one of the best-conserved polypeptides amongst the *Baculoviridae* family with a sequence identity of ca. 90 % within Alphabaculovirus, 60 % within Betabaculovirus, and 50 % between the two groups. The atomic structure of the *Autographa californica* nucleopolyhedrovirus (AcMNPV), a Betabaculovirus, has been recently solved [9]. The polyhedrin molecule folds into a central β-sandwich core with an extended N-terminal α-helical projection and a long C-terminal tail. Polyhedrin molecules assemble into trimers where molecules are held together predominantly by hydrophobic interactions. Polyhedra are built by tetrahedral clusters formed by four polyhedrin trimeric units. Tetrahedral polyhedrin clusters are densely packed and stabilized by the docking of C-terminal molecular arms that fit into cavities of neighboring clusters [9].

Polyhedra exhibit an extremely high resistance to severe thermal and chemical conditions. However, they are readily dissolved in alkaline solutions that resemble conditions of the insect intestinal tract. Ingestion of polyhedra by susceptible hosts triggers the release of the occluded viral particles, thus facilitating the infection of midgut cells. Early during infection, non-occluded virions are released from infected midgut cells into the hemolymph promoting the establishment of a systemic infection. The formation of polyhedra takes place during the final stages of the virus replication cycle. Despite the wealth of structural and functional information, mechanisms governing both polyhedra assembly and incorporation of virus particles into growing baculovirus and cypovirus polyhedra remain poorly understood.

13.8 Perspectives and Conclusions

Virus morphogenesis is initiated by the establishment of weak interactions amid their structural subunits. This strategy secures the proper assembly of the virus building blocks but results in the formation of immature assembly products, known as provirions, completely lacking the capacity to initiate a productive infection. The maturation of provirions is an essential and ubiquitous process that transforms fragile, noninfectious assemblages into fully infectious virus particles. This process involves the proteolytic processing and rearrangement of structural elements that consolidate the structure of the virus particle.

The maturation process of enveloped viruses involves a further step consisting on the proteolytic cleavage of glycosylated envelope polypeptides. This event is essential to expose glycoprotein domains involved in the interaction of virus particles with specific cell receptors, and the subsequent fusion of their envelope with the host cell membrane.

Finally, a reduced number of viruses undergo and additional maturation step that results in the incorporation of a fraction of the newly assembled virions into a virus-derived proteinaceous matrix. This superstructures, known as occlusion bodies, play a key role ensuring the survival of infective particles for prolonged periods of time under harsh environmental conditions.

Acknowledgements We thank Dolores Rodríguez and Ana Oña for critical readings of the manuscript. This work was supported by grant from the Spanish Ministry of Economy and Competitiveness AGL2011-24758 to JFR.

References and Further Reading

1. Agrawal DK, Johnson JE (1995) Assembly of the T = 4 Nudaurelia capensis omega virus capsid protein, post-translational cleavage, and specific encapsidation of its mRNA in a baculovirus expression system. Virology 207:89–97
2. Bilimoria SL, Arif BM (1979) Subunit protein and alkaline protease of entomopox virus spheroids. Virology 96:596–603

3. Briggs JA, Kräusslich HG (2011) The molecular architecture of HIV. J Mol Biol 410:491–500
4. Brown JC, Newcomb WW (2011) Herpesvirus capsid assembly: insights from structural analysis. Curr Opin Virol 1:142–149
5. Chan DC, Fass D, Berger JM, Kim PS (1997) Core structure of gp41 from the HIV envelope glycoprotein. Cell 89:263–273
6. Chan WE, Chen SS (2006) Downregulation of human immunodeficiency virus type 1 Gag expression by a gp41 cytoplasmic domain fusion protein. Virology 348:418–429
7. Chertova E, Chertov O, Coren LV, Roser JD, Trubey CM, Bess JW Jr, Sowder RC 2nd, Barsov E, Hood BL, Fisher RJ, Nagashima K, Conrads TP, Veenstra TD, Lifson JD, Ott DE (2006) Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. J Virol 80:9039–9052
8. Chiu E, Coulibaly F, Metcalf P (2012) Insect virus polyhedra, infectious protein crystals that contain virus particles. Curr Opin Struct Biol 22:234–240
9. Coulibaly F, Chiu E, Gutmann S, Rajendran C, Haebel PW, Ikeda K, Mori H, Ward VK, Schulze-Briese C, Metcalf P (2009) The atomic structure of baculovirus polyhedra reveals the independent emergence of infectious crystals in DNA and RNA viruses. Proc Natl Acad Sci U S A 106:22205–22210
10. Daniels R, Kurowski B, Johnson AE, Hebert DN (2003) N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin. Mol Cell 1179–1190
11. Davison AJ, Eberle R, Ehlers B, Hayward GS, McGeoch DJ, Minson AC, Pellett PE, Roizman B, Studdert MJ, Thiry E (2009) The order Herpesvirales. Arch Virol 154:171–177
12. Ganem D, Prince AM (2004) Hepatitis B virus infection–natural history and clinical consequences. N Engl J Med 350:1118–11129
13. Ganser-Pornillos BK, Yeager M, Sundquist WI (2008) The structural biology of HIV assembly. Curr Opin Struct Biol 18:203–217
14. Gao M, Matusick-Kumar L, Hurlburt W, DiTusa SF, Newcomb WW, Brown JC, McCann PJ 3rd, Deckman I, Colonno RJ (1994) The protease of herpes simplex virus type 1 is essential for functional capsid formation and viral growth. J Virol 68:3702–3712
15. Gething MJ, McCammon K, Sambrook J (1986) Expression of wild-type and mutant forms of influenza hemagglutinin: the role of folding in intracellular transport. Cell 46:939–950
16. Hanzlik TN, Gordon KH (1997) The tetraviridae. Adv Virus Res 48:101–168
17. Hebert DN, Foellmer B, Helenius A (1995) Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. Cell 81(3):425–433
18. Heermann KH, Goldmann U, Schwartz W, Seyffarth T, Baumgarten H, Gerlich WH (1984) Large surface proteins of hepatitis B virus containing the pre-s sequence. J Virol 52:396–402
19. Helgstrand C, Munshi S, Johnson JE, Liljas L (2004) The refined structure of nudaurelia capensis omega virus reveals control elements for a T = 4 capsid maturation. Virology 318:192–203
20. Herniou EA, Jehle JA (2007) Baculovirus phylogeny and evolution. Curr Drug Targets 8:1043–1050
21. Heymann JB, Cheng N, Newcomb WW, Trus BL, Brown JC, Steven AC (2003) Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy. Nat Struct Biol 10:334–341
22. Howard AR, Weisberg AS, Moss B (2010) Congregation of orthopoxvirus virions in cytoplasmic a-type inclusions is mediated by interactions of a bridging protein (A26p) with a matrix protein (ATIp) and a virion membrane-associated protein (A27p). J Virol 84:7592–7602
23. Huovila AP, Eder AM, Fuller SD (1992) Hepatitis B surface antigen assembles in a post-ER, pre-Golgi compartment. J Cell Biol 118:1305–1320
24. Johnson JE (2010) Virus particle maturation: insights into elegantly programmed nanomachines. Curr Opin Struct Biol 20:210–216
25. Klenk HD, Wagner R, Heuer D, Wolff T (2002) Importance of hemagglutinin glycosylation for the biological functions of influenza virus. Virus Res 82:73–75
26. Lu X, Mehta A, Dadmarz M, Dwek R, Blumberg BS, Block TM (1997) Aberrant trafficking of hepatitis B virus glycoproteins in cells in which N-glycan processing is inhibited. Proc Natl Acad Sci U S A 94:2380–2385

27. Martelli GP, Russo M (1977) Plant virus inclusion. Adv Virus Res 21:175–266
28. Matsui T, Lander G, Johnson JE (2009) Characterization of large conformational changes and autoproteolysis in the maturation of a T = 4 virus capsid. J Virol 83(2):1126–1134
29. Mehta A, Lu X, Block TM, Blumberg BS, Dwek RA (1997) Hepatitis B virus (HBV) envelope glycoproteins vary drastically in their sensitivity to glycan processing: evidence that alteration of a single N-linked glycosylation site can regulate HBV secretion. Proc Natl Acad Sci U S A 94:1822–1827
30. Merry T, Astrautsova S (2010) Alternative approaches to antiviral treatments: focusing on glycosylation as a target for antiviral therapy. Biotechnol Appl Biochem 56:103–109
31. Munshi S, Liljas L, Cavarelli J, Bomu W, McKinney B, Reddy V, Johnson JE (1996) The 2.8 A structure of a T = 4 animal virus and its implications for membrane translocation of RNA. J Mol Biol 261:1–10
32. Nobusawa E, Aoyama T, Kato H, Suzuki Y, Tateno Y, Nakajima K (1991) Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza a viruses. Virology 182:475–485
33. Payne CC, Mertens PPC (1983) Cytoplasmic polyhedrosis viruses. In: Joklik WK (ed) The reoviridae. Plenum Press, New York
34. Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, Klein CA, Swanstrom R (1994) The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectiousvirions. J Virol 68:8017–8027
35. Rixon FJ (2008) A good catch: packaging the virus genome. Cell Host Microbe 3:120–122
36. Roberts PC, Garten W, Klenk HD (1993) Role of conserved glycosylation sites in maturation and transport of influenza a virus hemagglutinin. J Virol 67:3048–3060
37. Rohrmann GF (1986) Polyhedrin structure. J Gen Virol 67:1499–1513
38. Schmitt S, Glebe D, Alving K, Tolle TK, Linder M, Geyer H, Linder D, Peter-Katalinic J, Gerlich WH, Geyer R (1999) Analysis of the pre-S2 N- and O-linked glycans of the M surface protein from human hepatitis B virus. J Biol Chem 274:11945–11957
39. Schreiber G, Keating AE (2011) Protein binding specificity versus promiscuity. Curr Opin Struct Biol 21:50–61
40. Shinya K, Makino A, Kawaoka Y (2010) Emerging and reemerging influenza virus infections. Vet Pathol 47:53–57
41. Skehel JJ, Wiley DC (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem 69:531–569
42. Steven AC, Heymann JB, Cheng N, Trus BL, Conway JF (2005) Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. Curr Opin Struct Biol 15:227–236
43. Stiasny K, Fritz R, Pangerl K, Heinz FX (2011) Molecular mechanisms of flavivirus membrane fusion. Amino Acids 41:1159–1163
44. Taylor DJ, Krishna NK, Canady MA, Schneemann A, Johnson JE (2002) Large-scale, pH-dependent, quaternary structure changes in an RNA virus capsid are reversible in the absence of subunit autoproteolysis. J Virol 76:9972–9980
45. Telesnitsky A (2010) Retroviruses: molecular biology, genomics and pathogenesis. Future Virol 5:539–5343
46. Thomsen DR, Roof LL, Homa FL (1994) Assembly of herpes simplex virus (HSV) intermediate capsids in insect cells infected with recombinant baculoviruses expressing HSV capsid proteins. J Virol 68:2442–2457
47. Vigerust DJ, Shepherd VL (2007) Virus glycosylation: role in virulence and immune interactions. Trends Microbiol 15:211–218
48. Werr M, Prange R (1998) Role for calnexin and N-linked glycosylation in the assembly and secretion of hepatitis B virus middle envelope protein particles. J Virol 72:778–782
49. Whitt MA, Manning JS (1988) A phosphorylated 34-kDa protein and a subpopulation of polyhedrin are thiol linked to the carbohydrate layer surrounding a baculovirus occlusion body. Virology 163:33–42

50. Wilson IA, Skehel JJ, Wiley DC (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 289:366–373
51. Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, Kuiken C, Korber B (2004) Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. Glycobiology 14:1229–1246

Further Reading

References [8, 13, 21, 24, 35, 41, 42, 43, and 47] *listed above are especially recommended for further reading.*

Chapter 14
Virus Morphogenesis in the Cell: Methods and Observations

Cristina Risco and Isabel Fernández de Castro

Abstract Viruses carry out many of their activities inside cells, where they synthesise proteins that are not incorporated into viral particles. Some of these proteins trigger signals to kidnap cell organelles and factors which will form a new macro-structure, the virus factory, that acts as a physical scaffold for viral replication and assembly. We are only beginning to envisage the extraordinary complexity of these interactions, whose characterisation is a clear experimental challenge for which we now have powerful tools. Conventional study of infection kinetics using virology, biochemistry and cell biology methods can be followed by genome-scale screening and global proteomics. These are important new technologies with which we can identify the cell factors used by viruses at different stages in their life cycle. Light microscopy, electron microscopy and electron tomography, together with labelling methods for molecular mapping *in situ*, show immature viral intermediates, mature virions and recruited cell elements in their natural environment. This chapter describes how these methods are being used to understand the cell biology of viral morphogenesis and suggests what they might achieve in the near future.

Keywords Virus factory • Virus assembly • Virus morphogenesis • Genomics • Proteomics • Light microscopy • Electron microscopy • Tomography • 3D electron microscopy • Super-resolution microscopy • Correlative microscopy • Molecular mapping • Immunofluorescence • Immuno gold • Immuno electron microscopy • Clonable tags

C. Risco (✉) • I. Fernández de Castro
Cell Structure Laboratory, Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: crisco@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_14,

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
ASFV	African swine fever virus
CLEM	Correlative light and electron microscopy
DNA	Deoxyribonucleic acid
ESCRT	Endosomal sorting complex required for transport
ET	Electron tomography
GFP	Green fluorescent protein
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
IEM	Immunoelectron microscopy
IF	Immunofluorescence
LM	Light microscopy
RC	Replication complex
RNA	Ribonucleic acid
SEM	Scanning electron microscopy
TEM	Transmission electron microscopy
VV	Vaccinia virus
Y2H	Yeast two-hybrid

14.1 Introduction: Cell Biology of Virus Morphogenesis and the Concept of the Virus Factory

The idea of viruses as inert molecular entities has progressively been transformed since scientists began to discover the myriad of interactions that occur during the intracellular phase of virus life. Restricted by their own limited genetic repertoire, viruses need to use a number of cell factors for genome replication and morphogenesis. Identification of these factors is essential for understanding the virus morphogenetic processes that often take place in intracellular structures known as viral inclusions, virosomes or viral factories. Viruses are generally thought to build these structures to recruit and concentrate viral and cell factors needed for replication and assembly; they are able to modify a variety of cell organelles and to create new inter-organelle contacts (Fig. 14.1). In the mammalian cell, virus assembly can start inside the nucleus, in association with components of the secretory pathway (Golgi, endoplasmic reticulum or ER), at different points in the endocytic pathway (endosomes, multivesicular bodies or MVB) or at the plasma membrane. Mitochondria and cytoskeletal elements are present in the factories built by many different viruses.

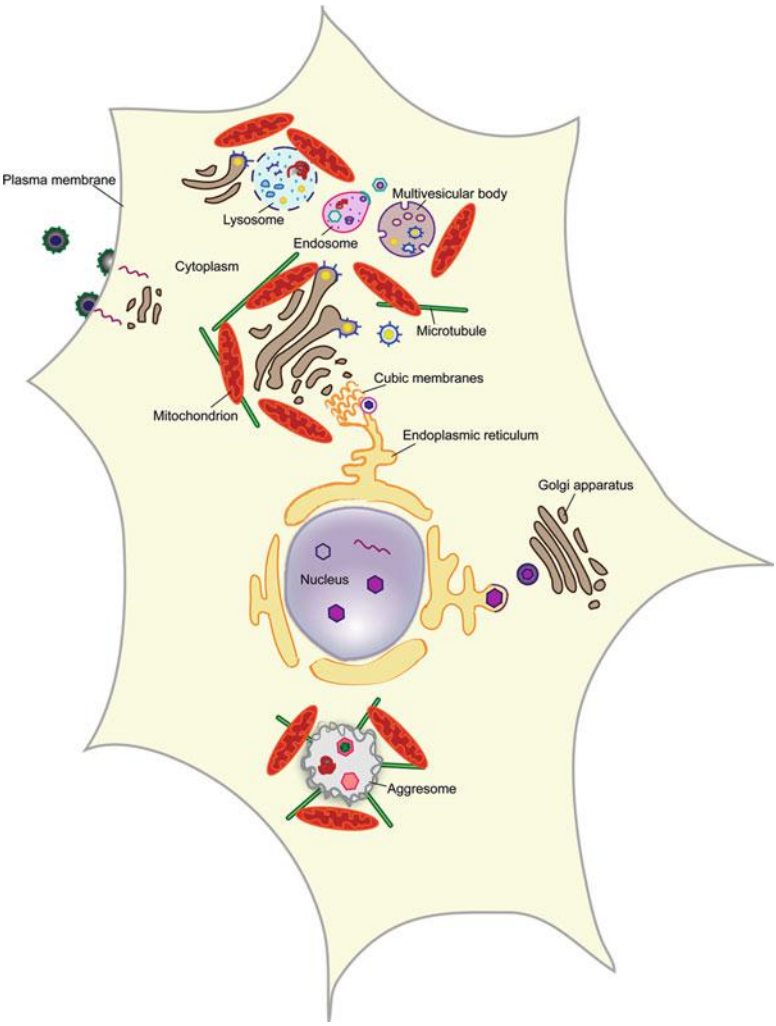

Fig. 14.1 Cell organelles used for viral morphogenesis. The cell nucleus is used by herpes- and papillomaviruses; corona-, bunya- and flaviviruses use components of the secretory pathway such as the endoplasmic reticulum and Golgi; togaviruses and the human cytomegalovirus use components of the endo-lysosomal pathway such as endosomes, multivesicular bodies and lysosomes; the African swine fever virus (ASFV) and poxviruses assemble in aggresome-like structures; assembly sites of retroviruses at the plasma membrane might be connected with a cytoplasmic factory. Viruses modify endomembranes, recruit mitochondria and cytoskeleton, and create new inter-organelle contacts

Virus factories are usually rather large, several microns in diameter, and are very dynamic, as their structure changes over time in accordance with virus needs [1]. This strategy appears to allow viruses to generate progeny with maximum efficiency in the use of cell resources. It is also thought that viruses can use these

scaffolds to protect themselves from cellular antiviral defences [2]. Viral factories were first described for large DNA viruses such as the African swine fever virus (ASFV) and the poxviruses; we now know that many DNA and RNA viruses build factories in the cytosol, in the nucleus, or both. Due to our limited knowledge of cell nucleus organisation, characterisation of nuclear viral factories is only beginning. In contrast, viral factories assembled in the cytosol are being studied extensively and there are many reports on their architecture and activities. Inside these cytosolic factories, morphogenesis of new viruses begins with the transport of replicated genomes from the structures that harbour the replication complexes (RC) to the assembly sites. Depending on the virus, a number of maturation steps will follow (see Chap. 13 for details) before new viral particles are ready for delivery and propagation.

This chapter will describe fundamental tools for diving inside the infected cell and understanding virus assembly. We also describe relevant examples of virus-cell interactions during virus morphogenesis that are being discovered using *in situ* techniques.

14.2 Methods for the Study of Virus-Cell Interactions During Morphogenesis

By identifying essential cell factors, studies based on recent advances in molecular biology, genomics and proteomics are broadening our knowledge of viral morphogenesis. To confirm and understand the role of potentially interesting genes in viral assembly, the information acquired must be subsequently analysed at a more complex level in infected cells. Exploring existing databases on cell pathways which combine information on genetic, metabolic and signal networks based on the literature can be a first step in further work that on many occasions will also include imaging with a variety of microscopy technologies.

14.2.1 Identification of Essential Cell Factors

Characterisation of infection kinetics is the first step in the study of viral morphogenesis. Conventional virology, biochemistry and cell biology methods allow us to determine optimal experimental conditions, including times post-infection (p.i.) and the most appropriate cell types. Key factors can be identified using two main groups of methods, (1) those that analyze gene expression patterns, and (2) those that study protein-protein interactions.

A number of new techniques can be applied to study interactions between viruses and cells on a genomic scale [3]. Microarrays are being used to identify mRNA transcription patterns in different phases of the virus life cycle. DNA gene

chips detect differences in gene expression between uninfected and virus-infected cells and at different infection stages. Gene data bases are then useful for associating the genes identified with specific cell pathways [4, 5]. High-throughput screening based on RNA interference (RNAi) is another category of methods that analyse gene expression patterns. RNAi is an RNA-dependent gene silencing process within living cells that is often exploited to study the function of genes. This emerging technology is used to study how viruses interact with their hosts at the molecular level. Analysis at various times post-infection has identified a number of cell factors potentially involved in viral morphogenesis, for example for dengue [6], influenza [7] and retroviruses [8]. RNAi may be used for large-scale screens that systematically shut down each gene in the cell. Studies using this approach have shown a requirement for cell factors such as the ESCRT machinery for assembly of the human immunodeficiency virus, HIV [9].

The group of techniques termed proteomics includes powerful methods to study protein-protein interactions. The yeast two-hybrid (Y2H) assay system remains one of the most amenable techniques and is widely used to search for virus-host interactions. Y2H works by expressing two candidate proteins in the yeast cell. Bait and prey proteins are fused either to a promoter-specific DNA-binding domain or to a transcription activation domain. Interaction between the two proteins in the yeast nucleus brings both domains together so that they can initiate expression of a reporter gene [10]. Individual bait proteins can be screened for interaction with a library of prey proteins. Genome-scale Y2H studies were used to identify 314 virus-host interactions for HCV [11] 109 interactions for vaccinia virus [12] and nine for HIV-1 [13]. Similar techniques, such as the yeast three-hybrid system, can be used to study interactions between nucleic acids and proteins [14, 15]. This group of methods also includes pull-down and immunoprecipitation assays, and tandem affinity purification (TAP) tagging approaches, as well as protein identification by quantitative and semi-quantitative mass spectrometry. Whereas Y2H usually detects transient interactions, affinity-tag purification mass spectrometry shows stable, stoichiometric complexes. Since cell proteins often incorporate into viral particles, these techniques can be applied to the study of protein-protein interactions in the infected cell and in purified viral assembly intermediates [16]. This is the case of clathrin, for example, which was found in retrovirus particles; clathrin was only recently identified as one of the cell factors that facilitate accurate morphogenesis of several retroviruses [17]. Y2H technology also detected cell proteins that interfere with virus assembly and viral proteins that block them. This is the case of tetherin, first detected by proteomics and mass spectrometric protein identification as a cell factor that restricts retrovirus assembly [18], and later confirmed as a restriction factor for a wide variety of enveloped viruses [19]. Viruses have several anti-tetherin proteins to counteract the effect of this factor [20]. The current challenge of high-throughput technologies is to develop more efficient informatics tools to accurately analyse the vast amount of information they can provide [21].

The next step in the characterisation of in-cell virus assembly pathways would be to study the specific roles of the factors identified in their natural environment; to

do this, we must of course return to the infected cell. Functions can be tested by protein depletion or overexpression, by mutagenesis or by protein targeting with tags. To visualise key factors in infected cells, there are a variety of classical and novel microscopy techniques that will be described in the following sections.

14.2.2 Studying Viral Morphogenesis in Situ with Light and Electron Microscopy

Microscopy has played an essential role in our understanding of cell architecture and viral assembly (see Chap. 3). Light microscopy (LM) and transmission electron microscopy (TEM) provide different types of information about viral infection, ranging from general events that involve the whole cell to the detailed imaging of nascent and maturing viral particles in specific cell compartments [22] (Fig. 14.2). With resolutions in the 100–500 nm range, LM shows organelle recruitment and modification in the assembly compartment and, in the case of the largest viruses, individual new viral particles as well [23]. In immunofluorescence assays using antibodies to viral proteins and cell compartments, we can see where viral structural and scaffolding proteins (Chap. 11) concentrate to create the assembly sites. Light microscopy shows, for example, that the same virus can build distinct factories in different cell types, depending on specific characteristics of the cell (Fig. 14.2a, b). Functional viruses that express proteins fused with clonable fluorescent tags such as the green fluorescent protein (GFP) (see Sect. 14.3 for details) can be followed in live cells [24]. Video microscopy facilitates dynamic characterisation of the biogenesis of the factory and virus assembly in real time. Time course experiments with antibody-labelled permeabilised cells or video microscopy studies with GFP fusions in living cells are essential for selecting specific conditions, such as the best times p.i., for more detailed, higher resolution study by TEM.

Electron microscopy generally uses ultra-thin sections of cells previously embedded in plastic resins after conventional fixation and dehydration; alternatively, cells are processed at low temperature after or upon fixation for optimal preservation of ultrastructure (see Chap. 3 for sample preparation details), or subjected to criofixation procedures such as high-pressure freezing prior to freeze substitution and embedding, or freeze-fracture analyses. With resolutions in the range of a few nanometers, cell TEM can show changes in shape and size of virus assembly intermediates in specific intracellular compartments [25, 26] (Fig. 14.2c, d). TEM of infected cells shows that mitochondria, endomembranes and cytoskeleton often participate in organising the structure that supports viral assembly. To complete maturation and become infectious, immature viral particles must often travel within the factory to find specific cell factors. At late times post-infection, once virus progeny have been produced and must find their way out of the cell, the factory can be dismantled [1].

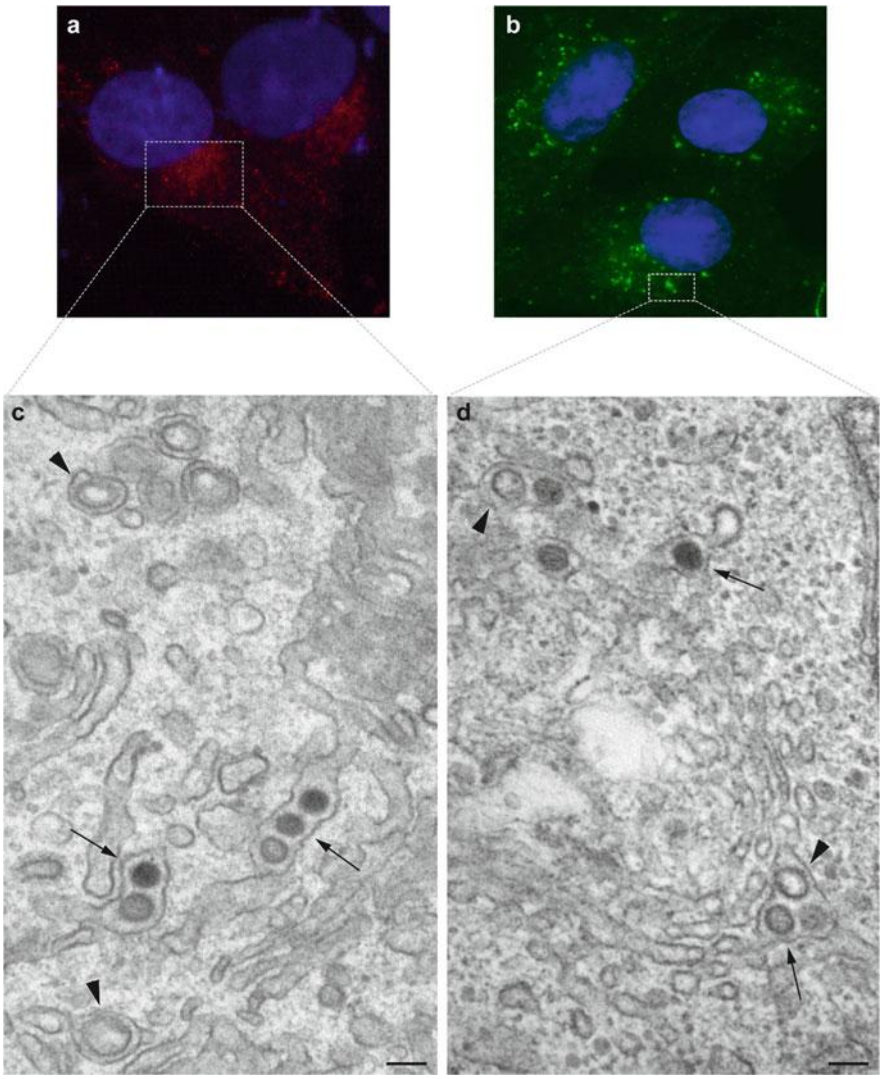

Fig. 14.2 Studying virus assembly by light and electron microscopy. (**a**, **b**) Immunofluorescence microscopy of BHK-21 (**a**) and Vero cells (**b**) infected with a bunyavirus at 10 h p.i. (h.p.i.). Cells were labelled with an antibody specific for one of the viral structural proteins that concentrates at the assembly sites. A single large perinuclear factory is formed in BHK-21 cells, whereas many mini-factories are seen in Vero cells. (**c**, **d**) TEM of BHK-21 (**c**) and Vero cells (**d**) at 10 h.p.i. In both cases, similar spherules, the structures that harbour the RC (*arrowheads*) [22] and viral intermediates (*arrows*) are distinguished in Golgi membranes. Scale bars, 100 nm

TEM of cells in two dimensions provides a first glimpse of the viral morphogenetic pathways *in situ*. The size of cells (several microns in diameter) and the thickness of ultra-thin sections (50–100 nm) nonetheless constitute a major

limitation, because we are actually studying single planes of much larger structures. In these conditions, many elements and contacts can be missed. Three-dimensional TEM avoids this limitation; the variety of methods available is described in the following sections.

14.2.3 Visualising Virus Assembly in Three Dimensions I: 3D Reconstructions of Serial Sections, TEM and SEM

In conventional TEM, electrons must traverse the sample to generate a projection image. Samples must therefore be thin, around 50–100 nm, thinner than many viruses. Given that eukaryotic cells are several microns in diameter, the ultra-thin sections are single planes of much larger structures. For conventional TEM, cells are fixed on culture plates and collected by low speed centrifugation, followed by sectioning of the pellet. In these conditions, cells preserve their morphology but present a variety of orientations (Fig. 14.3a); analysis is therefore restricted to random, unique sections of cells, and scarce or non-randomly distributed elements can be missed completely. Oriented serial sections solve this problem (Fig. 14.3b). By collecting all serial sections from each cell, all intracellular elements can be detected and studied. This strategy guarantees a complete analysis without missing any intracellular event of interest; immunofluorescence can assist by localising where viruses are assembled within a cell and thus, where cells should be sectioned for TEM [22]. The study of oriented serial sections is very informative and has revealed unreported contacts between cell organelles, RC and assembly sites, as well as the relationships of different viral intermediates with specific cell elements (Fig. 14.3c). After image processing and segmentation to assign identities to all individual structures in the images, serial sections can be aligned and combined in 3D reconstructions [27]. With resolution values of ~5 nm in the X and Y axes, these 3D reconstructions can show viruses in factories with considerable detail (Fig. 14.3d, e). Resolution in the Z axis is limited, however, with values of ~50–100 nm due to imperfections in the alignment process.

Scanning electron microscopy (SEM) can also be used to visualise virus morphogenesis. These microscopes use a focused beam of high-energy electrons to generate a variety of signals at the surface of solid specimens. Samples are covered by a thin layer of metal and scanned with a primary source of electrons; secondary electrons are then released from the sample surface and collected by a detector. The signals derived from electron-sample interactions furnish information about the sample, including external morphology in 3D. Modern scanning electron microscopes can now provide resolutions as precise as 2–5 nm, near that achieved by cellular electron microscopy [28]. SEM images can show morphological changes in large virus particles during maturation *in situ*; this is the case, for example, of the giant mimivirus (Fig. 14.4a, b). In lysed cells and isolated virus factories visualised by SEM, immature viruses are clearly distinguished at short

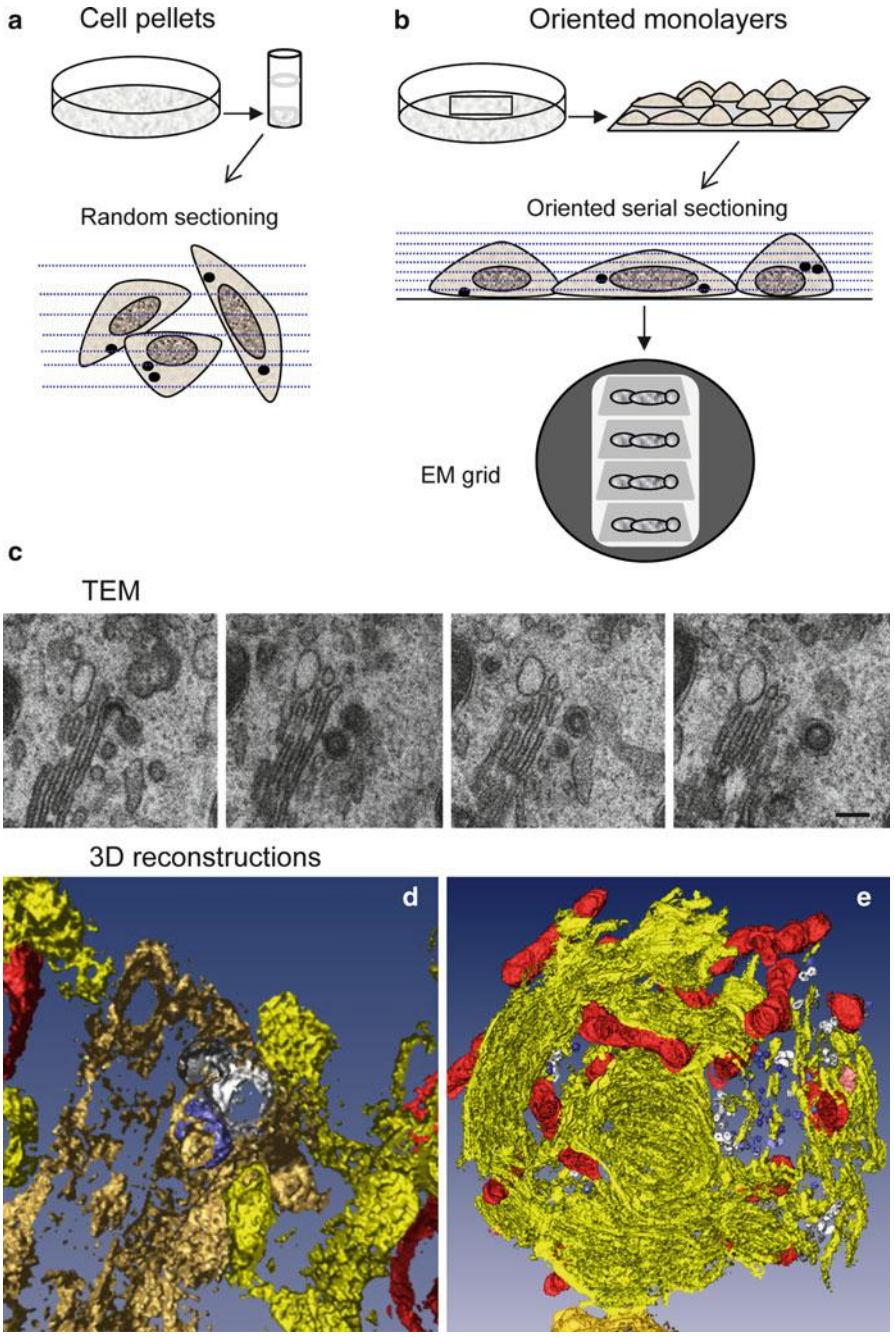

Fig. 14.3 Serial sections, TEM and 3D reconstructions. (**a**, **b**) Summary of the principles and differences between conventional ultra-thin sectioning and oriented serial sectioning. (**c**) Serial sections in TEM and (**d**) 3D reconstruction showing the interaction between a spherule, the structure that harbours the RC (*white*) and a viral particle (*blue*) in Golgi membranes (*beige*). Mitochondria are segmented in *red* and rough endoplasmic reticulum (RER) in *yellow*. (**e**) 3D reconstruction of a viral factory from a different cell. In this case, 15 serial sections were used. The Golgi complex has been removed to improve visualisation of RC and viral particles. Scale bar, 100 nm

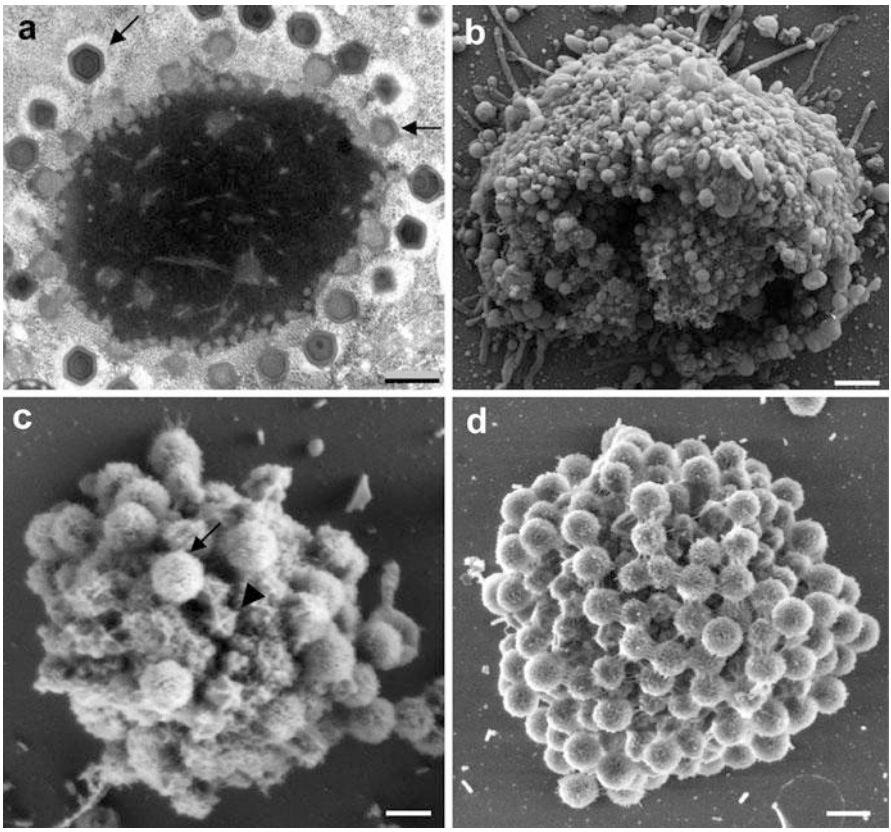

Fig. 14.4 Mimivirus factory in 2D and 3D by TEM and SEM. (**a**) 2D views of a virus factory as visualised by TEM, showing virus particles (*arrows*) at various assembly stages. (**b**) 3D views of a viral factory within an amoeba cell lysed at 8 h.p.i. and visualised by SEM. (**c**) SEM of a factory isolated at 8 h.p.i. Viral particles are seen at various assembly stages. The *arrow* indicates a mature virus particle and the *arrowhead*, an immature particle. (**d**) SEM of a virus factory isolated at 10 h. p.i. Only mature viruses can be detected. Scale bars, 500 nm in (**a**) and (**d**); 2 μm in (**b**); 300 nm in (**c**) (Reproduced with permission from [29])

times p.i. (Fig. 14.4c), whereas mature virions accumulate in the factory at longer times p.i. (Fig. 14.4d) [29].

TEM in combination with freeze-fracture and metal shadowing allows higher resolution 3D views of viruses inside intact cells. Frozen cells are fractured mechanically and covered with a thin layer of platinum and carbon in a vacuum chamber, the metal replicas are then washed to eliminate organic materials and mounted on EM grids. The main limitation of this approach derives from the unpredictable patterns of the fracture planes and the difficulty in interpreting the images. It can nevertheless be useful when studying the morphogenesis of large enveloped viruses (Chap. 11) and their interactions with cell endomembranes during assembly and maturation [30].

14.2.4 Visualizing Virus Assembly in Three Dimensions II: Electron Tomography

Electron tomography (ET) of infected cells shows virus assembly in 3D with resolutions of ~3-5 nm in all three axes, X, Y, and Z (see Chap. 3). Higher resolutions of ~1 nm have been reported, as was the case of the budding HIV viral particles visualised by cryo-electron microscopy on the surface of intact human cells [31]. Both normal and aberrant budding events were visualised on the cell surface, suggesting that cellular and/or viral factors control the quality of virus assembly and maturation. A limitation of cellular tomography is that the maximum thickness of the sample must be <0.5 μm, whereas virus factories are larger; however, conventional TEM methods like those described above can help to select specific elements within cells for more detailed ET structural analysis. Thanks to ET, elusive structures have been visualised for the first time, such as the transfer of viral genomes from RC to assembly sites in dengue virus-infected cells [32] (Fig. 14.5a, b). The morphogenesis of poxviruses is another good example; vaccinia virus (VV) assembly and architecture were the subject of numerous studies, but the organisation and biogenesis of immature and mature VV particles were not understood until the first ET studies were carried out [33, 34]. These analyses showed unprecedented remodelling of cell endomembranes during VV particle assembly (Fig. 14.5c, d). Tomograms are analysed in detail in computational slices of 1–2 nm extracted from the original volumes (Fig. 14.5a). As with serial sections in conventional TEM, the relevant information is contained in these single planes, whereas 3D representations are models used to summarise the most relevant features in the tomograms (see Chap. 3 for details).

In summary, two main groups of methods are available for 3D studies of viral assembly. 3D reconstructions of serial sections by TEM or analysis of surface morphology by SEM will show the general organisation of the virus factory, the inter-organelle contacts, and changes in the cell compartments where viruses are formed. In contrast, electron tomography is more appropriate for studying individual viral particles and to obtain fine details of their maturation *in situ*.

14.3 Molecular Mapping of Viral Morphogenesis

The discipline known as histochemistry includes a great variety of techniques to visualise molecules in biological samples [35]. There are several histochemical methods to localise nucleic acids, lipids, sugars, and other molecules, some of which have been used to label viruses in cells. Nonetheless, protein-labelling techniques are by far the most developed; because of their importance in the study of virus morphogenesis, in this section we will focus on methods that detect proteins in light and electron microscopy.

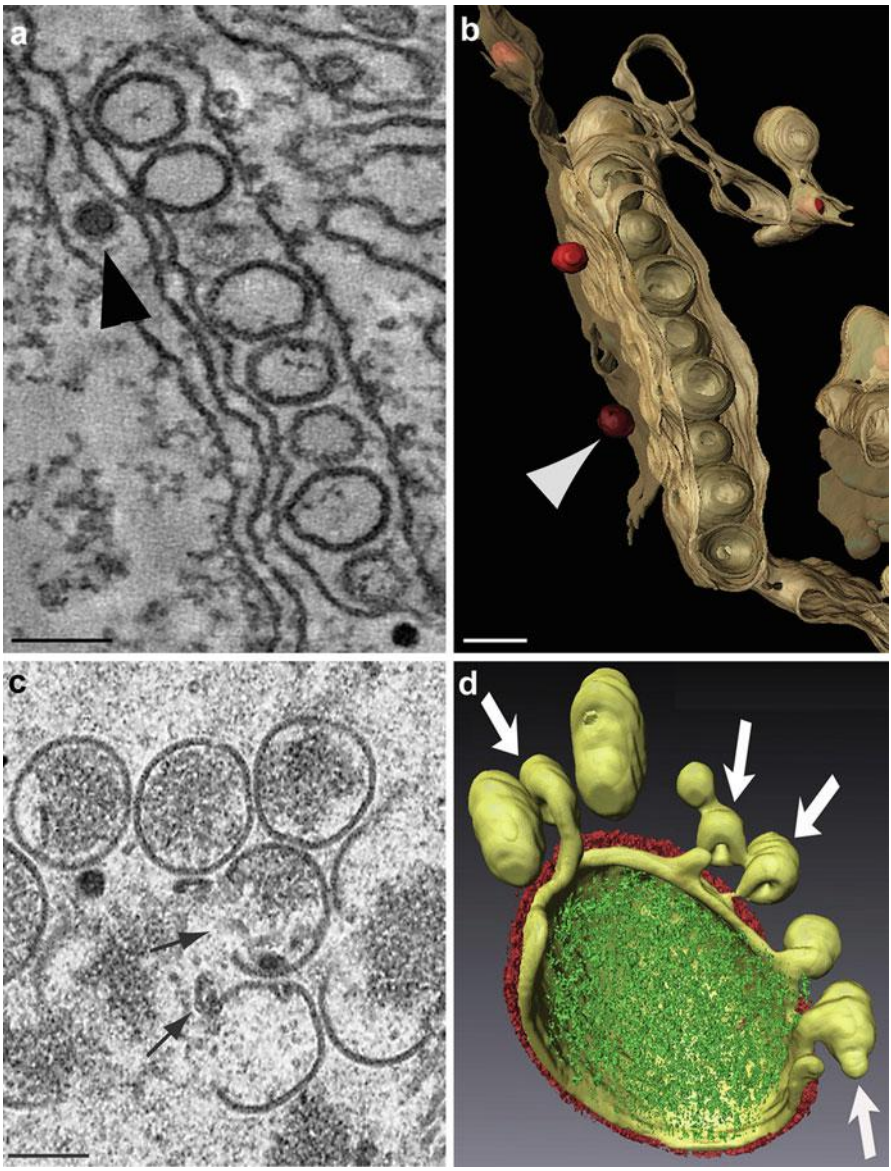

Fig. 14.5 3D electron tomography of virus assembly. (**a**) Computational slice and (**b**) 3D reconstruction of the dengue virus factory as visualised by ET. Nascent viral particles (*arrowheads*) face the spherules that harbour the RC. (**c**, **d**) 2D TEM and 3D ET, respectively, of immature VV particles in the process of assembly from cell membranes (*arrows*). ET shows how the viral envelope is connected to a collection of open membrane structures and how these membranes contribute to envelope formation. Scale bars, 100 nm in (**a**) and (**b**); 250 nm in (**c**) (Reproduced with permission from [32] (**a**) and (**b**), and [33] (**d**))

14.3.1 Antibodies, Immunofluorescence and Immunoelectron Microscopy

Previous sections of this chapter have shown how to identify key factors for virus morphogenesis and how to visualise virus assembly in cells in two and three dimensions. To link the information from both methods and to fully understand viral morphogenesis, we also need methods for molecular mapping *in situ*. Molecules of interest, which are viral and cellular factors, can be localised in cells with specific antibodies in immunolabelling assays or with clonable tags (Fig. 14.6).

Antibodies detect proteins with high specificity and variable sensitivity. Primary antibodies generated against proteins of interest are detected with secondary antibodies conjugated to a fluorescent probe for LM visualisation, or to an electron-dense colloidal gold particle, which is easy to detect by EM in immunogold assays (Fig. 14.6a). Antibodies have been and are still fundamental tools in electron microscopy [36, 37]. When immunolabelling proteins on cryosections, where cells have not been dehydrated and maintain their proteins in a natural hydrated state, the sensitivity of antibody detection is usually higher than that obtained when labelling sections of dehydrated, resin-embedded cells. Since the introduction of cryosectioning, the method has been improved and perfected [38, 39]. This approach allows colocalisation of nascent and maturing viral particles in specific cell compartments and the proteins being incorporated into assembling virus particles (Fig. 14.6b). Although antibodies are usually very specific, information derived from these experiments is later confirmed in biochemical assays.

Due to their large size, antibodies that recognise internal structures must be used on cell sections. Alternatively, cells can be permeabilised to label intracellular compartments (Fig. 14.6c), although this is incompatible with preservation of fine ultrastructure. Certain permeabilisation protocols use the bacterial exotoxin streptolysin O (SLO) to open pores in the plasma membrane while leaving intracellular membranes untouched. This was used in pre-embedding immunogold assays to follow viral and cell proteins during VV assembly from intracellular membranes [40].

An important limitation of antibodies is their variable sensitivity due to epitope loss during sample preparation and to macromolecular interactions inside cells that often mask the protein epitopes *in vivo*. This is particularly problematic when the proteins of interest are part of densely packed macromolecular complexes, such as those involved in virus assembly. The use of clonable tags can overcome these limitations.

14.3.2 Clonable Tags

Jellyfish green fluorescent protein (GFP), its mutants and homologues have caused a true revolution in cell biology. If proteins fused to a fluorescent tag maintain their

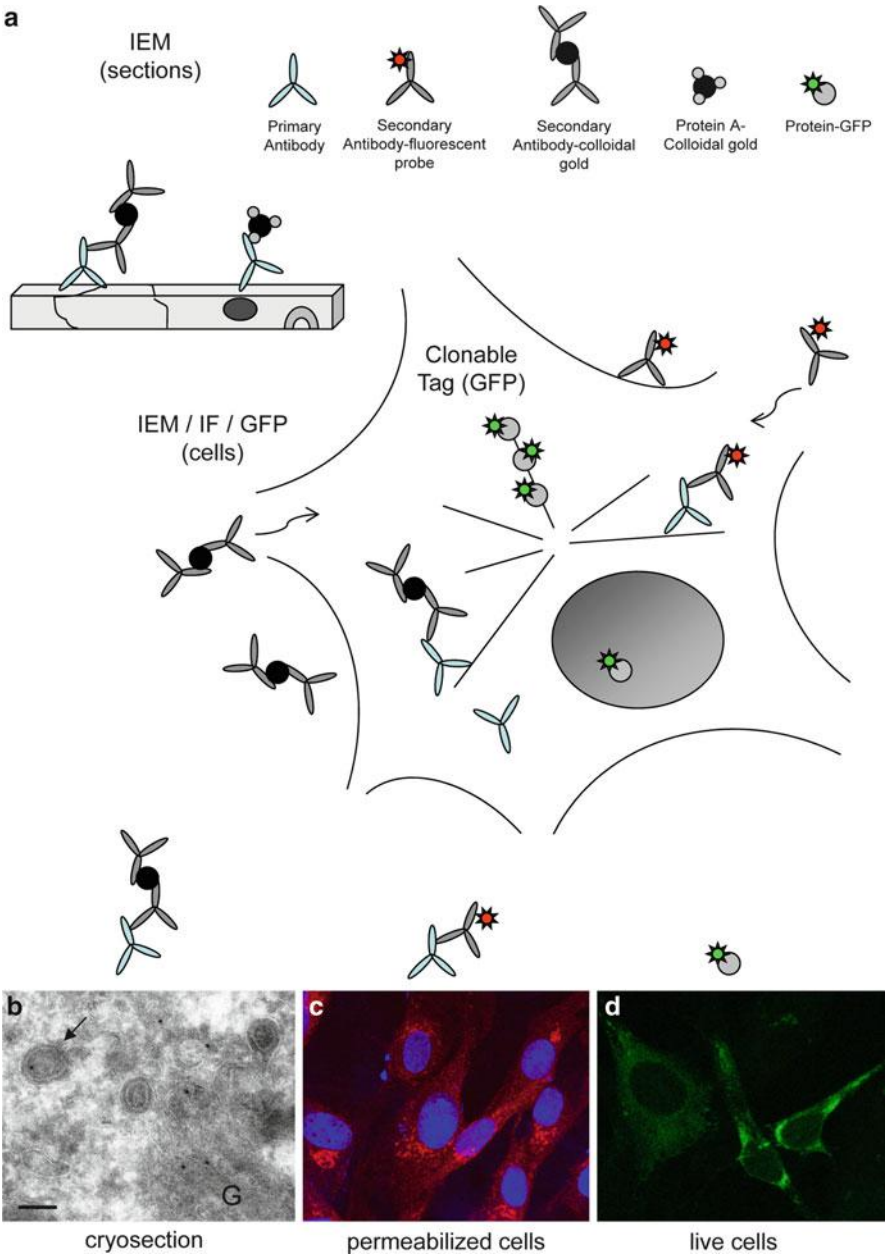

Fig. 14.6 Molecular mapping of virus assembly with antibodies and clonable tags. (**a**) Schemes showing the principles of labelling on thin sections, whole permeabilised cells and intact cells. (**b**) Immunogold detection on a cryosection from a bunyavirus-infected cell. Cells were labelled with a primary antibody specific for a viral scaffolding protein, followed by secondary antibodies conjugated to 10 nm colloidal gold particles. The protein is detected in Golgi membranes (G) and

normal functions, their movements and interactions with intracellular structures can be tracked in living cells (Fig. 14.6d). In conjunction with the new generation of LM methods, we can even follow the assembly of individual virus particles at real time [41]. GFP is a relatively large molecule (27 kDa) and, since viral proteins have very strict structural requirements, fusion to GFP can produce loss of function. If this is the case, probes smaller than GFP could be a viable option. The group of smaller probes includes the fluorescent flavoprotein known as miniSOG, which is less than half the size of GFP [42], or the smaller tetra-cysteine motifs, which are visualised after incubation with fluorescent biarsenical molecules such as ReAsH or FlAsH [43].

The use of clonable tags in TEM and ET would be a considerable advance for high resolution studies of virus assembly. The first clonable tags for EM validated in cells have yet to be used in virus morphogenesis studies, but show good prospects and will be listed in Sect. 14.5.3.

14.4 The Search for Signalling Pathways

The initiation of virus morphogenesis usually requires major reorganisation of cell membrane compartments and the cytoskeleton. Many laboratories are applying the methods described in previous sections to understand how viruses take control of cells and recruit all the necessary materials for their morphogenesis. The complexity of the interactions inside infected cells can be extraordinary, as illustrated by the viruses known as virophages that parasitise the factory built by other viruses. Some images of cells co-infected by the giant mimivirus and the virophage "Sputnik" suggest that the small virophage not only uses the materials recruited by the mimivirus for its own replication, but might even hide inside the mimivirus particles to exit the cell [44]. We still do not know how organelles and materials are recruited or how replicated viral genomes are transported from RC to assembly sites. We do know, however, that viruses target their proteins to specific cell compartments, that some viruses use a cellular defence mechanism termed the aggresome response, and that signals related to organelle movement on cytoskeletal tracks could also be involved [1, 2].

Fig. 14.6 (continued) viral particles (*arrow*). (**c**) Immunofluorescence detection of the same scaffolding protein in permeabilised cells. (**d**) Still image from a video recorded in a fluorescence microscope equipped for live cell imaging. Cells were infected with a recombinant virus that expresses the same scaffolding protein fused with GFP. Scale bar, 100 nm

14.4.1 DNA Viruses and Cell Aggresomes

Aggresomes are a defence response of cells to protein misfolding and aggregation. These inclusions form at the microtubule-organising centre (MTOC), where they enclose potentially toxic protein aggregates within vimentin cages. The large DNA viruses of animals such as the African swine fever virus (ASFV), the poxviruses, and the iridovirus frog virus three build factories that closely resemble cell aggresomes (Fig. 14.7a) [1, 2, 45, 46]. Virus factories and aggresomes both assemble at the MTOC, recruit mitochondria and cell chaperones, build a cage with vimentin filaments, and are maintained by the activity of dynein motors on microtubules. EM shows membranes, ribosomes, viral intermediates and fully assembled viruses inside the vimentin cage of the ASFV aggresome-like factory (Fig. 14.7b). It is suggested that viruses kidnap the aggresome pathway to avoid being recognised as foreign, or alternatively, to be mistaken for a misfolded protein by the cell, thus triggering the aggresome response [2].

A similar strategy might be used by viruses that replicate and assemble in the cell nucleus and associate with structures known as POD (potential oncogenic domains). POD are nuclear aggresomes used by herpes-, papilloma-, adeno- and parvoviruses. Recent observations suggest that some RNA viruses use aggresomes to build their factories. Since the common feature of factories built by RNA viruses is the remodelling and recruitment of cell endomembranes, however, different signalling pathways must be triggered in this case.

14.4.2 RNA Viruses and Membrane Remodelling

DNA viruses usually build distinct, and even distant factories for genome replication and morphogenesis. Herpesviruses, for example, must connect the first steps of assembly inside the nucleus with subsequent incorporation of proteins and membranes in the cytosol; poxviruses must coordinate replication in cytosolic mini-nuclei with primary assembly in aggresome-like structures and final wrapping in the Golgi apparatus [1]. In contrast, RNA viruses often induce the construction of a single sophisticated membranous web in which replication complexes and assembly sites are located near each other. Expression of viral replicase complexes is usually sufficient to trigger membrane remodelling and organelle recruitment [47, 48].

Viruses take control of cell endomembranes by interfering with lipid metabolism, protein regulation and transport. The secretory pathway is the most common target for this virus-induced membrane remodelling, while the endocytic pathway also participates in some cases (Fig. 14.1). The two pathways are closely related, in fact, and converge at the trans-Golgi compartment. Described in the literature with many different names, the membranous tubuloreticular structures (TBS) often detected in infected cells are indeed cubic membranes that consist of highly curved, 3D-folded lipid bilayers. Alterations in cholesterol metabolism are linked to the

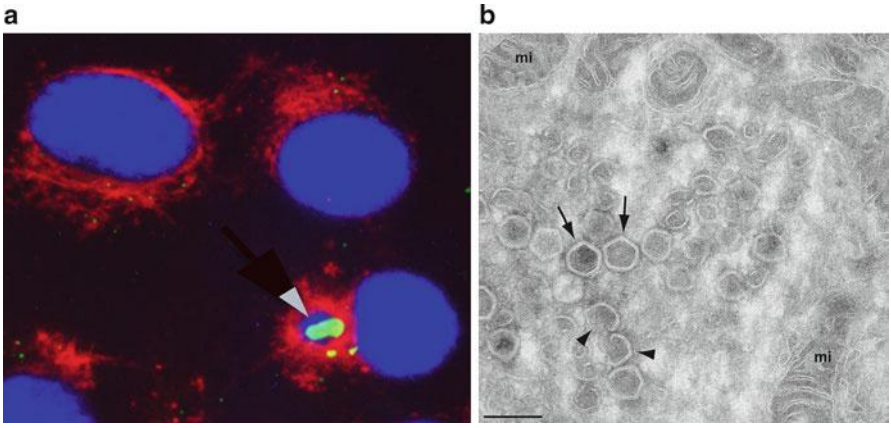

Fig. 14.7 Virus assembly in cell aggresomes. (**a**) The factories of the ASFV resemble cell aggresomes, as shown by immunofluorescence. Viral proteins (*green*) co-localise with DNA (*blue*) inside vimentin cages (*red*). (**b**) TEM of an ASFV factory, showing mitochondria (mi), viral intermediates (*arrowheads*) and mature viruses (*arrows*). Scale bar, 250 nm. Images kindly provided by Dr. Germán Andrés (CBMSO-CSIC, Spain)

biogenesis of cubic membranes, used for assembly by some viruses such as the SARS (severe acute respiratory syndrome) coronavirus [49]. Other viruses depend on phosphatidylcholine synthesis, fatty acid synthesis or geranylation. The highly curved nature of cubic membranes, possibly induced by multiple interactions between replicase proteins, might assist virus budding. Virus replication often slows secretion, however, which means that viruses can interfere with membrane trafficking proteins and their regulation by small GTPases [2, 47]. Additional host factors involved in virus-induced membrane remodelling include the SNARE (soluble NSF attachment protein receptor) proteins, which are mediators of vesicle fusion used by the human cytomegalovirus (HCMV), the VAMP-associated proteins (VAP) that bind to SNARE and are used by hepatitis C virus, and the ESCRT (endosomal sorting complex required for transport), which are essential for HIV assembly in the plasma membrane [19, 47, 50].

Inside membranous webs, virus assembly is precisely coordinated with transport of replicated viral genomes from the RC; the specific machinery remains to be identified. Lessons from plant viruses and their movement proteins involved in viral genome transport inside cells and between cells through plasmodesmata can provide some clues for the study of animal viruses [51].

Finally, signalling related to cytoskeleton-mediated organelle movement, particularly in the case of mitochondria, can be essential for factory assembly and virus morphogenesis. Mitochondria attach to membranous webs and are thought to provide energy for viral factory activities. They can have additional roles, as in infected cells it was observed that some mitochondrial proteins abandon the organelle and are integrated into the factories, where they interact with viral proteins [52, 53].

14.5 Perspectives and Conclusions

Some of the following new methodologies have not yet been applied to the study of viral morphogenesis, but their capacity to track macromolecules in cells suggests great potential for the study of virus assembly.

14.5.1 Super-Resolution Fluorescence Microscopy

Limited by the diffraction barrier, conventional light microscopy methods reach resolutions of ~100–500 nm. Many viruses and cell substructures are smaller than this and cannot be solved by LM. In recent years, several laboratories have developed a group of optical microscopy methods, termed super-resolution microscopy, that have improved spatial resolution by an order of magnitude over the diffraction limit [54]. Future improvements in microscopes, fluorescent probes and labelling chemistry will further refine the resolution of these methods, considerably narrowing the gap between light and electron microscopy. Using a variety of technical strategies, these new technologies have begun to provide insights into cell biology and virology. In particular, real-time imaging methods that track individual virus particles in living cells are being used to study virus assembly, and super-resolution microscopy has already defined interactions with cell factors during HIV-1 budding [55, 56]. In the near future, these microscopies will have an increasing impact in the field of virus morphogenesis.

14.5.2 Correlative Microscopy: From Live Cells to High Resolution

Electron microscopy has contributed more than any other method to our understanding of virus assembly in the cell, although its static nature nonetheless makes it difficult to characterise highly dynamic processes. Correlative light and electron microscopy (CLEM) combines the advantages of live cell imaging with the high resolution of EM. A number of procedures have been reported, and the method of choice depends on the research question [37]. Basically, with CLEM we can select individual live cells with interesting features for a detailed, high-resolution study in TEM and ET. Finding adequate probes for CLEM will be the main technical challenge of these studies.

14.5.3 Clonable Tags for Electron Microscopy and Tomography

Genetically clonable tags for TEM and ET would supply new strategies for the ultrastructural characterisation of virus assembly. To date, two types of approaches have been reported, using photoconversion of fluorescent tags or metal-binding proteins. Diaminobenzidine (DAB) can be photoconverted by production of singlet oxygen from fluorescing proteins to generate dense osmiophilic precipitates that are visible in the electron microscope [57]. The resulting signals are diffuse and lack the resolution of particulate probes, although reasonable results have been obtained for proteins concentrated in cell organelles and in electron tomography [37, 42].

Metal-binding proteins were recently validated as clonable tags for EM in bacteria [58, 59] and mammalian cells [60]. These methods allow detection of intracellular proteins at high sensitivity and molecular scale resolution. To track protein molecules in CLEM, proteins can be double-tagged with a metal-binding peptide and a fluorescent probe. This will give us a new way to look inside cells and visualise where and how individual macromolecules come together to build viral particles.

14.5.4 Conclusions

New technologies have recently begun to offer access to analysis of viruses in cells in unprecedented detail. The complexity of the interaction networks established in these contexts is changing our concept of viruses from that of inert molecular organisms to "live" entities able to carry out a wide variety of activities inside cells.

The different technologies and their integration for the study of virus-cell interactions are summarised in Fig. 14.8. From the conventional techniques used to characterise infection kinetics to the new developments in proteomics, genomics, bioinformatics and microscopy, research in this field is generating a vast amount of information about how viruses manage to assemble all the machinery needed to build infectious particles from viral macromolecules and cell materials. Relevant examples on what is being learned on the participation of host cell factors in virus assembly have been described in this chapter by focusing on studies with a few representative model systems. Future work must be accompanied by studies of how cells position their proteins and regulate organelle shape, size and movement. By understanding how viruses manipulate these processes, we will not only learn about viruses but also about cell architecture and compartmentalisation of functions.

Acknowledgements Our gratitude to Drs German Andrés, Abraham Minsky, Ralf Bartenschlager, Petr Chlanda, Jacomine Krijnse-Locker, Laura Sanz, José Jesús Fernández, Juan Fontana, and Noelia López Montero for providing images, and to Catherine Mark for editorial assistance. Work in C.R.'s laboratory was funded by a grant from the Ministry of Science and Innovation of Spain (BIO2009-07255). I.F.C. is recipient of a fellowship from the FPI Program of the Spanish Ministry of Economy and Competitiveness.

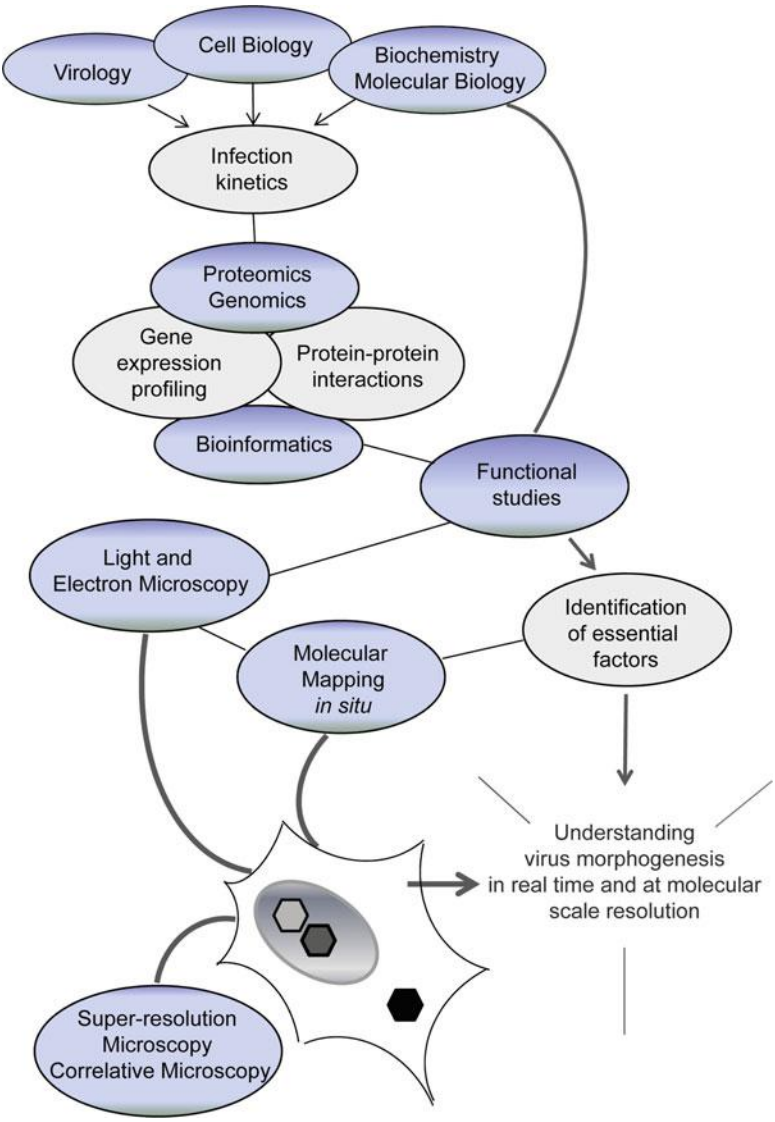

Fig. 14.8 Various technologies and their integration in the study of virus-cell interactions during viral morphogenesis. See text for a description

References and Further Reading

1. Novoa RR, Calderita G, Arranz R, Fontana J, Granzow H, Risco C (2005) Virus factories: associations of cell organelles for viral replication and morphogenesis. Biol Cell 97:147–172
2. Netherton CL, Wileman T (2011) Virus factories, double membrane vesicles and viroplasm generated in animal cells. Curr Opin Virol 1:381–387

3. Nagy PD, Pogany J (2010) Global genomics and proteomics to identify host factors as targets to induce resistance against *tomato bushy stunt virus*. Adv Virus Res 76:123–177
4. Alkhalil A, Hammamieh R, Hardick J, Ichou MA, Jett M, Ibrahim S (2010) Gene expression profiling of monkeypox virus-infected cells reveals novel interfaces for host-virus interactions. Virol J 7:173
5. Friedel CC, Haas J (2011) Virus-host interactomes and global models of virus-infected cells. Trends Microbiol 19:501–508
6. Sessions OM, Barrows NJ, Souza-Neto JA, Robinson TJ, Hershey CL, Rodgers MA, Ramirez JL, Dimopoulos G, Yang PL, Pearson JL, Garcia-Blanco MA (2009) Discovery of insect and human dengue virus host factors. Nature 458:1047–1050
7. Meliopoulos VA, Andersen LE, Birrer KF, Simpson KJ, Lowenthal JW, Bean AG, Stambas J, Stewart CR, Tompkins SM, van Beusechem VW, Fraser I, Mhlanga M, Barichievy S, Smith Q, Leake D, Karpilow J, Buck A, Jona G, Tripp RA (2012) Host gene targets for novel influenza therapies elucidated by high-throughput RNA interference screens. FASEB J 26:1372–1386
8. Pache L, König R, Chanda SK (2010) Identifying HIV-1 host cell factors by genome-scale RNAi screening. Methods 53:3–12
9. Morita E, Sandrin V, McCullough J, Katsuyama A, Hamilton IB, Sundquist WI (2011) ESCRT-III protein requirements for HIV-1 budding. Cell Host Microbe 9:235–242
10. MacFarlane SA, Uhrig JF (2008) Yeast two-hybrid assay to identify host-virus interactions. Methods Mol Biol 451:649–672
11. De Chassey B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, Agaugué S, Meiffren G, Pradezynski F, Faria BF, Chantier T, Le Breton M, Pellet J, Davoust N, Mangeot PE, Chaboud A, Penin F, Jacob Y, Vidalain PO, Vidal M, André P, Raourdin-Combe C, Lotteau V (2008) Hepatitis C virus infection protein network. Mol Syst Biol 4:230
12. Zhang L, Villa NY, Rahman MM, Smallwood S, Shattuck D, Neff C, Dufford M, Lanchbury JS, Labaer J, McFadden G (2009) Analysis of vaccinia virus-host protein-protein interactions: validations of yeast two-hybrid screenings. J Proteome Res 8:4311–4318
13. König R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, Ireland JT, Chiang CY, Tu BP, De Jesus PD, Lilley CE, Seidel S, Opaluch AM, Caldwell JS, Weitzman MD, Kuhen KL, Bandyopadhyay S, Ideker T, Orth AP, Miraglia LJ, Bushman FD, Young JA, Chanda SK (2008) Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell 135:49–60
14. Zhang B, Kraemer B, Sengupta D, Fields S, Wickens M (1999) Yeast three-hybrid system to detect and analyze interactions between RNA and protein. Methods Enzymol 306:93–113
15. Lee E-G, Linial ML (2004) Basic residues of the retroviral nucleocapsid play different roles in Gag-Gag and Gag-Z RNA interactions. J Virol 78:8486–8495
16. Jäger S, Gulbahce N, Cimermancic P, Kane J, He N, Chou S, D'Orso I, Fernandes J, Jan G, Frankel AD, Alber T, Zhou Q, Krogan NJ (2011) Purification and characterization of HIV-human protein complexes. Methods 53:13–19
17. Zhang F, Zang T, Wilson SJ, Johnson MC, Bieniasz PD (2011) Clathrin facilitates the morphogenesis of retrovirus particles. PLoS Pathog 7:e1002119
18. Bartee E, McCormack A, Früh K (2006) Quantitative membrane proteomics reveals new cellular targets of viral immune modulators. PLoS Pathog 2:e107
19. Martin-Serrano J, Neil SJD (2011) Host factors involved in retroviral budding and release. Nat Rev Microbiol 9:519–531
20. Kuhl BJ, Cheng V, Wainberg MA, Liang C (2011) Tetherin and its viral antagonists. J Neuroimmune Pharmacol 6:188–201
21. Gonzalez O, Zimmer R (2011) Contextual analysis of RNAi-based functional screens using interaction networks. Bioinformatics 27:2707–2713
22. Fontana J, López-Montero N, Elliott RM, Fernández JJ, Risco C (2008) The unique architecture of bunyamwera virus factories around the Golgi complex. Cell Microbiol 10:2012–2028

23. Katsafanas GC, Moss B (2007) Colocalization of transcription and translation within cytoplasmic poxvirus factories coordinates viral expression and subjugates host functions. Cell Host Microbe 2:221–228
24. Sivaraman D, Biswas P, Cella LN, Yates MV, Chen W (2011) Detecting RNA viruses in living mammalian cells by fluorescence microscopy. Trends Biotech 29:307–313
25. Bozzola JJ, Russell LD (1999) Electron microscopy. Principles and techniques for biologists. Jones and Bartlett Publishers, Inc, Boston
26. Goldsmith CS, Miller SE (2009) Modern uses of electron microscopy for detection of viruses. Clin Microbiol Rev 22:552–563
27. Fiala JC (2005) Reconstruct: a free editor for serial section microscopy. J Microsc 218:52–61
28. Schatten H (2011) Low-voltage high-resolution SEM (LVHRSEM) for biological structural and molecular analysis. Micron 42:175–185
29. Zauberman N, Mutsafi Y, Halevy DB, Shimoni E, Klein E, Xiao C, Sun S, Minsky A (2008) Distinct DNA exit and packaging portals in the virus *Acanthamoeba polyphaga* mimivirus. PLoS Biol 6:e114
30. Risco C, Rodríguez JR, López-Iglesias C, Carrascosa JL, Esteban M, Rodríguez D (2002) Endoplasmic reticulum-Golgi intermediate compartment membranes and vimentin filaments participate in vaccinia virus assembly. J Virol 76:1839–1855
31. Briggs JAG, Kräusslich H-G (2011) The molecular architecture of HIV. J Mol Biol 410:491–500
32. Welsch S, Miller S, Romero-Brey I, Merz A, Bleck CKE, Walther P, Fuller SD, Anthony C, Krijnse-Locker J, Bartenschlager R (2009) Composition and three-dimensional architecture of the dengue virus replication and assembly sites. Cell Host Microbe 5:365–375
33. Chlanda P, Carbajal MA, Cyrklaff M, Griffiths G, Krijnse-Locker J (2009) Membrane rupture generates single open membrane sheets during vaccinia virus assembly. Cell Host Microbe 6:81–90
34. Cyrklaff M, Risco C, Fernández JJ, Jiménez MV, Esteban M, Baumeister W, Carrascosa JL (2005) Cryo-electron tomography of vaccinia virus. Proc Natl Acad Sci USA 102:2772–2777
35. Kiernan JA (2008) Histological and histochemical methods. Theory and practice. Scion Publishing Ltd, Oxfordshire
36. Hayat MA (2002) Microscopy, immunohistochemistry, and antigen retrieval methods. Kluwer Academic//Plenum Publishers, New York
37. Van Weering JRT, Brown E, Sharp TH, Mantell J, Cullen PJ, Verkade P (2010) Intracellular membrane traffic at high resolution. Methods Cell Biol 96:619–648
38. Tokuyasu KT (1973) A technique for ultracryotomy of cell suspensions and tissues. J Cell Biol 57:551–565
39. Slot JW, Geuze HJ (2007) Cryosectioning and immunolabeling. Nat Protoc 2:2480–2491
40. Sodeik B, Doms RW, Ericsson M, Hiller G, Machamer CE, van't Hof W, van Meer G, Moss B, Griffiths G (1993) Assembly of vaccinia virus: role of the intermediate compartment between the endoplasmic reticulum and the Golgi stacks. J Cell Biol 121:521–541
41. Jouvenet N, Simon SM, Bieniasz PD (2011) Visualizing HIV-1 assembly. J Mol Biol 410:501–511
42. Shu X, Lev-Ram V, Deerinck TJ, Qi Y, Ramko EB, Davidson MW, Jin Y, Ellisman MH, Tsien RY (2011) A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues and organisms. PLoS Biol 9:e1001041
43. Gaietta G, Deerinck TJ, Admas SR, Bouwer J, Tour O, Laird DW, Sosinsky GE, Tsien RY, Ellisman MH (2002) Multicolor and electron microscopic imaging of connexin trafficking. Science 296:503–507
44. La Scola B, Desnues C, Pagnier I, Robert C, Barrasi L, Fournous G, Merchat M, Suzan-Monti M, Forterre P, Koonin E, Raoul D (2008) The virophage as a unique parasite of the giant mimivirus. Nature 455:100–104
45. Carvalho ZG, Alves de Matos AP, Rodrigues-Pousada C (1988) Association of African swine fever virus with the cytoskeleton. Virus Res 11:175–192

46. Rojo G, Chamorro M, Salas ML, Viñuela E, Cuezva JM, Salas J (1998) Migration of mitochondria to viral assembly sites in African Swine fever virus-infected cells. J Virol 72:7583–7588
47. Pierini R, Cottam E, Roberts R, Wileman T (2009) Modulation of membrane traffic between endoplasmic reticulum, ERGIC and Golgi to generate compartments for the replication of bacteria and viruses. Sem Cell Dev Biol 20:828–833
48. Fontana J, Tzeng WP, Calderita G, Fraile-Ramos A, Frey TK, Risco C (2007) Novel replication complex architecture in rubella replicon-transfected cells. Cell Microbiol 9:875–890
49. Deng Y, Almshwerqi ZA, Ng MML, Kohlwein SD (2010) Do viruses subvert cholesterol homeostasis to induce host cubic membranes? Trends Cell Biol 20:371–379
50. Liu STH, Sharon-Friling RS, Ivanova P, Milne SB, Myers DS, Rabinowitz JD, Brown HA, Shenk T (2011) Synaptic vesicle-like lipidome of human cytomegalovirus virions reveals a role for SNARE machinery in virion egress. Proc Natl Acad Sci USA 108:12869–12874
51. Harries P, Ding B (2011) Cellular factors in plant virus movement: at the leading edge of macromolecular trafficking in plants. Virology 411:237–243
52. Fontana J, López-Iglesias C, Tzeng WP, Frey TK, Fernández JJ, Risco C (2010) Three-dimensional structure of rubella virus factories. Virology 405:579–591
53. Ilkow CS, Weckbecker D, Cho WJ, Meier S, Beatch MD, Goping IS, Herrmann JM, Hobman TC (2010) The rubella virus capsid protein inhibits mitochondrial import. J Virol 84:119–130
54. Huang B, Babcock H, Zhuang X (2011) Breaking the diffraction barrier: super-resolution imaging of cells. Cell 143:1047–1058
55. Jouvenet N, Zhadina M, Bieniasz PD, Simon SM (2011) Dynamics of ESCRT protein recruitment during retroviral assembly. Nat Cell Biol 13:394–401
56. Lehmann M, Rocha S, Mangeat B, Blanchet F, Uji-I H, Hofkens J, Piguet V (2011) Quantitative multicolor super-resolution microscopy reveals tetherin HIV-1 interaction. PLoS Pathog 7: e1002456
57. Monosov EZ, Wenzel TJ, Luers GH, Heyman JA, Subramani S (1996) Labeling of peroxisomes with green fluorescent protein in living *P. pastoris* cells. J Histochem Cytochem 44:581–589
58. Diestra E, Fontana J, Guichard P, Marco S, Risco C (2009) Visualization of proteins in intact cells with a clonable tag for electron microscopy. J Struct Biol 165:157–168
59. Wang Q, Mercogliano CP, Löwe J (2011) A ferritin-based label for cellular electron cryotomography. Structure 19:147–154
60. Risco C, Sanmartín-Conesa E, Tzeng WP, Frey TK, Seybold V, de Groot RJ (2012) Specific, sensitive, high-resolution detection of protein molecules in eukaryotic cells using metal-tagging transmission electron microscopy. Structure 20:759–766

Further Reading

Baumgärtel V, Ivanchenko S, Dupont A, Sergeev M, Wiseman PW, Kräusslich HG, Bräuchle C, Müller B, Lamb DC (2011) Live-cell visualization of dynamics of HIV budding site interactions with an ESCRT component. Nat Cell Biol 13:469–474

Claverie JM, Abergel C (2009) Mimivirus and its virophage. Annu Rev Genet 43:49–66

Erickson KD, Bouchet-Marquis C, Heiser K, Szomolanyi-Tsuda E, Mishra R, Lamothe B, Hoenger A, Garcea RL (2012) Virion assembly factories in the nucleus of polyomavirus-infected cells. PLoS Pathog 8:e100263

Fogarty KH, Zhang W, Grigsby IF, Johnson JL, Chen Y, Mueller JD, Mansky LM (2011) New insights into HTLV-1 particle structure, assembly and Gag-Gag interactions in living cells. Viruses 3:770–793

Fu C, Johnson J (2011) Viral life cycles captured in three-dimensions with electron microscopy tomography. Curr Opin Virol 1:125–133

Iwasaki K, Omura T (2010) Electron tomography of the supramolecular structure of virus-infected cells. Curr Opin Struct Biol 20:632–639

Mutsafi Y, Zauberman N, Sabanay O, Minsky A (2010) Vaccinia-like cytoplasmic replication of the giant mimivirus. Proc Natl Acad Sci USA 107:5978–5982

Nagy PD, Pogany J (2012) The dependence of viral RNA replication on co-opted host factors. Nat Rev Microbiol 10:137–149

Otto A, Bernhardt J, Hecker M, Becher D (2012) Global relative and absolute quantitation in microbial proteomics. Curr Opin Microbiol 15:1–9

Rust M, Lakadamyali M, Brandenburg B, Zhuang X (2008) Single-virus tracking in live cells. In: Selvin PR, Ha T (eds) Single molecule techniques: a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Van Vliet K, Mohamed MR, Zhang L, Villa NY, Werden SJ, Liu J, McFadden G (2009) Poxvirus proteomics and virus-host protein interactions. Microbiol Mol Biol Rev 73:730–749

Also especially recommended for further reading are references [1, 2, 3, 5, 19, 37, 49, 54] *listed above*.

Chapter 15
Virus-Receptor Interactions and Receptor-Mediated Virus Entry into Host Cells

José M. Casasnovas

Abstract The virus particles described in previous chapters are vehicles that transmit the viral genome and the infection from cell to cell. To initiate the infective cycle, the viral genome must therefore translocate from the viral particle to the cytoplasm. *Via* distinct proteins or motifs in their outermost shell, the particles attach initially to specific molecules on the host cell surface. These virus receptors thus mediate penetration of the viral genome inside the cell, where the intracellular infective cycle starts. The presence of these receptors on the cell surface is a principal determinant of virus host tropism. Viruses can use diverse types of molecules to attach to and enter into cells. In addition, virus-receptor recognition can evolve over the course of an infection, and virus variants with distinct receptor-binding specificities and tropism can appear. The identification of virus receptors and the characterization of virus-receptor interactions have been major research goals in virology for the last two decades. In this chapter, we will describe, from a structural perspective, several virus-receptor interactions and the active role of receptor molecules in virus entry.

Keywords Virus-host • Virus tropism • Virus attachment • Virus-receptor • Virus structure • Virus entry • Virus neutralization • Capsid dynamics • Uncoating • Membrane penetration • Endocytosis • Cell surface • Cell adhesion • Cell surface molecules • Membrane proteins • Glycoproteins • Carbohydrates • Protein interactions • Molecular recognition • Crystallography • Cryo-EM

J.M. Casasnovas (✉)
Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: jcasasnovas@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_15,

Abbreviations

Ad	Adenovirus
APN	Aminopeptidase N
CAR	Coxsackievirus-adenovirus receptor
cryo-EM	Cryo-electron microscopy
CV	Coxsackievirus
D	Domain
DAF	Decay-accelerating factor
DC	Dendritic cells
DC-SIGN	(DC-specific ICAM-3-grabbing nonintegrin)
EFN	Ephrin
EV	Echovirus
FMDV	Foot-and-mouth disease virus
g/gp	Glycoprotein
H	Hemagglutinin
HA	Influenza A hemagglutinin
HeV	Hendra virus
HIV-1	Human immunodeficiency virus-type 1
HN	Hemagglutinin neuraminidase
HRV	Human rhinovirus
HS	Heparan sulphate
HSV	Herpes simplex virus
ICAM-1	Intercellular adhesion molecule-1
IgSF	Immunoglobulin superfamily
LDLR	Low-density lipoprotein receptor
MCP	Membrane cofactor protein
MV	Measles virus
N	Neuraminidase
NAG	N-acetyl-glucosamine
NDV	Newcastle disease virus
NiV	Nipah virus
PM	Plasma membrane
PV	Poliovirus
PVR	Poliovirus receptor
SCR	Short consensus repeats
SLAM	Signalling lymphocytic activation molecule
VP	Viral capsid protein.

15.1 Introduction: Virus Entry into Host Cells, the Recognition of Cell Surface Molecules

The viral particles formed in infected host cells (see Chaps. 10, 11, 12, 13, and 14) are metastable structures that transmit the viral genome and the infection from cell to cell. Viruses must therefore penetrate host cells to initiate the replicative infective cycle by exploiting the cell machinery. In the extracellular transit stage of the viral cycle, animal viruses and bacteriophages attach to specific cell surface molecules (virus receptors) suited for host cell entry following virus-specific entry pathways. Virus receptors must be distinguished from attachment factors, surface molecules to which some viruses can bind but that do not themselves promote virus entry into host cells [1]. Virus binding to attachment factors concentrates virus particles onto the cell surface, which can help viruses to encounter specific entry receptors that mediate genome translocation into the cytoplasm. The virus receptor molecules are not just required for initial virus binding to host cells, but also for the transfer of the viral genome through cellular membranes [1]. Virus-receptor interactions can trigger changes in the virus particles that initiate genome translocation, or alterations in the cell, such as signalling events that facilitate virus entry [2]. Viruses enter host cells at the cell *surface* or after endocytosis (Fig. 15.1). Multivalent binding of the virus particles to receptors on the cell surface can mediate uncoating or release of the viral genome in non-enveloped viruses and/or fusion of the virus and cell membranes (Fig. 15.1). Moreover, viruses bound to cell surface molecules can be internalized by following different endocytic pathways [3], where exposure to low pH, enzymatic modification or other cellular factors leads to the delivery of the genome into the cytoplasm (Fig. 15.1).

Viruses evolve to recognize specific cell surface receptor molecules appropriate for productive entry and infection of host cells, which frequently determines the host tropism or the cell type a virus can infect. Selection of cell entry receptor by viruses appears to be determined by subtle interactions that regulate the specificity and affinity necessary for efficient cell attachment. Virus-receptor interaction can nonetheless be a highly dynamic process. A single virus can recognize one or several cell entry molecules, which can also differ among virus variants or during the course of an infection [4, 5]. Virus recognition of receptors is under continuous evolutionary pressure to increase their infection efficiency, which can lead to the emergence of virus variants with altered infectivity or tissue tropism.

In this chapter, several examples of animal virus-receptor interactions will be presented, together with a description of known models of receptor-switching viruses. The chapter presents a structural view of some virus-receptor interactions that have been characterized by structures of complexes. We will also describe how the viral genome exits the capsid (uncoating) in some non-enveloped animal viruses, illustrating the role of cell surface receptor molecules in the entry process. Membrane penetration events in enveloped viruses are discussed in Chap. 16. Specific aspects of receptor recognition and injection-mediated genome uncoating by some bacteriophages are described in Chap. 17.

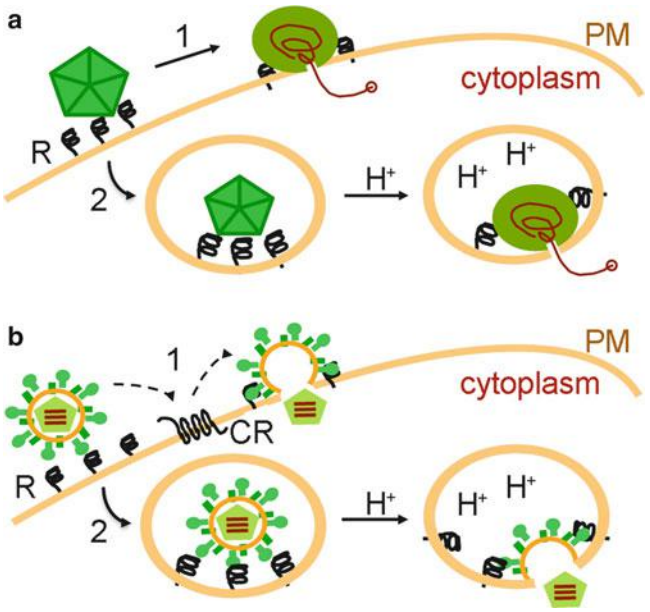

Fig. 15.1 Virus entry into host cells. Scheme illustrating two cell entry pathways by non-enveloped (**a**) or enveloped viruses (**b**). pH-independent entry at the cell surface (*pathway 1*): Genome uncoating and penetration into cytoplasm (**a**) or virus-cell membrane fusion (**b**) at the plasma membrane (PM). Virus binding to a single receptor (R) or to additional co-receptor molecules (CR) at neutral pH triggers genome translocation. pH-dependent entry in endosomes (*pathway 2*): Receptor-mediated endocytosis of viruses prior to penetration into the cytoplasm. In the endosomes, viruses firmly bound to multiple receptor molecules are exposed to progressively lower pH, which triggers genome uncoating and penetration (**a**) or fusion (**b**). The entry reaction is mediated by acidification of endosomal compartments, with or without receptor priming

15.2 Virus-Receptor Interactions and Receptor Specificity Switch

The known number and diversity of cell surface molecules exploited by viruses to enter host cells is still increasing [6]. Viruses recognize a variety of cell surface molecules specifically, including glycolipids, carbohydrates and proteins with very distinct folding structures. There are viruses specific for a single receptor molecule, whereas others bind to several structurally distinct receptors for host cell entry [2]. The virus particles use certain proteins in their outermost shell to attach to the cell surface molecules. The multivalent nature of the particles allows viruses to bind with very high avidity to the host cell surface, even though monomeric virus-receptor interactions are usually of moderate affinity (0.1–1 μM) [7].

Some non-enveloped viruses have proteins specialized in attachment to cell surface receptors; however, in other cases unique motifs on the naked capsids are engaged in receptor recognition. Here we discuss both receptor recognition modes,

providing a description of adenovirus and picornavirus binding to cell surface molecules. The enveloped viruses bear membrane-bound glycoproteins that bind to the receptor and trigger virus-cell membrane fusion during the entry process, as will be described in Chap. 16. We present the diverse types of cell surface receptors recognized by human immunodeficiency virus and paramyxoviruses, which have been characterized. In the last part of this section, we illustrate the importance of cell surface carbohydrates in virus attachment and infection of host cells.

15.2.1 Virus-Receptor Interactions in Non-Enveloped Viruses

In this subsection, we describe receptor recognition by two families of non-enveloped viruses (picornaviruses and adenoviruses) for which virus-receptor interactions have been characterised in detail by structural studies. Picornaviruses use their naked capsid to attach to receptors, whereas adenoviruses have fibres that protrude from the capsid for attachment to several cell surface molecules.

Picornaviruses and Their Receptors

The picornaviruses are a large family of non-enveloped viruses responsible for numerous human and animal diseases. Picornavirus particles are formed by an icosahedral protein capsid built by three external viral capsid proteins (VP1 to VP3) and the internal protein VP4 packed inside with a single-stranded RNA genome (Fig. 15.2a). The capsid is composed of 60 basic subunits or protomers arranged as 12 pentamers (see Chap. 10). The members of this virus family bind to distinct types of receptor molecules suited for entry into host cells [8].

The poliovirus receptor (PVR) was one of the first picornavirus receptors to be characterized [9]. PVR is a type I membrane protein and a member of the immunoglobulin superfamily (IgSF), which has three Ig-like domains (D1 to D3) at the extracellular region (Fig. 15.2a). The receptor for most (90 %) identified human rhinovirus (HRV) serotypes, the major group of HRV, is intercellular adhesion molecule-1 (ICAM-1) [10, 11], another IgSF member with five Ig-like extracellular domains (Fig. 15.2a). PV and HRV bind similarly to the N-terminal membrane distal domains of the receptor molecules (Fig. 15.2b) [8]. Both viruses use a depressed surface, or canyon, formed by two neighbouring protomers around the five-fold icosahedral vertices of the capsid (Figs. 15.2a, b). Cryo-electron microscopy (cryo-EM; see Chap. 3) and binding studies show some differences in the way PV and HRV bind to their IgSF receptors. Kinetics for monomeric receptor binding to the virus particles showed more rapid kinetic binding rates to PV than to HRV [12]. Moreover, cryo-EM structures of virus-receptor complexes showed that PV uses more exposed residues on the canyon walls than HRV [8, 12].

The use of the canyon for binding to cell surface receptors was also described for other picornaviruses such as the coxsackieviruses (CV) A21 and B3, and echovirus 1 (EV1) [8, 13]. CVA21 binds to ICAM-1, CVB3 to the coxsackievirus-adenovirus receptor (CAR), and echovirus 1 (EV1) to the $\alpha_2\beta_1$ integrin. The CAR protein also

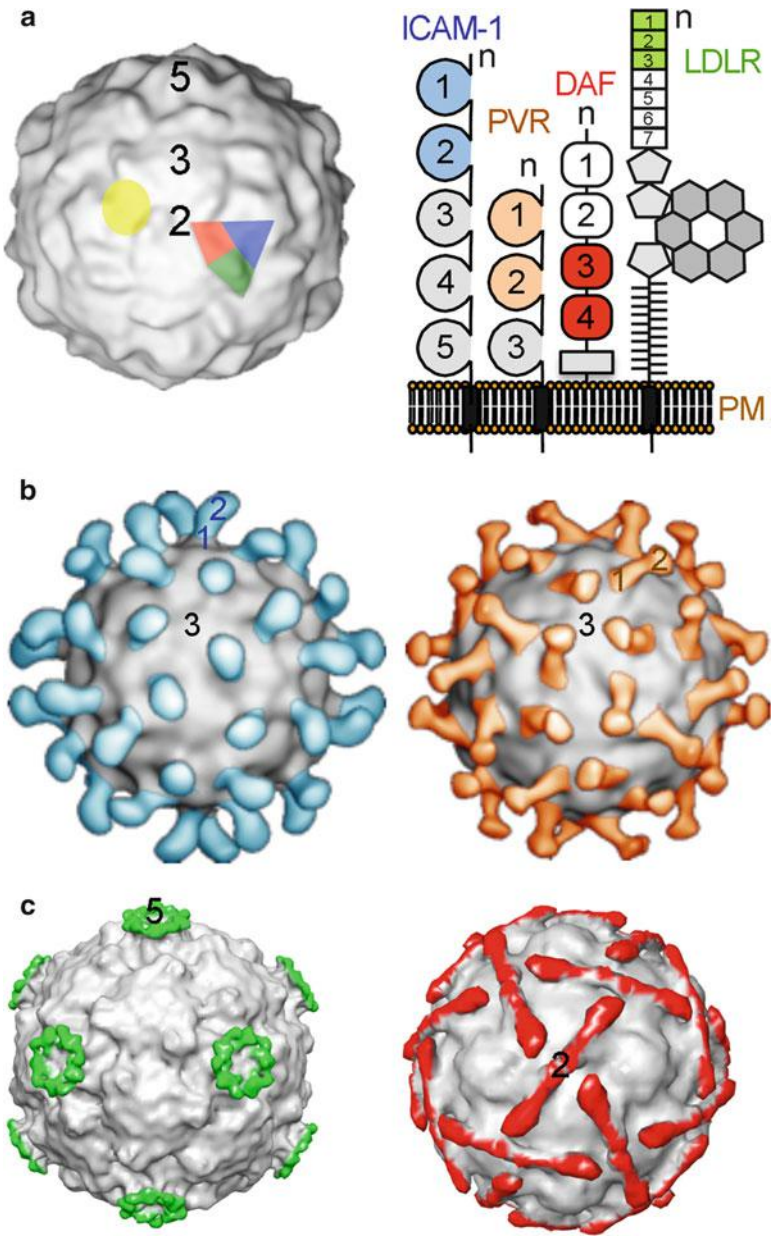

Fig. 15.2 Virus-receptor interactions in picornaviruses. (**a**) *Left*: Surface representation of the poliovirus (PV) particle computed from the crystal structure of PV1/M (PDB ID 1ASJ). Numbers indicate location of the icosahedral two-, three- and five-fold axes. The *triangle* illustrates the approximate location of a capsid building block or protomer, assembled from the external VP1 (*blue*), VP2 (*green*) and VP3 (*red*). The canyon region is marked with a *yellow circle*. *Right*: Scheme of picornavirus receptors ICAM-1, PVR, DAF and LDLR. The N-terminus (n) and

belongs to the IgSF, but the $\alpha_2\beta_1$ integrin is a non-IgSF member that fits into the picornavirus canyon. Cryo-EM structural studies showed that the I-domain of the integrin penetrates the EV1 capsid canyon [13], as described for PV, HRV and CVA21 binding to IgSF receptor molecules. Thus, the canyon is a well-defined receptor-binding region in picornaviruses [8]. The concave nature of the canyon is thought to be suited for hiding conserved receptor-binding residues from neutralizing antibodies, which cannot penetrate the depressed surface as efficiently as relatively narrower receptor molecules. This hypothesis was nonetheless challenged by studies showing an antibody that penetrates the canyon [14], although it interacts mostly with the walls rather than the bottom of the canyon. Receptor-binding residues in PV are more exposed on the canyon walls and can thus be targeted by antibodies elicited by PV vaccines [8, 12]. The use of recessed virus surfaces in receptor recognition is relatively common in virus-receptor interactions, and could be a viral strategy to protect some receptor-binding residues from immune surveillance and/or might have arisen to increase surface contact area and binding energy for cell receptor molecules.

Other picornaviruses do not use the canyon for attachment to cell entry receptors; viruses of this family, such as foot-and-mouth disease virus (FMDV), some HRV and echoviruses (EV), use exposed regions of the capsid. FMDV exploits an RGD motif exposed in the VP1 GH loop to bind to integrin receptors [15]. This receptor-binding motif is accessible and is a major antigenic site; however, virus can escape from immune neutralization by mutating some residues adjacent to the RGD, which prevents detection by some antibodies while preserving cell attachment activity and viability. Approximately 10 % of HRV serotypes, the minor group, do not bind to ICAM-1, and use members of the low density lipoprotein receptor (LDLR) family to enter host cells (Fig. 15.2a) [16]. Even though minor and major group HRV are closely related, the mode by which they bind their respective receptors is strikingly different (Figs. 15.2b, c). Minor group HRV do not use the canyon, and contact the receptor through a protruding region close to the capsid five-fold axis (Fig. 15.2c, left) [17]. These HRV bear a conserved Lys residue in VP1 that contacts an acidic cluster and a Trp residue in the N-terminal ligand binding domains of the LDLR proteins. This lysine is absent in major group HRV that bind to ICAM-1, and is therefore likely to be a major determinant of the distinct receptor-binding specificity described for rhinoviruses.

Fig. 15.2 (continued) extracellular domains are labelled. Receptor domains used to determine structures of virus-receptor complexes shown below are coloured. (**b**) Cryo-EM structure of HRV3-ICAM-1 (*left*) and PV1-PVR (*right*), picornaviruses that use the canyon for receptor binding. The complexes were prepared in solution with purified virus particles and two-domain (1 and 2) fragments of the receptor molecules. Location of the domains is shown. D1 penetrates the canyon, whereas D2 does not contact the virus and protrudes from the capsid. Images provided by Holland Cheng and Li Xing, adapted from references [12, 79]. (**c**) Cryo-EM structure of HRV2-LDLR (EMD-1049) and EV12-DAF (EMD-1057) complexes, representative of picornaviruses that do not use the canyon for receptor binding. Ligand binding repeats 1–3 of the very low-density lipoprotein receptor or domains 3–4 of DAF were used to prepare complexes. Surfaces of bound receptors are coloured as in (**a**). Images prepared with Chimera (cgl.ucsf.edu/chimera) from cryo-EM maps

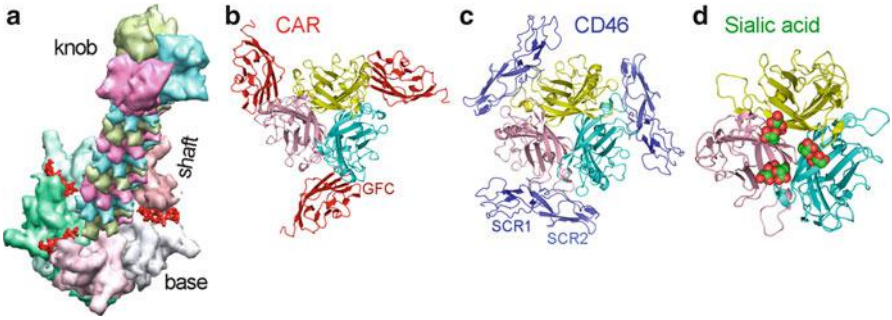

Fig. 15.3 Virus-receptor interactions in adenoviruses. Representation of the receptor-binding adenovirus fibre and penton base (**a**) (see also Chap. 11), and crystal structures of the trimeric Ad knob in complex with CAR (**b**), CD46 (**c**), or sialic acid (**d**). Ribbon drawings are shown for the complexes prepared with the N-terminal (D1) of CAR (*red*) and Ad12 (PDB ID 1KAC), or with the two N-terminal short consensus repeats (SCR1, 2) of CD46 (*blue*) and Ad11 (PDB ID 2O39). The Ad-binding GFC β-sheet of the CAR domain is indicated. Sialic acids in complex with the Ad37 knob (PDB ID 1UXA) are shown with carbons (*green*) and oxygens (*red*) as spheres. Figure (a) was provided by Carmen San Martín and figures (b–d) were prepared with PyMOL (pymol.org)

A picornavirus-receptor interaction different from those described above was reported for EV using decay-accelerating factor (DAF, CD55) as a cell entry receptor [18]. These viruses bind to DAF through a capsid protrusion at the southern rim of the canyon, distant from the five-fold axis and near the capsid two-fold axis (Fig. 15.2c, right) [19, 20]. The receptor-binding region in EV11 protrudes and is more exposed than the canyon. The kinetic association rate for the monomeric EV11-CD55 interaction is thus one to two orders of magnitude higher than those shown for picornaviruses that use the canyon for receptor recognition, which probably reflects the distinct nature of the virus-receptor interactions [21].

The structural studies of picornavirus-receptor complexes discussed here show that viruses of this family can use diverse capsid regions for recognition of cell surface molecules and for productive entry into host cells.

Adenoviruses, Non-enveloped Viruses That Bind to Several Cell Surface Molecules

Adenoviruses (Ad) are non-enveloped viruses with icosahedral capsids built by hexon and penton capsomers (see Chap. 2 for definition of capsomers and Chap. 11 for additional information and references). The pentons locate at each of the 12 icosahedral vertices of the capsid, and forms the base of protruding trimeric fibres (Fig. 15.3a). The penton base and the associated fibres form a complex that recognises cell surface receptor molecules [22, 23]. The most capsid-distal, globular region of the fibre, the knob (Fig. 15.3a), mediates initial Ad attachment to the cell, whereas the penton base is used for subsequent tight attachment of the virus to a secondary receptor that mediates virus internalization and host cell entry [22]. Receptor-bound Ad are transported to endosomes, where acidification triggers capsid disassembly and virion penetration into the cytoplasm.

Cell tropism and receptor recognition is well documented for Ad. This group of non-enveloped viruses can use at least three distinct types of receptor molecules for initial attachment to the cell surface through the fibre proteins [23], which is mediated by the most distal trimeric knob structure (Fig. 15.3). Subgroups A, C, D, E and F attach to the CAR receptor, a member of the IgSF found in epithelial tight junctions. CAR has two Ig-like domains at the extracellular region, and mediates homotypic cell-cell interactions through the same face (GFC β-sheet) of the N-terminal domain engaged by the Ad fibres (Fig. 15.3b) [24, 25]. Fibre binding to CAR interferes with homotypic cell-cell adhesion, which destabilizes epithelial cell layers and facilitates virus release to the airway lumen and spreading to a new host [26].

Most subgroup B Ad use the ubiquitous membrane cofactor protein (MCP, CD46) for initial attachment to the host cell [22]. The Ad knob contacts the two N-terminal short consensus repeats (SCR) of the CD46 molecule (Fig. 15.3c) [27]. Although all Ad use the fibre knob for receptor binding, the mode of contact of the knobs with the CAR or the CD46 proteins is distinct (Fig. 15.3). The switch in receptor-binding specificity between CAR and CD46-binding Ad appear to be related to the distinct conformation of the exposed loops in the periphery of the knob. CD46-binding Ad have an extended knob AB-loop that prevents binding to the CAR receptor, as well as specific structural features in the loops that contact the CD46 molecule [5]. Certain Ad can also attach to cell surface carbohydrates, such as heparin or sialic acid [22, 28, 29]. Subgroup D adenoviruses 8, 19 and 37, which recognize sialic acid, are responsible for epidemic keratoconjunctivitis, a very contagious ocular disease. The sialic acid binding site has been mapped also at the fibre knob, but it locates closer to the centre of the trimer than the sites used to bind to the other receptors (Fig. 15.3c) [28, 29]. The Ad-receptor interaction described here illustrate how different Ad evolved to use distinct knob surfaces to attach to different cell surface receptor molecules; this certainly translates in cell tropism and pathogenesis diversity among the member of this virus family.

Following initial fibre attachment to cell surface receptors, the Ad particles engage cell surface integrin molecules using RGD motifs exposed in loops at the capsid penton base, which is a necessary step for host cell entry by endocytosis [30]. The multivalent interaction of the Ad particle with the integrins mediates its clustering, triggering intracellular signals, rearrangement of actin, and clathrin-mediated endocytosis of the virions [22, 23]. During endocytosis, the virion disassembles and subsequently penetrates the membrane. Ad are thus an example of non-enveloped viruses that use two distinct capsid structures to bind different cell surface receptor molecules for attachment to or penetration of host cells.

15.2.2 Virus-Receptor Interactions in Enveloped Viruses

Enveloped viruses bear membrane-bound glycoproteins specialized in the recognition of cell surface molecules and in subsequent fusion of the viral and cell membranes. The receptor-binding and fusion proteins can be one or two distinct

polypeptides, in many cases resulting from the cleavage of a single precursor protein. Both of these proteins associate in the viral envelope, forming the envelope spikes. Among enveloped viruses, the fusion proteins are less diverse than receptor-binding proteins and will be presented in Chap. 16. Here we will discuss the interactions of some viral envelope spikes with cell surface receptor molecules, described by structures of virus-receptor complexes.

The Human Immunodeficiency Virus Cell Attachment Process

Attachment of the human immunodeficiency virus-type 1 (HIV-1) to cell entry receptors and subsequent fusion have been described in great detail. The HIV-1 particle bears two non-covalently bound viral glycoproteins, gp120 and gp41, which associate to form trimeric envelope spikes [31, 32]. There are around 15 spikes per particle [32]. The gp120 glycoprotein mediates attachment of the HIV-1 particles to cell surface receptors, whereas gp41 catalyses fusion of the virus and the cell membrane, and the release of the nucleoprotein to the cell cytoplasm [31, 33] (see Chap. 16). The HIV-1 entry process is relatively complex [33]. gp120 requires engagement of two distinct cell surface molecules to trigger membrane fusion, which occurs at the cell plasma membrane [34]. gp120 initially engages the lymphocyte-specific CD4 [33], an IgSF cell surface protein. CD4 is a type I membrane protein composed of four Ig-like domains in the extracellular region that forms part of the T cell receptor complex in a subset of T lymphocytes. A depressed groove or pocket in gp120 surrounded by hypervariable loops engages the CD4 N-terminal Ig-like domain with high affinity (Fig. 15.4) [35]. A Tyr residue protruding from the CD4 domain penetrates deeply into the gp120 pocket. The CD4-binding site is surrounded by long hypervariable loops and glycans, which further hide the depressed site from immune surveillance [36].

Binding of gp120 to CD4 triggers conformational changes in the HIV-1 protein that expose a conserved structure (bridging sheet) that binds to a second molecule termed co-receptor [35]. Engagement of the co-receptor molecule is essential for virus-cell membrane fusion and for HIV-1 penetration into the host cell. Two distinct HIV-1 co-receptors have been identified, CXCR4 and CCR5, which are chemokine receptors containing seven transmembrane segments [37]; HIV-1 interaction with CCR5 and CXCR4 is strain-specific. HIV-1 strains using CCR5 (R5 viruses) are mainly associated to sexual transmission. They can also evolve to infect T cells by acquiring the ability to use the CXCR4 receptor (X4 viruses). HIV-1 viruses that use both co-receptors are termed R5X4 viruses.

HIV-1 particles are also able to attach specifically to cell surface molecules that do not mediate cell entry and infection, but are used instead to present the virus to CD4-expressing T cells. This process, termed trans-infection, is very efficient and occurs when antigen-presenting dendritic cells (DC) encounter T cells [38]. HIV-1 can therefore bind to DC in the periphery, which transport the virus to lymphoid organs, where it is transferred to T cells that become infected. The C-type lectin DC-SIGN (DC-specific ICAM-3-grabbing nonintegrin) was the first HIV-1 receptor in DC to be characterized as a mediator of trans-infection [39]. DC-SIGN is specific for simple high-mannose glycans and fucose-containing glycosylations

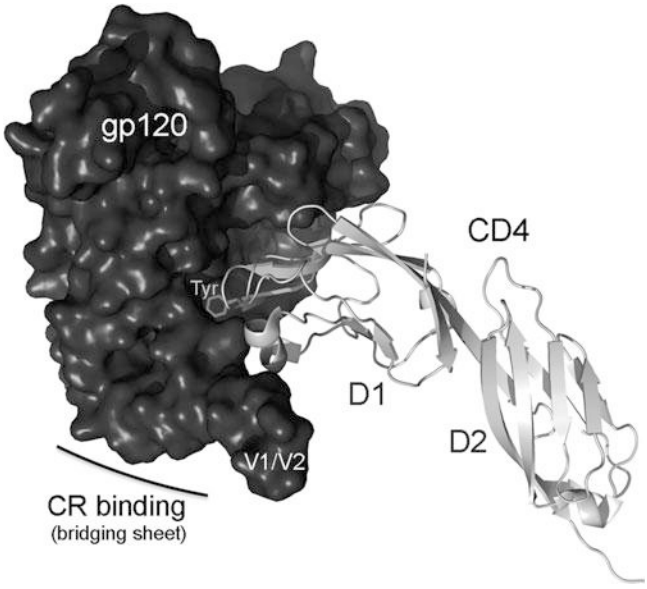

Fig. 15.4 HIV-1 receptor recognition. Crystal structure of the HIV-1 gp120 protein (*dark surface*) bound to the CD4 receptor (*ribbon drawing in grey*) (PDB ID 1GC1). The N-terminal domain of CD4 (D1) fits into a concave gp120 surface, with a deeply penetrating Tyr residue. The location is indicated of the truncated V1/V2 loop and the bridging sheet region that binds to HIV-1 entry co-receptors (CR). CR binding is necessary for virus-cell fusion at the plasma membrane. This and following Figs. 15.5 and 15.6 were prepared with PyMOL (pymol.org)

[40, 41]. gp120 is heavily glycosylated; its glycan composition is mostly of the high-mannose type, which is recognized by DC-SIGN. The glycan shield of the HIV-1 spike is thus used for attachment to DC that trans-infect T cells. It was recently shown that, in addition to the spike glycosylations, HIV-1 also uses charged glycans in its envelope lipidome to attach to DC [42, 43]. The HIV-1 envelope membrane is enriched in certain glycosphingolipids called gangliosides. A specific sialic acid-containing subset of gangliosides appears to recognize receptors on the DC surface [42, 43]. HIV-1 thus exploits different envelope components to attach to distinct cell types, which leads to spread of the infection.

Cell Receptor Recognition in Paramyxovirus

The paramyxoviruses are enveloped, negative-stranded RNA viruses that include serious human and animal pathogens [44]. In viruses of this family, cell attachment and virus-cell membrane fusion are mediated by two distinct membrane glycoproteins (see also Chap. 16). The paramyxovirus attachment proteins are type II membrane proteins anchored to the virus envelope by a single transmembrane domain [44, 45]. Their extracellular region can be divided into an N-terminal stalk region that serves as a spacer and a C-terminal globular domain with receptor binding activity. The C-terminal globular domains of paramyxovirus attachment

glycoproteins all fold into similar six-bladed β-propeller structures [44, 45]. In the virus envelope, the attachment proteins are present as disulphide-linked homodimers, with indications of tetramer formation in some cases [45]. The attachment proteins form complexes with the fusion proteins, which are homotrimers in the viral envelope. Receptor binding triggers rearrangements in these heterocomplexes that alter the structure of fusion proteins, resulting in fusion of viral and cell membranes at neutral pH [45–47] (see Chap. 16).

There is certain diversity in cell surface receptor usage among paramyxoviruses. Rubulaviruses (mumps virus), avulaviruses (Newcastle disease virus) and respiroviruses (Sendai virus) bind to cell surface sialic acids *via* the hemagglutinin neuraminidase (HN) attachment glycoprotein, a bifunctional protein engaged in recognition and hydrolysis of sialic acids. Neither activity is found in the haemagglutinin (H) of morbilliviruses (measles virus), or in the attachment glycoprotein G of henipaviruses (Hendra and Nipah viruses) or pneumoviruses (respiratory syncytial virus) [44]. Measles virus H (MV-H) and the Hendra (HeV-G) and Nipah G (NiV-G) proteins lack the conserved residues that mediate sialic acid binding and hydrolysis [5]. These viruses do not bind to sialic acid carbohydrates, but rather attach to cell surface proteins. HeV and NiV-G interact with the ephrin-B2 or -B3 receptors (EFNB2, EFNB3) [48–50], whereas MV-H can bind CD46 [51, 52], signalling lymphocytic activation molecule (SLAM) [53] or nectin-4 [54, 55], depending on the MV strain. All strains of MV bind to SLAM expressed on macrophages, DC, and lymphocytes, cells where infection starts and develops. MV uses nectin-4 to infect epithelial cells and cross the airway epithelium for transmission to new hosts [54, 55].

Structural studies defined the receptor binding modes of several paramyxoviruses. Both sialic acid and the EFNB2/EFNB3 receptors are recognized by overlapping sites at the recessed center of the β-propeller domain of the paramyxovirus attachment proteins (Fig. 15.5a) [44, 45, 56–58]. Protruding hydrophobic residues at the long GH loop of EFNB2 interact with residues in NiV-G that lie very close to the sialic acid binding site in HN at the central cavity of the β-propeller domain (Fig. 15.5a). The receptor binding mode in measles virus (MV) nonetheless differs from that of other paramyxoviruses. The structures of CD46 and SLAM in complex with MV-H protein show that MV uses the side of the β-propeller domain to bind to cell surface receptor molecules (Figs. 15.5a, b) [59, 60]. The recessed center of the β-propeller, the site of sialic acid binding in several paramyxoviruses, is closed off by a glycan in the MV-H protein. The receptor-binding region in MV-H includes a groove formed by the blades β4 and β5 in the β-propeller, a region with the largest structural difference between MV-H and the other paramyxovirus proteins [59]. Therefore, the receptor-binding regions in paramyxoviruses preserve the recessed nature shared by many receptor-binding sites in virus proteins (Figs. 15.5c, d). In addition, the receptor-binding regions contain a hydrophobic socket that is particularly recessed. This socket accommodates sialic acid in the paramyxovirus HN proteins or receptor-specific features in MV-H (Fig. 15.5e) [56, 59]. In the case of the MV-H, the recessed receptor-binding surface is more extended than in other paramyxovirus attachment protein using the centre of the β-propeller for attachment to cell surface receptors (Figs. 15.5c, d). This MV-H surface can recognise three distinct receptor molecules,

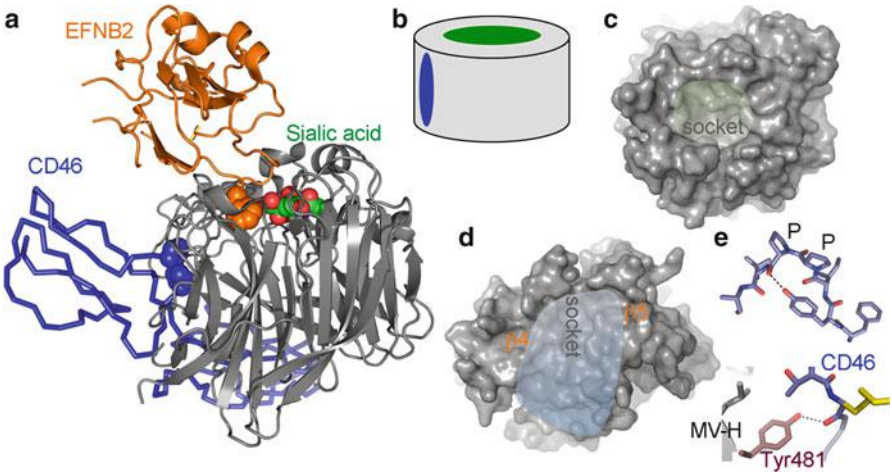

Fig. 15.5 Cell receptor recognition in paramyxovirus. (**a**) Ribbon drawing of the β-propeller domain of a representative paramyxovirus attachment protein (NDV-HN structure; *grey*), bound to three paramyxovirus receptors: Sialic acid (carbons (*green*) and oxygens (*red*) as spheres), the EFNB2 receptor (*orange*) and CD46 (*blue*). Two virus binding Pro residues in CD46 and one Phe in EFNB2 are shown as spheres. The figure was prepared with the crystal structures of complexes of NDV-HN with sialic acid (PDB ID 1E8T), of NiV-G with EFNB2 (PDB ID 2VSM), and of MV-H with CD46 (PDB ID 3INB) after superimposition of the virus proteins. NiV-G and MV-H structures are not shown. (**b**) Cylinder representing a β-propeller domain and location of main receptor binding surfaces in paramyxovirus attachment proteins, based on the complexes shown in (**a**); centre (*green*) and side (*blue*) of the propeller. (**c**) Top view of the NDV-HN structure showing the recessed sialic acid binding surface on the centre of the β-propeller (*light green*), with the socket in which the sialic acid binds (indicated). (**d**) Side view of the MV-H structure with the extended concave surface (*light blue*) use to bind to the SLAM, CD46 and nectin-4 receptors. The β-propeller blades β4 and β5 that form the receptor-binding surface and the socket on the surface are indicated. (**e**) Key interactions for MV recognition of CD46. *Top*. Stick diagram of protruding CD46 loop with two contiguous Pro residues that penetrates the MV-H socket. *Bottom*. Hydrogen bond (*dashed black lines*) of MV-H Tyr481 with a main chain carbonyl of CD46, required for MV binding to CD46 [59]. Oxygens are *red* and nitrogens *dark blue*

CD46, SLAM and nectin-4. As shown for the MV-H/CD46 complex [59], viral residues within the extended receptor-binding groove can show some variability, whereas less accessible residues are conserved; this leads to increasing binding affinity to alternative receptor molecules, which has an important implications on MV tropism and pathogenesis. Single residue mutations in the MV-H groove, such as the Asn481 to Tyr481 switch, is sufficient for MV binding to the ubiquitous CD46 receptor, which extends its cell tropism. Tyr481 can hydrogen bond to CD46 (Fig. 15.5e), an interaction responsible of the virus-receptor binding specificity. The use of an extended receptor-binding surface might explain MV recognition of multiple receptor molecules, which must be related to the efficient transmission of MV from host to host.

15.2.3 Carbohydrates as Viral Receptors

The cell surface displays a large variety of oligosaccharides linked to glycoproteins, proteoglycans and glycolipids, some of which are used by viruses to attach to host cells [61]. Most viral carbohydrate receptors are negatively charged and terminate the glycan moiety, features important for virus recognition. In some cases, however, viruses recognize neutral glycans such as histo-blood group antigens. Monomeric virus-carbohydrate interactions are usually of low affinity (mM range), however, virus particles have many recognition sites that easily engage several carbohydrate molecules, and thus attach to the cell surface with high avidity (Fig. 15.6a). Virus recognition of carbohydrates is associated also with functions other than cell attachment. Orthomyxo-, Paramyxo-, and Coronaviruses express envelope sialic acid-destroying glycoproteins (sialidases and esterases) that are essential for *in vivo* host infection. These enzymes can prevent re-attachment of newly released viruses from infected cells, can remove cell-bound viruses that fail to enter host cells, and can inhibit virus aggregation during budding. Sialidase inhibitors have proven to be useful anti-viral drugs for the treatment of some viral infections.

Sialic acid residues linked to glycoproteins and glycolipids act as receptors for many viruses. Sialic acids are derived from N-acetyl-neuraminic acid and mainly occupy the terminal position of a glycan chain, bound to a penultimate galactose through an (2–3) or (2–6) linkage [61], which renders them easily accessible. Moreover, sialic acids have a larger number of functional groups than other monosaccharides, and can thus participate in a network of polar and non-polar interactions with virus proteins (Fig. 15.6b) [62]. Sialic acids are receptors for distinct viruses, such as influenza, corona-, paramyxo-, toro-, adeno-, noro-, rota-, picorna-, parvo-, polyoma- and reovirus, some of which are important human pathogens [61]. Crystal structures of virus-sialic acid complexes showed that viruses use relatively recessed binding surfaces that interact mostly with the sialic acid face that bears a negatively charged carboxylate group (Figs. 15.6b, c, d) [62]. The mode of sialic acid recognition by distinct viruses can be similar, although viruses use structurally diverse proteins in receptor recognition. In some cases, however, related viruses such as murine polyomavirus and SV40 use similar capsid surface areas to recognize distinct motifs in the sialic acid molecules [5]. The virus-glycan contacts are thus not necessarily conserved (Fig. 15.6c, d).

Viral sialic acid receptors feature numerous cell-specific modifications that determine virus tropism, cell-to-cell transmission and pathogenicity. There are several examples of viruses responsible for serious diseases, such as adenovirus (Ad8, Ad19, Ad37) and enterovirus, which bind specifically to the (2–3)-linked sialic acid. The use of this sialic acid variant, which forms part of a branched glycan linked to the CD1a ganglioside, is responsible for the eye-tropism of those viruses, whose infection cause severe ocular diseases [29, 61]. Influenza A virus transmissibility and pathogenicity in humans also correlates with the recognition of specific sialic acid molecules by the envelope haemagglutinin (H) and neuraminidase (N)

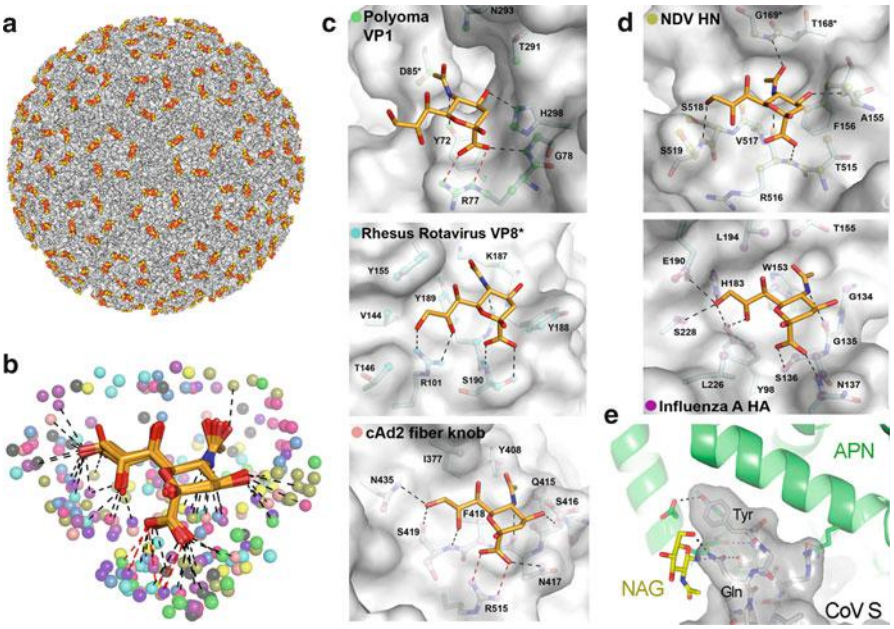

Fig. 15.6 Carbohydrates as viral receptors. (**a**) Multimeric binding of polyomavirus to its sialic acid receptor. Surface representation of the crystal structure (PDB ID 1SIE) of a polyomavirus particle (*surface in grey*) in complex with a sialic acid receptor fragment (*carbons in yellow*). (**b**) Possible contacts of viral proteins with terminal sialic acid, based on crystal structures [62]. Sialic acid is shown as sticks with carbons in orange. Viral atoms within 4.0 Å of the carbohydrate receptors in crystal structures of virus-sialic acid complexes are shown as spheres. Hydrogen bonds and salt bridges are indicated by *black* and *red lines*, respectively. (**c**) Non-enveloped viruses binding to sialic acid receptors. Detail of sialic acid binding interactions with polyoma VP1 (PDB ID 1VPS), rhesus rotavirus VP8* (PDB ID 1KQR) and cAd2 fibre knob (PDB ID 2WBV). (**d**) Enveloped viruses binding to sialic acid receptors. Network of interactions for the NDV-HN (PDB ID 1USR) and influenza HA protein (PDB ID 1HGG). (**e**) Coronavirus (CoV) binding to a glycan-containing region of its aminopeptidase N (APN) receptor. Detail of the interaction revealed by the crystal structure (PDB ID 4F5C) of a fragment of a porcine CoV spike protein (CoV S) in complex with the pig APN. CoV S is shown as *grey surface* and stick representations, the APN as *green ribbons*, with the key virus-binding N-acetyl-glucosamine (NAG) residue as sticks and *carbons in yellow*. The viral Tyr and Gln residues that contact the glycan in the CoV S protein are labelled. Hydrogen bonds are *dashed black lines*. Oxygens are *red* and nitrogens *dark blue* in all figures (Figures (**b-d**) were provided by U. Neu and T. Stehle, adapted with permission from [62])

proteins. The H (H1 to H15) and N (N1 to N9) protein variants identified in distinct species differ in their affinities for the (2–3)- or (2–6)-linked sialic acids, which is a key determinant in virus transmissibility [61]. Aquatic birds are the natural reservoir of influenza A, but the avian viruses are periodically transmitted to mammals, causing flu pandemics with significant numbers of deaths [63]. The H protein of avian influenza A (HA) binds preferentially to (2–3)-linked sialic acid receptors, which are not present in the human tracheal cells initially infected by influenza A [64];

transmission of avian influenza A from birds to humans or within the human population is therefore quite inefficient. Nonetheless, variants of some influenza A strains (H5N1) can infect humans, causing sporadic outbreaks with high mortality rates. Alterations in the virus receptor-binding specificity for (2–6)-linked sialic acids can cause the emergence of a pandemia such as the 1918 Spanish flu. In addition to these examples, other viruses can recognise specific substitutions in the sialic acid molecule [61].

Viruses can also bind to other negatively charged sulphated or neutral carbohydrates. Human herpesviruses are an example of viruses that attach to cell surface heparan sulphate (HS), and this interaction has been studied extensively for human herpes simplex viruses (HSV). HSV cell entry is complex and is mediated by several envelope glycoproteins (gC, gB, gD, gH and gL) [65]. The initial interaction of the virus particles with the cell is mediated by the gC and gB glycoproteins, which bind to HS. This interaction is nevertheless insufficient for virus entry, which also requires gD binding to cell surface proteins for virus-cell membrane fusion, catalysed by gB and the gH/gL complex. HS can thus be considered an HSV attachment factor that facilitates subsequent interactions with other receptors for cell entry.

Noroviruses exemplify a group of viruses that use neutral cell surface carbohydrate receptors, known as histo-blood group antigens [66]. Recognition is strain-specific and different virus strains recognize distinct types of antigens. These variations affect the tropism and pathogenesis of norovirus infections, as susceptibility to infection is dependent on the histo-blood group antigen receptors, which differ among individuals, who can therefore be resistant to infection by certain norovirus strains and susceptible to others. It is of interest to note that viruses can also recognize cell surface structures comprised by carbohydrates and amino acid residues, as recently described for some coronaviruses (CoV) [67]. A protruding loop with a tyrosine residue in the envelope spike of a subset of CoV docks between a neutral N-acetyl-glucosamine residue and an alpha helix exposed in the ectodomain of its receptor, the aminopeptidase N (APN) (Fig. 15.6e). This N-linked glycan is essential for virus binding to APN, and is a major determinant of the host range of some CoV [67].

Cell surface carbohydrates are ubiquitous, accessible and relatively variable "hooks" for cell attachment by highly distinct virus families. Carbohydrate-derived compounds can thus be used to inhibit many types of virus infections. Carbohydrates are major cell entry receptors for some viruses, and can be alternative or secondary entry receptors for others. Viruses that attach to protein receptors can evolve to use carbohydrates, thus expanding host cell tropism [4, 16, 68]. Enveloped viruses can also exploit the glycans linked to their envelope proteins to attach to cell surface lectins. Some of these lectins are specific for certain glycosylation patterns associated to virus membrane glycoproteins, as discussed above for DC-SIGN and HIV-1. Many enveloped viruses attach to host cells *via* DC-SIGN [6], a cell surface receptor for multiple pathogens. Glycans are thus intimately linked to virus-cell interaction.

15.3 Non-Enveloped Virus Entry into Host Cells: The Uncoating Process

15.3.1 Receptor-Mediated Uncoating of Picornaviruses

After binding to cell surface receptors, the viral genome must translocate from the particle to the cell cytoplasm (Fig. 15.1). In enveloped viruses, this transfer may be rather simple and occurs after fusion of the virus and cell membranes, as will be illustrated in Chap. 16. In non-enveloped viruses, mechanisms for genome penetration into the cytoplasm have not yet been characterized in detail, although several pieces of the process have been described. Particularly well-studied models include some viruses of the picornavirus family, PV and HRV, which will be discussed here.

The initial event during membrane penetration of picornaviruses is the opening of the viral capsid and the exit of the viral genome, a process known as uncoating. The capsids of PV and HRV are very similar and are characterized by a sedimentation coefficient of about 160S for PV and 150S for HRV; after RNA exit, the density of the empty capsids decreases markedly and their sedimentation coefficient is 80S [69–71]. Binding of soluble receptor molecules to the viral particles decreases their sedimentation coefficient to approximately 120S [72]. In addition, uncoating intermediates that lack the internal VP4 protein and have a sedimentation coefficient of about 135S have been described [69, 73, 74]. The distinct sedimentation behaviour of native virions and empty capsids allowed analysis of the uncoating process by ultracentrifugation in PV and HRV [70–72, 75, 76]. The structures of some of these capsid forms have been determined, and have helped to propose models for the structural rearrangements and dynamic processes associated with genome uncoating in these viruses (see below, Subsection 15.3.2).

The opening of the capsid in PV and HRV can be mediated by receptor binding, by low pH, or by the cooperative effect of both factors [71, 74–76]. Initial studies with PV and soluble poliovirus receptor (PVR) showed that receptor binding at physiological temperature (37 °C) mediates RNA exit from the capsid interior with no additional factors [70]. PVR binding to PV generates an intermediate particle with a sedimentation coefficient of 135S that lacks VP4, which is subsequently converted to empty 80S capsids. These particles are also observed during initial cell infection by PV [74], and thus represent entry intermediates. These findings show that PVR has an active role in PV entry; it is not just a “hook” used for attachment, but also an “unzipper” that mediates viral RNA uncoating [77]. Subsequent studies with HRV showed similar behaviour for the HRV receptor ICAM-1, which triggers RNA uncoating after binding to certain HRV serotypes at physiological temperatures [71, 72]. ICAM-1 binding to HRV serotypes 3 (HRV3) and 14 (HRV14) promotes RNA exit and formation of empty 80S capsids at neutral pH. This process is temperature dependent and requires temperatures over 25–30°C, necessary to overcome the high activation energy of the uncoating process (~45 kcal/mol) [72]; this energy is provided by receptor binding, an endothermic process in viruses sensitive to

receptor binding, such as PV and HRV3 [12]. PVR binding to PV is more endothermic than ICAM-1 binding to HRV3, which indicates that PVR is more efficient triggering uncoating than ICAM-1. As PV can enter host cells in the presence of agents that prevent endosomal acidification [74], RNA uncoating and transfer to the cytoplasm must therefore be mediated by receptor binding. PV infects hosts through the gastrointestinal tract; it is thus very stable at the low pH to which viruses are exposed during endocytosis. PV uses a receptor molecule suited to efficient uncoating at neutral pH and host cell entry.

The HRV are a group of viruses composed of around 100 different serotypes that bind to ICAM-1 (major group) or LDLR (minor group), as described above (Subsection 15.2.1). There is also certain diversity among serotypes in terms of the physiological factors that trigger HRV uncoating. Minor group HRV are sensitive to mildly acidic pH (5.5–6.0), which can trigger virus uncoating at physiological temperatures. Binding to LDLR does not mediate uncoating, and minor group HRV are thus dependent on endocytosis and endosome acidification for uncoating and host cell entry [16, 75]. The major group of HRV, which bind ICAM-1, comprises viruses that differ in stability and sensitivity to receptor binding [12, 71, 78]. Serotypes such as HRV3 are uncoated efficiently by ICAM-1 binding at neutral pH and physiological temperatures, whereas HRV16 are relatively stable and HRV14 have intermediate sensitivity. These differences in sensitivity to receptor binding are not related to affinity, but to binding energy. HRV3 binding to ICAM-1 is more endothermic than HRV16 binding [12]; the HRV3-receptor complex is thus less stable and is therefore primed for uncoating. On the contrary, HRV16-receptor complexes are more stable and require additional factors for uncoating, as explained below. In accordance with these findings, structural studies of HRV-receptor complexes prepared at physiological temperature show expansion of HRV3 following receptor binding and lack of expansion for HRV16 [79]. The energy absorbed by the HRV3-receptor interaction translates into capsid expansion, a metastable state primed for RNA exit.

The low pH at which cell surface receptor-bound viruses are exposed during endocytosis is a relevant factor that mediates virus uncoating (Fig. 15.1) [3]. Indeed, the mildly acidic pH (5.5–6.0) in endosomal compartments is optimal for uncoating of and infection by minor group HRV [16, 75, 80]. Exposure to low pH is also necessary for efficient uncoating and cell entry by some major group HRV [16, 76], although certain serotypes require only ICAM-1 binding, after which they can infect cells, even in the presence of agents that prevent endosome acidification. In the case of relatively stable HRV serotypes with pocket factor molecules, such as HRV16, both receptor binding and low endosome pH are needed for entry into host cells. The uncoating rate of HRV16 after binding ICAM-1 or exposure to mildly acidic pH (5.5–6.0) is very slow, and the virus can remain intact for several hours [71, 76]. After binding to ICAM-1, however, HRV16 particles become sensitive to low pH, which triggers rapid RNA release from the capsid [76]. Receptor binding appears to alter the HRV16 particles, priming them for low pH-mediated uncoating during endocytosis. This cooperative effect of the receptor and low pH in virus uncoating must be necessary to polarize RNA exit at the receptor-bound region of

the virus particle in endosomes, which must be relevant for efficient penetration of the cell membrane and infection.

The diverse receptor-mediated uncoating observed among major group rhinoviruses has been linked to the inherent stability and dynamics of the capsid. Biochemical studies of HRV indicate dynamic capsid behaviour [78, 81]; internal capsid polypeptides in the crystal structures, such as VP4 and the N terminus of VP1, are very sensitive to proteases, indicating that they are exposed to the environment at physiological temperatures. The dynamic behaviour of the capsids was termed "breathing" (see Chap. 6), and is likely to be mediated by the uncoordinated expansive-contractive movement of capsid protomers. Capsid breathing in HRV appears to be restricted by small molecules bound to a pocket in the VP1 [78, 81]. Natural molecules or pocket factors were identified as fatty acids in the capsids of the major group HRV16 and minor group HRV2. A group of molecules known as WIN compounds also bind to the VP1 pocket, restricting capsid dynamics and preventing uncoating. Moreover, the receptor-sensitive HRV3 serotype lacks pocket factor, which is present in the receptor-resistant HRV serotypes HRV16 and HRV2. These data suggest that capsid breathing is necessary for triggering receptor-mediated uncoating at physiological temperatures [79, 82].

15.3.2 The Structural Bases of Receptor-Mediated Virus Uncoating in Picornaviruses

Several structural studies over the last decade have examined the uncoating of PV and HRV intermediate particles by cryo-EM. One analysis determined the structure of the 135S PV particles, which lack VP4 but preserve the RNA inside, and compared this structure with 160S native virion and 80S empty capsids formed after RNA exit [73]. The 135S particle showed VP rearrangements relative to the native particle and 4 % expansion. Major rearrangements appear in the two-fold axes of the capsid and in the canyon region; however, the conformation of the five-fold axis, which was considered the RNA exit port, was similar to that of the 160S native PV capsid. HRV3 and ICAM-1 complexes prepared at 37 °C and in the process of uncoating were analysed by cryo-EM [79]; the structure also showed HRV3 capsid expansion not seen in stable HRV16-ICAM-1 complexes, indicating that capsid expansion must precede uncoating. The conformation of the RNA-containing HRV3 capsid in the virus-receptor complexes is similar to that of the native HRV capsid. Nevertheless, the empty HRV3 capsid is quite different, showing major rearrangements after RNA release; this has now been described in detail in a high-resolution structural study of empty HRV capsids [83]. Major differences between full and empty capsids in the HRV-receptor complexes appear at the five-fold axes; however, the structure of the five-fold axis regions are almost identical to those of native viruses in RNA-containing HRV3-receptor complexes [79]. These structural insights challenged the hypothesis of RNA exit through the

capsid five-fold axis, suggesting other exit ports. Recent studies with PV showed that RNA exit from the capsid near a two-fold axis, at a site that extends toward the receptor-binding canyon region [84].

The studies discussed here for PV and HRV show that receptors participate in the virus entry process, mediating not only virus-cell attachment but also RNA release, alone or with the contribution of low pH. The uncoating process requires multivalent receptor binding and capsid dynamics, which must explain the temperature dependence of this process. At physiological temperatures, capsids are dynamic entities whose protomers can move randomly, opening and closing holes at interprotomer junctions. This rapid, random movement does not allow for RNA release, probably because of the transient nature of the holes. After receptor binding at physiological temperatures, the capsid remains expanded and holes remain open for RNA release. PV and HRV recognize receptors *via* the canyon, a fragmented region at the junction of two protomers. PVR and ICAM-1 receptors can therefore act as wedges by binding at protomer junctions to maintain capsid expansion [12, 79], thus catalysing the uncoating process by locking the capsid in an intermediate, open state that allows RNA exit. The function of these two receptor molecules in virus uncoating is probably closely linked to the way they are recognized by PV and HRV.

The externalized genome in non-enveloped viruses either on the cell surface or in endosomes must translocate into the cell cytoplasm to initiate the intracellular phase of the life cycle. The study of the transfer of the genome from the virus particle to the cytoplasm is less understood in non-enveloped than in enveloped viruses. Structures of isolated envelope fusion proteins have delineated the process of membrane penetration or fusion of virus-cell membranes, described in Chap. 16. However, the analysis of viral genome penetration process in non-enveloped viruses requires the study of virus infecting host cells, which is methodologically challenging. Research carried out to date indicates two possible routes of penetration, either by membrane rupture or pores through which genome moves to the cytoplasm [16]. These disruptions in the membrane can be caused by hydrophobic capsid motif exposed during the uncoating process. The PV and HRV particles discussed here externalise the hydrophobic N-terminal region of VP1 that enable attachment of the particles to cell membranes in endosomes or on the cell surface. Moreover, the VP4 protein that is externalized together with the RNA bears a myristoyl group that can also interact with the cell membrane. The interaction of capsid motifs with the membrane can mediate its disruption for the transfer of the uncoated genome to the cytoplasm.

15.4 Perspectives and Conclusions

Viruses subvert cell surface molecules for cell entry and dissemination of infection. Receptor accessibility and cell expression patterns are important factors that underlie virus selection of specific surface receptors. Most of these receptors promote translocation of the viral genome into the cell. In some cases, however, viruses

attach to cellular factors that concentrate infectious particles on the plasma membrane and facilitate virus recognition of entry receptors. In other cases, viruses use cell surface molecules to spread the infection throughout the body; they can attach to migratory cells or induce signalling events that loosen cell-cell contacts and facilitate virus transmission. Cell surface molecules can thus have diverse functions in the dissemination of viral infections; the study of virus-receptor interactions is therefore of great relevance for understanding virus evolution, host tropism and pathogenesis.

The virus entry receptors have been the main focus of this chapter. These molecules are not just a "hook" for virus attachment, but can also participate in translocation of the viral genome into host cells, on the cell surface, or within endosomal compartments. Some virus receptors catalyse the entry process alone or in combination with other cell factors, such as the mildly acidic pH found in endosomes. Interaction with certain receptors transforms virus particles into metastable entities, priming delivery of their genomes to the cytoplasm. Alternatively, genome translocation can be mediated by low pH, but this requires endocytosis of receptor-bound viruses. The process of receptor-mediated virus entry continues to be studied. In some picornaviruses, discussed here, the energy absorbed in the virus-receptor interaction translates into virus particle expansion and genome exit. In other non-enveloped viruses, capsid expansion triggered by cell factors might be a necessary intermediate step for genome uncoating. Receptor binding also primes virus-cell membrane fusion in enveloped viruses, mediated by conformational changes in virus fusion proteins (Chap. 16). Cell surface molecules are thus key players in virus entry into host cells.

Virus-receptor interactions appear to be highly specific. There are many examples of virus selection of only a single member of closely-related cell surface molecules. This specificity is based on the recognition of certain unique structural features of a cell surface molecule, as well as on key polar interactions such as those described here for MV and its CD46 receptor. Virus-receptor surface complementarity must be important for virus recognition of certain receptor molecules. Viral proteins are shaped to dock into structural features of their receptors. Many viruses use concave surfaces for binding to cell surface receptors, which optimises receptor contact area and hides receptor-binding residues from antibodies. Nevertheless, viruses can also have protruding receptor-binding surfaces for receptor recognition. In both cases, viruses can escape from antibody neutralisation by mutating non-essential residues in these receptor-binding surfaces. Immune system pressure on viruses is a major determinant for the switch in receptor recognition observed in many virus groups, although this diversity can also be linked to opportunities to spread infection using ubiquitous or more accessible cell surface molecules. In addition to its high degree of specificity, virus-receptor recognition can be dynamic, and viruses can evolve to use alternative or distinct receptors, as illustrated here with several examples.

Preventing virus binding to receptors can efficiently block virus infection and cell damage. Receptor binding regions in virus proteins are relatively invariant and could potentially be targeted by immune responses to prevent infection. A large

number of structural studies have characterised viral antigenic sites targeted by neutralising antibodies. Many of these sites overlap receptor-binding regions, showing that blockade of virus binding to cell surface receptors is a major neutralisation mechanism. In viruses for which efficient vaccines have been developed (MV and PV), the receptor-binding residues are relatively well-exposed and accessible to antibodies. Viruses have nonetheless developed ways to evade neutralisation by hiding receptor-binding regions from antibodies; the use of concave and poorly accessible receptor-binding surfaces is one well-defined strategy of this type. In some viruses, these relatively inaccessible surfaces are also surrounded by variable loops and glycans, which further prevent antibody binding and neutralisation. Blocking infection of these viruses requires the design of improved vaccines that elicit immune responses focused on relatively inaccessible sites. These therapies require a deep understanding of virus-receptor interactions, including the determination of complex structures. These studies also open avenues for the development of molecules that prevent virus entry into cells. Soluble multimeric receptor molecules that impede virus binding to cells and infection were one of the first therapeutics developed for viruses such as HIV or HRV. Small molecule drugs have been developed, such as sialic acid analogues to treat influenza virus infection, or molecules that prevent HIV gp120 binding to its receptor, CD4 (see Chap. 20). The characterization of virus-receptor interactions is thus of considerable interest for the development of antiviral therapies.

Acknowledgements I thank U. Neu, T. Stehle, L. Xing and H.R. Cheng for providing figures included in this chapter and C. Mark for editorial assistance. Grant support by the National Institutes of Health (NIH P01-AI054456) and the Spanish Ministry of Science and Innovation (MICINN; BFU2011-23940) is acknowledged.

References and Further Reading

1. Marsh M, Helenius A (2006) Virus entry: open sesame. Cell 124:729–740
2. Grove J, Marsh M (2011) The cell biology of receptor-mediated virus entry. J Cell Biol 195:1071–1082
3. Mercer J, Schelhaas M, Helenius A (2010) Virus entry by endocytosis. Annu Rev Biochem 79:803–833
4. Baranowski E, Ruiz-Jarabo CM, Domingo E (2001) Evolution of cell recognition by viruses. Science 292:1102–1105
5. Stehle T, Casasnovas JM (2009) Specificity switching in virus-receptor complexes. Curr Opin Struct Biol 19:181–188
6. Backovic M, Rey FA (2012) Virus entry: old viruses, new receptors. Curr Opin Virol 2:4–13
7. Wang J-h (2002) Protein recognition by cell surface receptors: physiological receptors versus virus interactions. Trends Biochem Sci 27:122–126
8. Rossmann MG, He Y, Kuhn RJ (2002) Picornavirus-receptor interactions. Trends Microbiol 10:324–331
9. Mendelsohn CL, Wimmer E, Racaniello VR (1989) Cellular receptor for poliovirus: molecular cloning, nucleotide sequence, and expression of a new member of the immunoglobulin superfamily. Cell 56:855–865

10. Greve JM, Davis G, Meyer AM, Forte CP, Yost SC, Marlor CW, Kamarck ME, McClelland A (1989) The major human rhinovirus receptor is ICAM-1. Cell 56:839–847
11. Staunton DE, Merluzzi VJ, Rothlein R, Barton R, Marlin SD, Springer TA (1989) A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. Cell 56:849–853
12. Xing L, Tjarnlund K, Lindqvist B, Kaplan G, Feigelstock D, Cheng RH, Casasnovas JM (2000) Distinct cellular receptor interactions in poliovirus and rhinoviruses. EMBO J 19:1207–1216
13. Xing L, Huhtala M, Pietiainen V, Kapyla J, Vuorinen K, Marjomaki V, Heino J, Johnson MS, Hyypia T, Cheng RH (2004) Structural and functional analysis of integrin alpha2I domain interaction with echovirus 1. J Biol Chem 279:11632–11638
14. Smith TJ, Chase ES, Schmidt TJ, Olson NH, Baker TS (1996) Neutralizing antibody to human rhinovirus 14 penetrates the receptor-binding canyon. Nature 383:350–354
15. Hewat EA, Verdaguer N, Fita I, Blakemore W, Brookes S, King A, Newman J, Domingo E, Mateu MG, Stuart DI (1997) Structure of the complex of an Fab fragment of a neutralizing antibody with foot-and-mouth disease virus: positioning of a highly mobile antigenic loop. EMBO J 16:1492–1500
16. Fuchs R, Blaas D (2010) Uncoating of human rhinoviruses. Reviews in Medical Virology 20:281–297
17. Verdaguer N, Fita I, Reithmayer M, Moser R, Blaas D (2004) X-ray structure of a minor group human rhinovirus bound to a fragment of its cellular receptor protein. Nat Struct Mol Biol 11:429–434
18. Bergelson JM, Chan M, Solomon KR, St John NF, Lin H, Finberg RW (1994) Decay-accelerating factor (CD55), a glycosylphosphatidylinositol-anchored complement regulatory protein, is a receptor for several echoviruses. Proc Natl Acad Sci USA 91:6245–6248
19. He Y, Lin F, Chipman PR, Bator CM, Baker TS, Shoham M, Kuhn RJ, Medof ME, Rossmann MG (2002) Structure of decay-accelerating factor bound to echovirus 7: a virus-receptor complex. Proc Natl Acad Sci USA 99:10325–10329
20. Pettigrew DM, Williams DT, Kerrigan D, Evans DJ, Lea SM, Bhella D (2006) Structural and functional insights into the interaction of echoviruses and decay-accelerating factor. J Biol Chem 281:5169–5177
21. Casasnovas JM, Markarian S, Hammar L (2004) Sensor surface interactions in the study of macromolecular assemblies. In: Cheng R, Hammar L (eds) Conformational proteomics of macromolecular architecture. World Scientific, Singapore
22. Zhang Y, Bergelson JM (2005) Adenovirus receptors. J Virol 79:12125–12131
23. Berk AJ (2007) Adenoviridae: the viruses and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (eds) Fields virology, vol 2. Lippincott, Williams & Wilkins, Philadelphia, pp 2355–2394
24. Bewley MC, Springer K, Zhang Y-B, Freimuth P, Flanagan JM (1999) Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. Science 286:1579–1583
25. van Raaij MJ, Chouin E, van der Zandt H, Bergelson JM, Cusack S (2000) Dimeric structure of the coxsackievirus and adenovirus receptor D1 domain at 1.7 A resolution. Structure 8:1147–1155
26. Walters RW, Freimuth P, Moninger TO, Ganske I, Zabner J, Welsh MJ (2002) Adenovirus fiber disrupts CAR-mediated intercellular adhesion allowing virus escape. Cell 110:789–799
27. Persson BD, Reiter DM, Marttila M, Mei Y-F, Casasnovas JM, Arnberg N, Stehle T (2007) Adenovirus type 11 binding alters the conformation of its receptor CD46. Nat Struct Mol Biol 14:164–166
28. Burmeister WP, Guilligay D, Cusack S, Wadell G, Arnberg N (2004) Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. J Virol 78:7727–7736
29. Nilsson EC, Storm RJ, Bauer J, Johansson SMC, Lookene A, Ångström J, Hedenström M, Eriksson TL, Frängsmyr L, Rinaldi S, Willison HJ, Domellöf FP, Stehle T, Arnberg N (2010)

The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. Nat Med 17:105–109
30. Wickham TJ, Mathias P, Cheresh DA, Nemerow GR (1993) Integrins αvβ3 and αvβ5 promote adenovirus internalization but not virus attachment. Cell 73:309–319
31. Wyatt R, Sodroski J (1998) The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. Science 280:1884–1888
32. Zhu P, Liu J, Bess J, Chertova E, Lifson JD, Grisé H, Ofek GA, Taylor KA, Roux KH (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature 441:847–852
33. Freed RO, Martin MA (2007) HIVs and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (eds) Fields virology, vol 2. Lippincott, Williams & Wilkins, Philadelphia, pp 2107–2185
34. Berger EA (1998) HIV entry and tropism. When one receptor is not enough. Adv Exp Med Biol 452:151–157
35. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393:648–659
36. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, Sodroski JG (1998) The antigenic structure of the HIV gp120 envelope glycoprotein. Nature 393:705–711
37. Berger EA, Murphy PM, Farber JM (1999) Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol 17:657–700
38. Baribaud F, Pöhlmann S, Doms RW (2001) The role of DC-SIGN an DC-SIGNR in HIV and SIV attachment, infection, and transmission. Virology 286:1–6
39. Geijtenbeek TBH, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GCF, Middel J, Cornelissen IL, Nottet HS, KewalRamani VN, Littman DR, Figdor CG, van Kooyk Y (2000) DC-SIGN, a dendritic cell-specific HIV-1 binding protein that enhances trans-infection of T cells. Cell 100:587–597
40. Feinberg H, Mitchell DA, Drickamer K, Weis WI (2001) Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. Science 294:2163–2166
41. Guo Y, Feinberg H, Conroy E, Mitchell DA, Alvarez R, Blixt O, Taylor ME, Weis WI, Drickamer K (2004) Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. Nat Struct Mol Biol 11:591–598
42. Puryear WB, Yu X, Ramirez NP, Reinhard BM, Gummuluru S (2012) HIV-1 incorporation of host-cell-derived glycosphingolipid GM3 allows for capture by mature dendritic cells. Proc Natl Acad Sci USA 109:7475–7480
43. Izquierdo-Useros N, Lorizate M, Contreras FX, Rodriguez-Plata MT, Glass B, Erkizia I, Prado JG, Casas J, Fabrias G, Kräusslich H-G, Martinez-Picado J (2012) Sialyllactose in viral membrane gangliosides is a novel molecular recognition pattern for mature dendritic cell capture of HIV-1. PLoS Biol 10:e1001315
44. Lamb RA, Parks GD (2007) Paramyxoviridae: the viruses and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (eds) Fields virology, vol 1. Lippincott, Williams & Wilkins, Philadelphia, pp 1449–1496
45. Lamb RA, Paterson RG, Jardetzky TS (2006) Paramyxovirus membrane fusion: lessons from the F and HN atomic structures. Virology 344:30–37
46. Plemper RK, Brindley MA, Iorio RM (2011) Structural and mechanistic studies of measles virus illuminate paramyxovirus entry. PLoS Pathog 7:e1002058
47. Bose S, Zokarkar A, Welch BD, Leser GP, Jardetzky TS, Lamb RA (2012) Fusion activation by a headless parainfluenza virus 5 hemagglutinin-neuraminidase stalk suggests a modular mechanism for triggering. Proc Natl Acad Sci USA 109:E2625–E2634
48. Negrete OA, Levroney EL, Aguilar HC, Bertolotti-Ciarlet A, Nazarian R, Tajyar S, Lee B (2005) EphrinB2 Is the entry receptor for nipah virus, an emergent deadly paramyxovirus. Nature 436:401–405

49. Bonaparte MI, Dimitrov AS, Bossart KN, Crameri G, Mungall BA, Bishop KA, Choudhry V, Dimitrov DS, Wang LF, Eaton BT, Broder CC (2005) Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus. Proc Natl Acad Sci USA 102:10652–10657
50. Negrete OA, Wolf MC, Aguilar HC, Enterlein S, Wang W, Muhlberger E, Su SV, Bertolotti-Ciarlet A, Flick R, Lee B (2006) Two key residues in ephrinB3 are critical for its use as an alternative receptor for Nipah virus. PLoS Pathog 2:e7
51. Dorig RE, Marcil A, Chopra A, Richardson CD (1993) The human CD46 molecule is a receptor for measles virus (Edmonston strain). Cell 75:295–305
52. Naniche D, Varior-Krishman G, Cervoni F, Wild TF, Rossi B, Rabourdin-Combe C, Gerlier D (1993) Human membrane cofactor protein (CD46) acts as a cellular receptor for measles virus. J Virol 67:6025–6032
53. Tatsuo H, Ono N, Tanaka K, Yanagi Y (2000) SLAM (CDw150) is a cellular receptor for measles virus. Nature 406:893–897
54. Muhlebach MD, Mateo M, Sinn PL, Prufer S, Uhlig KM, Leonard VHJ, Navaratnarajah CK, Frenzke M, Wong XX, Sawatsky B, Ramachandran S, McCray PB, Cichutek K, von Messling V, Lopez M, Cattaneo R (2011) Adherens junction protein nectin-4 is the epithelial receptor for measles virus. Nature 480:530–533
55. Noyce RS, Bondre DG, Ha MN, Lin L-T, Sisson G, Tsao M-S, Richardson CD (2011) Tumor cell marker PVRL4 (nectin 4) is an epithelial cell receptor for measles virus. PLoS Pathog 7: e1002240
56. Crennell S, Takimoto T, Portner A, Taylor G (2000) Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. Nat Struct Biol 7:1068–1074
57. Bowden TA, Aricescu AR, Gilbert RJ, Grimes JM, Jones EY, Stuart DI (2008) Structural basis of nipah and hendra virus attachment to their cell-surface receptor ephrin-B2. Nat Struct Mol Biol 15:567–572
58. Xu K, Rajashankar KR, Chan YP, Himanen JP, Broder CC, Nikolov DB (2008) Host cell recognition by the henipaviruses: crystal structures of the Nipah G attachment glycoprotein and its complex with ephrin-B3. Proc Natl Acad Sci U S A 105:9953–9958
59. Santiago C, Celma ML, Stehle T, Casasnovas JM (2010) Structure of the measles virus hemagglutinin bound to the CD46 receptor. Nat Struct Mol Biol 17:124–129
60. Hashiguchi T, Ose T, Kubota M, Maita N, Kamishikiryo J, Maenaka K, Yanagi Y (2011) Structure of the measles virus hemagglutinin bound to its cellular receptor SLAM. Nat Struct Mol Biol 18:135–141
61. Olofsson S, Bergström T (2005) Glycoconjugate glycans as viral receptors. Ann Medicine 37:154–172
62. Neu U, Bauer J, Stehle T (2011) Viruses and sialic acids: rules of engagement. Curr Opin Struct Biol 21:610–618
63. Lipatov AS, Govorkova EA, Webby RJ, Ozaki H, Peiris M, Guan Y, Poon L, Webster RG (2004) Influenza: emergence and control. J Virol 78:8951–8959
64. Matrosovich MN, Matrosovich TY, Gray T, Roberts NA, Klenk H-D (2004) Human and avian influenza viruses target different cell types in cultures of human airway epithelium. Proc Natl Acad Sci USA 101:4620–4624
65. Navaratnarajah CK, Miest TS, Carfi A, Cattaneo R (2012) Targeted entry of enveloped viruses: measles and herpes simplex virus I. Curr Opin Virol 2:43–49
66. Huang P, Farkas T, Marionneau S, Zhong W, Ruvoen-Clouet N, Morrow AL, Altaye M, Pickering LK, Newburg DS, LePendu J, Jiang X (2003) Noroviruses bind to human ABO, Lewis, and secretor histo-blood group antigens: identification of 4 distinct strain-specific patterns. J Infect Dis 188:19–31
67. Reguera J, Santiago C, Mudgal G, Ordoño D, Enjuanes L, Casasnovas JM (2012) Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. PLOS Pathog 8:e1002859
68. Li F (2012) Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. J Virol 86:2856–2858

69. Rueckert RR (1996) Picornaviridae: the viruses and their replication. In: Fields BN, Knipe DM, Howley PM, Melnick JL, Chanock RM, Roizman B, Monath TP (eds) Virology, vol 1. Raven, New York, pp 609–654
70. Kaplan G, Freistadt MS, Racaniello VR (1990) Neutralization of poliovirus by cell receptors expressed in insect cells. J Virol 64:4697–4702
71. Hoover-Litty H, Greve JM (1993) Formation of rhinovirus-soluble ICAM-1 complexes and conformational changes in the virion. J Virol 67:390–397
72. Casasnovas JM, Springer TA (1994) The pathway of rhinovirus disruption by soluble intercellular adhesion molecule 1 (ICAM-1): An intermediate in which ICAM-1 is bound and RNA is released. J Virol 68:5882–5889
73. Belnap DM, Filman DJ, Trus BL, Cheng N, Booy FP, Conway JF, Curry S, Hiremath CN, Tsang SK, Steven AC, Hogle JM (2000) Molecular tectonic model of virus structural transitions: the putative cell entry states of poliovirus. J Virol 74:1342–1354
74. Hogle JM (2002) Poliovirus cell entry: common structural themes in viral cell entry pathways. Ann Rev Microbiol 56:677–702
75. Prchla E, Kuechler E, Blaas D, Fuchs R (1994) Uncoating of human rhinovirus serotype 2 from late endosomes. J Virol 68:3713–3723
76. Nurani G, Lindqvist B, Casasnovas JM (2003) Receptor priming of major group human rhinoviruses for uncoating and entry at mild low-pH environments. J Virol 77:11985–11991
77. Racaniello VR (1996) The poliovirus receptor: a hook, or an unzipper? Structure 4:769–773
78. Lewis JK, Bothner B, Smith TJ, Siuzdak G (1998) Antiviral agent blocks breathing of the common cold virus. Proc Natl Acad Sci USA 95:6774–6778
79. Xing L, Casasnovas JM, Cheng RH (2003) Structural analysis of human rhinovirus complexed with ICAM-1 reveals the dynamics of receptor-mediated virus uncoating. J Virol 77:6101–6107
80. Schober D, Kronenberger P, Prchla E, Blaas D, Fuchs R (1998) Major and minor receptor group human rhinoviruses penetrate from endosomes by different mechanisms. J Virol 72:1354–1364
81. Roy A, Post CB (2012) Long-distance correlations of rhinovirus capsid dynamics contribute to uncoating and antiviral activity. Proc Natl Acad Sci USA 109:5271–5276
82. Casasnovas JM (2000) The dynamics of receptor recognition by human rhinoviruses. Trends Microbiol 8:251–254
83. Garriga D, Pickl-Herk A, Luque D, Wruss J, Caston JR, Blaas D, Verdaguer N (2012) Insights into minor group rhinovirus uncoating: the X-ray structure of the HRV2 empty capsid. PLoS Pathog 8:e1002473
84. Bostina M, Levy H, Filman DJ, Hogle JM (2011) Poliovirus RNA is released from the capsid near a twofold symmetry axis. J Virol 85:776–783

Further Reading

Especially recommended for further reading are the following references listed above: References [1–3] present a general overview of virus-receptor recognition and virus entry into host cells; references [4, 5] describe virus switch of receptor specificity; reference [6] is part of a *Current Opinion in Virology* issue describing recent research on virus entry. I also recommend the following references related to specific chapter sections: 15.2.1, references [8, 16] (picornavirus), and [22] (adenovirus); 15.2.2, references [34, 35] (HIV-1), and [45, 46] (paramyxovirus); 15.2.3, references [61, 62]; 15.3, references [16, 74, 79].

Chapter 16
Entry of Enveloped Viruses into Host Cells: Membrane Fusion

Vicente Más and José A. Melero

Abstract Viruses are intracellular parasites that hijack the cellular machinery for their own replication. Therefore, an obligatory step in the virus life cycle is the delivery of the viral genome inside the cell. Enveloped viruses (*i.e.*, viruses with a lipid envelope) use a two-step procedure to release their genetic material into the cell: (i) they first bind to specific surface receptors of the target cell membrane and then, (ii) they fuse the viral and cell membranes. This last step may occur at the cell surface or after internalization of the virus particle by endocytosis or by some other route (*e.g.*, macropinocytosis). Remarkably, the virus-cell membrane fusion process goes essentially along the same intermediate steps as other membrane fusions that occur for instance in vesicular fusion at the nerve synapsis or cell-cell fusion in yeast mating. Specialized viral proteins, fusogens, promote virus-cell membrane fusion. The viral fusogens experience drastic structural rearrangements during fusion, liberating the energy required to overcome the repulsive forces that prevent spontaneous fusion of the two membranes. This chapter describes the different types of viral fusogens and their mode of action, as are currently known.

Keywords Class I fusion protein • Class II fusion protein • Class III fusion protein • Enveloped virus • Fusion pore • Glycoprotein • Membrane • Membrane fusion intermediate • Post-entry events • Viral fusogen • Virus entry

V. Más • J.A. Melero (✉)
Biología Viral, Centro Nacional de Microbiología and CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Majadahonda, 28220 Madrid, Spain
e-mail: jmelero@isciii.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_16,

Abbreviations

FP	Fusion peptide
G	Glycoprotein
GPI	Glycosyl phosphotidylinositol
HA	Haemagglutinin
HIV	Human immunodeficiency virus
hMPV	Human metapneumovirus
HN	Haemagglutin-neuraminidase
HRA and HRB	Heptad repeat sequences A and B, respectively
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
RSV	Respiratory syncytial virus
SFV	Semliki forest virus
SNARE	Soluble N-ethylamine sensitive factor attachment receptor protein
TBV	Tick-borne encephalitis virus
TM	Transmembrane
VSV	Vesicular stomatitis virus

16.1 Introduction

Enveloped viruses are characterized by having a lipid bilayer (envelope) surrounding the virus particle (virion) (see Chap. 11). One or several virus-encoded glycoproteins are inserted into the envelope. These proteins are exposed at the virion surface and are responsible therefore of the initial interactions of the virus with the target cell, leading to "virus entry". In fact, it is only the cargo inside the envelope layer and not the virus itself what is actually discharged inside the cell. The virus glycoproteins are also the main targets of the neutralizing antibody response produced by the host in its defense against the virus.

To infect a new cell, the virus particle must first attach to the cell surface through non-covalent interactions of one or more of the viral glycoproteins anchored into the lipid bilayer with specific cell surface receptors. These interactions are described in detail in Chap. 15. Suffice to say here that virus-receptor binding is one of the factors that can influence virus tropism; *i.e.*, which cell types are actually infected by the virus.

For enveloped viruses, fusion of the viral and cell membranes is an obligatory step that follows virus binding to cells. Virus-cell fusion is therefore the step at which the virus particle loses its individuality. Membrane fusion may proceed at the cell surface or alternatively after internalization of the virus particle, generally by endocytosis. In either case, fusion is driven by specialized viral glycoproteins (fusogens) which are activated (triggered) by specific events occurring either at the cell surface or inside the endosome. The viral fusogens are in metastable

conformations in the virus particle. Once triggered, they initiate a series of conformational changes (in most cases irreversibly) that facilitate approximation of the two membranes, followed by fusion. At the end of the fusion process, the viral fusogens adopt highly stable conformations. The free energy liberated during the transition from the metastable pre-fusion to the highly stable post-fusion conformation drives the fusion process.

16.2 General Principles of Membrane Fusion

16.2.1 Protein-Free Membrane Fusion

Lipid mixing occurs spontaneously in monolayers but several forces prevent the spontaneous mixing of lipids between bilayer membranes [1]. Most important among them are: (i) hydrophobic effects that seek to minimize solvent-exposed apolar surfaces, (ii) elastic forces that prevent monolayer deformation and (iii) electrostatic repulsions between negatively charged phospholipids. Nevertheless, fusion between protein-free lipid bilayers (*e.g.*, liposomes) can be induced under certain conditions. For instance, certain phospholipids (*e.g.*, phosphatidylcholine, PC) induce positive curvature of the lipid monolayer whereas others (*e.g.*, phosphatidylethanolamine, PE) induce negative curvature. The distribution of PC and PE between the two leaflets of the lipid bilayer can either promote or inhibit spontaneous protein-free membrane fusion. Also direct dehydration between bilayers promotes fusion by bringing the two membranes into very close contact. However, under most physiological conditions, specialized proteins are needed to overcome the repulsive forces that prevent membrane fusion.

Independently of the driving machinery, fusion of two lipid bilayers occurs in a stepwise manner that includes the formation of an hourglass-like structure, known as the lipid stalk (Fig. 16.1a) [2]. This stalk is then expanded forming what is called the hemifusion diaphragm, in which lipids of the two distal leaflets of the bilayers are now in direct contact. Finally, rupture of the hemifusion diaphragm leads to formation of the fusion pore that is then expanded to complete membrane fusion and content mixing of the two compartments. However, opening of the fusion pore may be a reversible step and does not always lead to full fusion [3]. Energy is therefore required through all steps of the fusion process, including expansion of the fusion pore. Hence, proteins when present must operate from the initial stages of membrane deformation to final merging of the two membranes.

16.2.2 Virus-Induced Membrane Fusion

Enveloped viruses contain specialized surface glycoproteins that mediate: (i) initial binding of virus to the cell surface and (ii) fusion of the virus and cell membranes,

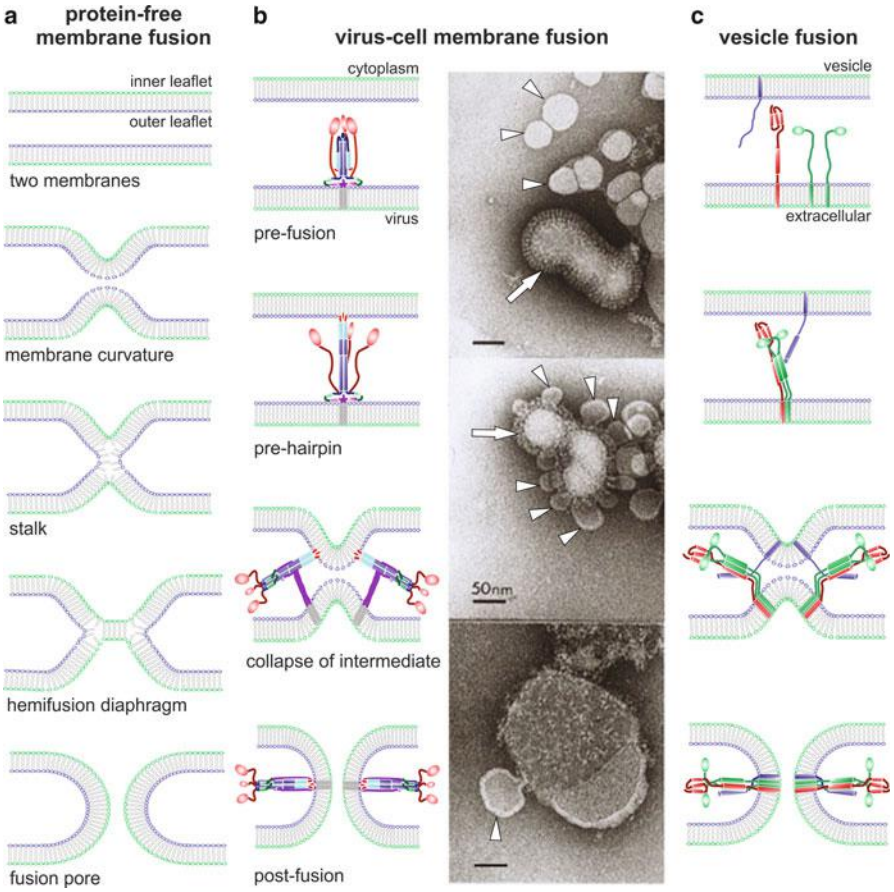

Fig. 16.1 Steps of the membrane fusion process. **(a)** Diagram of the fusion steps between two protein-free lipid bilayers. From *top* to *bottom*: the lipids (represented by heads and tails) of the two bilayers are initially curved into nipple-like structures that approach the two membranes. This is followed by formation of the stalk in which the two proximal leaflets are fused. This stalk is then expanded forming a hemifusion diaphragm in which lipids of the distal leaflets of the bilayers are now in direct contact. Finally, rupture of the hemifusion diaphragm leads to formation of the fusion pore. **(b)** Diagram of the virus-cell fusion process: As an example, fusion mediated by the influenza HA is illustrated. *Left panels:* HA is a homotrimer that initially binds to the cell surface (not shown) by interactions of each subunit head with sialic acid. Then, the virus is internalized by endocytosis. For simplicity, only 1 HA trimer in the pre-fusion conformation is shown, anchored into the viral membrane. After endosome acidification, the HA globular head falls apart, allowing refolding of the molecule to produce three long α-helices. The fusion peptides, placed at the N-terminal end of each α-helix, insert into the target membrane. This intermediate, dubbed pre-hairpin, refolds to bring the two membranes into proximity leading to formation of the lipid stalk followed by formation of the hemifusion diaphragm (not shown). Finally, the fusion pore is formed by the concerted action of several HA molecules that adopt a very stable post-fusion conformation. *Right panels:* The upper panel shows a mixture of an influenza virus particle (strain X31, H3N2, *white arrow*) and liposomes (some of them indicated by *white arrowheads*) made with lipids commonly found in cell membranes, incubated at neutral pH. Note the glycoprotein spikes

either at the cell surface or after endocytosis [4]. The initial attachment of the virus to the cell surface may involve only one type of viral glycoprotein (*e.g.*, influenza virus) or it may require the combined action of several viral glycoproteins (*e.g.*, herpesvirus). However, the actual process of membrane fusion is driven by a specialized type of viral glycoproteins (the viral fusogens) which may or may not have participated additionally in attachment.

Although the sequences of viral fusion proteins vary considerably, they all share certain structural characteristics and are subject to analogous structural rearrangements during the fusion process. As an example, Fig. 16.1b (left panels) illustrates the structural changes that the influenza virus haemagglutin (HA) undergo during fusion of the viral and the endosomal membrane. In this case, influenza HA also mediates the initial interaction of the virus with sialic acid of glycoproteins or glycolipids at the cell surface (see Chap. 15). After this initial binding, the virus is internalized by endocytosis. Acidification of the endosome (probably by fusion with lysosomes) triggers HA to start the conformational changes depicted in Fig. 16.1b (left panels) and described next.

Independently of the triggering event, all viral fusion proteins undergo conformational changes upon activation that lead to the formation of an extended unstable intermediate, dubbed pre-hairpin (Fig. 16.1b, "pre-hairpin"). Formation of the pre-hairpin intermediate involves very large-scale structural rearrangements in the viral fusogens with exposure of hydrophobic segments or loops (the fusion peptide, FP). Since the hydrophobic fusion peptide cannot be exposed to a hydrophilic environment it inserts into the target membrane. At this point the viral and target membranes are bridged by two separate segments of the same polypeptide; one is the fusion peptide bound to the target membrane and the other is the transmembrane (TM) region of the viral fusogen inserted into the viral membrane. The pre-hairpin intermediate may have a relatively long half-life; for the human immunodeficiency

Fig. 16.1 (continued) (mostly haemagglutinin, HA) sticking out of the viral membrane in contrast with the smooth surface of liposomes. The middle panel shows the same virus/liposome mixture after incubation for 5–10 s at pH 5.0 followed by neutralization. Note binding of liposomes to the virus surface and initiation of virus-liposome fusion. The lower panel shows the virus/liposome mixture after incubation for 5 min at pH 5.0 followed by neutralization. Note that the virus has fused with several liposomes, yielding a large vesicle with viral glycoproteins disperse throughout the surface and with a small liposome still in the process of fusion. The HA spikes also have changed morphology after exposure to low pH and fusion. (Courtesy of L.J. Calder and S.A. Wharton, Division of Virology, MRC National Institute for Medical Research, London, UK). (**c**) Diagram of vesicle fusion at the synaptic junction: Initially one of the SNARE proteins (synaptobrevin, *blue*) is inserted into the vesicle membrane, while three other SNAREs (two SNAP 25, *green* and one syntaxin, *red*) are inserted in the plasma membrane. After an initial interaction, refolding of the SNAREs leads to formation of a bundle of four parallel α-helices that drives approximation of the two membranes and formation of the stalk and hemifusion intermediates (not shown). Completion of the SNARE complex results in formation of the fusion pore. In the two lower panels of parts (b) and (c), two HAs and two SNARE complexes are shown surrounding the fusion pore, although the actual number of molecules involved in fusion pore formation is likely to be higher

virus (HIV) gp41 protein, the half-life seems to be several minutes [5], but in other cases, it may only be a few seconds [6]. The pre-hairpin bridge then collapses bringing into close apposition the viral and target membranes, which are distorted probably into protein-free nipple-like configurations (Fig. 16.1b, "collapse of intermediate"). This is followed by the formation of a lipid stalk and a hemifusion diaphragm (in analogy with the protein-free membrane fusion path), which allows lipid mixing between the two proximal leaflets of the viral and target membranes. Finally, the hemifusion diaphragm opens to form a transient fusion pore that may flicker open and closed until it expands [7], leaving the viral fusogen in a highly stable post-fusion hairpin conformation inserted into the target membrane (Fig. 16.1b, "post-fusion").

Influenza HA dependent membrane fusion can be reproduced in the test tube with purified virus and liposomes and observed by electron microscopy, as illustrated in Fig. 16.1b (right panels). After mixing of virions (arrows) and liposomes (arrowheads) they remain separated (upper panel), since the former lack the influenza virus receptor (sialic acid). However, a brief pulse at low pH (middle panel) exposes the fusion peptide of the influenza HA which is then inserted into the membrane of multiple liposomes. Longer pulses of low pH result in fusion of the virus with multiple liposomes leading to formation of large vesicles (lower panel; note two vesicles, one small and one large, caught in the process of fusion).

The fusion peptides probably insert only into the outer leaflet of the cell target membrane. Due to the large number of fusogen molecules present at the viral surface, multiple fusion peptides may interact with the external leaflet of the target membrane upon formation of the pre-hairpin intermediate, potentially initiating membrane deformation. This suggests that cooperativity between several viral fusogens may be required for membrane fusion. In fact, fusion mediated by the influenza HA is positively affected by protein density [8]. It is estimated that 4–6 HA molecules are required for fusion, forming a protein ring at the periphery of the fusion pore. Also, electron microscopy (Chap. 3) and X-ray crystallography (Chap. 4) results indicate that the E1 glycoprotein of Semliki Forest alphavirus interacts cooperatively during membrane insertion and fusion [9].

Despite the above arguments in favor of cooperativity, calculations of the energy barrier that must be overcome en route to a hemifusion diaphragm is estimated to be about 40–50 kcal.mol^{-1}. A free energy of roughly this magnitude could be recovered from the collapse of one or two pre-hairpin intermediates, depending on the interactions driving such collapse. In fact, experiments with HIV suggest that only one or two active envelope glycoproteins are sufficient for fusion [10], although later estimates have increased this number [11]. It may be that the fusion proteins of HIV and other retroviruses have evolved to manage with a single fusion protein, as the number of envelope glycoproteins in the virus particle (estimated 15–20, in contrast to hundreds in other viruses) is rather sparse.

Formation of the pre-hairpin structure and refolding of this intermediate entails some of the most drastic protein rearrangements ever found in biology. Pre-hairpin collapse involves folding back of the membrane proximal domain of the viral

fusogen onto a trimeric core whose distal end from the viral membrane is inserted into the target membrane (Fig. 16.1b). Zippering together of these two domains brings the membranes into close proximity. Dehydration of the initial contact site induces monolayer rupture resulting in lipid stalk formation and hemifusion. However, formation of the fusion pore requires further structural rearrangements, including interactions between regions adjacent to the fusion peptide and the transmembrane region [12, 13] and, probably, additional contacts between these two hydrophobic regions that are now inserted into the same membrane. For instance, membrane fusion by the influenza HA with a glycosyl phosphotidylinositol (GPI) anchor replacing the TM region halts at the hemifusion stage [14].

Finally, enlargement of the initial fusion pore is probably the most energy demanding step and requires the coordinated action of several fusogen molecules that surround the early nipple-like fusion intermediate [15].

16.2.3 Vesicle and Cell-Cell Fusion

This topic is brought here only to emphasize the analogies and differences between membrane fusions promoted by unrelated proteins. Vesicle fusion is required for essential biological processes, such as exocytosis and synaptic transmission. Cell-cell fusion is involved in hypodermal cell fusion in *C. elegans*, sperm-egg fusion, yeast mating (mating of two haploid yeast cells to produce a diploid cell), placenta formation in mammals, and muscle and bone formation.

In all cases, membrane fusion follows the same steps already described in previous sections; *i.e.*, deformation and approximation of the two membranes, formation of the stalk and hemifusion intermediates and finally formation and enlargement of the fusion pore (Fig. 16.1c). In analogy with virus-cell fusion, vesicle and cell-cell fusion requires formation of highly stable protein assemblies that provide the energy necessary to overcome the repulsive forces of membranes in close proximity [16]. Also, vesicle and cell-cell fusion, as viral fusion, requires higher order multimerization of the fusogens that delineate the hemifusion diaphragms and the fusion pores [17].

The main difference between virus-cell fusion and vesicle or cell-cell fusion is that in the former process the protein fusogen is present only in the viral membrane. In contrast, the proteins involved in vesicle fusion and cell-cell fusion are initially inserted in the two membranes predestined to fuse (Fig. 16.1c). In synaptic vesicles, the main proteins responsible of membrane fusion are the so-called SNARE (soluble N-ethylamine sensitive factor attachment receptor protein) proteins [18] which share a conserved 60–70 amino acid motif. These proteins, when they find each other refold into a highly stable four-helix parallel coiled-coil bundle that resembles the six-helix bundle formed by the heptad repeats (structural motifs with a repeating pattern of seven amino acids) of certain viral fusogens (see below). Formation of the four-helix bundle leads to membrane apposition and hemifusion, as with the collapse of the pre-hairpin intermediate of viral fusogens. However, a

unique characteristic of vesicle fusion is that the protein machinery involved in the process is disassembled, once fusion is finished to be reused in subsequent fusion events. This is accomplished by the ATPase (adenosine triphosphatase) N-ethylmaleimide sensitive factor (NSF) [19].

In contrast to vesicle fusion, cell-cell fusion entails the same set of fusion proteins in the two membranes. For instance, the exceptional process of hypodermal cell fusion in *C. elegans* to form a large multinucleated syncytium of all skin cells is driven by the EFF-1 protein [20]. Unlike SNAREs and viral fusogens, EFF-1 has a homotypic fusion machinery in the opposite membrane. In other words, both membranes must have EFF-1 for fusion to occur. Nevertheless, cell-cell fusion is a multistep process that goes along the same lipid intermediates as viral and vesicle fusion.

16.3 Viral Fusion Proteins

Based on biosynthetic and structural characteristics, viral fusogens have been classified into three categories (Table 16.1). Class I fusion glycoproteins are characterized by being synthesized as inactive precursors that require proteolytic processing to become fusion-competent. They are all homotrimers that upon fusion refold into hairpins containing a long central coiled-coil core structure (formed by helices that are coiled together). Class II fusion glycoproteins are derived from longer polyprotein precursors that are proteolytically processed during biosynthesis. The class II fusion proteins form icosahedral scaffolds of protein dimers at the viral surface. During fusion, these proteins undergo an oligomeric rearrangement, converting the metastable prefusion dimer into a stable hairpin homotrimer composed of β-sheet structures. Finally, class III glycoproteins are not proteolytically processed. Their post-fusion hairpin trimer displays a central α-helical coiled-coil, as class I glycoproteins, but the fusion domain exposes two fusion loops located at the tip of an elongated β-sheet, revealing a striking convergence with class II fusion proteins.

16.3.1 Class I Viral Fusion Proteins

The first atomic structure of any viral or cellular glycoprotein was determined by X-ray crystallography and reported in 1981 by the laboratories of Wiley and Skehel [21]. It was the structure of the influenza haemagglutinin (HA) trimeric ectodomain (the domain that protrudes from the plasma membrane), as released from the virus particles by bromelain treatment, which cleaves the HA polypeptides near the TM region.

Influenza HA is synthesized in the infected cell as a polypeptide precursor (HA0) of about 550 amino acids that is cleaved proteolytically to generate the HA1 (roughly the N-terminal two thirds) and HA2 (the C-terminal third) chains that

Table 16.1 Classification of viral fusion proteins

Class	Virus family	Representative	Viral fusogen	Involved in attachment
I	*Orthomyxoviridae*	Influenza virus	Haemaglutinin (HA)	Yes
	Retroviridae	Human immunodeficiency virus (HIV)	Envelope glycoprotein; gp 41 subunit	Yes
	Filoviridae	Ebola virus	GP glycoprotein	Yes
	Coronaviridae	Severe acute respiratory syndrome (SARS) virus	S glycoprotein	Yes
	Paramyxoviridae	Sendai virus	F glycoprotein	No
II	*Alphaviridae*	Semliki Forest Virus	E1 glycoprotein	No
	Flaviviridae	Dengue virus	E glycoprotein	Yes
III	*Rhabdoviridae*	Vesicular stomatitis virus	G glycoprotein	Yes
	Baculoviridae	Baculovirus	Gp64 glycoprotein	Yes
	Herpesviridae	Herpes simplex virus	gB glycoprotein	No

remain covalently linked by a disulfide bond. At the newly created HA2 N-terminus there is a stretch of hydrophobic amino acids, called the fusion peptide, which is inserted into the target membrane during fusion. The overall structure of the influenza HA is that of an elongated spike sticking out of the membrane. The distal head, formed exclusively by HA1 sequences, bears the receptor (sialic acid) binding site, formed by a shallow pocket exposed on its outward-forming surface. The stem, made largely by HA2 amino acids, is a trimeric α-helical coiled-coil. The structural rearrangements of the influenza HA during membrane fusion are shown in Fig. 16.1b.

As influenza, other viruses also contain class I fusion glycoproteins that have both receptor and membrane fusion activities (Table 16.1). For instance, the envelope glycoprotein of HIV that is also proteolitically processed and that binds to protein receptors (CD4) and chemokine co-receptors before engaging in membrane fusion at the cell surface. Similarly, the receptor-binding proteins of filovirus and coronavirus mediate additionally viral-cell membrane fusion.

In contrast, the attachment and fusion activities reside in two different surface glycoproteins of paramyxoviruses. The attachment protein (named HN, H or G) is required for the initial interaction of the virus with the cell surface (see Chap. 15 for virus receptor usage). Once the virus is bound to the cell, the other major viral glycoprotein (called F, for fusion) is triggered to promote fusion of the viral and cell membranes. Structure determination of prototypic paramyxovirus F proteins in the pre-fusion metastable conformation [22] and in the post-fusion state [23] by X-ray crystallography, as well as identification of fusion intermediates [24], has provided the most complete picture of the membrane fusion process driven by class I fusion glycoproteins, as depicted in Fig. 16.2.

The paramyxovirus F protein, as other class I glycoproteins, is synthesized as an inactive precursor (F0) that is translocated co-translationally to the lumen of the endoplasmic reticulum where it assembles into a trimer. Each F protein subunit is

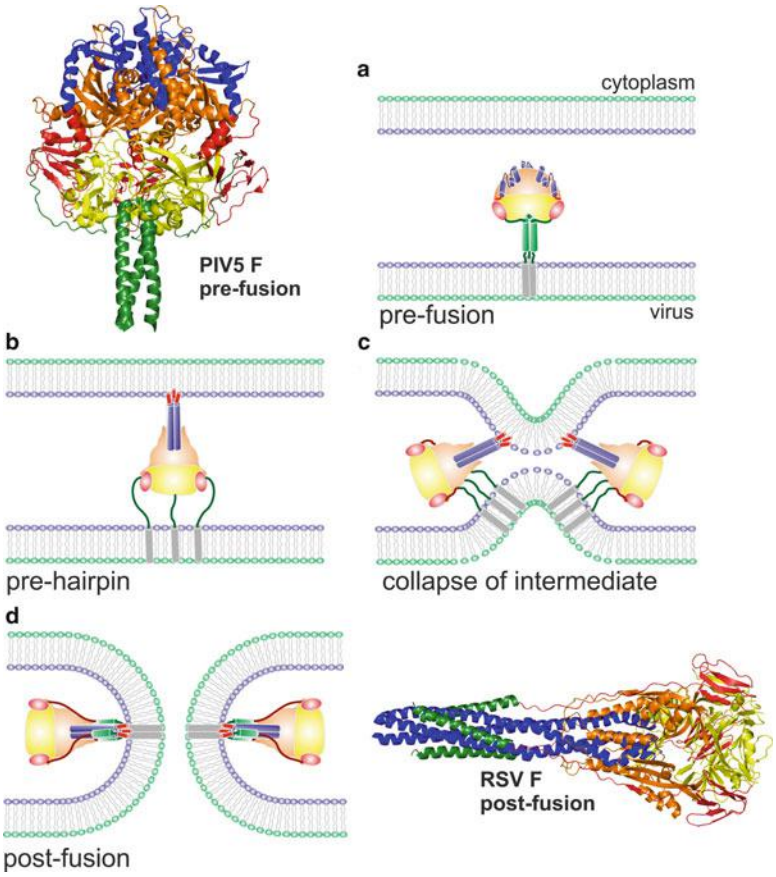

Fig. 16.2 Membrane fusion mediated by a class I fusion protein (Paramyxovirus). The atomic structures of the pre-fusion form of Parainfluenza virus type 5 (PIV5) [22] (*upper left*) and the post-fusion form of Respiratory Syncytial Virus (RSV) [53] (*lower right*) F proteins are shown as ribbons. The same protein regions are highlighted with identical colors in the two conformations. **(a–d)** Diagram of the fusion process denoting: **(a)** the pre-fusion paramyxovirus F protein trimer inserted in the viral membrane before activation, **(b)** formation of the pre-hairpin structure which includes refolding of the long central HRA α-helices (*blue*) with the fusion peptide (*red*) inserted into the cell membrane, **(c)** collapse of the pre-hairpin to approach the two membranes, and **(d)** formation of the fusion pore and stabilization of the F trimer in the post-fusion conformation. In the last two steps, two F protein molecules are represented to indicate the cooperation needed to drive the fusion process

proteolytically cleaved during transport to the cell surface, generating two chains, F2 N-terminal and F1 C-terminal that remain linked by one or more disulfide bonds. F1 (equivalent to the HA2 chain of influenza virus) has a hydrophobic fusion peptide at the N-terminus and two heptad repeat sequences (HRA and HRB) in its ectodomain. HRA is adjacent to the fusion peptide and HRB is proximal to the transmembrane (TM) region, which is placed near the F1 C-terminus.

The pre-fusion three-dimensional (3D) structure of the parainfluenza virus type 5 (PIV5) F protein contains a large globular head connected to a short trimeric coiled-coil made by the HRB region [22] (Fig. 16.2). Comparison with the post-fusion structure of the F ectodomain from other paramyxovirus (for instance respiratory syncytial virus (RSV (Fig. 16.2)) and functional studies using peptide inhibitors [24] provided the following model of membrane fusion: Upon virus binding to cells through the attachment protein, F is activated and initiates a series of conformational changes, including separation of the HRB coils and refolding of HRA sequences to form a very elongated trimeric coiled-coil. The fusion peptides -now at the N-terminus of the HRAs- insert into the target cell membrane, resulting in formation of the pre-hairpin intermediate. This step is followed by zipping of the C-terminal part of the molecule along the core coiled-coil to bring together the two membranes and the fusion and TM domains, in analogy with the process described before for influenza HA. However, in the case of the paramyxovirus F the HRB sequences wrap around the HRA coiled-coil forming an extremely stable six-helix bundle (6HB) in the post-fusion hairpin. Formation of this 6HB provides most of the energy required to overcome membrane repulsion. The 6HB structure is shared by other class I fusion glycoproteins, such as the gp41 chain of the HIV envelope glycoprotein.

While activation of influenza HA requires exposure to the endosomal low pH (probably by protonation of key amino acid residues), the event that triggers paramyxovirus F proteins is still ill-defined. Cell-cell fusion of transfected cells that express paramyxovirus F requires in most cases co-expression of the homotypic attachment protein, suggesting that an interaction of the two proteins is needed for membrane fusion. Two alternative models ("clamp" and "provocateur") have been proposed to explain the requirement of the attachment protein for fusion:

1. The *clamp* model postulates that HN (or the equivalent attachment protein depending on the virus) is complexed with F in the virus particle, retaining the latter in the metastable configuration. Conformational changes in HN upon receptor binding release F from the complex to initiate membrane fusion.
2. Alternatively, the *provocateur* model postulates that HN and F do not interact in the virus before contacting the cell. Concomitantly to the structural changes induced in HN upon receptor binding, HN binds to F and this interaction triggers F for fusion [25].

Intriguingly, the F protein of viruses belonging to the *Pneumovirinae* subfamily of paramyxoviruses (*e.g.*, RSV and human metapneumovirus, hMPV) do not require co-expression of the attachment protein (G) for cell-cell fusion [26]. Furthermore, deletion mutant viruses have been obtained in which the entire G gene is obliterated. These mutants still infect cells *in vitro*, although less efficiently than the wild type virus and are attenuated in animal models of infection [27]. Activation of the F protein of those deletion mutants cannot rely on interactions with the G protein and therefore alternative regulatory mechanisms should control membrane fusion. Of note, a unique characteristic of the RSV F protein is the presence of two proteolytic cleavage sites (instead of one, as in all other

paramyxovirus) in the F0 protein precursor [26]. The presence of a double cleavage site in F has been found to influence membrane fusion activation by a still poorly understood G independent mechanism [28].

16.3.2 Class II Viral Fusion Proteins

In contrast to class I fusion proteins, the so-called class II fusion proteins (Table 16.1) are derived from a polyprotein precursor that is cleaved during biosynthesis to generate the E1 protein of alphaviruses (*e.g.*, Semliki Forest virus (SFV)) or the E protein of flaviviruses (*e.g.*, dengue virus and tick-borne encephalitis virus). Both proteins fold co-translationally with a companion or regulatory protein, termed p62 for alphaviruses and prM for flaviviruses [9].

In alphavirus, the p62-E1 complex is transported to the plasma membrane where they are incorporated into new budding icosahedral virus particles as dimers of p62-E1. p62 is then proteolitically processed (and then named E2) but remains bound to the virus where it covers most of the fusion protein E1 and specially its fusion loop. E2 mediates binding to the cell surface receptor.

In contrast, flavivirus particles bud into the endoplasmic reticulum as immature virions containing prM-E protein complexes. The immature viruses are then transported to the exterior through the exocytic pathway where prM is processed and separated from E [29]. The latter protein is then arranged in E-E homodimers at the virion surface with icosahedral symmetry. The flavivirus fusion E protein is additionally responsible for receptor binding.

The first structure of any class II glycoprotein, solved by X-ray crystallography, was that of the tick-borne encephalitis (TBE) flavivirus E protein ectodomain [30], solubilized from virions by limited trypsin digestion. Similar structures have now been solved for the E ectodomain of dengue virus types 2 and 3 [9]. The polypeptide chain of the E protein follows a complex path, resulting in three globular domains, essentially constituted by β-sheets (Fig. 16.3). The first domain is a β-barrel with up-and-down topology (red). Two adjacent strands in domain I are extended, forming domain II (yellow) which is a long "finger-like" structure that runs parallel to the viral membrane. At the tip of domain II is the hydrophobic fusion loop which remains buried in the virion from the hydrophilic environment by interaction with domain III (blue) of the adjacent monomer in the E-E dimer. Domain I is also connected to domain III which bridges the E ectodomain with the so-called stem region that extends to the TM region of the protein.

Unlike the class I fusion proteins, which are trimeric in their pre- and post-fusion conformations, class II fusion glycoproteins undergo major oligomeric transformations during fusion. As in the case of influenza virus, the flavivirus E protein first binds to a cell surface receptor which induces endocytosis of the virion. Once in the acidic endosome, the E-E homodimer dissociates, resulting in disassembly of the icosahedral scaffold. The individual subunits swing outward by the hinge region that connects domains I and II, and the fusion loops insert into the target membrane. Lateral

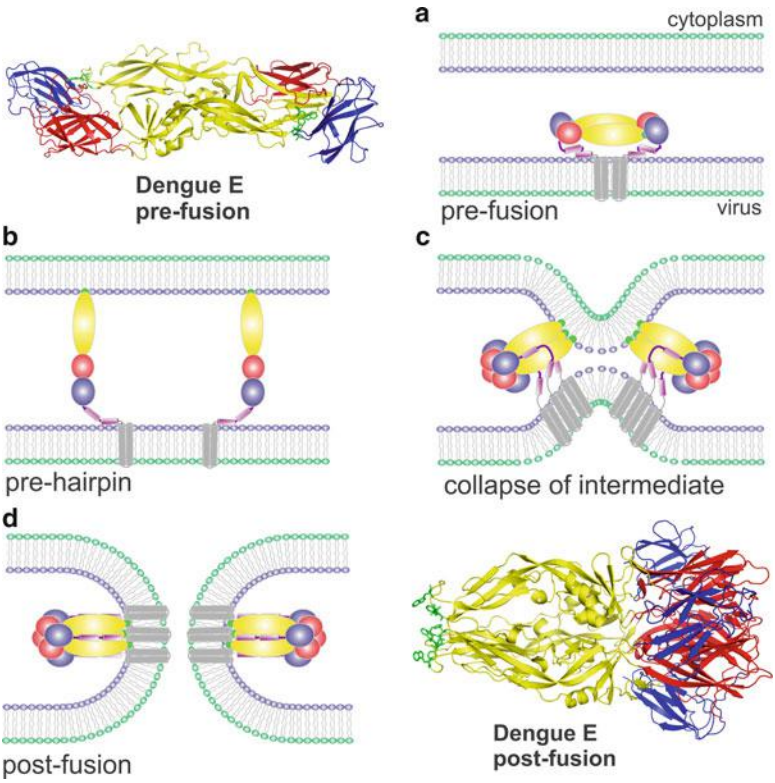

Fig. 16.3 Membrane fusion mediated by a class II fusion protein (Flavivirus). Ribbon representation of the atomic structures of the dengue virus E protein dimer in the pre-fusion conformation [54] (*upper left*) and the E protein trimer in the post-fusion conformation [55] (*lower right*). Domains I, II and III of the E glycoprotein are colored *red*, *yellow* and *blue*, respectively. **(a–d)** Diagram of the fusion process denoting: **(a)** the structure of the flavivirus E glycoprotein dimer already in the endosome before activation, **(b)** dissociation of the E protein subunits, refolding of the fusion domain (*yellow*) and insertion of the fusion loop (*green*) into the endosomal membrane, **(c)** formation and refolding of the E protein trimer to approach the two membranes, and **(d)** formation of the fusion pore and stabilization of the E trimer in the post-fusion conformation. In the last two steps, two E protein molecules are represented to indicate the cooperation needed to drive the fusion process

interactions between monomers facilitates reclustering into trimers [31]. These rearrangements lead to the formation of an extended trimeric structure, analogous to the pre-hairpin intermediate of class I fusion glycoproteins, in which two different regions of each E polypeptide are inserted into the two membranes to be fused. Collapse of the extended intermediate can proceed by rotation of domain III in each subunit about the segment that links it to domain I and zipping up of the stem alongside the clustered domains II. This refolding brings the two membranes together to initiate formation of the lipid stalk, the hemifusion diaphragm and the fusion pore.

The structure of the fusion E1 glycoprotein of alphaviruses (SFV) was found unexpectedly very similar to that of the flavivirus E protein, despite the lack of

detectable sequence conservation. E1 has also three discernible domains, equivalent to those of flavivirus E. The only significant difference is the association of E1 with E2 in the virus particle. E2 interacts with the cell surface receptor to initiate the endocytic internalization of the SFV virion [32]. In the acidic endosome, E2 separates from E1 and it is probably degraded. Upon low pH exposure, E1 undergoes similar conformational changes to those of the flavivirus E protein, leading to fusion of the viral and endosomal membranes. Electron microscopy and X-ray crystallography results provide support for interactions between adjacent E1 trimers when the fusion loops are inserted in the target membrane to produce rings of five or six trimers. It has been postulated that these fivefold interactions would act at the fusion site to induce the formation of a nipple-like curvature in the viral and target membranes, favoring membrane fusion [33]. Although there is no direct evidence, it is likely that the flavivirus E protein forms similar rings of trimers during fusion.

16.3.3 Class III Viral Fusion Proteins

The best characterized members of the so-called class III fusion viral glycoproteins are the rhabdovirus (*e.g.*, vesicular stomatitis virus, VSV) G glycoprotein, the herpesvirus gB glycoprotein and the baculovirus gp64 glycoprotein. The pre-fusion [34] and post-fusion [35] structures of the VSV_G glycoprotein ectodomain have been solved by X-ray crystallography while only the post-fusion conformations of gB [36] and gp64 [37] are known.

Class III fusion glycoproteins are expressed from individual mRNAs and do not require proteolytic processing of either a protein precursor (as in class I proteins) or an accompanying protein (as in class II proteins) for activity. Class III proteins are trimeric before and after fusion and share structural characteristics with both class I and class II fusion glycoproteins, as described below.

The rhabdovirus G protein possesses both receptor binding and fusion promoting activities. As in the case of influenza virus, binding of rhabdovirus G to a poorly characterized receptor at the cell surface induces endocytosis of the virus particle. Acidification of the endosome triggers G for membrane fusion. However, and in contrast with all other fusion proteins, the low pH inactivation of rhabdovirus G is reversible. Thus, virions inactivated by prolonged incubation at pH <6 can be reactivated by raising the pH to neutral [38]. This reversibility may be required to allow G to be transported through the acidic Golgi apparatus and to recover its native fusion-competent state when incorporated to new virions [39]. Given this reversibility, it is believed that the energy released during the structural transition of a single trimer from the pre-fusion to the post-fusion conformation is probably small, compared with the energetic barrier of the fusion reaction. In agreement with this hypothesis, the estimated number of rhabdovirus spikes required for fusion is higher (at least 15 trimers) than for other enveloped viruses.

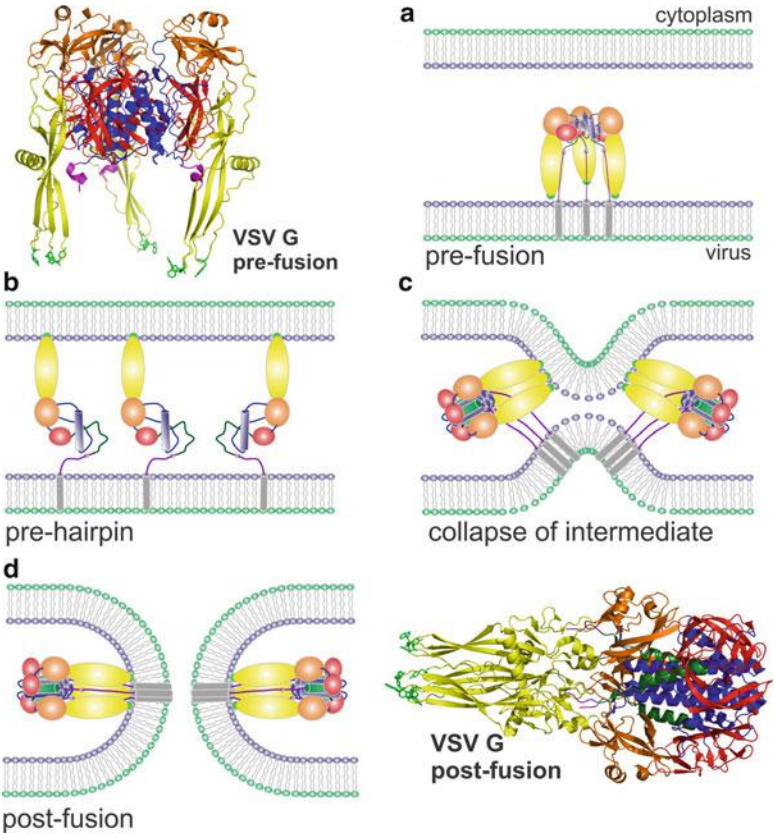

Fig. 16.4 Membrane fusion mediated by a class III fusion protein (Vesicular Stomatitis Virus, VSV). Ribbon representations of the VSV glycoprotein (G), in the pre- (*upper left*) and post-fusion (*lower right*) conformations. Domains are colored similarly in all images. The fusion domain is colored in *yellow* and the fusion loops in *green*. **(a–d)** Diagram of the fusion process denoting: **(a)** the structure of the VSV G glycoprotein trimer already in the endosome before activation, **(b)** dissociation of the G protein subunits, refolding of the fusion domain (*yellow*) and insertion of the fusion loop (*green*) into the endosomal membrane, **(c)** formation and refolding of the G protein trimer to approach the two membranes, and **(d)** formation of the fusion pore and stabilization of the G primer in the post-fusion conformation. In the last two steps, two G protein molecules are represented to indicate the cooperation needed to drive the fusion process

A soluble ectodomain of the VSV_G glycoprotein, released from purified virions treated with thermolysin, was used to solve the structures of the pre-fusion and post-fusion conformations, after exposure to high and low pH, respectively [34, 35]. Several domains could be observed in both structures that are rearranged in their relative orientations during transit from the pre- to the post-fusion structure (Fig. 16.4). In the pre-fusion conformation, the fusion domain contains two fusion loops reminiscent of class II proteins that are oriented downward towards the viral membrane. After low pH exposure, the fusion domain moves upward by flipping

relative to the central core of the trimer. Thus, an intermediate equivalent to the pre-hairpin structure of class I proteins is formed. This is followed by the reversal of the molecule around a central rigid block formed by lengthening of the central helix and refolding of the three C-terminal segments into helices that position themselves in the grooves of the central core in an anti-parallel manner. This six-helix bundle has obvious resemblance with that of the class I proteins.

It is likely that the transition of VSV_G from the pre-fusion to the post-fusion conformation involves disassembly of the trimer into monomers and reassembly into trimers upon interaction of the fusion loops with the target membrane [40]. It is also likely that cooperativity between G glycoproteins is needed to overcome the energy barrier, as mentioned above. As for class II glycoproteins, lattices of G proteins have been observed in virions, particularly in the planar base of the rhabdovirus bullet-shape particle, which may act to induce nipple-like deformations in the viral and target membranes.

Although it has not been reported, it is likely that low pH exposure also leads to reversible inactivation of the baculovirus gp64 glycoprotein. This protein, like VSV_G, is involved in both receptor binding and membrane fusion after endosome acidification, since baculoviruses also use an endocytic route of entry [41]. In contrast, membrane fusion mediated by herpesvirus gB can occur either at the plasma membrane or in endosomes, depending on the virus and the target cell type. In either case, attachment of herpesvirus to host cells follows a complex mechanism in which several viral glycoproteins interact with cell surface molecules. Some of these interactions trigger fusion, whereas others simply serve to tether the virus to the cell and are dispensable for fusion. In any case, the gB protein, shared by all viruses of the *Herpesviridae* family, is responsible for fusion [42].

16.3.4 Other Viral Fusion Proteins

Poxviruses (vaccinia virus is the best known member) represent an extreme case among enveloped viruses, regarding the number of viral glycoproteins required for entry. As for herpes virus, entry can occur by fusion at the plasma membrane or in a low pH-dependent manner from within an intracellular particle, depending on the virus strain and the cell type. Vaccinia virus internalization is believed to occur by macropinocytosis (a type of non-specific endocytosis). Four vaccinia virus proteins are involved in attachment to cell surface proteoglycans or laminin [43]. Eleven or 12 other relatively small glycoproteins, ranging in size between 35 and 377 amino acids, form the so-called entry fusion complex (EFC) that mediates membrane fusion [44]. These proteins have N- or C-terminal transmembrane domains but no sequence similarity with the fusion peptide of other viral fusion glycoproteins has been found in any of them. Therefore, the actual mechanism of vaccinia virus membrane fusion remains to be elucidated but it seems to be different from that of other enveloped viruses. By using conditional lethal mutants of each of the 11 proteins that make the EFC, it was found that eight of them were required to reach

the hemifusion step and the other three were needed for completion of virus entry [44]. It is likely that hydrophobic regions of several proteins may assemble in the EFC to form a hydrophobic surface that could bind to the target membrane and drive membrane fusion by some novel mechanism.

Finally, the fusion-associated small transmembrane (FAST) proteins of reoviruses are brought here -despite not being involved in virus entry and reovirus being a non-enveloped virus- because they induce cell-cell fusion and therefore facilitates dissemination of virus to neighboring cells. The FAST proteins are small non structural proteins (98–148 amino acids, depending on the viral strain) that are expressed on the surfaces of virus-infected cells, where they induce cell-cell fusion and syncytia (multinucleate cells) formation. Purified FAST proteins, when reconstituted into liposome membranes, induce fusion indicating that they are *bona fide* fusogens [45]. The orientation of the FAST polypeptides in the cell membrane is also unique among viral fusogens, with a relatively short N-terminal ectodomain followed by a transmembrane region and a long C-terminal cytoplasmic tail. Although they lack a fusion peptide, a relatively hydrophobic region near the N-terminus which is additionally myristoylated seems to insert into the target membrane to drive membrane fusion, at least for certain FAST proteins [46] .

16.4 Early Post-Entry Events

Once membrane fusion has been completed, the viral genome -generally in complex with other proteins or inside a viral nucleocapsid (see Chaps. 2 and 11)- is found for a second time in a cytoplasmic environment. The first time is when the genome assembles in the cytosol of the infected cell or when it is trafficking from the nucleus to the cell exterior, depending on the virus. However, the fate of the incoming genome is now very different and characteristic for each virus.

Most RNA viruses replicate in the cell cytoplasm, although there are exceptions like influenza virus or borna virus that do so in the nucleus. If the RNA is of positive polarity, like in flavivirus, the genome may act as mRNA to be translated by the cell protein synthesis machinery. In most cases the primary translational product is a polyprotein that matures into the different viral gene products by proteolytic processing [47]. In the case of negative-stranded RNA viruses, like paramyxoviruses, the first step after entry is the transcription of the viral genome to yield the different mRNAs that are translated into the distinct viral gene products [48].

For RNA viruses that replicate in the nucleus, the nucleoprotein complex of the viral genome and associated proteins has to be transported to the cell nucleus for transcription. Most of the viral proteins required in the nucleus have their own nuclear localization signal (NLS), which is a cluster of basic amino acids. However, actual import of the viral ribonucleoprotein into the cell nucleus may require additional interactions with certain host factors. For instance, the NLS of the influenza nucleoprotein interacts with karyopherin α and this in turn with karyopherin β which

mediates interactions with the proteins of the nuclear pore to promote nuclear import of the viral ribonucleoproteins by an energy-dependent process [49].

Retroviruses represent a special case in which a RNA viral genome (diploid), packed in a capsid inside the virus envelope has to be transcribed to DNA before integration into the cell host genome. In this case, fusion of the virus and cell membrane delivers the capsid into the cell cytoplasm, where it interacts with cytoskeleton and other cell components for transport to the vicinity of the nucleus where reverse transcription and uncoating takes place. Then, the resulting pre-integration complexes are transported through the nuclear pore inside the nucleus for integration into the host genome [50].

Most DNA viruses replicate in the nucleus, with exceptions like poxviruses (*e.g.*, vaccinia virus). In the case of herpesviruses, the capsids that contain the viral genome are transported from the site of release to the nuclear pore where importin β promotes nuclear import of the viral DNA by an energy-dependent mechanism [51]. In contrast, the virus core of vaccinia virus is released into the cytoplasm. The virus core has all the machinery for transcription of the early genes that ensues further progress on DNA replication and transcription of the remaining genome [52].

16.5 Perspectives and Conclusions

As noted, viral-cell membrane fusion is the last extracellular step of enveloped viruses and therefore still amenable to inhibition by chemical or biological products without the drawback of the membrane permeability barrier. Hence, it is not surprising that viral fusogens have received recently very much attention as ideal targets for the development of effective antivirals, some of them already in clinical use (*e.g.*, the T20 peptide for HIV or the Synagis antibody for respiratory syncytial virus) (see also Chap. 20). The development of high throughput technologies for screening of large libraries of chemical or biological molecules (*e.g.*, antibodies) should provide in the near future a plethora of drugs to fight some of the important diseases caused by enveloped virus.

A critical step in the process of viral induced membrane fusion is the activation of the viral fusogen. While most viruses that enter the cell through low pH endocytosis rely on protonation of certain residues (mostly histidines) of the viral fusogen to trigger fusion, the activating step in the case of other viruses is still ill-defined. Better understanding of fusion triggering may therefore bring new possibilities for the manipulation of virus entry. Finally, determination of the structures of certain fusion intermediates could provide additional targets for antivirals.

Learning how virus-cell membrane fusion proceeds may also be relevant for other spheres of Biology. For instance, regulation of vesicle fusion could benefit from knowledge acquired in the field of Virology and find applications in studies of synaptic transmission and Neurobiology.

Acknowledgements Current research activities in the Biología Viral laboratory are funded by grants GR09/0039 from Instituto de Salud Carlos III and SAF2009-11632 from Plan Nacional de I + D + i.

References and Further Reading

1. Chernomordik LV, Kozlov MM (2008) Mechanics of membrane fusion. Nat Struct Mol Biol 15:675–683
2. Yang L, Huang HW (2002) Observation of a membrane fusion intermediate structure. Science 297:1877–1879
3. Chanturiya A, Chernomordik LV, Zimmerberg J (1997) Flickering fusion pores comparable with initial exocytotic pores occur in protein-free phospholipid bilayers. Proc Natl Acad Sci U S A 94:14423–14428
4. Harrison SC (2008) Viral membrane fusion. Nat Struct Mol Biol 15:690–698
5. Munoz-Barroso I, Durell S, Sakaguchi K, Appella E, Blumenthal R (1998) Dilation of the human immunodeficiency virus-1 envelope glycoprotein fusion pore revealed by the inhibitory action of a synthetic peptide from gp41. J Cell Biol 140:315–323
6. Floyd DL, Ragains JR, Skehel JJ, Harrison SC, van Oijen AM (2008) Single-particle kinetics of influenza virus membrane fusion. Proc Natl Acad Sci U S A 105:15382–15387
7. Zimmerberg J, Blumenthal R, Sarkar DP, Curran M, Morris SJ (1994) Restricted movement of lipid and aqueous dyes through pores formed by influenza hemagglutinin during cell fusion. J Cell Biol 127:1885–1894
8. Danieli T, Pelletier SL, Henis YI, White JM (1996) Membrane fusion mediated by the influenza virus hemagglutinin requires the concerted action of at least three hemagglutinin trimers. J Cell Biol 133:559–569
9. Kielian M, Rey FA (2006) Virus membrane-fusion proteins: more than one way to make a hairpin. Nat Rev Microbiol 4:67–76
10. Yang X, Kurteva S, Ren X, Lee S, Sodroski J (2006) Subunit stoichiometry of human immunodeficiency virus type 1 envelope glycoprotein trimers during virus entry into host cells. J Virol 80:4388–4395
11. Magnus C, Rusert P, Bonhoeffer S, Trkola A, Regoes RR (2009) Estimating the stoichiometry of human immunodeficiency virus entry. J Virol 83:1523–1531
12. Chen J, Skehel JJ, Wiley DC (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. Proc Natl Acad Sci U S A 96:8967–8972
13. Buzon V, Natrajan G, Schibli D, Campelo F, Kozlov MM, Weissenhorn W (2010) Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. PLoS Pathog 6:e1000880
14. Kemble GW, Danieli T, White JM (1994) Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion. Cell 76:383–391
15. Chernomordik LV, Kozlov MM (2003) Protein-lipid interplay in fusion and fission of biological membranes. Annu Rev Biochem 72:175–207
16. Fasshauer D, Otto H, Eliason WK, Jahn R, Brunger AT (1997) Structural changes are associated with soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor complex formation. J Biol Chem 272:28036–28041
17. Rickman C, Hu K, Carroll J, Davletov B (2005) Self-assembly of SNARE fusion proteins into star-shaped oligomers. Biochem J 388:75–79
18. Rizo J, Rosenmund C (2008) Synaptic vesicle fusion. Nat Struct Mol Biol 15:665–674
19. Sollner T, Bennett MK, Whiteheart SW, Scheller RH, Rothman JE (1993) A protein assembly-disassembly pathway *in vitro* that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. Cell 75:409–418

20. Mohler WA, Shemer G, del Campo JJ, Valansi C, Opoku-Serebuoh E, Scranton V, Assaf N, White JG, Podbilewicz B (2002) The type I membrane protein EFF-1 is essential for developmental cell fusion. Dev Cell 2:355–362
21. Wilson IA, Skehel JJ, Wiley DC (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature 289:366–373
22. Yin HS, Wen X, Paterson RG, Lamb RA, Jardetzky TS (2006) Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. Nature 439:38–44
23. Yin HS, Paterson RG, Wen X, Lamb RA, Jardetzky TS (2005) Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein. Proc Natl Acad Sci USA 102:9288–9293
24. Russell CJ, Jardetzky TS, Lamb RA (2001) Membrane fusion machines of paramyxoviruses: capture of intermediates of fusion. EMBO J 20:4024–4034
25. Connolly SA, Leser GP, Jardetzky TS, Lamb RA (2009) Bimolecular complementation of paramyxovirus fusion and hemagglutinin-neuraminidase proteins enhances fusion: implications for the mechanism of fusion triggering. J Virol 83:10857–10868
26. Gonzalez-Reyes L, Ruiz-Arguello MB, Garcia-Barreno B, Calder L, Lopez JA, Albar JP, Skehel JJ, Wiley DC, Melero JA (2001) Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion. Proc Natl Acad Sci U S A 98:9859–9864
27. Biacchesi S, Skiadopoulos MH, Yang L, Lamirande EW, Tran KC, Murphy BR, Collins PL, Buchholz UJ (2004) Recombinant human Metapneumovirus lacking the small hydrophobic SH and/or attachment G glycoprotein: deletion of G yields a promising vaccine candidate. J Virol 78:12877–12887
28. Rawling J, Cano O, Garcin D, Kolakofsky D, Melero JA (2011) Recombinant sendai viruses expressing fusion proteins with two furin cleavage sites mimic the syncytial and receptor-independent infection properties of respiratory syncytial virus. J Virol 85:2771–2780
29. Li L, Lok SM, Yu IM, Zhang Y, Kuhn RJ, Chen J, Rossmann MG (2008) The flavivirus precursor membrane-envelope protein complex: structure and maturation. Science 319:1830–1834
30. Rey FA, Heinz FX, Mandl C, Kunz C, Harrison SC (1995) The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. Nature 375:291–298
31. Bressanelli S, Stiasny K, Allison SL, Stura EA, Duquerroy S, Lescar J, Heinz FX, Rey FA (2004) Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. EMBO J 23:728–738
32. Lescar J, Roussel A, Wien MW, Navaza J, Fuller SD, Wengler G, Wengler G, Rey FA (2001) The fusion glycoprotein shell of Semliki forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. Cell 105:137–148
33. Gibbons DL, Vaney MC, Roussel A, Vigouroux A, Reilly B, Lepault J, Kielian M, Rey FA (2004) Conformational change and protein-protein interactions of the fusion protein of Semliki forest virus. Nature 427:320–325
34. Roche S, Rey FA, Gaudin Y, Bressanelli S (2007) Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. Science 315:843–848
35. Roche S, Bressanelli S, Rey FA, Gaudin Y (2006) Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. Science 313:187–191
36. Backovic M, Longnecker R, Jardetzky TS (2009) Structure of a trimeric variant of the Epstein-Barr virus glycoprotein B. Proc Natl Acad Sci U S A 106:2880–2885
37. Kadlec J, Loureiro S, Abrescia NG, Stuart DI, Jones IM (2008) The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines. Nat Struct Mol Biol 15:1024–1030
38. Roche S, Gaudin Y (2002) Characterization of the equilibrium between the native and fusion-inactive conformation of rabies virus glycoprotein indicates that the fusion complex is made of several trimers. Virology 297:128–135
39. Albertini AA, Baquero E, Ferlin A, Gaudin Y (2012) Molecular and cellular aspects of rhabdovirus entry. Viruses 4:117–139

40. Albertini AA, Merigoux C, Libersou S, Madiona K, Bressanelli S, Roche S, Lepault J, Melki R, Vachette P, Gaudin Y (2012) Characterization of monomeric intermediates during VSV glycoprotein structural transition. PLoS Pathog 8:e1002556
41. Hefferon KL, Oomens AG, Monsma SA, Finnerty CM, Blissard GW (1999) Host cell receptor binding by baculovirus GP64 and kinetics of virion entry. Virology 258:455–468
42. Connolly SA, Jackson JO, Jardetzky TS, Longnecker R (2011) Fusing structure and function: a structural view of the herpesvirus entry machinery. Nat Rev Microbiol 9:369–381
43. Ho Y, Hsiao JC, Yang MH, Chung CS, Peng YC, Lin TH, Chang W, Tzou DL (2005) The oligomeric structure of vaccinia viral envelope protein A27L is essential for binding to heparin and heparan sulfates on cell surfaces: a structural and functional approach using site-specific mutagenesis. J Mol Biol 349:1060–1071
44. Laliberte JP, Weisberg AS, Moss B (2011) The membrane fusion step of vaccinia virus entry is cooperatively mediated by multiple viral proteins and host cell components. PLoS Pathog 7: e1002446
45. Top D, de Antueno R, Salsman J, Corcoran J, Mader J, Hoskin D, Touhami A, Jericho MH, Duncan R (2005) Liposome reconstitution of a minimal protein-mediated membrane fusion machine. EMBO J 24:2980–2988
46. Top D, Read JA, Dawe SJ, Syvitski RT, Duncan R (2012) Cell-cell membrane fusion induced by p15 fusion-associated small transmembrane (FAST) protein requires a novel fusion peptide motif containing a myristoylated polyproline type II helix. J Biol Chem 287:3403–3414
47. Lindenbach BD, Thiel H-J, Rice CM (2012) Flaviviridae: the viruses and their replication. In: Knipe DM, Howley PM (eds) Fields Virology, 5th edn. Lippincott, Williams and Wilkins, Philadelphia
48. Lamb RA, Parks GD (2012) Paramyxoviridae: the viruses and their replication. In: Knipe DM, Howley PM (eds) Fields virology, 5th edn. Lippincott, Williams and Wilkins, Philadelphia
49. Melen K, Fagerlund R, Franke J, Kohler M, Kinnunen L, Julkunen I (2003) Importin alpha nuclear localization signal binding sites for STAT1, STAT2, and influenza A virus nucleoprotein. J Biol Chem 278:28193–28200
50. Arhel N (2010) Revisiting HIV-1 uncoating. Retrovirology 7:96
51. Roizman B, Knipe DM, Whitley RJ (2012) Herpes simplex viruses. In: Knipe DM, Howley PM (eds) Fields virology, 5th edn. Lippincott, Williams and Wilkins, Philadelphia
52. Broyles SS (2003) Vaccinia virus transcription. J Gen Virol 84:2293–2303
53. McLellan JS, Yang Y, Graham BS, Kwong PD (2011) Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes. J Virol 85:7788–7796
54. Modis Y, Ogata S, Clements D, Harrison SC (2003) A ligand-binding pocket in the dengue virus envelope glycoprotein. Proc Natl Acad Sci U S A 100:6986–6991
55. Modis Y, Ogata S, Clements D, Harrison SC (2004) Structure of the dengue virus envelope protein after membrane fusion. Nature 427:313–319

Further Reading

Palfreyman MT, Jorgensen EM (2009) *In vivo* analysis of membrane fusion. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Chichester. doi:10.1002/9780470015902.a0020891

Grove J, Marsh M (2011) The cell biology of receptor-mediated virus entry. J Cell Biol 7:1071–1082

Helenius A (2007) Virus entry and uncoating. In: Knipe DM, Howley PM (eds) Fields virology, 5th edn. Lippincott, Williams and Wilkins, Philadelphia

Also especially recommended for further reading are references [4, 9, 15, 42, 44] *listed above*

Chapter 17
Bacteriophage Receptor Recognition and Nucleic Acid Transfer

Carmela Garcia-Doval and Mark J. van Raaij

Abstract Correct host cell recognition is important in the replication cycle for any virus, including bacterial viruses. This essential step should occur before the bacteriophage commits to transfer its genomic material into the host. In this chapter we will discuss the proteins and mechanisms bacteriophages use for receptor recognition (just before full commitment to infection) and nucleic acid injection, which occurs just after commitment. Some bacteriophages use proteins of the capsid proper for host cell recognition, others use specialised spikes or fibres. Usually, several identical recognition events take place, and the information that a suitable host cell has been encountered is somehow transferred to the part of the bacteriophage capsid involved in nucleic acid transfer. The main part of the capsids of bacteriophages stay on the cell surface after transferring their genome, although a few specialised proteins move with the DNA, either forming a conduit, protecting the nucleic acids after transfer and/or functioning in the process of transcription and translation.

Keywords Adsorption • Attachment • Bacterium • Bacteriophage • Baseplate • Capsid • Cell wall • Flagellum • Nucleic acid conduit • Contractile tail • Gram-negative • Gram-positive • Lipopolysaccharide • Membrane • Peptidoglycan • Pilus • Receptor • Tail fibre • Tailspike • Teichoic acid

C. Garcia-Doval • M.J. van Raaij (✉)
Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: mjvanraaij@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_17,

Abbreviations

DNA	Deoxyribonucleic acid
EMDB	Electron microscopy data bank
gp	Gene product
Hoc	Highly antigenic outer capsid protein
LPS	Lipo-polysaccharide
nm	Nanometer
Omp	Outer membrane porin
ORF	Open reading frame
PDB	Protein structure data bank
RNA	Ribonucleic acid
Soc	Small outer capsid protein
stf	Side tail fibre
tfa	Tail fibre assembly

17.1 Introduction

Bacteriophages, or phages, are viruses that infect bacteria and were independently discovered by both Félix d'Herelle and Frederick Twort. Bacteriophages are now known to be the most numerous replicating biological entities on earth; for every bacterium there are thought to be on average ten phages [1]. The observation that different bacteriophages specifically recognise and kill different bacterial species and strains has led to multiple applications [2]. Félix d'Herelle himself pioneered phage therapy, although with the advent of modern antibiotics it is not commonly used anymore. This situation may change now as more and more bacterial strains develop resistance against many or all useful antibiotics. Bacteriophages and the lytic enzymes they produce are also investigated for the control of microbes in the food industry, of water-borne pathogens and pathogenic bacteria in hospital settings. A much-used application of bacteriophages is phage typing, in which a bacterial species can often be identified at the strain level based on its susceptibility to a library of different bacteriophages.

Phages may interfere with industrial fermentation processes such as those in dairy plants and have important negative economic consequences. Bacteriophages, especially lysogenic ones, may also transfer DNA encoding virulence factors to bacteria, making them more pathogenic. Therefore, bacteriophages to be used in the food industry and medical applications should be well characterised and proven to be free of sequences encoding virulence factors.

In the laboratory, phages have been used in many fundamental experiments in microbiology, genetics, biochemistry and molecular biology since Max Delbrück

and Salvador Luria started investigating them in the middle of the twentieth century. Nowadays, bacteriophages are used as cloning and phage display vectors, DNA encoding phage T7 RNA polymerase is part of protein expression vectors [3] and phage enzymes such as DNA polymerases and ligases are used in routine DNA manipulations. The ease with which many bacteriophages can be grown, their stability and relatively facile genetic modification has led to the development of advanced applications such as their use as labelled ground tracers to follow the flow of underground water, as innocuous test particles for filters designed to remove dangerous viruses and phage-based diagnostics to detect low amounts of dangerous bacteria [2]. Some phages are being used as nanoparticles for nanotechnological developments (see Chap. 22).

Bacteriophages can be divided into virulent, or obligatory lytic, phages on one hand and temperate, or lysogenic phages on the other. Upon infection of their host cell, temperate phages may take one of two paths, lytic or lysogenic. In the lytic pathway, the phage takes over the host cell and redirects its biosynthetic pathways towards producing many daughter phages and ultimately, lysis of the host cell to release its progeny. Alternatively, a temperate phage may integrate its DNA into the host genome – this phage form is called a prophage. In this way, its genome replicates with that of the bacterium in conditions favourable for bacterial growth (some phages even encode virulence factors that favour bacterial growth inside a eukaryotic host). When conditions cease to be favourable for bacterial growth and the bacterial SOS system is activated, the phage may switch to the lytic pathway, produce daughter phages, lyse the bacterium and "escape". Virulent phages do not have the ability to integrate into the host cell genome and can only follow the lytic pathway. Filamentous phages (see below) are an exception and follow an intermediate pathway. When they infect bacteria, they redirect the bacterial biosynthetic systems to produce phages, but phage egress does not lead to lysis. Instead, the bacterium turns into a phage factory, continuously extruding phages. The bacterium continues to grow and divide, albeit at a slower rate.

Bacteriophages can be divided into several different types based on their size and nature of their genetic material: large DNA phages, small DNA phages and RNA phages. Large DNA phages have a double-stranded DNA genome of 20 kilobases or more and can infect both Gram-negative and Gram-positive bacteria. Probably all large DNA phages belong to the *Caudovirales* order, as no phages with genomes greater than 15 kilobases have been found without a tail. Their characteristic tails function in efficient host cell recognition and infection and can be long, straight and contractile (*Myoviridae* family), long, flexible and non-contractile (*Siphoviridae* family), or short and stubby (*Podoviridae* family). Empty phage particles stay on the cell surface (unlike most other viruses). Well-known examples of myoviruses are the virulent *Escherichia coli* phage T4 (Fig. 17.1a) and the temperate *Escherichia coli* phage μ, while examples of *Escherichia coli* siphoviruses are the virulent phage T5 and the temperate phage λ (Fig. 17.1b). Examples of podoviruses are the virulent *Escherichia coli* phage T7, the temperate *Salmonella* phage P22 (Fig. 17.1c) and the virulent *Bacillus subtilis* phage ϕ29. Phages with long tails invariably contain a tape

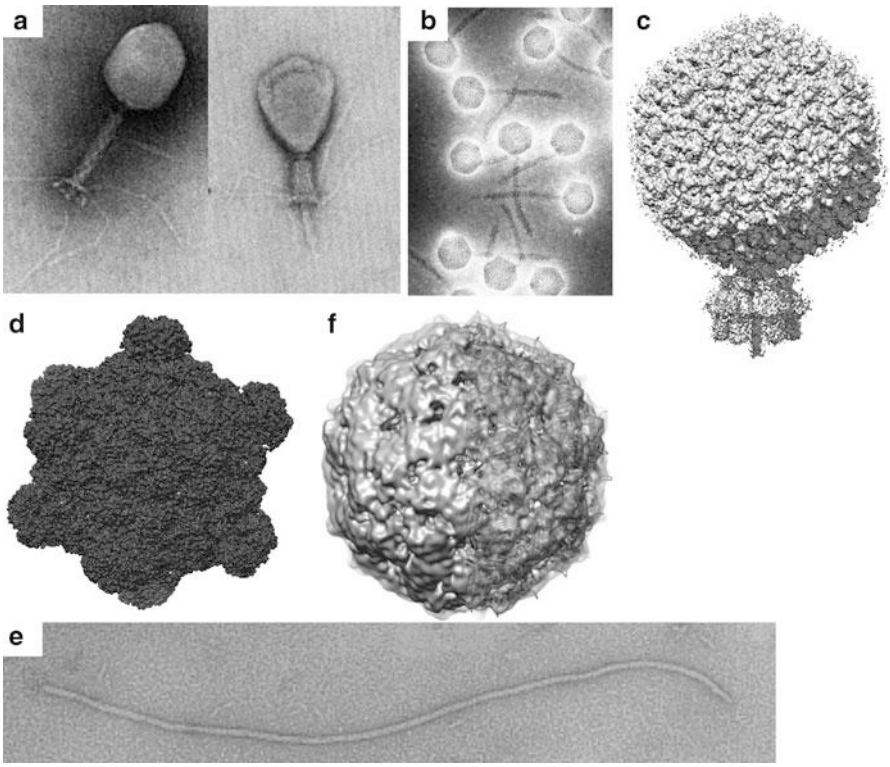

Fig. 17.1 Diverse members of the bacteriophage family. (**a**) Electron micrographs of extended (*left*) and contracted (*right*) T4 phage; the capsid is about 86 nm wide (Reproduced from [4]. With permission). (**b**) Electron micrograph of bacteriophage λ; the capsid width is 550 nm (Reproduced from [5]. With permission). (**c**) Cryo-EM density of phage P22 [6]; here the capsid has a width of 70 nm (figure produced from EMDB entry 5348). (**d**) Crystal structure of phage ϕX174; the capsid diameter is 33 nm (PDB entry 2BPA). (**e**) Electron micrograph of the filamentous phage M13; the filament is 8 nm wide (Reproduced from [7]. With permission). (**f**) Cryo-EM density of bacteriophage MS2 with a capsid of 28 nm in diameter (Reproduced from [8]. With permission)

measure protein, around which the tail is assembled and which thus determines the length of the tail.

A family of intermediate size DNA phages are the *Tectiviridae*. These bacterial viruses have the peculiarity of an internal membrane located below the outer protein shell and can be divided in two subgroups – those infecting Gram-positive and those infecting Gram-negative bacteria. Bacteriophage PRD1 is the most well known example [9]. It is a non-tailed icosahedral virus with a double-stranded DNA genome of 15 kilobases. PRD1 has a broad host range and infects several species of Gram-negative bacteria including *Escherichia coli* and *Salmonella*. Susceptibility of bacteria for PRD1 depends on the presence in the bacteria of a conjugative IncP, IncN or IncW plasmid, which codes for the receptor [10].

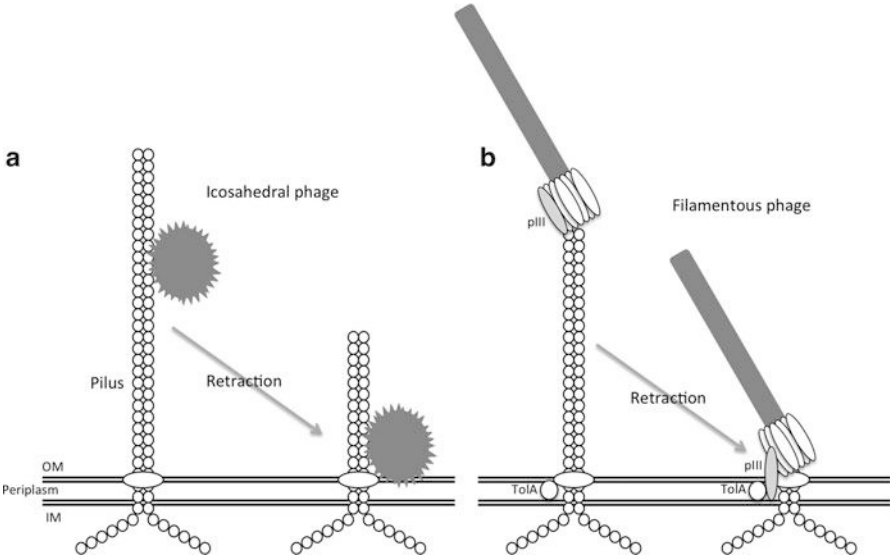

Fig. 17.2 Receptor binding of pilus-binding bacteriophages. (**a**) Icosahedral RNA phage (such as MS2 or Qβ) bound to the side of the pilus. After retraction of the pilus, the phage is close to the membrane and transfer of the RNA into the host takes place. (**b**) Filamentous RNA phage (such as M13 or fd) bound to the end of the pilus. After retraction of the pilus, pIII interacts with the secondary receptor TolA and transfer of the DNA into the host takes place

Small DNA phages or isometric phages have a covalently closed circular plus-strand DNA of up to 10 kilobases. They have an icosahedral capsid of around 30 nm. The prototype is ϕX174 (Fig. 17.1d), for which the different steps of its replication cycle (infection, DNA replication, capsid assembly and host lysis) have been studied [11, 12].

Filamentous phages also contain a circular single-stranded plus-strand DNA molecule that is folded back onto itself in the elongated phage particle. Prototypes of this family of phages are fd and M13 (Fig. 17.1e). They are useful cloning vectors for producing single-stranded DNA due to the lack of a packaging limit – larger circular DNA molecules simply lead to longer phage particles.

RNA phages are small, have icosahedral capsids and contain a single linear single-stranded RNA of about 4 kilobases, coding for only four genes (maturation protein, capsid protein, lysis protein and replicase protein). A well-studied RNA phage is MS2 (Fig. 17.1f) [13]. Like the filamentous phages, they attach to the F-coded pilus (Fig. 17.2) and can thus only infect bacteria containing the F-plasmid. However, they enter bacterial cells by attachment to the side of the F-pilus (Fig. 17.2a), rather than the end as filamentous phages do (Fig. 17.2b).

Phages may change their host range by mutation of their primary receptor-binding proteins or adoption of new receptor-binding proteins *via* horizontal gene transfer, and bacterial mutants that are resistant to infection by a particular bacteriophage have generally lost or modified the primary phage receptor. In this chapter, we will discuss

the specialised proteins used by bacteriophages for the very first step of infection, receptor recognition. We mention what is known about the mechanism different bacteriophages use to decide a correct host cell has been encountered and the phage should commit itself and proceed to the following step, DNA transfer. We will also revise what is known about the DNA transfer step for the different phages.

17.2 Proteins Used in Receptor Recognition and Nucleic Acid Transfer

Phages take advantage of host physiology for their infection process and replication. Like eukaryotic viruses, some bacteriophages recognise their host cell receptors with proteins of the capsid itself. Others have specialised proteins for this function. In this section, we discuss the receptor-binding mechanism of one or several examples of each type of phage. We will also discuss nucleic acid transfer for the different types of phages, although for many phages the mechanism of nucleic acid transfer is still unknown. Naming of bacteriophage proteins may be confusing to readers unfamiliar with phages. In many cases gp (gene product) followed by letters (capital or not) or numbers (arabic or roman) is used, *i.e.* gp37 for phage T4 or gpF in case of ϕX174. The single letter p or P (protein) may also be used, *i.e.* pIII of bacteriophage M13. In the case of bacteriophage T5, pb (protein band) is used, *i.e.* pb1. Proteins may be numbered or lettered according to the order of discovery, size, location on the genome or other scheme the discoverers fancied. Other names or the designation ORF (open reading frame) followed by a number may also be used.

17.2.1 RNA Phages

RNA phages contain 90 dimers of capsid protein, together forming a T = 3 icosahedral capsid [13]. They bind to the bacterial pilus, which is made up of multiple copies of the pilin protein, in the first step of infection. For MS2 (Fig. 17.1f), it is known that one copy of the pilin-binding maturation protein is located at one of the fivefold vertices. Cryo-EM reconstruction at 2 nm resolution on pilus-bound MS2 phage showed bound phages are attached at a slight angle, consistent with the symmetry-mismatch of a single copy of the maturation protein bound to one of the fivefold vertices [14]. Information at the atomic level on how the maturation protein interacts with pilin is not available. Retraction of the pilus presumably brings the phage to the bacterial membrane (Fig. 17.2). It is not known if specific secondary receptor interaction of the phage with membrane components takes place.

Apart from binding pilin on the outside of the capsid, the maturation protein is thought to bind the 5′ and 3′ ends of the viral RNA on the inside. Arrival of the phage at the membrane may be a signal for unpacking of the RNA and transferring it into the cell, but details of this process are unknown. At the vertex contacting the pilus, density that presumably includes contributions from both RNA and maturation protein is poised near the channel that goes through the fivefold vertex to the outside. This suggests that the RNA-maturation protein complex leaves the capsid as the first step of the nucleic acid transfer process [14], presumably through the hole observed at the fivefold vertex.

17.2.2 Filamentous Phages

Filamentous bacteriophages (Fig. 17.1e) are thin and long particles (about 8 by 900 nm) that are assembled around a circular single stranded DNA, they can be significantly longer if larger DNA molecules are encapsulated. The end of the virus that assembles first is formed by proteins pVII and pIX, this end is also called the distal end. The major capsid protein is protein pVIII, a small α-helical protein that assembles in helical fashion around the encapsulated DNA molecule. The proximal end is formed by proteins pVI and pIII. The receptor-binding protein pIII has three domains, N1, N2 and a C-terminal domain that anchors the protein to the phage. Infection initiates by absorption of protein pIII to the distal terminus of F-pilus (Fig. 17.2b), *via* the N2 domain of pIII. As both the tip and sides of the conjugative pilus are formed by the protein pilin, filamentous phages must recognise a different epitope than the small RNA phages – an epitope that is not available on the sides of the pilus.

The C-terminal domain of the *Escherichia coli* periplasmic protein TolA acts as co-receptor for filamentous phages [15]. Upon retraction of the pilus by the bacterium, the phage protein pIII approaches the membrane and a normally shielded epitope of domain N1 of pIII binds to TolA (Fig. 17.2b). Details of subsequent steps are obscure, although it has been postulated that pIII may participate in forming a channel through the bacterial membrane [15]. It is not known whether only naked phage DNA enters the bacterium or whether phage capsid proteins also enter.

17.2.3 Small DNA Phages

Small DNA phages, also called isometric phages, recognise their receptor by the capsid proper. In the case of ϕX174 (Figs. 17.1d and 17.3), primary, reversible receptor interaction with the terminal N-acetyl-glucosamine at the non-reducing end of the core polysaccharide of bacterial lipo-polysaccharide (LPS; [11]) is mediated by the capsid protein gpF. The binding site has been localised to the

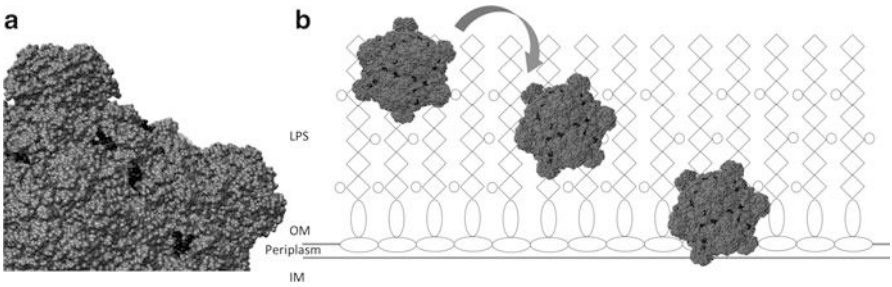

Fig. 17.3 Receptor binding of bacteriophage ϕX174. (**a**) Location of the primary, reversible, receptor binding site (in *black*) on the phage capsid (*grey*) formed by protein gpF. (**b**) Steps in attachment and DNA transfer. Phage ϕX174 binds reversibly to the non-reducing ends (*grey circles*) of the core polysaccharide of bacterial lipo-polysaccharide. ϕX174 may roll over the cell surface until a suitable, unknown, secondary receptor is encountered. After irreversible binding to this receptor, transfer of the pilot protein gpH bound to phage DNA into the host takes place

side of a depression on gpF [16]. Spike proteins gpG and gpH are necessary for irreversible cell attachment to a second factor [12], which may be encoded by the *phxB* gene of *Escherichia coli*. Usually, two to three spikes become embedded in the cell wall and the phage submerges to about one-half of its diameter (Fig. 17.3). Tailed bacteriophages are thought to diffuse laterally along the cell surface after reversible attachment until they encounter their secondary receptor and bind irreversibly. Bacteriophage ϕX174 cell attachment may proceed similarly. However, instead of walking along the surface, ϕX174 may roll (Fig. 17.3b).

Adsorbed ϕX174 particles are often located at regions of adhesion between the cell wall and the inner membrane. This suggests that the secondary receptor is also located here and indicates that the DNA genome may be transferred directly into the cytoplasm. DNA eclipse from the phage has a high activation energy barrier and requires the LPS lipid A moiety of the host. DNA transfer into the cell requires protein gpH of the phage (also called the DNA pilot protein) and an active host cell metabolism [11]. This protein also enters into the cell and directs stage I DNA synthesis. It has been proposed that DNA synthesis is necessary for complete DNA transfer, in the same way that active transcription is necessary for transfer of phage T7 DNA (see below). In the atomic structure of ϕX174, diffuse density is located at each of the fivefold vertices of the virion. This density has been attributed to the DNA pilot protein [16], suggesting that DNA exits through one of the fivefold vertices. After entering into the bacterium with the DNA, GpH promotes phage replication, perhaps by stimulating transcription, being involved in DNA replication or just stabilising the DNA and protecting it against degradation by host cell enzymes.

17.2.4 Lipid-Containing Icosahedral Phages

The best studied and prototype of the *Tectiviridae*, the virulent bacteriophage PRD1 (Fig. 17.4f), first reversibly binds to its receptor, followed by an irreversible step

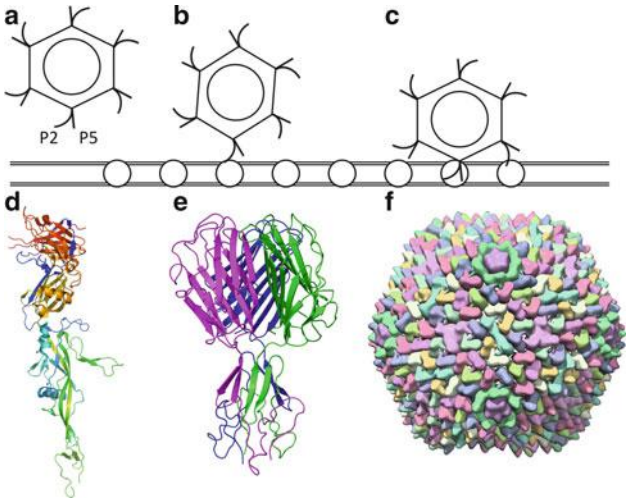

Fig. 17.4 Receptor binding of bacteriophage PRD1. Schematic drawings of (**a**) phage PRD1 and its spike proteins P2 and P5, (**b**) phage PRD1 binding to its receptor using the P2 protein and (**c**) opening of the vertex for membrane penetration and DNA ejection. (**d**) Crystal structure of the monomeric receptor-binding protein P2 (PDB entry 1n7v). (**e**) Crystal structure of the carboxy-terminal distal part of the trimeric vertex protein P5 (PDB entry 1yq8). (**f**) Structure of bacteriophage PRD1 (based on PDB entry 1W8X)

(Fig. 17.4a–c). The vertex of bacteriophage PRD1 shows two different types of spikes associated with the pentameric P31 capsid protein, composed of the proteins P2 (Fig. 17.4d) and P5 (Fig. 17.4e) [10]. Specific binding to the cell surface receptor depends on the protein P2. The cell surface receptor is a component of the conjugative system coded for by a plasmid, as bacteria without these plasmids are resistant to PRD1 infection. The entire conjugative apparatus coded by the plasmid is necessary for infection, and the TraF protein is considered to be the phage receptor protein. Receptor binding of the P5 protein has not been demonstrated, but it might give the virus a selective advantage in the wild by binding to an unknown, alternative receptor.

After binding to the receptor, P2 is thought to trigger conformational changes within the spike complex that are transmitted to the internal membrane. Information about receptor binding is somehow transferred to the DNA delivery apparatus, which contains at least the proteins P7, P11, P14, P18 and P32. This leads to considerable conformational changes in the vertex structure. The removal of the spike complex creates an opening in the vertex, which enables an appendage to protrude that penetrates the outer membrane of the host. The protruding tube probably contains protein P11, but may also contain an extension of the phage lipid membrane, as isolated phage membrane forms small vesicles with tubular extensions [9]. The lytic transglycosylase (P7) is thought to assist in genome entry by locally degrading the peptidoglycan layer. The appendage formed extends some 35 nm (the thickness of the cell envelope) penetrating the cytoplasmic membrane.

DNA transfer is dependent on active membrane tube formation and reduction of the membrane vesicle volume assisted by at least the proteins P14, P18 and P32. The force required for DNA transfer may result in part from the pressure release of the tightly packed DNA in the capsid, but it is tempting to speculate that contraction of the lipid membrane of the phage also plays a role. After all, in the phage the membrane is forced into a spherical form, while in isolation it prefers to assume a tubular conformation. After DNA injection, transcription and protein-primed genome replication take place. It is not known whether, like in the case of bacteriophage T7 (see below), transcription may also play a role in DNA transfer. The phage protein shell stays associated with the host cell surface.

17.2.5 Large DNA Phages

Large double-stranded DNA phages of the *Caudovirales* order possess a tail for efficient host cell recognition and DNA transfer (Fig. 17.1a, b, and c). In general, in a first step reversible attachment is achieved, with a subsequent irreversible attachment, after which the phage is fully committed to DNA transfer into the selected host cell. Proteins on the outside of the DNA-containing capsid may also be involved in highly reversible attachment to the host, augmenting the local phage concentration near the bacterium and favouring efficient infection. In the following three sections, we will discuss what is known about receptor binding and DNA transfer for the most known examples of the *Podoviridae* (P22 (Fig. 17.1c), T7 and ϕ29), *Siphoviridae* (T5, λ (Fig. 17.1b), SPP1 and lactobacillus phages) and *Myoviridae* (T4 (Fig. 17.1a), P2, SPO1 and spore-binding phage 8a).

Podoviridae

Podoviruses have a tail that is not long enough to span the periplasmic space of Gram-negative bacteria or the cell wall of Gram-positive bacteria. Here we will discuss the host recognition and DNA transfer steps of three podoviruses as examples, two that infect the Gram-negative bacteria *Salmonella enterica* (phage P22) and *Escherichia coli* (phage T7), and one that infects the Gram-positive bacterium *Bacillus subtilis* (phage ϕ29).

Bacteriophage P22. P22 (Fig. 17.1c) infects the host cell using a three-step mechanism (Fig. 17.5a–c): (i) the virus binds through its tailspikes to LPS; (ii) the tail needle contacts with the secondary receptor and/or pierces the membrane; (ii) the tail proteins change their conformation and the DNA is ejected. As in many other podoviruses, the P22 tail (Fig. 17.5d) contains six tailspikes (Fig. 17.5e). Each one is a stable homotrimer of gp9 [17], and can be divided into three parts: an amino-terminal phage-binding domain, a central parallel β-helix domain and a carboxy-terminal, highly interdigitated part important for trimerisation and thermostability. The amino-terminal phage-binding domain is more evolutionary conserved between related phages, while the sequence of the carboxy-terminal domains has evolved more, apparently depending on the bacteria to infect. In phage P22 a hinge is present between the amino-terminal domain and central

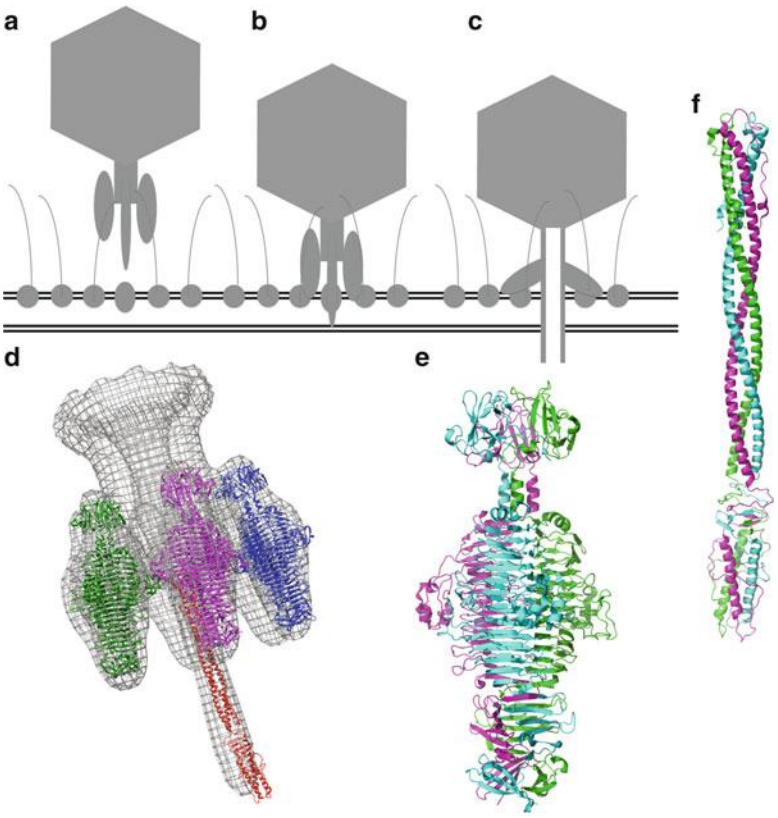

Fig. 17.5 Bacteriophage P22. (**a**) The bacteriophage binds through its tailspikes to LPS. (**b**) The tail needle contacts with the secondary receptor and/or pierces the membrane. (**c**) The tail proteins change their conformation and the DNA is ejected. (**d**) Cryo-EM reconstruction (EMDB: 1119) of the tail with crystal structures of tailspikes and tail needle fitted. (**e**) Crystal structure of the tailspike (gp9, PDB: 2XC1). (**f**) Crystal structure of the tail needle (gp26, PDB: 2POH)

domain, which allows flexibility between the two. Folding and trimeric assembly of the tailspikes are dependent on their interdigitated carboxy-terminal domains. The tailspike possesses receptor-destroying endorhamnosidase enzymatic activity. It cleaves the α(1,3)-O-glycosidic bond between rhamnose and galactose of the bacterial LPS O-antigen repeats. This results in cleavage of the LPS O-antigen receptor during the adsorption of the phage to the cell surface, rendering the binding reversible. More importantly, this activity may be used by the phage to clear a path through the bacterial outer LPS region (the capsule) to gain access to the secondary receptor on the cell surface. The enzymatic activity may also serve to release progeny phage particles from the cellular debris at the end of the lytic cycle.

From the centre of the tail, the needle protein projects, which is a thin trimer of gp26 [18] (Fig. 17.5f). The tip of the needle projects beyond the tailspikes (Fig. 17.5d) and is thus likely the first protein to contact the cell surface. Whether

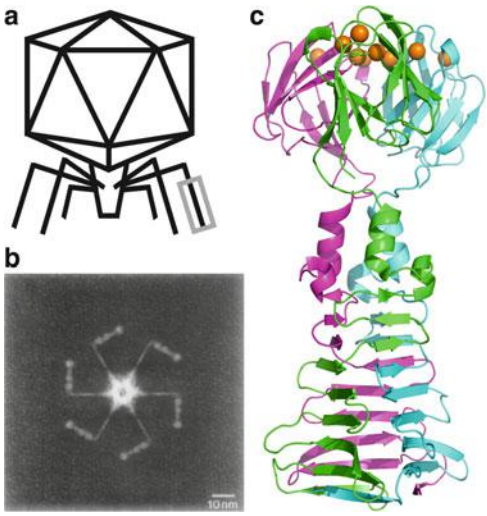

Fig. 17.6 Bacteriophage T7. (**a**) Schematic drawing of phage T7 (Reproduced from [21]. With permission). (**b**) Bottom view of the tail, which clearly shows the six fibres (Reproduced from [22]. With permission). (**c**) Crystal structure of the carboxy-terminal domain of the fibre (PDB entry 4A0T) with residues that affect host range marked with orange balls

this interaction is specific, *i.e.* with a putative secondary receptor, or unspecific, for example with the membrane, is currently unknown. At the same time, gp26 is the plug that keeps the DNA in the P22 particle; which suggests that this interaction is somehow connected to DNA release. Bacteriophage P22 contains three so-called ejection proteins: gp7, gp16 and gp20 [19]. These proteins are likely involved in DNA transfer from the phage into the host bacterium. However, the mechanism is still unknown.

Another *Salmonella* phage that has been studied quite extensively is ε15. Electron cryotomographic analysis of the infection process of *Salmonella anatum* by ε15 has shown that in a first stage, the tailspikes of ε15 attach to LPS on the surface of the bacterium. The tail hub of ε15 then attaches to an unknown cell receptor and forms a tube. Transferred core proteins and cellular proteins may contribute to the tube formation, which spans the periplasmic space. DNA transfer then takes place through the tube. The tube allows the DNA to directly enter the cytoplasm and protect it from periplasmic nucleases. Once the DNA has been injected into the cell, the tube and portal seals, and the empty bacteriophage remains at the cell surface [20].

Bacteriophage T7. Not all podoviruses contain enzymatically active receptor binding tailspikes. Instead, T7 contains six thin, kinked, tail fibres (Fig. 17.6). The fibres, homotrimers of the protein gp17, are responsible for the first specific, but still reversible, attachment to *Escherichia coli* LPS. The kinked fibres are comprised of an amino-terminal tail attachment domain, a slender shaft, a flexible kink, and a carboxy-terminal domain composed of several nodules [22]. The structure of the carboxy-terminal region comprising residues 371–554 is known [23]. The structure revealed a β-stranded pyramid domain and a globular carboxy-terminal receptor

binding domain, in which residues known from mutational studies to be involved in receptor binding are located in loops at the end of the trimer (Fig. 17.6c).

Binding of the T7 fibre to the inner core of the LPS mediates the initial contact with the target cell, but does not yet trigger the DNA release. This initial and reversible binding may position the phage for irreversible binding and DNA ejection. Whether a specific receptor is involved in the secondary, irreversible, binding, is currently unknown. If so, the first, reversible binding may allow the phage to stay associated with the bacterial surface while it diffuses two-dimensionally until the putative secondary receptor is encountered. For T7, these secondary interactions may be mediated by its outer tail proteins gp11 and/or gp12. Penetration of the outer membrane of a Gram-negative bacterium by a podovirus must involve forming a channel for DNA transport across the cytoplasmatic membrane [19]. The T7 tail is too short to span the periplasmic space of *Escherichia coli* and a channel needs to be made to allow the phage genome to travel from the virion to the cytoplasm. It is thought T7 extends its tail by ejecting the internal head proteins (gp14, gp15, gp16) into the cell prior to DNA transfer into cytoplasm. These three proteins are known to form a cylindrical structure inside the phage head, and they are ejected from the infected virion into the cell envelope before the phage genome. They should disaggregate from their structure in the mature virion and also almost completely unfold in order to leave the head and pass through the head-tail connector and form a trans-envelope channel that connects the virion tail tip to the cell cytoplasm. Bacteriophage T7 has therefore been described as having an extensible tail.

After phage adsorption and protein ejection, approximately one kilobase of the 40-kilobase genome is normally internalised by the cell [24], perhaps as a result of the release of the pressure of the DNA encapsulated in the capsid (see Chaps. 12, 18 and 19). Subsequent to internalisation of the leading first kilobase, translocation is coupled to transcription. The leading three kilobases include mainly promoters for *Escherichia coli* RNA polymerase. The RNA polymerase moves along T7 DNA until it encounters the cytoplasmatic membrane but continues to transcribe, thereby pulling DNA from the capsid into the cell. Transcription by *Escherichia coli* RNA polymerase produces messenger RNA for the T7 proteins located at the leading end of the genome, including the single-subunit T7 RNA polymerase and an inhibitor of bacterial restriction enzymes that might otherwise degrade the T7 DNA. The leading end of the genome does not contain restriction sites, while the rest of the genome does. At 30 °C, the T7 genome is internalised at a constant forty base pairs per second across the genome, the same rate as transcription of ribosomal RNA operons. *Escherichia coli* RNA polymerase is one of the strongest motor proteins known and it is thus not surprising that it can pull a DNA molecule into the cell cytoplasm. Of course, once T7 RNA polymerase is produced, it may well contribute to DNA internalisation, although it is not absolutely necessary for this function.

Many encapsulated bacteria are infected by specialised bacteriophages that carry host-capsule degrading hydrolases. These usually tailspike-associated enzymatic activities enable the phages to penetrate the capsular layer and are important determinants of the bacteriophage host range. Podovirus K1F infects *Escherichia coli* K1, a polysialic acid encapsulated pathogen. K1F has degrading tailspikes known as endosialidases that break the α2,8-glycosidically linked sialic acid

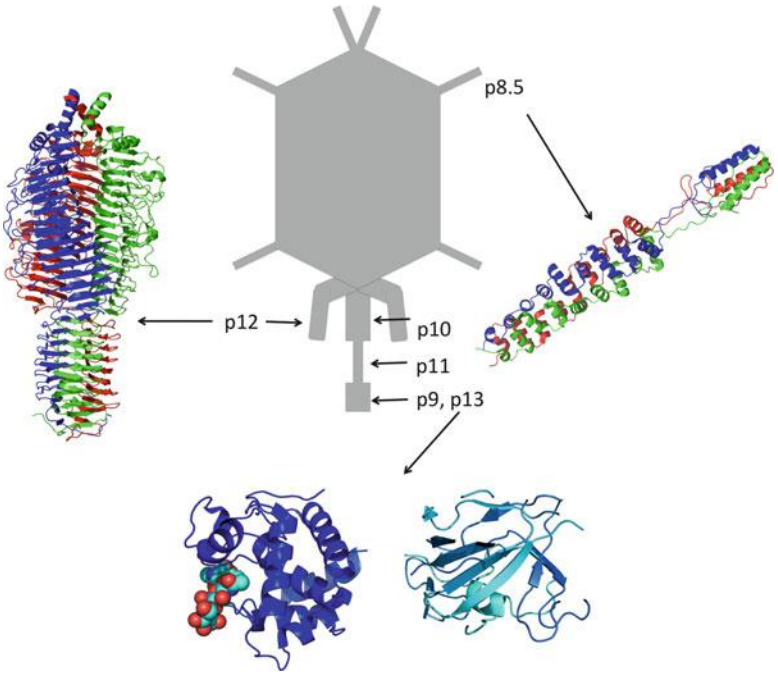

Fig. 17.7 Bacteriophage ϕ29 and its constituent proteins important in the first steps of infection. A schematic drawing of phage ϕ29 is presented with p8.5, p10, p12, p11, p9 and p13 indicated. The phage is surrounded with crystal structures of the relevant proteins. Crystal structures are shown of the head fibre p8.5 (PDB entry 3QC7), the carboxy-terminal part of the appendage p12 (PDB entry 3GQ7) and p13. The amino-terminal lysozyme-like domain and the carboxy-terminal endopeptidase-like domain of p13 are shown (*left* and *right*, respectively), the amino-terminal domain with the co-crystallised substrate tri-N-acetyl-D-glucosamine (PDB entries 3CSZ and 3CSQ, respectively). P13 is located at the end of the phage tail

oligomers or polymers. The tailspike endoNF permits the phage to adsorb to the polysialic capsule, depolymerise the capsule to tunnel towards the outer bacterial membrane and to mediate the adhesion to the membrane. Infection may then proceed as for other podoviruses. There are only some isolated phages capable of degrading biofilms. The T7-related bacteriophage ϕ15 is able to degrade biofilms of two *Pseudomonas putida* strains indicating exopolysaccharide depolymerase activity. The ϕ15 tailspike protein is responsible for exopolysaccharide degrading activity and this exopolysaccharide is probably a primary and essential receptor for ϕ15.

Bacteriophage ϕ29. As bacteriophage ϕ29 infects a Gram-positive host, it is perhaps not surprising it has some specific characteristics compared to phages infecting Gram-negative hosts. The phage capsid is decorated with 55 head fibres attached to the quasi-threefold symmetry positions (Fig. 17.7). The head fibres are trimers of the protein p8.5. Although phage particles without head fibres are infective, the head fibres might enhance attachment of the virions onto the host cell wall, interacting with the cell teichoic acids [25].

The tail of bacteriophage ϕ29 consists of the dodecameric portal protein p10, 12 radially arranged trimeric appendages formed by p12 and an extension formed by proteins p11, p9 and p13 (Fig. 17.7). Protein p12 can cleave the teichoic acid of the bacterial cell wall and is thus the functional equivalent of the P22 tailspike in ϕ29 in mediating the first specific, but reversible binding of the phage to the bacterial host. The p12 appendages are kinked; an amino-terminal cylindrical outward-pointing arm links them to the lower collar region of the ϕ29 tail and a carboxy-terminal spindle-shaped domain contains the polyglycerol phosphate degrading active site [26]. Most of the carboxy-terminal enzymatic domains point downwards in a cryo-EM reconstructed structure of the whole phage, but two point horizontally outwards. The movement of the appendages may allow the phage to move downwards through the teichoic acid layer to the bacterial membrane [27].

Protein p12 monomers are β-helical and assemble into a homotrimer in a sideways manner (Fig. 17.7). An intramolecular chaperone at the very carboxy-terminus helps the p12 monomers to trimerise and is then cleaved off by an autoproteolytic reaction to yield the mature protein. Proteins p11, p9 and p13 make up the central rod of the tail, with p13 being the most distal protein and thus the one likely to make the first irreversible contact with the bacterial cell surface. Its crystal structure revealed both a transglycolase domain and a metalloendoprotease domain. Both likely function in degrading the peptidoglycan layer before DNA transfer can occur [28].

Bacteriophage ϕ29 DNA transfer proceeds *via* a push-pull mechanism ([29] and references therein). The right end of the genome enters the bacterium first and the leading two-thirds enter the cell in the absence of phage protein synthesis, presumably by release of the pressure of the packaged DNA in the capsid (the push phase). Transfer of the remaining third of the genome depends on synthesis of the specific phage proteins p16.7 and p17 encoded by the part of the genome transferred in the push phase. Energy is also required for this pull phase, but the exact mechanism is unknown.

Siphoviridae

Siphoviruses have an apparent communication problem between the two ends of their long, flexible tail tube. Somehow, the fact that a suitable bacterium to infect has been encountered with the distal end of the tail has to be transmitted to the end most proximal to the head domain, where the packaged DNA must be released. A conformational change taking place in a domino effect, opening the inner diameter of the tail, has been proposed for this purpose [30], although a simpler mechanism may be that opening the bottom of the tail allows the tape measure protein to leave and DNA simply follows [31].

Bacteriophage T5. Like many phages that infect Gram-negative bacteria, bacteriophage T5 (Fig. 17.8a) has two different types of fibres: the outward-pointing L-shaped tail fibres and the central straight tail fibre [32]. The three L-shaped tail fibres are trimers of the protein pb1. They bind reversibly to the O-antigen of *Escherichia coli* LPS and accelerate the absorption [33]. However, loss of this first receptor does not lead to T5 resistance. The interaction of the L-shaped tail fibres with O-antigen accelerates the adsorption of the phage to the cell and keeps the fibred phage at the surface of the host for a considerably longer time than T5 phage

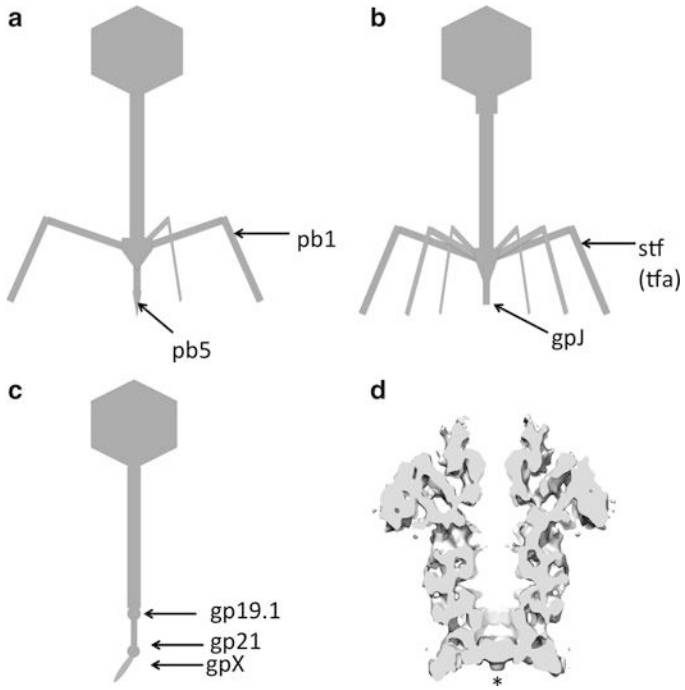

Fig. 17.8 Schematic drawings of some siphoviruses. (**a**) Bacteriophage T5 with the L-shaped tail fibres (pb1) and the FhuA-binding protein pb5 indicated. (**b**) Bacteriophage λ with the side tail fibres (stf, tfa) and the LamB-binding protein gpJ indicated. (**c**) Bacteriophage SPP1. The positions of the hub protein gp19-1, the tail protein gp21, which may have a glycan hydrolysing function, and the unknown receptor-binding tail tip, gpX, are indicated. (**d**) Cut-through of the neck region of phage SPP1 consisting of the portal protein gp6, gp15 and gp16. Electron microscopy density (EMDB 1021) is shown in *grey*. An *asterisk* indicates the stopper of gp16 that prevents DNA from exiting the phage head before infection

lacking fibres, making binding to the secondary receptor more likely the longer the phage can move along the cell surface. Efficient surface attachment occurs when more than one of the three tail fibres binds to an O-antigen at the same time [34]. The L-shaped tail fibres have a thin outward-pointing part followed by a kink and a thicker downward-pointing part consisting of consecutive bead-shaped domains. Pb1 is a large protein of 1,396 amino acids per monomer. The first 200 amino acids are rich in alanines and serines and are predicted to form a coiled-coil, which may correspond to the thin domain of the L-shaped tail fibres. The remaining part is more varied in sequence and contains large hydrophobic and aromatic residues, so likely forms the consecutive globular domains. The very carboxy-terminal part of the protein is an intramolecular chaperone and probably functions in correct trimerisation of the protein as it does for p12 of phage ϕ29, the carboxy-terminal 132 amino acids probably get autoproteolysed after trimerisation. No atomic structure for any of the domains of pb1 is known, but like other phage fibres it is likely to contain β-structure.

In a second adhesion step, pb5 binds irreversibly to the outer membrane ferrichrome transporter FhuA, a monomeric transporter for which the crystal structure is known. The gating loop of FhuA, consisting of residues 316–356, is the largest extracellular loop and is crucial for the binding of the ferrichrome-iron and T5 phages. After FhuA binding, the straight fibre tip of the viral tail traverses the lipid bilayer and undergoes a major conformational change [35]. A pore is formed in the bacterial membrane next to FhuA through which the DNA is transferred. In the mature phage particle the DNA is connected to the proximal end of the tail. Upon binding to the receptor a signal has to be transmitted from the receptor-binding protein along the tail to the head-tail connection to induce DNA penetration through the tail tube. Signal transmission may occur as a progressive conformational change of the major tail protein starting at the distal end of the tail tube, perhaps widening the inner volume to allow for DNA passage. Bacteriophage T5 transfers its double-stranded DNA in two steps. In a first step, 8 % of the total T5 chromosome is transferred, perhaps by simple pressure release of the tightly packed DNA in the capsid. After a pause of a few minutes during which proteins coded by this fragment are synthesised, the remaining part of the DNA is transferred [36]. The section of the T5 genome that is injected in the first step encodes proteins that degrade host cell DNA and shut down the synthesis of host proteins. How the remaining section of DNA is transferred into the bacterium is unknown.

Bacteriophage λ. The lysogenic bacteriophage λ (Figs. 17.1b and 17.8b) has traditionally been an important model system for molecular biology. Laboratory strains of λ appear to have lost their side tail fibres and can still infect their host *Escherichia coli* efficiently in laboratory conditions [37]. However, λ Ur (the original λ) does have side tail fibres (stf; Fig. 17.8b). The side tail fibres need the tail fibre assembly protein tfa for their correct trimerisation and assembly; it is not sure whether tfa remains associated with the side tail fibres in the mature phage particles. The side tail fibres probably recognise OmpC in a primary, reversible binding step [38]. In a second step, the carboxy-terminal end of gpJ at the tip of the tail binds to the *Escherichia coli* lamB gene product [39], a maltoporin. These two proteins form a very stable complex and this interaction probably leads to irreversible binding. Two types of complexes have been identified of the λ tail tip with LamB-containing membranes. In the first, the hollow tail remains at a distance of 17 nm from the membrane. In a subsequent step, the hollow tail attaches to the membrane for DNA transfer.

During DNA transfer, the linear phage genome is transferred past the *Escherichia coli* outer membrane. During this process, transmembrane channels are formed, which permit the entry and escape of small molecules, but not proteins [40]. The DNA then passes through a separate bacterial sugar transport protein, the mannose permease of the bacterial phosphotransferase system, in the cytoplasmic membrane [41]. The two hydrophobic subunits II-PMan and II-MMan alone are sufficient for penetration of λ DNA. It seems thus that bacteriophage λ has subverted the bacterial maltose entry system into a transfer system for its DNA. Once in the cytoplasm, the linear DNA circularises using the *cos* sites, which are the twelve-base G-C rich cohesive sticky ends of the genome.

Bacteriophage SPP1. The Gram-positive *Bacillus subtilis* phage SPP1 (Fig. 17.8c) has a single central fibre for host interaction. The tail tip of SPP1 binds to YueB, a membrane protein whose extracellular region crosses the thick cell wall to expose a receptor region at the bacterial surface. The tail tip also mediates degradation of the bacterial cell wall. The tip is 31 nm long and 10 nm wide at its largest diameter. It can be divided into three domains: a sphere-like region, a broad flattened domain and a terminal rod (Fig. 17.8c). These domains correspond to different proteins: gp19.1 being the sphere-like region nearest the phage, gp21 forming the central part and an as yet unknown distal receptor-binding protein. Gp19.1 is the baseplate hub protein and forms a hexameric ring with a central hole [42]. Gp21 has predicted β-helical structure and may resemble the P22 tailspike in structure and glycan hydrolysing function.

The tip does not have a DNA channel and therefore probably moves out of the way before DNA transfer. The binding of the tail fibre to YueB leads to the irreversible commitment of the virus particle to eject its DNA. In this case, no previous, reversible, interactions have been identified. The structure of the siphovirus SPP1 tail has been determined before and after DNA ejection [30]. The results revealed extensive structural rearrangements in the internal wall of the tail tube. A proposal was made that the binding of the tail tip to YueB triggers a conformational switch that is propagated as a cascade of conformational changes along the 160 nm-long helical tail structure to reach the head-to-tail connector. This would lead to the connector opening and allowing passage of DNA into the host cell through the tail tube. The structure of the connector region between the tail and the head of DNA-containing phage has been determined by electron microscopy and fitting of atomic structures. It consists of three dodecameric proteins: the portal protein gp6, the adaptor protein gp15 and a "stopper" protein, which prevents premature DNA release [43] (Fig. 17.8d).

***Lactococcus lactis* bacteriophages.** Phages of the Gram-positive *Lactococcus lactis* bacterium sport more compact receptor-binding proteins. The crystal and electron microscopy structures of the baseplates of two *Lactococcus lactis* phages have been determined (Fig. 17.9): that of lactococcal phage p2 (not to be confused with the Gram-positive bacteria-infecting myovirus P2) and TP901-1. The baseplate of phage p2 is composed of three protein species [44]. The central part of the baseplate is formed by a circular hexamer of ORF15 with a central hole. A trimer of ORF16 is located at the bottom of the baseplate, forming a closed dome that does not allow DNA passage. Six ORF18 trimers are attached to the central ring, each trimer interacting with a carboxy-terminal extension of an ORF16 monomer. ORF18 is the receptor binding protein and its structure shows an N-terminal β-sandwich shoulder domain, which binds to the baseplate. The remainder of the structure is a short triple β-helical neck and a carboxy-terminal head domain. In the unactivated baseplate, the six trimeric receptor-binding proteins point upwards, *i.e.* away from the bacterial membrane. Activation by calcium ions causes large conformational changes in p2 baseplate, leading to rotation of receptor-binding domains by 200° to point downwards towards the host cell. The head domains of ORF18 may recognise lipoteichoic acids, which are phospho-glycerol polymers. The conformational change of the baseplate also leads the three ORF16 monomers

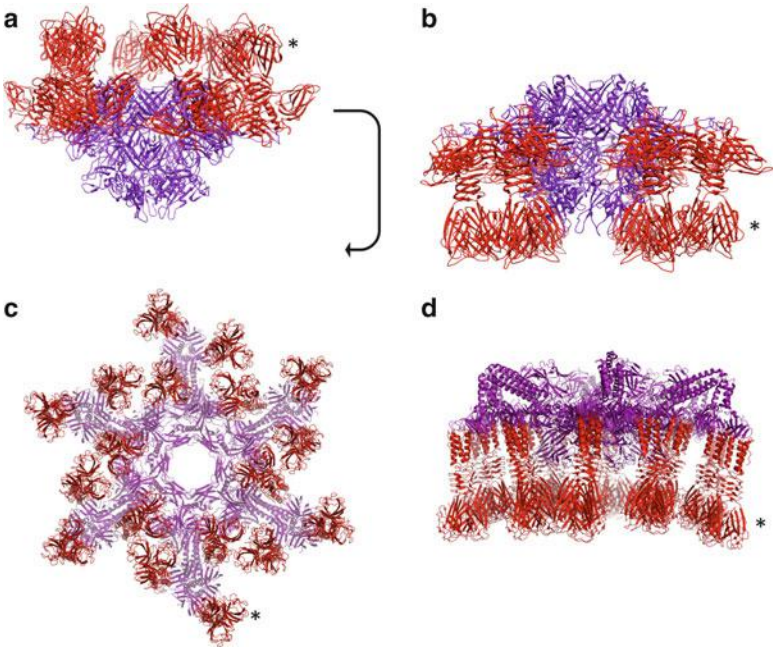

Fig. 17.9 Crystal structures of *Lactococcus lactis* bacteriophage baseplates. (**a**) The baseplate of phage p2 before activation, with the receptor-binding domains of the receptor-binding proteins (in *red*) pointing upwards, away from the bacterium (PDB entry 2WZP). The movement of 200° by the receptor-binding proteins that completely reverses their orientation is indicated by an *arrow*. (**b**) The baseplate of phage p2 after activation by calcium ions, with the receptor-binding domains of the receptor-binding proteins (in *red*) pointing downwards, towards from the bacterium (PDB entry 2X53). (**c**) The baseplate of phage TP901-1 (PDB entries 4DIV and 4DIW) viewed from below (*i.e.*, from the side of the bacterium) and viewed from the side (**d**). The 18 receptor-binding protein trimers are in *red*. One of the 18 receptor-binding domains in parts (**a**) and (**b**), as well as one of the 54 receptor-binding domains in parts (**c**) and (**d**), are indicated by *asterisks*

to separate, opening up a hole in the centre of the baseplate, presumably allowing passage of DNA for transfer into the host.

The baseplate of phage TP901-1 is composed of multiple copies of four different proteins [45]. The centre is a hexameric circular core formed by the Dit protein. The hole in this core is filled by a central tail fibre, Tal. From the core, six arms emanate, each arm being composed of a trimer of the BppU protein. The arms are α-helical up to the elbow. The rest of the arm points downwards and is an adaptor domain for the receptor-binding proteins. To each adaptor, three trimeric receptor binding proteins bind, leading to a total of 54 sites for the host cell envelope saccharides. The receptor-binding proteins point downwards, *i.e.* towards the host bacterium, ready for adhesion. No calcium ions are necessary for activation of TP901-1, so there are likely no conformational changes that have to take place before receptor binding. How receptor binding is related to DNA transfer in these phages is less clear, perhaps the strong binding with up to 54 receptor molecules pushes the

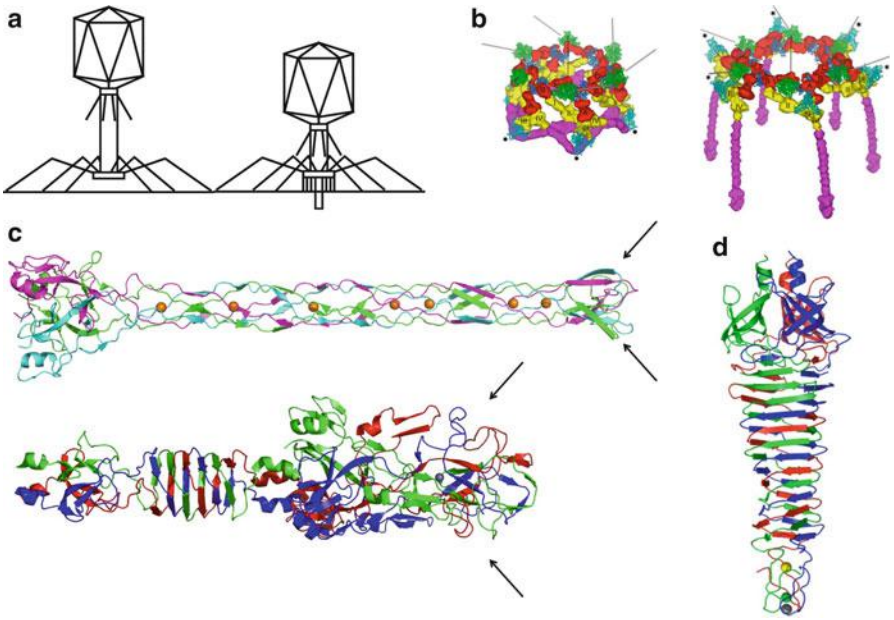

Fig. 17.10 Receptor recognition and DNA transfer by *Escherichia coli* myoviruses. (**a**) Schematic drawing of bacteriophage T4 before and after tail contraction (Adapted from [46], copyright by the National Academy of Sciences). (**b**) Structure of the baseplate in the hexagon and star conformations with the different proteins coloured differently. The short tail fibres are shown in *magenta*; folded away in the hexagon-state (extended tail sheath), and extended in the star-state (contracted tail sheath) (Reproduced from [47]). (**c**) Structures of the long tail fibre protein gp37 tip domain and the carboxy-terminal domain of gp12. *Arrows* indicate the receptor-binding regions. (**d**) The central spike of bacteriophage P2. The structure contains three ions at the tip: a calcium ion (*yellow*), a chloride ion (*green*) and an iron ion (*grey*)

central tail fibre against the cell wall and the force of the cell wall pushing against the tail fibre is sensed by the other end of the tail fibre, which opens a hole in the end of the tail tube.

Myoviridae

With their double-shafted tail, myoviruses (Fig. 17.1a) have resolved the host cell attachment and DNA transfer steps in perhaps the most advanced way, piercing the bacterial outer membrane and traversing the periplasmic space or peptidoglycan layer with the inner tail tube upon contraction of the outer tail tube (Fig. 17.10a). The tips of myovirus inner tail tubes even sport a spike for efficient membrane penetration [48] (Fig. 17.10d). This way, even large genomes of up to several hundred kilobases can be transferred efficiently into the bacterial cell. Here we discuss what is known about receptor attachment and DNA transfer of the most well known myoviruses.

Bacteriophage T4. Like the head fibres of phage ϕ29, immunoglobulin-like domains on the surface of bacteriophage T4 (Fig. 17.1a; see also Chap. 11) could interact with the cell surface and help to maintain the phage in the vicinity of the cell

surface by weak and non-specific interactions until the correct receptor is contacted. The bacteriophage T4 capsid is decorated with two non-essential outer capsid proteins: Hoc (highly antigenic outer capsid protein) and Soc (small outer capsid protein). Soc provides stability to T4 to survive under hostile environments; Hoc has immunoglobulin-like domains and may interact with the bacterial surface providing certain advantages [49].

The archetypal myovirus T4 (and presumably other myoviruses as well) infect their host bacteria efficiently compared with podovirus and siphoviruses, probably due to the sophistication of its contractile tail [50] (Fig. 17.10a). Infection of phage T4 starts with the interaction of the long tail fibres with cell surface receptors (see below). The long tail fibres of bacteriophage T4 are kinked structures of about 145 nm long with a variable width of up to 5 nm [47]. They can be divided into proximal and distal half-fibres, attached to each other at an angle of about 20°. Upon phage assembly in the previous host cell and in adverse conditions for phage multiplication, the long tail fibres are in a retracted conformation, lying against the tail sheath and head of the bacteriophage. Because the neck whisker protein (called gp*wac* or fibritin) is necessary for incorporation of the long tail fibres into the phage [51], we can suppose that this interaction is also responsible for maintaining the long tail fibres in a retracted state. In the extended conformation, only the proximal end of the fibre is attached to the baseplate and the fibre tips point outwards.

The long tail fibres are composed of four different proteins: gp34, gp35, gp36 and gp37 [47]. The proximal half-fibre (the thigh) is formed by a homotrimer of gp34. The amino-terminal part of gp34 is attached to the trimeric protein gp9 of the baseplate while the carboxy-terminal part interacts with the gp35. Gp35 (the knee) is a monomer and may be responsible for creating the angle between the proximal and distal half-fibres. The distal half-fibre is composed of a trimer of gp36 and a trimer of gp37. Gp36 forms the upper part of the shin and gp37 constitutes the rest of the shin, including the very distal receptor recognsing tip (the foot). Only the atomic structure of the receptor recognising tip is known [46]. It forms a globular collar domain, followed by a thin needle region in which seven iron atoms are coordinated and ends in a small intertwined receptor-binding domain (Fig. 17.10). A phage-encoded molecular chaperone, gp57A, is required for the correct trimerisation of short tail fibre protein gp12 and the long tail fibre proteins gp34 and gp37. A specific chaperone, gp38, is required for the correct trimeric assembly of gp37. In the extended conformation, the tips of the long tail fibres can interact reversibly with receptor molecules. The glucosyl-α-1,3-glucose terminus of LPS or outer membrane porin C (OmpC) can function as alternative receptors and both outer membrane protein C and the interaction region of LPS have to be absent or inaccessible for *Escherichia coli* to be resistant to infection by T4.

When at least three long tail fibres have bound a receptor molecule, a signal is transferred to the baseplate of the phage, which then changes conformation. The binding information transfer is likely related to the angle of attachment of the long tail fibre to the baseplate. In the free phage, this angle is variable and the fibres are flexible up to certain limits. When several fibres are attached to their receptor, external forces on the phage may force these angles to less or more than these

allowed angles, pushing proteins near the long fibre attachment site in different conformations and triggering the conformational change of base plate, from the hexagon conformation to the star conformation [47]. This conformational change is very extensive, involving changes of location and interaction partners for several baseplate proteins. Among others, it leads to extension of the short tail fibres, which are normally folded away in the baseplate (Fig. 17.10b). The short tail fibres are trimers of the protein gp12. When the baseplate is in its hexagon conformation, the short tail fibres are incorporated into the baseplate by interactions along their lengths. However, when the baseplate switches to its star conformation, only the proximal (amino-terminal) end of gp12 remains bound to the baseplate from which the rest of the molecule extends. The distal carboxy-terminal part of the short tail fibres comprising residues 317–517 is responsible for a tight, irreversible interaction with the core region of the LPS (Fig. 17.10). The short tail fibres (gp12) have been seen in thin sections of phage-infected bacteria, forming rigid connections of the baseplate to the host cell surface.

In a subsequent step, contraction of the tail sheath (a helical polymer of protein gp18) is initiated [50]. The sheath of T4 is 24 nm wide and 93 nm long and is composed of 138 gp18 subunits, which are arranged in a six-start helix. The pitch and twist of the helix are 4 nm and 17°, respectively. The contracted T4 sheath is 33 nm wide and 42 nm long. The six-start helix changes its pitch and twist to 1.6 nm and 33°, respectively. In the process of sheath contraction, gp18 subunits move as rigid bodies without refolding or significant domain shifts. Upon contraction, the contact area between gp18 molecules increases by about four times.

As a result of tail sheath contraction, the rigid inner tail tube (a helical polymer of gp19) is driven through the outer cell membrane. The trimeric protein gp5 at the tip of the tail tube contains three lysozyme domains that are activated when they enter in contact with the periplasmic peptidoglycan layer. They create a hole, which allows reaching the cytoplasmic membrane. When the tail tube interacts with the cytoplasmic membrane, the phage DNA is released through the tail tube. As for other phages, the pressure of the tightly packed DNA in the head domain probably accounts for the force that initiates DNA transfer, but its completion may need further mechanisms. It appears a proton-motive force over the inner membrane is necessary for successful DNA transfer [52]. Under favourable conditions, the 172 kilobase genome of phage T4 DNA enters the bacterium is less than one minute, making it a very efficient process.

Bacteriophage P2. P2 is a myovirus with a smaller genome than that of T4, namely 33 kilobases [53]. It also has a more simple tail, for which the synthesis of only eleven gene products is necessary. Phage P2 sports a baseplate with a single central spike, which is a trimer of gpV, and six tail fibres, which are formed by gpG. Little is known about the structure of the baseplate, but the structure of the central spike protein has been solved at high resolution [48]. Like the λ side tail fibre and bacteriophage T4 distal tail fibre protein gp37, the fibres have a specialised chaperone to aid them in their folding. In the first step of infection, the fibres attach to the core region of the LPS. Calcium ions have been shown to greatly improve

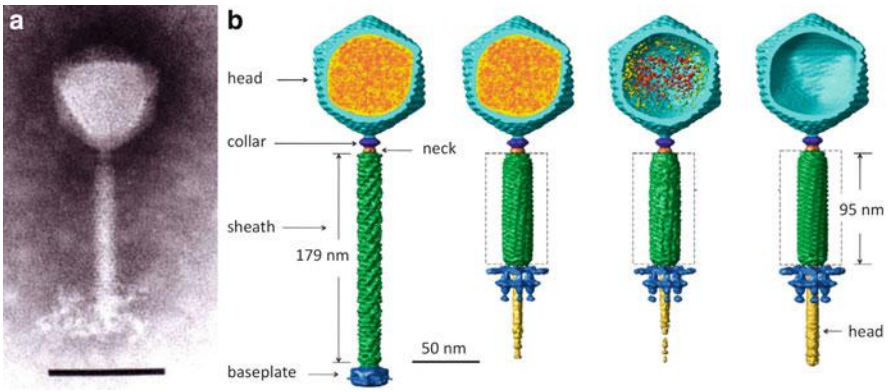

Fig. 17.11 Myoviruses of Gram-positive bacteria. (**a**) Electron micrograph of bacteriophage SPO1 (Reproduced from [54]. With permission). The complex baseplate is clearly visible, with multiple appendages that are presumably receptor-binding proteins. The bar indicates 100 nm. (**b**) *Bacillus anthracis* spore-binding phage 8a. Electron cryotomographic images of four stages on the pathway of DNA ejection are shown: the phage with extended sheath, the phage with contracted sheath but still containing DNA, the phage with contracted sheath in the process of ejecting its DNA and the phage with contracted sheath that has fully ejected all DNA (Reproduced from 55]. With permission). The conformational change of the baseplate from a disk-shaped structure to a structure with sideways- and downward-pointing extensions is very clear

adsorption. The baseplate then presumably changes conformation to allow the central tail tube to pass through it. Contraction of the tail sheath drives the tail tube through the baseplate and the bacterial outer membrane and periplasm, delivering the DNA exit site to the inner membrane. The tip of the tail tube consists of a triple-helical conical spike, of which the very tip is stabilised by three ions: iron, chloride and calcium (Fig. 17.10d).

Bacteriophage SPO1. An example of a myovirus that infects a Gram-positive bacterium is bacteriophage SPO1 [54] (Fig. 17.11a). SPO1 is a large, strictly virulent phage infecting *Bacillus subtilis*. It features a non-permuted double-stranded DNA genome of 146 kb, with redundant, invariable repeats of 13 kb at both ends. The long contractile tail consists of a complex baseplate and a sheath that is 140 nm long and 19 nm wide; the contracted sheath is 64 nm long and 27 nm wide. The baseplate, which functions as a receptor-recognition device and as a trigger for tail sheath contraction, undergoes structural rearrangement upon contraction, rearranging itself into a double-ringed structure. However, high-resolution studies on bacteriophage SPO1, its subcomplexes or proteins by cryo-electron microscopy or crystallography have not been performed.

Bacteriophage 8a. The structure of the *Bacillus anthracis* spore-binding phage 8a was determined by cryo-electron tomography (see Chap. 3) [55] (Fig.17.11b). This phage infects the Gram-positive *Bacillus anthracis* and, like the name suggests, it also binds *Bacillus anthracis* spores. Like other myoviruses, the tail consists of a six-

start helical sheath surrounding a central tail tube, but it has a baseplate different from other phages. The baseplate recognises and attaches to host cells. Cryo-electron tomographic analysis was performed on four different conformations that likely correspond to four states of the infection process: DNA-filled particles with extended tails, DNA-filled particles with contracted tails, partially DNA-filled particles with contracted tails and empty particles with contracted tails.

The baseplate has a claw-like structure, with six extensions in a closed conformation when the tail is extended, and in an open conformation when the tail is contracted. Primary receptor interaction probably takes place by the closed baseplate, although it is unknown which part of the baseplate is responsible, as phage 8a lacks long tail fibres or other obvious primary receptor binding modules. However, densities reminiscent of the bacteriophage T4 short tail fibres [47] are present at the periphery of the baseplate, in a folded-away state when the tail sheath is extended. Perhaps the folded-away short tail fibre-like densities present a sufficient receptor-binding site for interaction. In any case, the baseplate undergoes a conformational change from a closed to an open conformation upon binding to the host cell surface, because in the contracted state, these densities point downwards. The downward extension of the densities may bind host cell receptors more efficiently to irreversibly immobilise the baseplate on the host cell surface and cause its structural rearrangement. As the baseplate and tail sheath are physically connected, the structural rearrangement of the baseplate likely initiates contraction of the tail sheath.

The extended tail sheath is assembled around the 8 nm-diameter tail tube as a right-handed, six-stranded helical structure measuring 20 nm in diameter and 180 nm in length. Each helical strand consists of 45 subunits related to each other by a 4 nm helical pitch and a 22° twist. The 45 subunits in each helical strand of the contracted tail are related by a 2 nm pitch and a 31° twist; the contracted tail sheath is thus about half the length of the extended one. Like for bacteriophage T4, the structure of individual tail sheath subunits is similar in both conformations, indicating that tail contraction results from rigid body rotations and translations of entire subunits rather than by conformational changes within subunits. Contraction of the tail sheath forces the central tail tube downwards into the cell surface. A central spike structure is present at the distal end of the tail tube in extended tails, which may function as a cell-puncturing device as proposed for T4. No equivalent structure is seen in contracted tails, so it is likely the tail spike falls off during or after contraction to allow passage of the DNA.

Comparison of the neck regions in extended and contracted tails indicates that there are two conformations. In the extended state, the neck is assembled into a compact shape plugging the connection between the head and the tail tube. In the contracted state, the neck forms a central pore that is large enough to allow the passage of double stranded DNA. Thus, the neck protein may function as a gate that regulates the passage of DNA from the head into the tail, by undergoing a substantial conformational change upon tail contraction.

17.3 Perspectives and Conclusions

A bacteriophage that ejects its DNA spuriously or into an unsuitable host cell ceases to exist [31]. Evolution has therefore endowed bacteriophages with efficient host cell recognition and DNA transfer mechanisms. These mechanisms differ widely between phages, and it is therefore difficult to propose a general mechanism. Also, due to the research interests of the scientists studying particular phages, they have often concentrated on certain aspects of phage biology. Thus, much is known about the genetic regulation of phage λ, but the adsorption and DNA transfer processes have not been studied profoundly. On the other hand, the tail of bacteriophage T4 for example, has been analysed in great structural detail. However, given these restrictions, here we give a brief overview, attempt to draw some general conclusions and draw attention to some interesting phage facts and potential applications.

In general, bacteriophages attach to host cells using a two-step mechanism. In the first step, attachment is specific but reversible. This allows the phage to diffuse two-dimensionally over the bacterial surface until the second receptor is encountered. This second receptor is then bound irreversibly and at this point, the phage is immobilised and committed to DNA transfer. The sequences of receptor-binding proteins of phages are highly variable; and varying the receptor-binding properties is probably the main mechanism a phage has to adapt to hosts that have varied their cell wall structure or to adapt to a new bacterial host strain or species. Horizontal gene transfer between phages plays an important role in this variation, even between different phage types (horizontal gene transfer is probably most likely when a phage infects a bacterial cell containing a prophage). An example of relatively recent horizontal gene transfer is between the tip of the temperate siphovirus λ side tail fibre and the virulent myovirus T4 long tail fibre, in which residues that are structurally important are conserved, but the distal domain has evolved to bind a different receptor [46]. Another example of evolutionary recent horizontal gene transfer is the sequence and structural homology between the receptor-binding domains of the tailspikes of the podovirus P22 and the myovirus Det7 [56]. There are of course many phages still to be discovered, and their biological characterisation, the sequencing of their genomes and structure determination by electron microscopic (see Chap. 3) and crystallographic techniques (see Chap. 4) will likely reveal new receptor-recognition mechanisms.

Some phages encode several alternate tail fibre genes and may express them either alternately or simultaneously in the same particle. An example is the myovirus ϕ92, which encodes a Swiss army knife-like set of fibres to infect most laboratory strains of *Escherichia coli* (including those with a sialic acid capsule) and many *Salmonella* strains [57]. In the temperate myoviruses μ and P1 the specificity for the host receptor is altered upon inversion of a genome segment. The inversion results in a tail fibre with an alternative carboxy-terminal receptor-binding part. The amino-terminal part, which binds to the phage, is coded outside the inversion region and thus stays constant. In this way, in one form bacteriophage μ carries tail fibres that allow it to infect *Escherichia* and *Salmonella* bacteria and in a second form tail fibres that allow it to infect *Shigella*, *Enterobacter*, *Erwinia*, and *Citrobacter* [58]. *Bordetella* phage

BPP-1 has been shown to use an error-prone reverse transcriptase to generate variation in the distal knob of its tail fibre proteins [59].

As Gram-positive bacteria lack an outer membrane and instead have a thicker layer of peptidoglycan, one would expect their receptor-binding mechanisms to be different. Phages infecting Gram-negative hosts tend to use porins, transport proteins, enzymes or LPS available on the outer membrane as primary or secondary receptors. Phages infecting Gram-negative hosts tend to use peptidoglycan, teichoic acids or proteins that span the cell wall as receptors. One characteristic of tailed phages infecting Gram-positive hosts is that they appear to lack the side tail fibres many tailed phages infecting Gram-negative hosts possess. Phages may also bind to pili, flagella and capsules of both kinds of bacteria.

In most bacteriophages, much less is known about the mechanisms of nucleic acid transfer. Because the nucleic acid in phages is packaged under considerable pressure (see Chaps. 12, 18 and 19), it is likely initial transfer is at least helped by the partial release of this pressure when an appropriate transfer channel has been opened. Transfer of the remaining nucleic acid may then be promoted by phage proteins co-transferred into the cell, produced by transcription from the first region of nucleic acid transferred or by host factors. For a phage that always starts transfer with a specific end of the genome, like T7, proteins coded for at this end may be involved in successful genome transfer. For circularly permuted phages like T4, the entire nucleic acid is likely transferred by pressure and host factors, as the phage can not count on a specific protein being expressed from the part of the genome first transferred. Phages have also developed different mechanisms to protect the nucleic acid that is transferred into the cells. In general, nucleic acids are protected from nucleases in the periplasm by a physical barrier: the extensible tail tube of podoviruses and probably also siphoviruses, or the inner tail tube in myoviruses. Once in the cytoplasm, restriction enzymes are a major danger. Some phages modify the nucleotides of their DNA so they cannot be recognised by restriction enzymes. Other phages (T7, T5) only transfer a small part of their genome in a first step, and this part encodes proteins that inactivate restriction enzymes or otherwise protect DNA against restriction.

For nucleic acid transfer, phages infecting Gram-positive bacteria only have to cross the inner membrane. Therefore, if they can somehow get close to it by digesting the peptidoglycan cell wall, they may not need a to extend a tube like tailed phages infecting Gram-negative bacteria need to do to allow their genomes to safely traverse the nuclease-ridden periplasmic space. This may be an explanation for the fact that there appear to be relatively less myoviruses infecting Gram-positive bacteria [31]. Instead, podoviruses and siphoviruses can, at least in theory, digest their way straight to the cytoplasmic membrane of Gram-positive bacteria if their baseplates contain the appropriate enzymes. Alternatively, they can attach to the outer layer of the cell wall and extend a tube like phages that infect Gram-negative bacteria do.

Future structural studies of bacteriophage proteins, phage protein assemblies and whole bacteriophages by crystallography (Chap. 4) and electron microscopy (Chap. 3) will certainly resolve more details of phage receptor-binding and nucleic

acid transfer. Cryo-tomography (Chap. 3) using mini-cells of bacteria will play an important role in this, like performed for the *Escherichia coli* myovirus P1 [60]. Apart from structural studies, future studies will aim to measure individual interactions of phages and their proteins with their host cells. Single molecule techniques may be used to measure the force of binding of bacteriophage or their receptor-binding proteins to the receptor. In atomic force microscopy (see Chap. 8) a phage fibre may be bound to the tip and used to probe a surface containing its receptor, while optical tweezers (see Chap. 9) could also be used to measure these interactions.

The remarkable properties of bacteriophages have led to many ideas for applications (see Chap. 22). Now that information in atomic detail is becoming available on receptor-binding proteins of different bacteriophages, this information can be exploited to perform directed mutation and modulate their receptor-binding properties. Future knowledge of the nucleic transfer mechanisms may also lead to applications in designing nucleic acid delivery vehicles.

Acknowledgements Work in the authors' laboratory is supported by grant BFU2011-24843 from the Spanish Ministry of Economy and Competitiveness. Structure figures were made with PyMol (The PyMOL Molecular Graphics System, Schrödinger, LLC) or the UCSF Chimera package (developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco and supported by NIGMS 9P41GM103311).

References and Further Reading

1. Wommack KE, Colwell RR (2000) Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev 64:69–114
2. Marks R, Sharp R (2000) Bacteriophages and biotechnology: a review. J Chem Technol Biotechnol 75:6–17
3. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol 185:60–89
4. Mesyanzhinov VV (2004) Bacteriophage T4: structure, assembly, and initiation infection studied in three dimensions. Adv Virus Res 63:287–352
5. Nakanishi M, Akuta T, Nagoshi E, Eguchi A, Mizuguchi H, Senda T (2001) Nuclear targeting of DNA. Eur J Pharm Sci 13:17–24
6. Tang J, Lander GC, Olia AS, Li R, Casjens S, Prevelige P Jr, Cingolani G, Baker TS, Johnson JE (2011) Peering down the barrel of a bacteriophage portal: the genome packaging and release valve in p22. Structure 19:496–502
7. Specthrie L, Bullitt E, Horiuchi K, Model P, Russel M, Makowski L (1992) Construction of a microphage variant of filamentous bacteriophage. J Mol Biol 228:720–724
8. Toropova K, Basnak G, Twarock R, Stockley PG, Ranson NA (2008) The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. J Mol Biol 375:824–836
9. Bamford DH, Caldentey J, Bamford JK (1995) Bacteriophage PRD1: a broad host range dsDNA tectivirus with an internal membrane. Adv Virus Res 45:281–319
10. Butcher SJ, Manole V, Karhu NJ (2012) Lipid-containing viruses: bacteriophage PRD1 assembly. In: Rossmann MG, Rao V (eds) Viral molecular machines. Springer, New York

11. Hayashi M, Aoyama A, Richardson DL, Hayashi MN (1988) Biology of the bacteriophage PhiX174. In: Calendar R (ed) The bacteriophages. Plenum Publishing Corporation, New York
12. Fane BA, Brentlinger KL, Burch AD, Hafenstein SL, Moore E, Novak CR, Uchiyama A (2005) PhiX174 et al., the microviridae. In: Calendar RL (ed) The bacteriophages, 2nd edn. Oxford University Press, Oxford
13. Golmohammadi R, Valegard K, Fridborg K, Liljas L (1993) The refined structure of bacteriophage MS2 at 2.8Å resolution. J Mol Biol 234:620–639
14. Toropova K, Stockley PG, Ranson NA (2011) Visualising a viral RNA genome poised for release from Its receptor complex. J Mol Biol 408:408–416
15. Riechmann L, Holliger P (1997) The C-terminal domain of TolA is the co-receptor for filamentous phage infection of *E. coli*. Cell 90:351–360
16. McKenna R, Ilag LL, Rossmann MG (1994) Analysis of the single-stranded DNA bacteriophage PhiX174, refined at a resolution of 3.0Å. J Mol Biol 237:517–543
17. Betts S, King J (1999) There's a right way and a wrong way: *in vivo* and *in vitro* folding, misfolding and subunit assembly of the P22 tailspike. Structure 7:R131–R139
18. Olia AS, Casjens S, Cingolani G (2007) Structure of phage P22 cell envelope–penetrating needle. Nat Struct Mol Biol 14:1221–1226
19. Casjens SR, Molineux IJ (2012) Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. In: Rossmann MG, Rao V (eds) Viral molecular machines. Springer, New York
20. Chang JT, Schmid MF, Haase-Pettingell C, Weigele PR, King JA, Chiu W (2010) Visualizing the structural changes of bacteriophage Epsilon15 and its *Salmonella* host during infection. J Mol Biol 402:731–740
21. Garcia-Doval C, van Raaij MJ (2012) Crystallization of the C-terminal domain of the bacteriophage T7 fibre protein gp17. Acta Crystallogr Sect F Struct Biol Cryst Commun 68:166–171
22. Steven AC, Trus BL, Maizel JV, Unser M, Parry DA, Wall JS, Hainfield JF, Studier FW (1988) Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. J Mol Biol 200:351–365
23. Garcia-Doval C, van Raaij (2012) Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers. Proc Natl Acad Sci USA 109:9390–9395
24. Chang CY, Kemp P, Molineux IJ (2010) Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell. Virology 398:176–186
25. Xiang Y, Rossmann MG (2011) Structure of bacteriophage Phi29 head fibers has a supercoiled triple repeating helix-turn-helix motif. Proc Natl Acad Sci USA 108:4806–4810
26. Xiang Y, Leiman PG, Li L, Grimes S, Anderson DL, Rossmann MG (2009) Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. Mol Cell 34:375–386
27. Xiang Y, Morais MC, Battisti AJ, Grimes S, Jardine PJ, Anderson DL, Rossmann MG (2006) Structural changes of bacteriophage Phi29 upon DNA packaging and release. EMBO J 25:5229–5239
28. Xiang Y, Morais MC, Cohen DN, Bowman VD, Anderson DL, Rossmann MG (2008) Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage Phi29 tail. Proc Natl Acad Sci USA 105:9552–9557
29. Alcorlo M, Gonzalez-Huici V, Hermoso JM, Meijer WJ, Salas M (2007) The phage Phi29 membrane protein p16.7, involved in DNA replication, is required for efficient ejection of the viral genome. J Bacteriol 189:5542–5549
30. Plisson C, White HE, Auzat I, Zafarani A, Sao-Jose C, Lhuillier S, Tavares P, Orlova EV (2007) Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. EMBO J 26:3720–3728
31. Davidson AR, Cardarelli L, Pell LG, Radford DR, Maxwell KL (2012) Long noncontractile tail machines of bacteriophages. In: Rossmann MG, Rao V (eds) Viral molecular machines. Springer, New York

32. Letellier L, Boulanger P, Plancon L, Jacquot P, Santamaria M (2004) Main features on tailed phage, host recognition and DNA uptake. Front Biosci 9:1228–1339
33. Heller K, Braun V (1979) Accelerated adsorption of bacteriophage T5 to *Escherichia coli* F, resulting from reversible tail fiber-lipopolysaccharide binding. J Bacteriol 139:32–38
34. Heller K, Braun V (1982) Polymannose O-antigens of *Escherichia coli*, the binding sites for the reversible adsorption of bacteriophage T5 *via* the L-shaped tail fibers. J Virol 41:222–227
35. Boehm J, Lambert O, Frangakis AS, Letellier L, Baumeister W, Rigaud JL (2001) FhuA-mediated phage genome transfer into liposomes: a cryo-electron tomography study. Curr Biol 11:1168–1175
36. Guihard G, Boulanger P, Letellier L (1992) Involvement of phage T5 tail proteins and contact sites between the outer and inner membrane of *Escherichia coli* in phage T5 DNA injection. J Biol Chem 267:3173–3178
37. Hendrix RW, Duda RL (1992) Bacteriophage Lambda PaPa: not the mother of all Lambda phages. Science 258:1145–1148
38. Montag D, Schwarz H, Henning U (1989) A component of the side tail fiber of Escherichia coli bacteriophage Lambda can functionally replace the receptor-recognizing part of a long tail fiber protein of the unrelated bacteriophage T4. J Bacteriol 171:4378–4384
39. Berkane E, Orlik F, Stegmeier JF, Charbit A, Winterhalter M, Benz R (2006) Interaction of bacteriophage Lambda with its cell surface receptor: an *in vitro* study of binding of the viral tail protein gpJ to LamB (maltoporin). Biochemistry 45:2708–2720
40. Roessner CA, Ihler GM (1986) Formation of transmembrane channels in liposomes during injection of Lambda DNA. J Biol Chem 261:386–390
41. Erni B, Zanolari B, Kocher HP (1987) The mannose permease of *Escherichia coli* consists of three different proteins. Amino acid sequence and function in sugar transport, sugar phosphorylation, and penetration of phage Lambda DNA. J Biol Chem 262:5238–5247
42. Veesler D, Robin G, Lichiere J, Auzat I, Tavares P, Bron P, Campanacci V, Cambillau C (2010) Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1): a baseplate hub paradigm in Gram-positive infecting phages. J Biol Chem 285:36666–36673
43. Lhuillier S, Gallopin M, Gilquin B, Brasiles S, Lancelot N, Letellier G, Gilles M, Dethan G, Orlova EV, Couprie J, Tavares P, Zinn-Justin S (2009) Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. Proc Natl Acad Sci USA 106:8507–8512
44. Sciara G, Bebeacua C, Bron P, Tremblay D, Ortiz-Lombardia M, Lichière J, van Heel M, Campanacci V, Moineau S, Cambillau C (2010) Structure of lactococcal phage p2 baseplate and its mechanism of activation. Proc Natl Acad Sci U S A 107:6852–6857
45. Veesler D, Spinelli S, Mahony J, Lichiere J, Blangy S, Bricogne G, Legrand P, Ortiz-Lombardia M, Campanacci V, van Sinderen D, Cambillau C (2012) Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism. Proc Natl Acad Sci U S A 109:8954–8958
46. Bartual SG, Otero JM, Garcia-Doval C, Llamas-Saiz AL, Kahn R, Fox GC, van Raaij MJ (2010) Structure of the bacteriophage T4 long tail fiber receptor-binding tip. Proc Natl Acad Sci USA 107:20287–20292
47. Leiman PG, Arisaka F, van Raaij MJ, Kostyuchenko VA, Aksyuk AA, Kanamaru S, Rossmann MG (2010) Morphogenesis of the T4 tail and tail fibers. Virol J 7:355
48. Browning C, Shneider MM, Bowman VD, Schwarzer D, Leiman PG (2012) Phage pierces the host cell membrane with the iron-loaded spike. Structure 20:326–339
49. Rao VB, Black LW (2010) Structure and assembly of bacteriophage T4 head. Virol J 7:356
50. Leiman PG, Shneider MM (2012) Contractile tail machines of bacteriophages. In: Rossmann MG, Rao V (eds) Viral molecular machines. Springer, New York
51. Wood WB, Eiserling FA, Crowther RA (1994) Long tail fibers: genes, proteins, structure, and assembly. In: Karam JD (ed) Molecular biology of bacteriophage T4. ASM Press, Washington, DC

52. Goldberg EB, Grinius L, Letellier L (1994) Recognition, attachment and injection. In: Karam JD (ed) Molecular biology of bacteriophage T4. ASM Press, Washington, DC
53. Nilsson AS, Haggard-Ljungquist E (2005) The P2-like bacteriophage. In: Calendar RL (ed) The bacteriophages, 2nd edn. Oxford University Press, Oxford
54. Klumpp J, Lavigne R, Loesnner MJ, Ackermann HW (2010) The SPO1-related bacteriophages. Arch Virol 155:1547–1561
55. Fu X, Walter MH, Paredes A, Morais MC, Liu J (2011) The mechanism of DNA ejection in the *Bacillus anthracis* spore-binding phage 8a revealed by cryo-electron tomography. Virology 421:141–148
56. Walter M, Fiedler C, Grassl R, Biebl M, Rachel R, Hermo-Parrado XL, Llamas-Saiz AL, Seckler R, Miller S, van Raaij MJ (2008) Structure of the receptor-binding protein of bacteriophage Det7: a podoviral tail spike in a myovirus. J Virol 82:2265–2273
57. Schwarzer D, Buettner FF, Browning C, Nazarov S, Rabsch W, Bethe A, Oberbeck A, Bowman VD, Stummeyer K, Muhlenhoff M, Leiman PG, Gerardy-Schahn R (2012) A multivalent adsorption apparatus explains the broad host range of phage Phi92: a comprehensive genomic and structural analysis. J Virol 86:10384–10398
58. Paolozzi L, Ghelardini P (2006) The bacteriophage Mu. In: Calendar RL (ed) The bacteriophages, 2nd edn. Oxford University Press, Oxford
59. Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A, Maskell DJ, Simons RW, Cotter PA, Parkhill J, Miller JF (2002) Reverse-transcriptase mediated tropism switching in *Bordetella* bacteriophage. Science 295:2091–2094
60. Liu J, Chen CY, Siomi D, Niki H, Margolin W (2011) Visualization of bacteriophage P1 infection by cryo-electron tomography of tiny *Escherichia coli*. Virology 417:304–311

Further Reading

Calendar RL (2005) The bacteriophages, 2nd edn. Oxford University Press, Oxford
Campbell AM (1996) Bacteriophages. In: Fields BN, Knipe DM, Howley PM (eds) Fields virology, 3rd edn. Lippincott Williams & Wilkins, Philadelphia
Karam JD (1994) Molecular biology of bacteriophage T4. ASM Press, Washington, DC
Karam JD, Miller ES (2010) Bacteriophage T4 and its relatives: a series of critical reviews. Biomed Central, London
Rakhuba DV, Kolomiets EI, Dey ES, Novik GI (2010) Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. Pol J Microbiol 59:145–155
Rossmann MG, Rao VB (2012) Viral molecular machines. Adv Exp Med Biol, vol 726, Springer, New York

Also especially recommended for further reading are references [2, 3, 4, 9, 11, 17, 32, 54, 58] *listed above.*

Chapter 18
Mechanical Properties of Viruses

Pedro J. de Pablo and Mauricio G. Mateu

Abstract Structural biology techniques have greatly contributed to unveil the relationships between structure, properties and functions of viruses. In recent years, classic structural approaches are being complemented by single-molecule techniques such as atomic force microscopy and optical tweezers to study physical properties and functions of viral particles that are not accessible to classic structural techniques. Among these features are mechanical properties such as stiffness, intrinsic elasticity, tensile strength and material fatigue. The field of virus mechanics is contributing to materials science by investigating some physical parameters of "soft" biological matter and biological nano-objects. Virus mechanics studies are also starting to unveil the biological implications of physical properties of viruses. Growing evidence indicate that viruses are subjected to internal and external forces, and that they may have adapted to withstand and even use those forces. This chapter describes what is known on the mechanical properties of virus particles, their structural determinants, and possible biological implications, of which several examples are provided.

Keywords Virus • Capsid • Atomic Force Microscopy • Mechanical Properties • Elasticity • Stiffness • Spring Constant • Tensile Strength • Material Fatigue • Molecular Structure • Stability • Dynamics • Biological Materials • Virus Engineering • Biotechnology • Nanotechnology

P.J. de Pablo (✉)
Department of Physics of the Condensed Matter,
C03, Facultad de Ciencias, Universidad Autónoma de Madrid,
Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: p.j.depablo@uam.es

M.G. Mateu (✉)
Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM) and Department of Molecular Biology of the Universidad Autónoma de Madrid,
c/Nicolás Cabrera 1, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: mgarcia@cbm.uam.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_18,

Abbreviations

AFM	Atomic force microscopy
CCMV	Cowpea chlorotic mottle virus
ds	Double-stranded
EM	Electron microscopy
FEA	Finite-element analysis
FZ	Force *vs*. z-piezo displacement
HBV	Hepatitis B virus
HIV-1	Human immunodeficiency virus type 1
HSV-1	Herpes simplex virus type 1
MLV	Murine leukemia virus
MVM	Minute virus of mice
NV	Norovirus (Norwalk virus)
ss	Single-stranded
vdW	Van der Waals

18.1 Introduction

The mechanical properties of some biomolecular complexes are essential for their function. Indeed, forces at the nanoscale play a central role in biochemistry, from the myosin-actin system [1], which is the ultimate responsible of muscular action, to cellular or viral motor protein assemblies ([2], see Chaps. 9 and 12).

Viruses are striking examples of biomolecular complexes endorsed with specific material properties that may provide a basis to understand some aspects of their biophysical function, and their ability to endure a variety of environmental aggressions. Indeed, during the passive extracellular stage of the infection cycle the viral capsid can be considered as a container [3] (see Chap. 2) that protects the viral genome against physicochemical assaults [4]. This critical function of many viral capsids have imposed strong selection pressures on them, leading to structural stabilization. On the other hand, the need to uncoat the viral genome during infection of host cells has favored the evolution of many capsids as metastable, conformationally dynamic biological complexes. Thus, nature may have modulated some of the capsid material characteristics for both structural stability and dynamics (see Chap. 1). Different examples of the delicate balance between viral capsid structural stability and dynamics and its control during the viral cycle are provided in Chaps. 1, 10, 11, 13, 15 and 19.

Mechanical properties such as elasticity or tensile strength are among the material properties of any solid object including virus particles. The development of atomic force microscopy (AFM) (see [5] and Chap. 8) has enabled the study of the

mechanical properties of virions and virus capsids at the single-particle level [6]. In this chapter we describe current approaches to determine mechanical properties of virus particles that are inherent to materials science analyses. These properties include particle stiffness (by measuring spring constant values), intrinsic elasticity of the capsid material (Young's Modulus), brittleness and material fatigue. We then discuss the relationships between these mechanical properties and the molecular structure of viruses, possible biological implications and bio/nanotechnological applications of virus mechanics.

18.2 Mechanical Stiffness of Virus Particles Determined by AFM in Indentation Assays

The advent of single-molecule techniques, such as AFM (Chap. 8), optical tweezers (Chap. 9) and similar methods [7] opened the possibility of measuring the tiny forces (nN or pN) that take place in biological molecules and processes at the nanometer scale, such as conformational changes [8], protein folding [9] or the mechanochemical action of motor proteins [10] (see Chaps. 9 and 12). In particular, AFM is ideally suited to probe the mechanical elasticity of biological specimens in liquid, in close to physiological conditions (see also Chap. 8). AFM was applied early to cells [11] and microtubules [12], and it was just a matter of (a short) time that it was also applied to study virus elasticity. It may be illustrative to briefly refer here the origin of the first experiments on the mechanical properties of small, spherical virus particles in liquid using phage ϕ29 as a model [6]. Optical tweezers had been used to study the mechanochemical action of the ϕ29 DNA packaging motor [13] (see Chap. 9), and other experiments using optical tweezers on this same phage were later started by colleagues of one of the authors of this chapter (P.J.P.). Since these experiments were not working properly, we decided to check the virus integrity by AFM imaging. From that moment on, it was obvious that the elasticity experiments we and others had carried out by indenting (much larger) cells or (much longer) rod-like microtubules with the AFM tip could be readily adapted to quantitatively probe the mechanical features of even the smallest spherical viruses.

Chapter 8 describes the technique of AFM and some experimental setups regarding its use for virus imaging. In this section we will briefly describe some particular conditions required for measuring the mechanical stiffness of viral particles using AFM. The procedure requires first the strong adsorption of the specimens on a suitable surface, in order to minimize unwanted displacements during relatively deep indentations. In the case of viruses, a typical recipe is to silanize glass surfaces to enhance the hydrophobic interactions with the viral particles [6]. The virus suspension is diluted to get a few particles (3–5) per square micron on the surface and guarantee their convenient localization, while avoiding crowding (Fig. 18.1a). Jumping mode [14] has shown to be extremely successful for imaging individual viruses on surfaces (see Chap. 8). It is convenient to prewet the

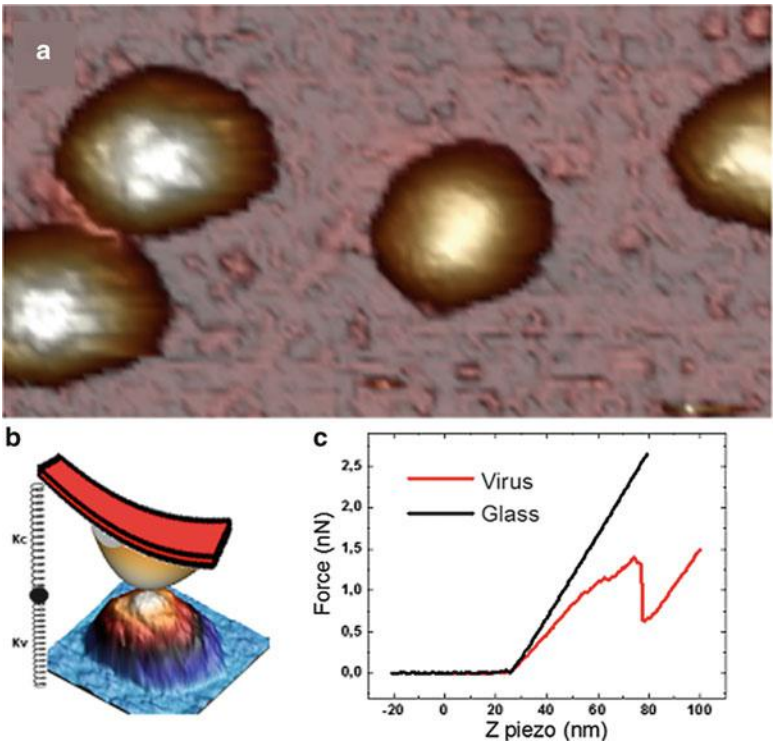

Fig. 18.1 Single indentation assay. (**a**) AFM topography of ϕ29 phage particles adsorbed on HOPG. (**b**) a sketch of an ideal nanoindentation of a viral particle and the interpretation of the mechanical result by considering the AFM cantilever and the virus particle as a system of two springs in series. **c** FZ (force *vs*. z-piezo displacement) curves on a hard substrate (*dark*) and a virus particle (*red*)

tip with ~20 μl of buffer to avoid undesired cantilever bending. A great amount of the work published on virus mechanics has been performed with AFM systems provided by *Nanotec Electrónica* [15]. Rectangular cantilevers RC800PSA, and Biolevers (BL-AC40TS) (*Olympus*) with nominal spring constants of 0.05 N/m and 0.03 N/m, respectively, are typically used. Cantilevers spring constants are typically calibrated using Sader's method [16]. Siffness determination is based in the monitored indentation of individual viral particles with the AFM tip. Therefore, in order to perform nano-indentations, single FZ (force *vs*. z-piezo displacement) experiments are performed by *pushing* on the top of a selected single virus particle. The particle is continuously zoomed in by reducing the x-y scanning size until the bump of the very top is under the whole piezo scan (about ~50 nm × 50 nm). Afterwards the FZ is executed by indenting the top of the particle, likely within a few nm of uncertainty (mainly provoked by thermal drift and the intrinsic non-linearity and creep of the piezo). This method has proven to be robust enough to

establish electrical contact with carbon nanotubes [17], which are much smaller in diameter than viral particles.

When the tip is in contact with the sample the tip-sample force F can be described by:

$$F = F_{contact} + F_{DLVO}$$

$F_{contact}$ is determined by the contact mechanics. In general, the expression of this force is a complicated solution of the elasticity equations, but there are two special and useful situations where it adopts a simple form:

i. *Hertzian contact* – for the gentle indentation of a spherical tip on an elastic solid half-space,

$$F_{contact} = 4/3E^* \sqrt{R} d^{3/2}$$

where R is the tip radius, d the indentation depth, and E^* is given by:

$$1/E^* = \left(1 - {\nu_t}^2\right)/E_t + \left(1 - {\nu_s}^2\right)/E_s$$

being ν_t, E_t, ν_s and E_s the Poisson's ratio and elastic modulus of the tip (t) and the sample (s) (defined below).

ii. *Linear contact* – when the effective response is like that of a spring with a constant k_{eff}. This situation occurs for small indentations of a thin shell-like structure [18], and

$$F_{contact} = -k_{eff} d$$

F_{DLVO} follows the Derjaguin-Landau-Verwey-Overbeek model (DLVO) [19] and accounts for the double Debye double-layer electrostatic force (F_{el}) and van der Waals (vdW) force (F_{vdW}). In the simplest case of two planes the forces are given by:

$$F_{DLVO} = F_{el} + F_{vdW} = \frac{2\sigma_s \sigma_p}{\varepsilon_0 \varepsilon} e^{\frac{-z}{\lambda_D}} - \frac{H_a}{6\pi z^3}$$

where σ_s and σ_t are the charge density of sample and tip, respectively; $\varepsilon\varepsilon_0$ is the dielectric constant times the permittivity of vacuum; z is the tip-sample distance, H_a is the Hammaker constant describing the van der Waals interactions between tip and sample, and λ_D is the Debye length that determines the range of the electrostatic forces.

We must consider two issues at this point. First, it is evident that to prime the measurement of sample elasticity it is necessary to decrease the Debye length. Since the Debye length decreases with the concentration c of salt in solution as $\lambda_D \sim 1/\sqrt{c}$ [20], when investigating viruses relatively high (but physiological) salt

concentrations are usually employed to screen as much as possible the electrostatic forces. In those circumstances, the tip-sample interactions are dominated by the contact forces and sample elasticity (Fig. 18.1b). Second, since in a first approximation most icosahedral virus capsids can be considered as hollow, relatively thin shells, it is tempting to use directly the linear approximation for modeling the tip-virus contact. Nevertheless, some caution is necessary because of two main reasons. On one hand, the linear model can only be applied if the total indentation of the virus capsid is at most the shell thickness (see [18] and Fig. 18.1c), which in most viruses is about a couple of nanometers. On the other hand, capsids or virions may be filled with either scaffolding proteins or nucleic acid, thus deviating from an ideal shell-like model. Even so, since these fillings cannot be considered as homogeneous solids in order to use the Hertzian contact approach, the linear model is still approximately valid and provides an easy and direct framework to interpret the experimental results.

In most AFM nanoindentation experiments on viruses it has been found that, for small deformations, the force is indeed linear with the indentation. In such cases, the spring constant of the virus k_v can be easily obtained by simply considering the virus like a spring in series with the cantilever (Fig. 18.1b):

$$k_v^{-1} = k_{eff}^{-1} - k_c^{-1}$$

where k_c is the spring constant of the cantilever (here the substrate is considered as non-deformable). For practical use, this equation is customarily rewritten in terms of the slope of the cantilever deflection on the glass, s_g (nm/V), and the slope of the cantilever deflection on the virus s_v (nm/V), as:

$$k_v = k_c \frac{s_g}{s_v - s_g}$$

A standard protocol of an AFM nanoindentation experiment is as follows. Each viral particle is indented with a sequence of about 5 FZ's. Following each FZ set, an image of the virus is taken to confirm its integrity, by checking that several effects, such as damage [21], collapse [12], or buckling [22] have not occurred; the image serves also to ascertain the particle position in order to correct for any drift, if needed, before performing the next FZ set. It is convenient to adjust the FZ speed to about ~60 nm/s in order to allow the water leave the virus when it is squeezed [23]. Even if the shell integrity is maintained, only those viral particles showing stable values for the spring constant along successive FZ measurements should be considered. Thus artifacts such as particle mobility effects that often occur when the particle is loosely bound to the surface can be mostly avoided. However, effects such as tip sliding or particle tilting are difficult to quantify, and techniques to rule them out remain to be developed. On the other hand, if breakage of the virus particle [24] is desired to determine tensile strength (rather than particle stiffness) the z-piezo elongation should be large enough to provoke a cantilever deflection similar

to the breaking force [6]. Several particles are normally used in the experiments, and the results are appropriately averaged.

It may be tempting to directly relate virus particle stiffness with structural stability against disruption into subunits. This interpretation may imply that "soft" virus particles are less resistant to dissociation than "hard" ones. However, very often elastic objects are more resilient than rigid ones. We can find examples in everyday life, *e.g.*, rubber *vs.* glass. Rubber is softer than glass, but glass do not support mechanical deformation without breaking as rubber does. Thus, other parameters such as tensile strength of virus particles need to be investigated.

18.3 Intrinsic Elasticity of Virus Capsids: Young's Modulus

In materials science, the Young's modulus of elasticity E provides a measure of the *intrinsic* stiffness of an elastic material. Let us first consider a solid piece of some material, say a cylinder of initial length L_0 and a cross-section with a surface A. If we apply a force F along the main axis, the cylinder elongates from its original length L_0 to $L_0+\Delta\ell$. For small deformations, the stress $\sigma = F/A$, *i.e.* the force per unit area A, is proportional to the relative deformation or strain ε, given by $\Delta\ell/L_0$ (Fig. 18.2a). Thus, $\sigma = E\varepsilon$, where the proportionality constant E is the Young's modulus. From the latter we can find an expression for the force [25]

$$F = \frac{AE}{L_0}\Delta\ell$$

The coefficient of $\Delta\ell$ has units of N/m, being the effective longitudinal spring constant k of the object (see Fig. 18.2a). Other important mechanical magnitude is the non-dimensional Poisson ratio parameter μ, which accounts for the ratio of transverse and longitudinal strains:

$$\mu = -\frac{\varepsilon_{trans}}{\varepsilon_{long}}$$

Transversal strain ε_{trans} is typically negative, since when a piece of material is pulled by the force F the transversal dimension decreases by Δw (Fig. 18.2b). There are exceptions like auxetic materials that, when stretched, become thicker perpendicularly to the applied force, and thus have negative values of the Poisson ratio. In general, the Poisson ratio μ value is constrained between -1 and 0.5, and steel and rigid polymers show values around 0.3 [26]. Actually, in studies of virus mechanics μ is assumed to be 0.3.

In the following we will go through the typical approaches followed in most published works to extract the Young's modulus of viral shells, *i.e.*, thin-shell theory and finite element analysis (see also Chap. 19). Certainly, from a physical

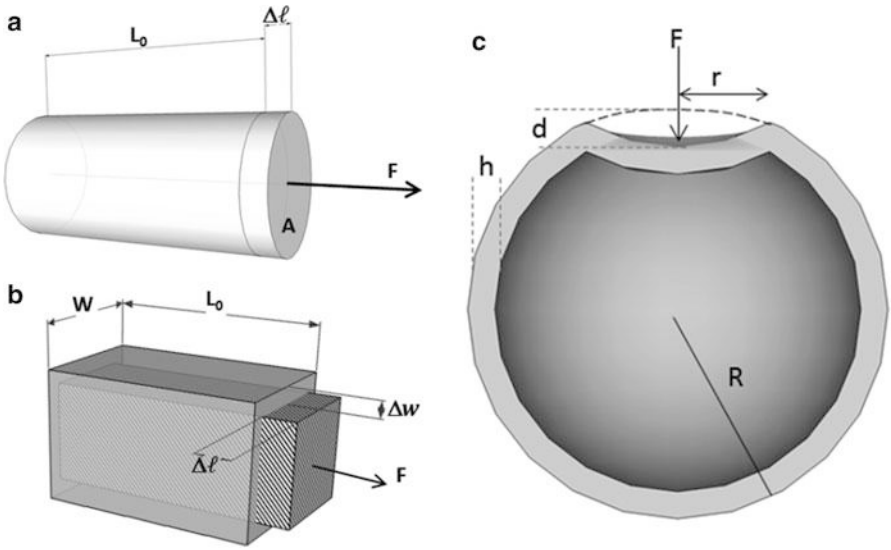

Fig. 18.2 Physical magnitudes of materials mechanics. (**a**, **b**) Geometrical parameters required for estimating the Young's Modulus (**a**) or the Poisson Ratio (**b**). (**c**) Main geometrical definitions for the application of thin shell theory (see text)

point of view, one outcome of most works published on virus mechanics so far is the estimation of the Young's modulus of a virus shell.

18.3.1 Shell Theory

It is well known that shells, *i.e.* thin plates which are curved in their undeformed state, exhibit special mechanical properties. For the sake of clarity, let us begin with a thin and straight rod. When supported at both ends, transverse forces cause bending and deflection. However, the same rod can withstand much greater axial load with unnoticeable deformation. Similar concepts apply to a thin flat shell. Perpendicular forces actuating on a flat plate would generate bending, resisted by bending moments in the cross-section. In-plane forces, in turn, would cause stretching, compression and/or shearing, resisted by in-plane forces on a cross-section. A closed curved shell would be subjected to both bending and stretching under the application of a perpendicular force, offering more mechanical strength than the flat plate that will essentially be bent without stretching. In the following we will analyze the details of these ideas in thin curved shells. When a spherical shell of radius R and thickness $h \ll R$ is deformed a distance d by an indenting point force F in the regime where $d \ll R$, the shell is locally both stretched and bent (Fig. 18.2c) [18]. As a consequence a bulge of size r develops right at the

application point of F. We can express the stretching and bending energies as [27]:

$$E_{str} \sim K\left(\frac{d}{R}\right)^2 r^2, \text{ and } E_b \sim \kappa\left(\frac{d}{r^2}\right)^2 r^2$$

where $K = \frac{Eh}{2(1-\mu)}$ and $\kappa = \frac{Eh^3}{12(1-\mu^2)}$ are the 2D stretching and bending modulus, respectively. Note that these refer to the three-dimensional properties of the material making up the shell. We can write the sum of bending and stretching energies as:

$$E_{total} \sim E\frac{h^2}{R}\left[\frac{hR}{r^2} + \frac{r^2}{hR}\right]d^2$$

By imposing the equilibrium condition $\frac{dE_{total}}{dr} = 0$ we can find that the size of the bent region scales as $r \sim \sqrt{hR}$. This means, for instance, that for two spherical shells with identical thickness h, doubling the radius R only increases the lateral deformation by a factor $\sqrt{2}$. By making the derivative of E_{total} with respect to the indentation d and using the obtained value of r, we find that the force is linear in the indentation with a spring constant of the shell that scales like $\sim E\frac{h^2}{R}$. To calculate the proportionality constant accurately requires to expand the shape of the deformed shell in spherical harmonics [12, 28], and for a thin shell it turns out to be $\frac{2}{\sqrt{3(1-\nu^2)}}$.

18.3.2 Finite Element Analysis

Another popular approximation for interpreting the results of experiments on virus mechanics is to use Finite Element Analysis (FEA). In simple terms, FEA analysis attempts to obtain a numerical solution of the complicated differential equations describing a physical problem concerning a physical object considered as a continuous medium [29]. The key-point is to divide the object in many small non-intersecting domains (finite elements) where the physical equations are discretized and solved. This numerical methodology can be applied for a variety of physical phenomena, including mechanical deformations, circulation of fluids or electrostatic problems.

In order to simulate the mechanical deformation of a virus capsid, an idealized, simplified geometrical model of the capsid is obtained from the three-dimensional structure provided by X-ray crystallography or cryo-electron microscopy (cryo-EM) (Fig. 18.3a, b) using commercially available software. Young's modulus and Poisson ratio values are established as material properties of the model. Afterwards, the proper boundary conditions are fixed and the FEA structure is deformed by punctual forces, Hertzian forces or spheres (Fig. 18.3c). The values of the forces used in FEA

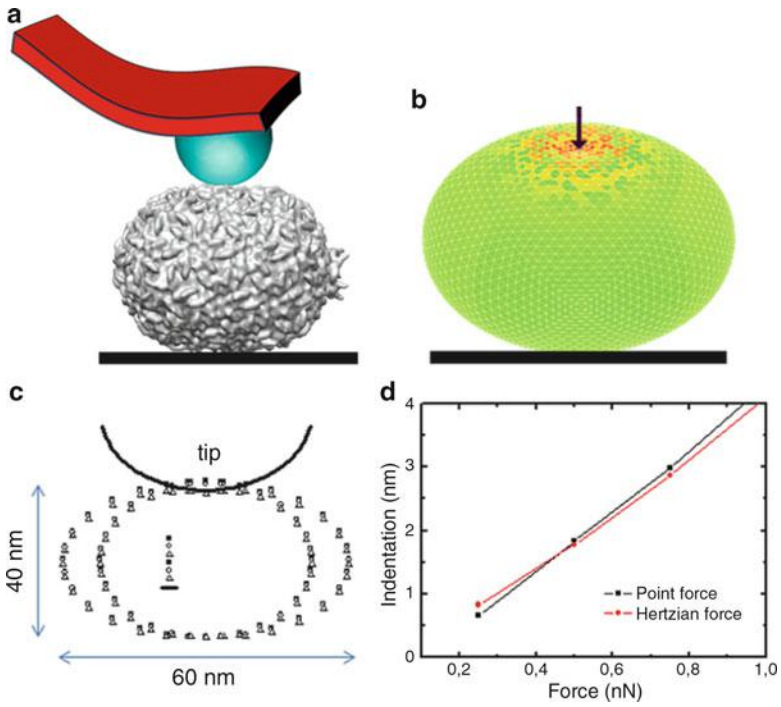

Fig. 18.3 Finite Elements Modeling. (**a**) Cartoon that depicts the actual geometry of a simple AFM tip and a morphologically complex virus (phage ϕ29). (**b**) Simplified FEA modeling of phage ϕ29 as used in ref. [6]. (**c**) Maximum and minimum cross-section variations of the FEA model as the tip deforms the virus model. (**d**) Typical FEA results of indentations by using point and Hertzian forces as a function of the indenting force

are the same than those actually used in AFM nanoindentation experiments. Therefore, deformation of the model is obtained as a function of the loading force and the indentation is depicted *vs.* force. Interestingly, for indentations of the order of the shell thickness, punctual and Hertzian forces result in similar deformations (Fig. 18.3d). Thus the typical approach is to vary the Young's modulus to match the experimental indentation in actual AFM experiments. Specifically, FEA models of viruses are designed in terms of the capsid general shape, diameter and thickness. However, it is difficult to capture the complexity of a virus capsid structure with FEA. The proteins forming the capsid and the bonds between them cannot be modeled just by using Young's modulus and Poisson ratio values. Nonetheless, FEA provides a coarse approach for understanding the mechanical properties of the capsid protein material [6, 30]. In order to refine the calculation, some studies have performed FEA using virus models that approach more closely cryo-EM structural models [21]. Other studies have contemplated coarse grained models that include as discrete pieces the capsid subunits or capsomers that form the capsid structures, and approximate the interactions between subunits using harmonic or Lennard Jones potentials [31].

18.4 Brittleness and Material Fatigue of Virus Particles

A material is brittle if it breaks without a significant deformation (strain) when subjected to mechanical deformation. This is the case of glass or an eggshell. Brittle materials usually do not present plastic deformation before breaking. On the other hand, material fatigue studies apply cycling loading forces below rupture strength of the object under examination [32], revealing not only the evolution with time and repeated application of force of its resistance to breakage, but also the location of the object's weakest parts. A fundamental requirement of these studies is monitoring the structural integrity of the sample along the experiment. The investigation of both mechanical characteristics of virus particles require a precise control of the force applied on the particle [33] to induce sequential disruption of the capsid subunits. These novel mechanical studies on viral particles may provide biologically relevant information about virus disassembly.

18.5 A Case Study: The Mechanics of Bacteriophage ϕ29

In this section we will exemplify the combination of AFM experiments and FEA modeling to extract some biophysical implications on the mechanical properties of virus particles, using bacteriophage ϕ29 as a model. First, we will discuss the values obtained for the spring constant of the ϕ29 prohead, or immature capsid (as a measure of capsid mechanical stiffness), and its relationship with the Young's modulus (as a measure of the intrinsic elasticity of the capsid material) by using Finite Element modeling. Second, we will explain the detection and evaluation of internal pressure in bacteriophage ϕ29. Finally, we will describe the identification of built-in mechanical stress in viral particles and its influence on capsid stiffness.

18.5.1 Mechanical Properties of the ϕ29 Capsid: Stiffness and Structural Stability Against Breakage

Indentation assays of individual ϕ29 proheads revealed the existence of a bimodal distribution of the spring constant [6], and yielded average values of 0.16 N/m and 0.31 N/m. However, further studies of the same proheads [31, 34] revealed a single distribution of the spring constant with an average value of about 0.3 N/m. It is possible to calculate analytically the deformation of a homogeneous, spherical, and elastic shell that is subjected to equal and opposite forces applied at the poles by expanding the shape of the deformed sphere in spherical harmonics [28]. The spring constant for the total indentation becomes $k = 2.25\frac{Eh^2}{R}$, and the Young's modulus of the shell can be estimated, by using the thickness h and the radius R of the phage particle (obtained from EM studies [35]), to be between 1.2 GPa and 1.6 GPa.

Further modeling with FEA (Sect. 18.3.2) results in a Young's modulus of 1.8 GPa. Phage ϕ29 proheads also showed a decrease of the spring constant after repeated indentations. Further experiments suggested that this capsid softening was probably due to capsid disruption [24, 36].

18.5.2 Internal Pressure in the ϕ29 Virion

It has been proposed that translocation of the double-stranded (ds) DNA genome of phage ϕ29 through the virus tail into the bacterium host is initiated by a *push* mechanism, followed by a *pull* mechanism mediated by bacterium proteins [37]. The push mechanism would be driven by the elastic energy provided by the internal pressure accumulated during the DNA packaging process (see also Chaps. 17 and 19). However, the very existence of pressure in ϕ29 has remained elusive, and single-molecule techniques have added valuable information. Specifically, AFM nanoindentation experiments on individual viral particles reveal that the ϕ29 virion is stiffer than the ϕ29 prohead (Fig. 18.4b). Since no major mechanically-relevant structural differences between the prohead and the empty mature head particles are apparent, it is reasonable to consider that the packaged dsDNA is the main responsible of the observed stiffening.

In order to distinguish between structural or pressure-induced reinforcement, further experiments were performed in the presence of spermidine^{3+}, a positively charged counter ion that can induce DNA condensation, even inside viruses [38], because it neutralizes the negative charges of the nucleic acid phosphates. Real-time indentation experiments probing the same individual particles before and during treatment with spermidine^{3+} showed that this compound induces a mechanical softening of the full virion that is made as elastic as the prohead, as determined by measuring spring constant values. Experiments carried out by adding and later removing spermidine^{3+} showed that this is a reversible process (Fig. 18.4c). Theoretical calculations using the model described in [39] yielded an internal pressure in the ϕ29 virion of about 30 ± 10 atm in those conditions, in the absence of spermidine^{3+}. The experimental evaluation of the internal pressure from the determination of spring constant values in the absence or presence of spermidine^{3+} and the use of FEA (see Sect. 18.3.2 and Chap. 19) yielded a pressure of 40 ± 10 atm, in good agreement with the theoretical predictions (Fig. 18.4d) [34].

18.5.3 Built-in Mechanical Stress in the ϕ29 Capsid

We have already commented on the mechanical advantage that curved shells present when subjected to deformation (*e.g.*, by indentation), as they undergo simultaneous stretching and bending. Nevertheless, there are other mechanisms which viral shells may take advantage of to improve their mechanical performance.

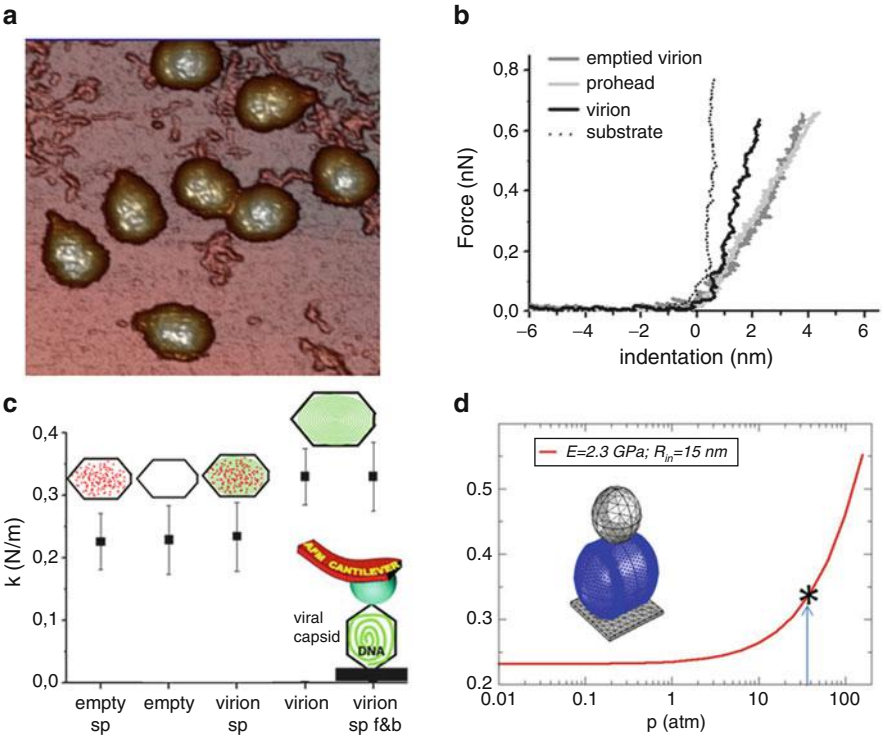

Fig. 18.4 Phage ϕ29 internal pressure. (**a**) AFM topography of mature ϕ29 particles; the tail is clearly visible. (**b**) Individual indentation curves on the substrate (*dotted*), DNA-filled ϕ29 virion (*black*), empty ϕ29 virion (*dark gray*) and ϕ29 prohead (*gray*). (**c**) Values of the spring constants obtained for the different types of ϕ29 particles analyzed in (**b**); sp, spermidine; sp f&b, spermidine added and later removed. (**d**) FEA results of the spring constant as a function of pressure for a phage ϕ29-like structure

Residual mechanical stress is the stress remaining after its cause has been removed. If we take a straight rod and simultaneously push from both ends (Fig. 18.5a), it bulges up in a stressed bump. This deformation quickly disappears when forces are removed, and the rod recovers the original straight shape. However, this stress can be conserved if both ends of the bulged rod are pinned.

The shell of bacteriophage ϕ29 prohead exhibits an intriguing anisotropic stiffness: Nanoindentations show that the prolate ϕ29 prohead is about twofold stiffer along the short axis (particle laying on its side) than along the long axis (upright particle) (Fig. 18.5b). The experimental and theoretical evidence indicates that this anosotropy is due to residual mechanical stress, as discussed next.

If the ϕ29 prohead is considered as a shell, continuum elasticity theory (see Sect. 18.3.1 and Chap. 19) predicts spring constant opposite trends along the short and long axes of the empty virus particle to those revealed by experiment. The ϕ29 capsid can be considered as a shell made by a cylindrical body closed by two

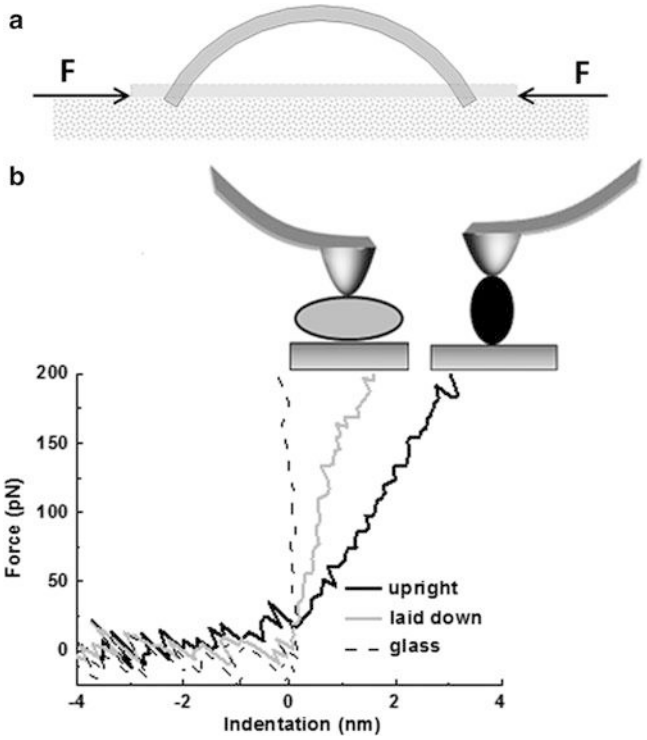

Fig. 18.5 Built-in mechanical stress. (**a**) The concept of residual stress. (**b**) Experimental results of indentations on ϕ29 particles laying on its side (*gray*) or upright (*dark*). Interpretation of these results indicate the presence of built-in mechanical stress in these particles (see text)

spherical caps. By considering independently the elastic response of the cylindrical and spherical parts of the virus particle one would expect that the laid-down virus would be mechanically softer than the upright one, since a cylindrical shell is easier to deform than a spherical shell of the same dimensions [18]. This is because a cylindrical shell, unlike a spherical one, can be bent without much stretching. We have already seen that, for a spherical shell, the effective spring constant is $k_{eff}^{sph} \sim Eh^2/R$. However, when indentation of a cylinder is considered [12], the spring constant equals $k_{eff}^{cyl} \sim Eh^{5/2}/R^{3/2}$. Therefore, the ratio of spring constants is $k_{eff}^{sph}/k_{eff}^{cyl} \sim \sqrt{R/h}$, which for the geometry of ϕ29 ($R = 21\ nm$ and $h = 1.6\ nm$ [35]) becomes approximately $k_{eff}^{sph}/k_{eff}^{cyl} \sim 3.6$. Hence, in the framework of continuum elasticity theory of shells, the cylindrical body of the virus is expected to be more than three times softer that the spherical caps. This clear disagreement with the experimental results can be attributed to the existence of residual stress on the equatorial zone of the shell imposed by the protein curvature, a hypothesis that was confirmed by coarse-grained simulations [31].

This built-in stress in the virus prohead could be an evolutionary strategy to provide extra mechanical strength. It is interesting to discuss prestress generation during the formation of the capsid. In the assembly process there is a competition between the tendency of proteins to aggregate at the preferred curvature and the need to minimize the rim area exposed in a partially assembled capsid by completing a closed structure [40]. If the curvature of the closed capsid is different from the spontaneous one, lateral stress will develop. Interestingly, not only ϕ29 [41], but other complex dsDNA bacteriophages have a scaffolding protein that co-assembles with the major capsid protein to produce a prohead with the correct shape and size [42] (see Chap. 11). Once the shell is built, the scaffolding proteins are released concomitant to DNA packaging. The possibility that some scaffolding protein subunits might be released from the proheads used in the experiments during previous storage cannot be excluded, but no change in prohead structure was in fact detected. Since protein-protein binding interactions in viral capsids are relatively weak, on the order of a few k_BT [43] (see Chap. 19), it is likely that scaffolding proteins help to impose a curvature in the capsid very different from the spontaneous one, by assisting the bending of the proteins at the junctions between subunits. This will generate a much larger in-plane stress, which will mechanically reinforce the capsid. In the absence of a scaffold, the stress generated during assembly could help to better tolerate DNA packaging, avoiding capsid disruption.

18.6 Differences and Variations in Virus Mechanical Properties

The studies on phage ϕ29 mechanics described in Sect. 18.5, and related studies carried out with a dozen other, structurally different model viruses so far (Fig. 18.6) [44] reveal that some fundamental mechanical concepts may apply to virus capsids in general. However, they also reveal that virus capsids may differ widely in specific mechanical features; as much as they may differ in structure, other properties, or the way they function during the viral life cycle. Comparison of the mechanical features of different viruses is helping in the development of theoretical models, and in the formulation of hypothesis on the structural foundations of the mechanical properties of viruses and their possible roles in virus biology. Such models and hypothesis are being subjected to experimental verification by using physical, chemical and/or molecular biology techniques including those described in this book (Chaps. 3, 4, 5, 6, 7, 8 and 9).

18.6.1 Differences in Stiffness

All viral capsids are made up of a same material, protein, and the capsids of about half the virus families can be roughly approximated to hollow spheres with a relatively thin crust made of protein tiles. Yet, determination of the spring constant k of the

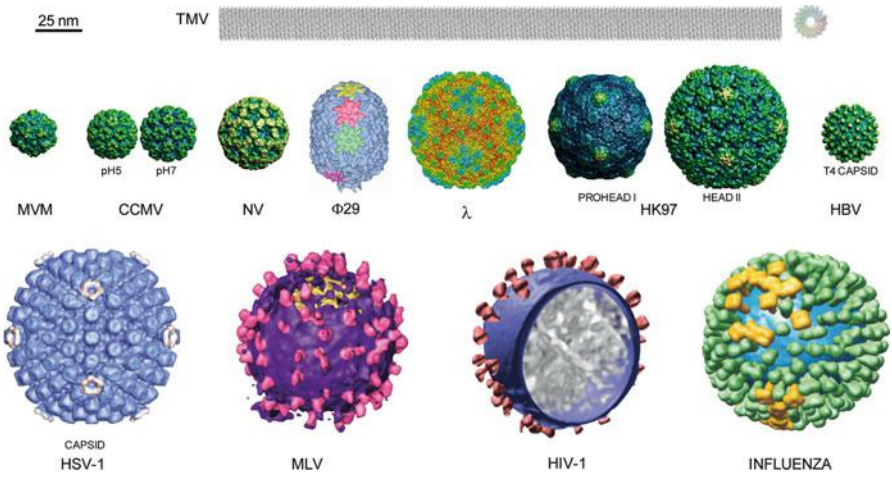

Fig. 18.6 Virus particles whose mechanical properties have been studied to date (see Sects. 18.5 and 18.6). All particles are reproduced at (approximately) the same scale, indicated by the bar at *top left*. Side and top views of TMV, native and swollen forms of CCMV, and prohead-I and head-II forms of HK97 are represented. Fibers in ϕ29 and tails in ϕ29, λ and HK97 virions are not shown. For HBV (T = 4 form) and HSV-1, mechanical properties were analyzed for naked capsids only, and these (not the enveloped virions) are represented. This figure is reproduced from [44] with permission from the publisher. The different viral particle images were originally taken from VIPERdb [45] or from the references indicated in the original figure in [44], and are reproduced with permission from the publishers

dozen or so spherical virus capsids experimentally studied to date using AFM reveals that their mechanical stiffness can differ by close to two orders of magnitude: from 0.018 N/m for the prohead-I of bacteriophage HK97 (as low as that of a "soapy" lipidic viral envelope) to as much as 1.3 N/m for some minute virus of mice (MVM) capsid mutants. How could we explain such large differences?

A quick look at the molecular structures of spherical viruses determined by cryo-EM (Chap. 3) or X-ray crystallography (Chap. 4) reveals that virus capsids are very different from each other in overall size, thickness and the number, size, shape and geometric arrangement of the protein tiles they are made of (Chap. 2 and Fig. 18.6). A thicker hollow sphere will be more difficult to deform by pushing with the tip of the AFM than a thinner one made of the same material. A smaller hollow sphere will be more difficult to deform than a larger one having the same thickness and composition. The measured mechanical stiffness of a virus capsid, or any other solid object, as given by the spring constant, depends not only on the material it is made of, but also on its dimensions and geometry (in the widest sense of the term).

18.6.2 Differences in Intrinsic Elasticity

It may be more useful for our discussion to consider not the spring constant of the capsid, but the *intrinsic* elasticity of the capsid material, the protein crust, as given

by its Young's modulus (see Sect. 18.3 and Chap. 19). Unfortunately, as already discussed, some simplifications are needed to calculate the Young's modulus of a virus capsid. Although many icosahedral virus capsids can be approximated to thin-shelled perfect spheres or icosahedrons, an inspection of the capsid structure reveals that virus capsids are ridden with protrusions and depressions, sometimes very conspicuous. Capsid thickness is not uniform, either. However, for some virus capsids, the approaches described in Sect. 18.3 or other approaches yielded reasonably similar Young's modulus values.

The average Young's moduli obtained for different spherical virus capsids still span over one order of magnitude, from 0.14 GPa for the cowpea chlorotic mottle virus (CCMV) capsid to 1.8 GPa for the phage Φ29 prohead (see Sect. 18.5.1), and up to 2.8 GPa for some MVM capsid mutants. From these values, some of the hollow quasi-spherical protein shells of viruses could mechanically resemble soft rubber balls, while others could be compared to hard acrylic plastic globes. The actual situation is even more heterogeneous. Some viral capsids (those of phage Φ29 or some MVM mutants) are mechanically anisotropic, and discrete regions in them may show up to severalfold differences in spring constant [31, 46]. Thus, a single value for the Young's modulus may not be appropriate to describe the capsid intrinsic elasticity. In addition, for a same viral capsid in different conformational states (phage HK97; [47]) or carrying different single mutations per subunit (MVM; [48]), the Young's modulus may vary up to severalfold. Thus, the protein material the nanoscopic viral capsids are made of, contrary to the material that constitutes many macroscopic solid-state objects, is able to substantially change its intrinsic elastic properties through very minor modifications in composition (*e.g.*, a single amino acid substitution per subunit) or the spatial rearrangement of (some of) its molecular components.

18.6.3 Differences in Brittleness and Resistance to Material Fatigue

As for mechanical stiffnesss, spherical capsids differ widely in their ability to withstand high mechanical loads (forces) without being physically disrupted, or even irreversibly deformed. In general, mechanical failure (breakage) of a virus capsid appears to occur when it is deformed to a certain extent [36, 49]. Some capsids (phage Φ29 prohead, mature phage λ capsid, phage HK97 prohead-II, the MVM capsid, or the Norwalk virus (Norovirus, NV) capsid at basic pH; Fig. 18.6) were fractured or disrupted when the tip of the AFM was used to apply forces that caused not very deep deformations of the protein shell, generally not above 10–15 % of the capsid diameter. In contrast, other capsids (phage HK97 prohead-I, hepatitis B virus (HBV), NV at neutral or acidic pH, CCMV at pH = 6; Fig. 18.6) withstood forces that elicited very large deformations of the particle. Some of the latter withstood even a wall-to-wall collapse under load; once the force was

removed and the particles were allowed to relax, they generally recovered their original dimensions, with no structural damage being detected. Thus, some viral protein shells can be regarded as relatively brittle, while others are considerably more resilient under mechanical loads. Differences between viral capsids regarding material fatigue have also been noted. Like stiffness, brittleness and resistance to material fatigue of a viral capsid have been observed to depend on the conformational state of the particle. However, it must be stressed that systematic, controlled and precise estimations of the brittleness or resistance to material fatigue of virus particles are still lacking.

Stiffness, brittleness and resistance to material fatigue in viral capsids do not appear to be correlated, and may also vary in complex ways (*e.g.*, as in phage HK97; [47]). A same capsid may be quite stiff but, at the same time, relatively brittle and fail under relatively low mechanical loads causing small deformations. Another capsid may substantially change its brittleness without any variation in stiffness, and so on and so forth. This aspect also requires further investigation.

18.6.4 A Relationship Between Molecular Structure and Mechanical Properties of a Virus Capsid

As viral capsids are all made of protein, what structural differences cause the substantial differences in intrinsic elasticity (Young's modulus), brittleness and resistance to fatigue among them, and even within them? The logical interpretation advanced by several researchers is that neither capsid protein subunits, nor the lattice formed by these subunits in the viral capsid, are really a continuous, mechanically homogeneous material. The intrinsic elasticity of a viral capsid may depend on the number, distribution, directionality, type and energy of the covalent and non-covalent interactions between atomic groups and amino acid residues within a same capsid protein subunit, or between neighboring subunits. Covalent interactions include the linkages between chemical groups and residues in each polypeptide chain and, in a few capsids, disulfide bonds or other covalent crosslinks between residues. Non-covalent interactions are extremely numerous and include very different types, such as van der Waals contacts, hydrogen bonds, ionic interactions, and the entropically driven hydrophobic effect. The above molecular interpretation has been invoked to explain, for example, the pH-dependent reduction in stiffness of the CCMV capsid [21, 50]; the large increase in stiffness of the phage HK97 capsid during its initial maturation stage (prohead-I to prohead-II) [47]; or the remarkable (up to 120 %) increases in mechanical stiffness caused by a single amino acid substitution (per protein subunit) in the MVM capsid [48].

In summary, virus capsids can vary widely in intrinsic stiffness, structural strength under load and/or, probably, resistance to material fatigue. Such mechanical differences may be due to either large, or sometimes small differences in molecular structure, including number, type, strength and distribution of inter- and intrasubunit covalent and non-covalent interactions.

What specific structural differences are responsible for the observed mechanical differences between two viral capsids? If we can answer this question at the molecular (atomic) level (Sect. 18.7), the roles of the mechanical features of a viral capsid during the virus life cycle could be explored and elucidated (Sect. 18.8). It could also facilitate the rational modification of the mechanical properties of viral capsids and other protein nanoparticles for biotechnological purposes (Sect. 18.9).

18.7 Structural Determinants of the Mechanical Properties of Viruses

Despite the complex relationship between molecular structure and mechanical properties of virus capsids, theoretical and experimental approaches, and combinations of them, are starting to unveil some aspects of that relationship. However, quite a lot remains to be done in this novel area of research. Yet, the few studies to date have already provided some facts and many testable hypothesis that hold the promise of a better understanding of virus particles from the points of view of fundamental physics and structural biology.

18.7.1 Structural Determinants of Capsid Mechanics: A Theoretical Approach

A very considerable number of physics-based theoretical analyses of virus mechanics have been undertaken in the last years to propose fundamental physical explanations of different experimentally observed mechanical features of virus particles, including some of those described in this chapter (see [49, 51, 52] and Chap. 19). In addition, theoretical analyses and modeling are leading to predictions related to virus mechanical properties and their materials foundations [53], most of which have not been experimentally addressed yet. Many such theoretical analyses are based on elasticity theory using simplified models, finite element or coarse-grained molecular dynamics (MD) simulations, and are described in detail in Chap. 19.

18.7.2 Structural Determinants of Capsid Mechanics: An Experimental Approach

Structural Variations and Changes in Mechanical Behavior

An experimental structure-function relationship approach has been followed by several research groups to investigate, at different levels of resolution, the structural

foundations of mechanical features in viral capsids. In this type of studies, a difference in mechanical behavior may be traced to a structural difference between two similar capsids.

So far, most of the yet scarce virus structure-mechanics comparisons of viral particles have been made between structurally or conformationally quite different particles. For example, between different global conformational states of the CCMV and HK97 capsids [21, 47, 50]; or between the NV or the herpes simplex virus type 1 (HSV-1) capsids with or without a large protein part removed [36, 54]. In such cases, the large structural differences make it very difficult to identify with high resolution the specific molecular foundation(s) of the mechanical difference considered.

Ideally, the higher the structural similarity between the capsids compared, the better the chance to find the true molecular cause of any change in a physical property or function. Either two evolutionarily very closely related capsids can be compared or, better yet, rationally chosen single amino acid substitutions can be introduced by site-directed mutagenesis in a given viral capsid, and their effect on capsid structure and mechanics analyzed. Unfortunately, only in very few cases this high-resolution structure-function approach has been undertaken so far to study the molecular basis of virus mechanics.

In a study on CCMV, a single amino acid substitution (K42R) at the N-terminal arm of the capsid protein was shown to introduce 660 new intersubunit interactions in the whole capsid; the new interactions were invoked to explain both the higher mechanical stiffness compared to the non-mutated capsid, and the increase in stability against capsid dissociation when high salt concentrations are present [55].

In another study, the individual replacement of several amino acid substitutions in the MVM capsid by alanine, which involved removal of a single chemical group per subunit, led to a drastic stiffening of the viral capsid [48]. The group removed was involved in very few noncovalent interactions within each capsid subunit, or between pairs of subunits, and the mechanical effect was tentatively attributed to a conformational rearrangement in a metastable capsid.

In a third study, the decreased mechanical stiffness of the NV capsid at acidic pH was attributed to deprotonation of a specific lysine residue that could introduce an electrostatic repulsion between certain aspartate residues, triggering a conformational rearrangement [56]. This prediction could be tested by site-directed mutagenesis of the residues putatively involved in the pH-dependent effect on mechanics.

Additional experimental structure-function studies, as well as all-atom MD simulations of whole viral capsids once the latter becomes more feasible (see Chap. 19), may help to ascertain at high resolution the structural determinants of the mechanical behavior of viral capsids, and provide insights into the relationship between mechanical, thermal and chemical stability in a capsid or any other nanoparticle.

Molecular Interactions and Capsid Fracture

As described in previous paragraphs, virus capsids can be mechanically irreversibly modified, even fractured, under a high enough and/or repeated load exerted, for example, by the AFM tip (Fig. 18.7). Such fractures are observed in

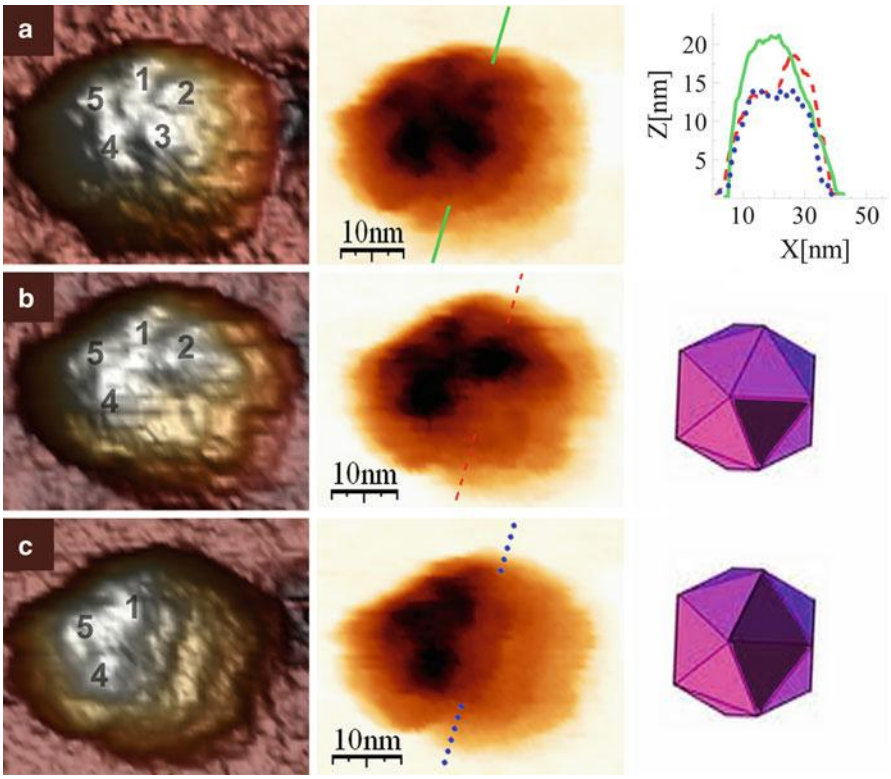

Fig. 18.7 Mechanical disassembly of a simple virus. AFM images of a same individual MVM capsid before indentation (**a**), after removal of one capsid building block (a trimer of capsid proteins) (**b**) and after removal of an additional trimer (**c**) *Left*, 3D images; *center*, negative images. *Top right*, height profiles of the intact capsid (*green line*), of the capsid after losing the first trimer (*red dashed line*), and of the capsid after losing an additional trimer (*blue dotted line*). *Bottom right*, schemes of the MVM capsid after losing one trimer (*top image*) or two trimers (*bottom image*) (Reproduced from [57]. With permission of the publisher)

nanoindentation experiments as nonlinearities in the FZ curves obtained during the indentation. In some cases (*e.g.*, NV capsid) these fractures may occur but the capsid recovers ("self-heal") after the load is removed [56]. In other cases, more or less permanent fractures may occur. Visualization of these fractures in different viral capsids [Φ29, HSV-1, MVM (Fig. 18.7), adenovirus] by AFM indicates that they may occur along the borders of capsid subunits or assemblies of capsid subunits that may act as intermediates of capsid assembly and/or disassembly [24, 36, 57]. In some cases, the dissipated energy during those fractures was estimated and found roughly comparable to the estimated association energy between capsid subunits [58].

The results obtained so far are scarce, but they suggest that mechanical failure (fracture) in virus capsids could generally start by the disruption of many (or most) noncovalent interactions at the energetically weakest interfaces between capsid subunits or assembly/disassembly intermediates, leaving the energetically more stable

building blocks intact. The intersubunit interactions thus disrupted may re-form relatively fast, leading to self-healing. Alternatively, further interfaces may be disrupted. If all of the interfaces between a single capsid building block and its neighboring blocks are eventually disrupted, the building block will be released (as observed with MVM (Fig. 18.7) and adenovirus). If additional interfaces are disrupted by a continuous or repeated high load, the capsid will collapse, also as repeatedly observed.

Loss of capsid subunits (Fig. 18.7) and self-healing of viral capsids by restoration of disrupted interfaces, have been predicted by thermodynamic-kinetic models [59] and coarse-grained MD simulations [60] (see Chap. 19) of reversible capsid (dis) assembly in solvent. According to these models and simulations, capsids that lose one capsomer under mechanical load or chemical intervention tend to reassociate with the released capsomers, acting as kinetic traps that prevent further dissociation of the capsid. Thus, thermodynamically unstable capsids that would dissociate at low particle concentrations, as found in the extracellular stage of the virus life cycle) are stabilized, and can preserve its physical integrity ([40, 43] and Chap. 19).

18.7.3 Modulation of Virus Particle Mechanics by Components Different from the Capsid

In the preceding paragraphs, only the mechanical properties of the protein capsids of viruses have been considered. What about the effect of other viral particle components on the mechanical behavior of the complete virion?

The Viral Nucleic Acid

As described above, the mechanical behavior of some viral capsids can be substantially modified if they contain a viral nucleic acid packaged inside, as in complete virions. In MVM, CCMV, and phages λ and ϕ29 (Sect. 18.5.2) the nucleic acid stiffens the capsid [30, 34, 46, 55, 61, 62], but in HSV-1 no mechanical stiffening was detected [36]. In many spherical virions, the hydrated nucleic acid occupies a very large fraction of the capsid interior, not infrequently reaching crystalline densities. Under such conditions, the molecular basis for the observed nucleic-acid mediated increase in rigidity of several virus capsids could, in principle, include steric effects, electrostatic repulsions between the phosphates, the high resistance to bending (given by the persistence length) of the nucleic acid molecule if double-stranded, hydration forces, etc. (see Chap. 19). Interestingly, the studies carried out reveal that the molecular determinants of the nucleic-acid mediated mechanical stiffening effects in different viruses can be quite different.

The high internal pressure (about 40–60 atm) exerted by the tightly packaged dsDNA inside phage Φ29 contributes to increase the stiffness of the particle, not unlike the stiffness of a football is increased by the air compressed inside ([34], see Sect. 18.5.2 for a description). Wild-type phage λ also showed a dsDNA-mediated increased stiffness. However, two mutant λ virions that contained somewhat

shortened dsDNA molecules, which would still exert a substantial internal pressure, showed no higher stiffness than the empty particle [52, 61]. Based on their results and calculations, Ivanovska et al. [61] proposed that, because of DNA-hydrating water molecules, an osmotic pressure exists in the λ capsid that increases rapidly when the DNA density approaches that in the wild-type phage, and it is this osmotic pressure what increases the stiffness of the particle.

The single-stranded (ss) DNA in MVM is also packaged in the capsid (Chap. 10) to densities that approach those found in dsDNA phages. However, the drastic ssDNA-mediated stiffening of MVM [30] is due not to internal pressure or hydration forces, but to a buttressing effect of short DNA segments that bind a number of equivalent sites at the capsid inner wall [46].

Unlike the above viruses, the ssRNA of CCMV is not packaged in a preformed capsid, but both are assembled together (Chap. 12), and the CCMV virion may not be pressurized. However, the ssRNA inside does increase the mechanical stiffness of the CCMV particle [55] by a yet unknown mechanism.

The Viral Envelope

Some mechanical features of three complete enveloped virions have been investigated. Again, remarkable differences were encountered. The influenza virion (Fig. 18.6) is extremely elastic, mechanically behaving almost like its very soft lipidic envelope [63, 64]. In contrast, the immature virions of two retroviruses, murine leukemia virus (MLV) and human immunodeficiency virus (HIV-1) (Fig. 18.6), are mechanically very stiff [65, 66], yielding estimated Young's moduli comparable to those of icosahedral capsids.

The available evidence suggests that the large stiffness differences between those viruses are not due to the lipid envelope, but depend on the structural organization of the protein layers contained within. In the influenza virion, the matrix M1 protein layer immediately below the envelope may be highly flexible and will not oppose any significant resistance to deformation. In contrast, the very thick Gag capsid below the envelope of retroviruses would be partly responsible for the high spring constants determined for the immature MLV and HIV-1 virions. In addition, the even higher stiffness of the immature HIV-1 compared to MLV has been traced to the Env protein embedded in the envelope, perhaps due to the presence of Env-Gag interactions that may not occur in the immature MLV, although this possibility remains to be further investigated.

The remarkable decrease observed in MLV and HIV-1 stiffness during virion maturation has been attributed, at least in part, to the disassembly of the thick immature Gag capsid, leaving a much thinner matrix (MA) protein layer under the envelope. The reassembled CA protein capsid has a conical shape and is much smaller than the virion; thus, the latter could be mechanically deformed without reaching contact with the loose, smaller capsid contained inside. In the HIV-1 case, maturation would additionally disrupt the proposed Env-Gag interaction, contributing to the extreme mechanical softening observed in this case [65, 66]. It may be noticed here that the three infectious enveloped viruses analyzed are mechanically very flexible, and this may not be a coincidence (see Sect. 18.8).

18.8 Mechanical Properties and Virus Biology

The solid-state physicist and the structural biologist/chemist may appreciate the value of the emerging results on the mechanical properties of viruses and their structural determinants for a better understanding of "soft" condensed matter and self-organizing molecular assemblies. However, the virologist may question whether the specific mechanical characteristics of virus particles are merely a consequence of their being solid objects, or are shaped by evolution because they confer some selective advantage to the virus. This question is still open, but growing evidence indicates that virus mechanical features may be biologically critical in more than one way.

If the mechanical features of viruses have been shaped by evolution, what are the selective pressures behind such evolution? Are viruses subjected to mechanical stress that may select for mechanically more robust variants? Do virus particles require a certain mechanical elasticity to complete some step in their life cycle? Indeed, virions are subjected to physical forces that may deform or break them either outside or inside cells. Such forces include shear forces while drifting in viscous fluids, or being extruded through nuclear pores or other openings; osmotic effects; pressurization due to packaged dsDNA; hydrostatic pressure; capillary forces on desiccation, etc. Viruses with a tubular, very elongated shape, such as tobacco mosaic virus (Fig. 18.6) may be particularly sensitive to mechanical stress, but any virus, because it is made of "soft" material, may seriously risk mechanical disruption at some point or other in its existence. Moreover, some virus particles must change its shape (undergo conformational rearrangements) to fulfill its biological function (see Chaps. 1, 13 and 15); thus, their ability to deform or withstand deformation, as determined in mechanical measurements, may be connected with their biologically relevant conformational dynamics.

We have stressed in previous sections the remarkable differences in mechanical features when different virus particles, or a same particle in different stages of the infection cycle, are compared. It seems reasonable to propose that viruses may have evolved different mechanical features in response to the selection pressures of the different forces acting on them, to withstand those forces, or even for using them to their own advantage. In this section we provide three case studies that strongly support different biological roles for different mechanical features found in virus particles.

18.8.1 Mechanical Stiffening to Withstand Pressurization

At least in some dsDNA viruses, including phages Φ29, λ and HK97, and the herpesvirus HSV-1 (Fig. 18.6), the genome is actively packaged inside a preformed capsid that is barely large enough to contain the hydrated full-length nucleic acid molecule. As a consequence of the very high packing density, the dsDNA inside exerts a very high internal pressure that tends to disrupt the capsid. Several authors

have noticed that the intrinsic stiffness of these viruses is relatively high (Young's modulus 1.0 GPa or higher). In contrast, in several (albeit not all) viruses that enclose their viral ssRNA (much less stiff than dsDNA) in a condensation process (Chap. 12), such as HBV, CCMV, influenza virus, and immature MLV, the Young's modulus is low (0.37 GPa or lower). One could ask why the pressurized dsDNA virus have not evolved to decrease their high internal pressure by slightly shortening its DNA molecule. In fact, pressurization in the dsDNA tailed phages (at least) has a biological function by driving injection of the viral DNA into the host bacterium [52, 67]; see Sect. 18.5.2 and Chaps. 17 and 19. Pressurization in HSV-1 may be likewise required for the translocation of its DNA into the cell nucleus through a nuclear pore [67].

The available evidence has led to the proposal that dsDNA phages and other viruses may have evolved a capsid of relatively high intrinsic stiffness as a biological adaptation to withstand the high internal pressure required for DNA injection [31, 49, 52, 61, 67]. During maturation (Chap. 13), the capsid of these phages expands and, as a consequence, is made thinner. Instead of making the capsid thicker, these phages may have solved the problem of mechanical weakening during maturation by making the capsid of a particularly strong material. This high intrinsic mechanical resistance may be due to strong interactions between capsid subunits. In addition and in agreement with this view, it has been observed that, during maturation, expansion and thinning is accompanied by the introduction of external reinforcements, like a chainmail of protein crosslinks (HK97) or cementing proteins (λ) [47], or even by prestressing the capsid (Φ29) [31]. The results suggest that successful maturation of at least some dsDNA viruses may critically depend on the mechanical features built-in in their capsids, and on the appropriate variations in mechanical resistance through the controlled introduction of certain structural modifications.

18.8.2 Mechanical Softening to Allow Entry into Cells

Immature MLV and HIV-1 virions enter host cells very inefficiently compared with the mature, infectious virions. Interestingly, maturation makes these viruses much less rigid (as determined by their spring constant) [65, 66]. In the case of HIV-1, it was also observed that a specific structural modification introduced in the laboratory, the truncation of the carboxy-terminus of the Env protein promoted the efficient entry of immature HIV-1 virions into cells and reduced the extreme stiffness of the immature virion to a value that approached that of the mature virion [66]. Based on those results, it has been hypothesized that effective fusion with the cell membrane to allow virus entry (see Chap. 16) may require a mechanically soft, flexible enough virion for extensive virus-cell contact. The structural changes effected during virus maturation, specifically a probable loss of Env-Gag interactions, would have as one consequence the mechanical softening of the virion, that would be required for virus entry. The high mechanical elasticity of a very different enveloped virus, influenza virus, is also consistent with the above

hypothesis. If this hypothesis is verified, changes in mechanical features of enveloped virus could effect a mechanical control of entry of enveloped viruses into host cells.

NV, a nonenveloped virus, must withstand the acidic conditions encountered during passage through the stomach, but must be able to dissociate to release its genome at weakly basic conditions during infection within the ileum. Remarkably, the NV capsid was found to be stiffer at neutral and acidic pH and less rigid at alkaline pH [56]. The authors suggested that the pH-dependent mechanical softening of the NV capsid could result from a putative, pH-triggered conformational rearrangement needed for infection, or even constitute a direct requirement to facilitate virus entry and RNA release. If this hypothesis is correct, mechanical features could have a biologically critical role also during the entry of some nonenveloped viruses into host cells (see Chap. 15).

18.8.3 Balancing Virus Stiffness to Prevent Inactivation Without Impairing Infection

In the MVM virion, segments of the viral ssDNA bound to equivalent sites at the capsid inner wall act like molecular buttresses that increase the mechanical stiffness of most regions in the virus particle. However, the regions around pores located at the fivefold symmetry (S5) axes of the icosahedral capsid are free from bound DNA, and remain mechanically as flexible (soft) in the DNA-filled virion as in the empty (DNA-free) capsid [30, 46]. Is this anisotropic distribution of mechanical stiffness in the virion due to mere chance, or has it evolved because it provides some biological advantage? The available evidence (summarized next) supports a biological role for the anisotropic stiffness of the MVM virion.

Previous studies had revealed that the capsid-bound DNA segments do provide the MVM virion with the biological advantage of increased resistance against thermal inactivation, which appears to be mediated by a conformational change in the virion; it was suggested that the interactions between the capsid and DNA segments may impair this conformational change [68]. In a series of different studies it was found that the MVM capsid regions around the S5 pores are structurally dynamic, as they participate in biologically relevant conformational rearrangements. These involve opening and closing of the capsid pores for translocation of peptide segments carrying molecular signals, and also the encapsidation (Chap. 10) and uncoating (Chap. 15) of the viral DNA. Mutations of residues at the base of the pores in the MVM capsid impair the biologically relevant conformational dynamics of the S5 (pore) regions and dramatically reduce virus infectivity at physiological temperature [69]; likewise, mutations of residues lining the S5 pores of adeno-associated virus type 2, a related virus, also impair local dynamics.

Those observations led to the hypothesis that in a virus capsid, high mechanical flexibility, or softness (as determined by AFM) and high conformational dynamism

(identified by other techniques like limited proteolysis/mass spectrometry or intrinsic fluorescence, see Chap. 6) are two descriptions of a same molecular feature, *structural flexibility*: If a region of a viral particle can be varied from its minimal free energy conformation without requiring too much energy, a structural biologist may detect a conformationally dynamic region prone to undergo some structural rearrangement (see Chap. 1); while a condensed matter physicist using AFM may detect a mechanically soft region with a relatively low spring constant. The phenomenon would be the same, but it would be detected in different ways by probing different properties. If this hypothesis is correct, regions of higher mechanical flexibility (softness) in a virus particle could match conformationally more dynamic regions.

Recently, the above prediction was verified by independently testing the effects of many single amino acid substitutions on capsid stiffness, the impairment of specific conformational rearrangements (as a measure of capsid dynamics) and virus infectivity. As predicted, mutations at the base of the capsid pores that impaired the local conformational dynamics invariably increased the mechanical stiffness of the regions around the pores [48]. Also, the two mutations tested so far that facilitated the heat-induced non-productive conformational rearrangement of the virion decreased the stiffness of most regions in the virion (but not the regions around the pores), nearly abolishing the DNA-mediated rigidification of the virion [46]. Mutations that had no effect on conformational dynamics had no effect on stiffness, and further mutations that restored a lost conformational transition also restored the lost mechanical softness [48], again as predicted by the hypothesis that high conformational dynamism and high mechanical flexibility in virus particles may be phenomenologically linked.

The above results provide strong experimental support for a biological role of the anisotropic distribution of mechanical elasticity in the MVM virion. MVM may have evolved DNA-binding sites in its capsid because the DNA-mediated structural rigidification of most capsid regions increases virion thermostability (by impairing a virion-inactivating conformational rearrangement). In contrast, the regions around the capsid pores may have been kept free of bound DNA by negative selection, because a rigidification of these regions would impair the local S5 dynamics required for infectivity. In short, the anisotropic mechanical stiffness in the MVM virion reflects the unequal distribution of structural flexibility/conformational dynamism in the viral particle, and may constitute a biological adaptation that impairs virion inactivation without impairing infection.

18.9 Engineering Mechanical Properties of Virus Particles

There is a rapidly growing interest in the use of viral particles for different biotechnological or nanotechnological applications (see Chaps. 21 and 22; [70]). However, it is envisaged that, for quite a few of such applications, natural viral particles may be not physically resistant enough. Potentially useful nanoparticles may be subjected to high

mechanical stress during purification, fabrication, storage and/or use. Such conditions are very demanding for "soft" biological materials and may lead to the disruption of natural viral particles. In addition, many natural viral particles are metastable and particularly sensitive to conformational rearrangements (see Chaps. 1 and 13; [71]). Conformational instability of a virus particle may be desirable for some bio/nanotechnological applications. For example, virus particles for gene therapy or targeted drug delivery may have to be capable of dynamic rearrangements required to deliver their cargo into cells. Even in those cases, improved mechanical stability may be highly desirable to withstand physical stresses during production and purification. On the other hand, structural rearrangements may lead to functional inactivation. For many potential applications including contrast agents, diagnostic agents, vaccines, building blocks for nanomaterials, etc., viral particles may not need to be conformationally dynamic. In such cases, the highest possible mechanical stiffness to withstand permanent deformations, and the highest structural strength against mechanical disruption, as well as thermal and chemical stability, may be critical.

The physical, chemical and/or genetic manipulation of virus particles, generally carried out to investigate the structural basis or the biological relevance of specific mechanical features, have led as a by-product to the mechanical destabilization of the manipulated capsids. More elastic and/or more brittle particles have been obtained by truncating a several-domain protein [54, 66]; eliminating some capsomers [36, 57]; eliminating or reducing the length of the nucleic acid molecule [61]; neutralizing the nucleic acid negative charge [62]; or eliminating capsid-nucleic acid interactions [46].

However, mechanical stabilization may be much more useful than destabilization when considering viral particles for most nanotechnological applications. Research on the structural elements that underlie the mechanical features of natural viral particles may provide useful guidelines for engineering mechanically more robust viral nanoparticles. For example, it was found that during maturation of phage λ, protein gpD attaches to the capsid, acting as a cement that stabilizes the particle [72]. In phage HK97, a chainmail of covalent crosslinks between capsid subunits contribute to increase the structural strength and resistance to material fatigue and physically stabilize the particle [47]. In CCMV, a single amino acid substitution biologically fixed in a variant of CCMV conferred resistance of the capsid to high salt concentrations and also increased its stiffness [55].

A rational approach based on biophysical and biological evidence has been successfully used to increase the mechanical stiffness of a viral capsid by protein engineering. Some amino acid substitutions in the MVM capsid that impaired conformational dynamics were predicted to increase mechanical stiffness (see Sect. 18.8.3). In addition, substitutions of amino acids surrounding conspicuous cavities in the MVM capsid were also designed to modify the size and shape of such cavities, and some of them were predicted to increase capsid stiffness. When tested, many of these mutations did increase, as predicted, the capsid mechanical stiffness (up to close to threefold higher spring constant and Young's modulus, compared with the natural empty capsid). Some of these engineered mutant capsids were, by far, stiffer (1.3 N/m, 2.8 GPa) than any other naked capsid analyzed to date [48]. These studies provide the only examples so far of the purposeful mechanical stabilization of

a viral particle by protein engineering in the laboratory. They constitute proof-of-principle that viral nanoparticles can be manipulated by protein engineering to adequately modify their mechanical characteristics, with the aim of improving their usefulness in bio/nanotechnology.

18.10 Perspectives and Conclusions

Fifteen years have passed since a pioneering study on virus mechanics in air using the rod-like tobacco mosaic virus as a model was published [73]; the first study on the mechanical properties of small spherical viruses in liquid appeared only 8 years ago [6]. In this short time, several outstanding mechanical features of virus particles have been discovered, including the high elasticity, resilience and resistance to material fatigue of some (albeit not all) virus capsids; the mechanical anisotropy of virus particles; the nucleic acid-mediated mechanical buttressing in some virions, and the different structural foundations of such effects; the drastic effects on capsid elasticity of single amino acid substitutions, capsid subunits or components; and the existence of built-in stress.

The results obtained on virus mechanics are relevant in materials science, structural and physical virology and bio/nanotechnology. From the point of view of materials science, "soft" biological materials such as protein layers are being mechanically characterized. However, physical interpretations of some results may be not without caveats. Mechanical characterization includes the estimation of some volumetric physical magnitudes such as the Young's modulus, based on homogenous models of continuum mechanics. However, does it make any sense to define a Young's modulus for proteins based on a continuum approach? The relative Young's modulus can be useful, for example as a parameter for mechanical comparisons, but proteins are not a homogeneous material, and virus capsids themselves have very complex shapes, and are made of discrete subunits connected by different intersubunit interactions. Can mechanical experiment extract information about capsid subunits and their reciprocal interactions? Further experiments are required to gain insights into the relationship between the discontinuous nature of virus capsids and their mechanical properties. Mechanical parameters extracted from the experiments, such as the spring constant, are appropriate for mechanical comparisons between virus particle mutants, different regions in a same capsid, capsids in different maturation stages, and empty capsids *vs.* nucleic acid-filled virions. Fatigue material in virus particles has been little explored so far; experiments aimed at the mechanical disassembly of viral particles may provide information in this respect. Controlled mechanical dissociation of virus particles may also reveal assembly/disassembly intermediates; however, these experiments should be carefully controlled to allow the characterization of the intermediates thus produced.

From the point of view of virus biology, growing experimental evidence is providing support for a functional relevance of the mechanical properties of viruses. Mechanical stiffening achieved through the establishment of capsid-nucleic acid interactions, built-in stress, specific noncovalent interactions or covalent crosslinks,

for example, may help virus particles to withstand mechanical forces acting on them in the extracellular environment or, in the case of dsDNA phages, as a consequence of the internal pressure exerted by the packaged nucleic acid. Changes in mechanical properties of retroviral virions during maturation could help entry into cells by facilitating virus-cell membrane fusion. The DNA-mediated anisotropic mechanical stiffness of the MVM virion reflects the unequal distribution of structural flexibility and conformational dynamism in the virus particle, which may have arisen as a biological adaptation to better resist inactivation without impairing infectivity.

From the point of view of bio/nanotechnology, studies on the molecular determinants of the mechanical stability of viral particles at single amino acid-resolution may provide insights for the engineering of virus-derived nanoparticles with increased physical resistance to undesirable structural deformations or disruption.

Acknowledgements We gratefully acknowledge J.L. Carrascosa, J. Gómez-Herrero, D. Reguera, C. San Martín and N. Verdaguer for collaboration on virus mechanics, advice and fruitful discussions; C. Carrasco, P.J.P. Carrillo, M. Castellanos, M. Hernando-Pérez, A. Llauró, A. Ortega-Esteban and R. Pérez for excellent work on virus mechanics; M.A. Fuertes and A. Rodríguez-Huete for great experimental assistance. M.A.F. kindly provided Fig. 18.6. Work in P.J.P.'s laboratory is funded by the Spanish Government (grants PIB2010US-00233 and FIS2011-29493). Work in M.G.M.'s laboratory is funded by the Spanish Government (grants BIO2009-10092 and BIO2012-37649) and Comunidad de Madrid (S-2009/MAT/1467), and by an institutional grant from Fundación Ramón Areces. M.G.M. is an associate member of the Institute for Biocomputation and Physics of Complex Systems, Zaragoza, Spain.

References and Further Reading

1. Shiroguchi K, Kinosita K (2007) Myosin V walks by lever action and brownian motion. Science 316:1208–1212
2. Moreno-Herrero F, de Jager M, Dekker NH, Kanaar R, Wyman C, Dekker C (2005) Mesoscale conformational changes in the DNA-repair complex Rad50/Mre11/Nbs1 upon binding DNA. Nature 437:440–443
3. Moody MF (1999) Geometry of phage head construction. J Mol Biol 293:401–433
4. Cordova A, Deserno M, Gelbart WM, Ben-Shaul A (2003) Osmotic shock and the strength of viral capsids. Biophys J 85:70–74
5. Binnig G, Quate CF, Gerber C (1986) Atomic force microscope. Phys Rev Lett 56:930–933
6. Ivanovska IL, de Pablo PJ, Ibarra B, Sgalari G, MacKintosh FC, Carrascosa JL, Schmidt CF, Wuite GJL (2004) Bacteriophage capsids: though nanoshells with complex elastic properties. Proc Natl Acad Sci USA 101:7600–7605
7. Gittes F, Schmidt CF (1998) Thermal noise limitations on micromechanical experiments. Eur Biophys J Biophys Lett 27:75–81
8. Rief M, Oesterhelt F, Heymann B, Gaub HE (1997) Single molecule force spectroscopy on polysaccharides by atomic force microscopy. Science 275:1295–1297
9. Fisher TE, Carrión-Vázquez M, Oberhauser AF, Li H, Marszalek PE, Fernández JM (2000) Single molecule force spectroscopy of modular proteins in the nervous system. Neuron 27:435–446
10. Svoboda K, Schmidt CF, Schnapp BJ, Block SM (1993) Direct observation of kinesin stepping by optical trapping interferometry. Nature 365:721–727

11. Alcaraz J, Buscemi L, Grabulosa M, Trepat X, Fabry B, Farre R, Navajas D (2003) Microrheology of human lung epithelial cells measured by atomic force microscopy. Biophys J 84:2071–2079
12. de Pablo PJ, Schaap IAT, MacKintosh FC, Schmidt CF (2003) Deformation and collapse of microtubules on the nanometer scale. Phys Rev Lett 91:98101
13. Smith DE, Tans SJ, Smith SB, Grimes S, Anderson DL, Bustamante C (2001) The bacteriophage ϕ29 portal motor can package DNA against a large internal force. Nature 413:748–752
14. Ortega-Esteban A, Horcas I, Hernando-Pérez M, Ares P, Pérez-Berná AJ, San Martín C, Carrascosa JL, de Pablo PJ, Gómez-Herrero J (2012) Minimizing tip-sample forces in jumping mode atomic force microscopy in liquid. Ultramicroscopy 114:56–61
15. Horcas I, Fernandez R, Gomez-Rodriguez JM, Colchero J, Gomez-Herrero J, Baro AM (2007) WSXMa software for scanning probe microscopy and a tool for nanotechnology. Rev Sci Instrum 78:013705
16. Sader JE, Chon JWM, Mulvaney P (1999) Calibration of rectangular atomic force microscope cantilevers. Rev Sci Instrum 70:3967–3969
17. de Pablo PJ, Gomez-Navarro C, Martinez MT, Benito AM, Maser WK, Colchero J, Gomez-Herrero J, Baro AM (2002) Performing current versus voltage measurements of single-walled carbon nanotubes using scanning force microscopy. Appl Phys Lett 80:1462–1464
18. Landau LD, Lifshizt E (1986) Theory of elasticity. Pergamon, London
19. Basak S, Raman A (2007) Dynamics of tapping mode atomic force microscopy in liquids: theory and experiments. Appl Phys Lett 91:064107, 064107-3
20. Israelachvili J (2002) Intermolecular and surface forces. Academic Press, London
21. Klug WS, Bruinsma RF, Michel JP, Knobler CM, Ivanovska IL, Schmidt CF, Wuite GJL (2006) Failure of viral shells. Phys Rev Lett 97:228101
22. Vliegenthart GA, Gompper G (2006) Mechanical deformation of spherical viruses with icosahedral symmetry. Biophys J 91:834–841
23. Zink M, Grubmuller H (2009) Mechanical properties of the icosahedral shell of southern bean mosaic virusa molecular dynamics study. Biophys J 96:1350–1363
24. Ivanovska IL, Miranda R, Carrascosa JL, Wuite GJL, Schmidt CF (2011) Discrete fracture patterns of virus shells reveal mechanical building blocks. Proc Natl Acad Sci USA 108:12611–12616
25. Sokolnikoff IS (1983) Mathematical theory of elasticity. Kriger Pub Co, Florida
26. Greaves GN, Greer AL, Lakes RS, Rouxel T (2011) Poisson's ratio and modern materials. Nat Mater 10:823–837
27. Helfer E, Harlepp S, Bourdieu L, Robert J, MacKintosh FC, Chatenay D (2001) Buckling of actin-coated membranes under application of a local force. Phys Rev Lett 87:088103
28. Niordson FI (1985) Shell theory. North Holland, Amsterdam
29. Hutton DV (2004) Fundamentals on finite element analysis. McGraw Hill, New York
30. Carrasco C, Carreira A, Schaap IAT, Serena PA, Gomez-Herrero J, Mateu MG, de Pablo PJ (2006) DNA-mediated anisotropic mechanical reinforcement of a virus. Proc Natl Acad Sci USA 103:13706–13711
31. Carrasco C, Luque A, Hernando-Pérez M, Miranda R, Carrascosa JL, Serena PA, de Ridder M, Raman A, Gómez-Herrero J, Schaap IAT, Reguera D, de Pablo PJ (2011) Built-in mechanical stress in viral shells. Biophys J 100:1100–1108
32. Schijve J (2009) Fatigue of structures and materials. Kluwer Academic Publishers, Dordrecht
33. Ortega-Esteban A, Pérez-Berná AJ, Menéndez-Conejero R, Flint SJ, San Martín C, de Pablo PJ (2013) Monitoring dynamics of human adenovirus disassembly induced by mechanical fatigue. Sci Rep 3:1434
34. Hernando-Pérez M, Miranda R, Aznar M, Carrascosa JL, Schaap IAT, Reguera D, de Pablo PJ (2012) Direct measurement of phage ϕ29 stiffness provides evidence of internal pressure. Small 8:2366–2370
35. Tao YZ, Olson NH, Xu W, Anderson DL, Rossmann MG, Baker TS (1998) Assembly of a tailed bacterial virus and its genome release studied in three dimensions. Cell 95:431–437

36. Roos WH, Radtke K, Kniesmeijer E, Geertsema H, Sodeik B, Wuite GJL (2009) Scaffold expulsion and genome packaging trigger stabilization of herpes simplex virus capsids. Proc Natl Acad Sci USA 106:9673–9678
37. Gonzalez-Huici V, Salas M, Hermoso JM (2004) The push-pull mechanism of bacteriophage ϕ29 DNA injection. Mol Microbiol 52:529–540
38. Podgornik R, Leforestier A, Siber A, Livolant F (2011) Protein-DNA interactions determine the shapes of DNA toroids condensed in virus capsids. Biophys J 100:2209–2216
39. Purohit PK, Kondev J, Phillips R (2003) Mechanics of DNA packaging in viruses. Proc Natl Acad Sci USA 100:3173–3178
40. Zandi R, van der Schoot P, Reguera D, Kegel W, Reiss H (2006) Classical nucleation theory of virus capsids. Biophys J 90:1939–1948
41. Choi KH, Morais MC, Anderson DL, Rossman MG (2006) Determinants of bacteriophage ϕ29 head morphology. Structure 14:1723–1727
42. Dokland T (1999) Scaffolding proteins and their role in viral assembly. Cell Mol Life Sci 56:580–603
43. Zlotnick A (2003) Are weak protein-protein interactions the general rule in capsid assembly? Virology 315:269–274
44. Mateu MG (2012) Mechanical properties of viruses analyzed by atomic force microscopy: a virological perspective. Virus Res 168:1–22
45. Carrillo-Tripp M, Sheperd CM, Borelli IA, Sangita V, Lander G, Natarajan P, Johnson JE, Brooks CL III, Reddy VS (2009) VIPERdb2an enhanced and web API enabled relational database for structural virology. Nucleic Acid Res 37:D436–D442
46. Carrasco C, Castellanos M, de Pablo PJ, Mateu MG (2008) Manipulation of the mechanical properties of a virus by protein engineering. Proc Natl Acad Sci USA 105:4150–4155
47. Roos WH, Gertsman I, May ER, Brooks CL III, Johnson JE, Wuite GJL (2012) Mechanics of bacteriophage maturation. Proc Natl Acad Sci USA 109:2342–2347
48. Castellanos M, Pérez R, Carrasco C, Hernando-Pérez M, Gómez-Herrero J, de Pablo PJ, Mateu MG (2012) Mechanical elasticity as a physical signature of conformational dynamics in a virus particle. Proc Natl Acad Sci USA 109:12028–12033
49. Roos WH, Ivanovska IL, Evilevitch A, Wuite GJL (2007) Viral capsids: mechanical characteristics, genome packaging and delivery mechanisms. Cell Mol Life Sci 64:1484–1497
50. Wilts BD, Schaap IAT, Young M, Douglas T, Knobler CM, Schmidt CF (2010) Swelling and softening of the CCMV plant virus capsid in response to pH shifts. Biophys J 98:656a
51. Roos WH, Wuite GJL (2009) Nanoindentation studies reveal material properties of viruses. Adv Mater 21:1187–1192
52. Roos WH, Bruinsma R, Wuite GJL (2010) Physical Virology. Nature Phys 6:733–743
53. Zandi R, Reguera D (2005) Mechanical properties of viral capsids. Phys Rev E 72:021917
54. Baclayon M, Shoemaker GK, Uetrecht C, Crawford SE, Estes MK, Prasad BVV, Heck AJR, Wuite GJL, Roos WH (2011) Prestress strengthens the shell of Norwalk virus nanoparticles. Nano Lett 11:4865–4869
55. Michel JP, Ivanovska IL, Gibbons MM, Klug WS, Knobler CM, Wuite GJL, Schmidt CF (2006) Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength. Proc Natl Acad Sci USA 103:6184–6189
56. Cuellar JL, Meinhoevel F, Hoehne M, Donath E (2010) Size and mechanical stability of norovirus capsids depend on pH: a nanoindentation study. J Gen Virol 91:2449–2456
57. Castellanos M, Pérez R, Carrillo PJP, de Pablo PJ, Mateu MG (2012) Mechanical disassembly of single virus particles reveals kinetic intermediates predicted by theory. Biophys J 102:2615–2624
58. Reddy VS, Johnson JE (2005) Structure-derived insights into virus assembly. Adv Virus Res 64:45–68
59. Singh S, Zlotnick A (2003) Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly. J Biol Chem 278:18249–18255
60. Rapaport DC (2008) Role of reversibility in viral capsid growth: a paradigm for self-assembly. Phys Rev Lett 101:1861012008

61. Ivanovska I, Wuite G, Jonsson B, Evilevitch A (2007) Internal DNA pressure modifies stability of WT phage. Proc Natl Acad Sci USA 104:9603–9608
62. Evilevitch A, Roos WH, Ivanovska IL, Jeembaeva M, Jönsson B, Wuite GJL (2011) Effects of salt on internal DNA pressure and mechanical properties of phage capsids. J Mol Biol 405:18–23
63. Eghiaian F, Schaap IA, Skehel JJ, Veigel C, des Georges A (2009) The influenza virus mechanical properties are dominated by its lipid envelope. Biophys J 96:15a
64. Li S, Eghiaian F, Sieben C, Herrmann A, Schaap IA (2011) Bending and puncturing the influenza virus envelope. Biophys J 100:637–645
65. Kol N, Gladnikoff M, Barlam D, Shneck RZ, Rein A, Rousso I (2006) Mechanical properties of murine leukemia virus particles: effect of maturation. Biophys J 91:767–774
66. Kol N, Shi Y, Tsvitov M, Barlam D, Shneck RZ, Kay MS, Rousso I (2007) A stiffness switch in human immunodeficiency virus. Biophys J 92:1777–1783
67. Gelbart WM, Knobler CM (2009) Pressurized viruses. Science 323:1682–1683
68. Reguera J, Grueso E, Carreira A, Sánchez-Martínez C, Almendral JM, Mateu MG (2005) Functional relevance of amino acid residues involved in interactions with ordered nucleic acid in a spherical virus. J Biol Chem 280:17969–17977
69. Reguera J, Carreira A, Riolobos L, Almendral JM, Mateu MG (2004) Role of interfacial amino acid residues in assembly, stability, and conformation of a spherical virus capsid. Proc Natl Acad Sci USA 101:2724–2729
70. Mateu MG (2011) Virus engineering: functionalization and stabilization. Prot Eng Des Sel 24:53–63
71. Johnson JE (2003) Virus particle dynamics. Adv Protein Chem 64:197–218
72. Lander GC, Evilevitch A, Jeembaeva M, Potter CS, Carragher B, Johnson JE (2008) Bacteriophage lambda stabilization by auxiliary protein gpD: timing, location and mechanism of attachment determined by cryo-EM. Structure 16:1339–1406
73. Falvo MR, Washburn S, Superfine R, Finch M, Brooks FP Jr, Chi V, Taylor RM II (1997) Manipulation of individual viruses: friction and mechanical properties. Biophys J 72:1396–1403

Further Reading

Stockley PG, Twarock R (eds) (2010) Emerging topics in physical virology. Imperial College Press, London

Roos WH (2011) How to perform a nanoindentation experiment on a virus. In: Peterman EJG, Wuite GJL (eds) Single molecule analysismethods and protocols, vol 783, Methods Mol Biol. Humana Press, Totowa, pp 251–264

Kurland NE, Drira Z, Yadavalli VK (2012) Measurement of nanomechanical properties of biomolecules using atomic force microscopy. Micron 43:116–128

Also especially recommended for further reading are references [6, 44, 51, 52, 67] *listed above*.

Chapter 19
Theoretical Studies on Assembly, Physical Stability and Dynamics of Viruses

Antoni Luque and David Reguera

Abstract All matter has to obey the general laws of physics and living matter is not an exception. Viruses have not only learnt how to cope with them, but have managed to use them for their own survival. In this chapter we will review some of the exciting physics behind viruses and discuss simple physical models that can shed some light on different aspects of the viral life cycle and viral properties. In particular, we will focus on how the structure and shape of the capsid, its assembly and stability, and the entry and exit of viral particles and their genomes can be understood using fundamental physics theories.

Keywords Biophysics • Virus • Bacteriophage • Capsid • DNA • RNA • Budding • Statistical Mechanics • Continuum elasticity theory • Nucleation • Self-assembly • Wrapping • Pressure • Electrostatics

Abbreviations

AFM	Atomic force microscopy
BMV	Brome mosaic virus
CBB	Capsid building block
CCMV	Chlorotic cowpea mottle virus

A. Luque
Department of Fundamental Physics, Universitat de Barcelona,
c/Martí i Franquès 1, 08028 Barcelona, Spain

Department of Chemistry, New York University, 31 Washington Place,
New York, NY 10003, USA

D. Reguera (✉)
Department of Fundamental Physics, Universitat de Barcelona,
c/Martí i Franquès 1, 08028 Barcelona, Spain
e-mail: dreguera@ub.edu

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*,
Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_19,

CK	Caspar and Klug
CNT	Classical nucleation theory
cryo-EM	Cryo-electron microscopy
DH	Debye-Hückel
dsDNA	Double-stranded DNA
FEA	Finite element analysis
FvK	Föppl-von Kármán number
HBV	Hepatitis B virus
HIV	Human immunodeficiency virus
HK97	Bacteriophage Hong Kong 97
HPV	Human papillomavirus
NMA	Normal mode analysis
ssRNA	Single-stranded RNA
STMV	Satellite tobacco mosaic virus
TMV	Tobacco mosaic virus
TST	Thin shell theory

19.1 Introduction

Viruses are an endless source of fascination for biophysicists. Contrary to most biological organisms, viral particles are made of a minimal number of relatively simple components that are not capable of any metabolic activity, except when their genome takes over the metabolism of the infected host to achieve the replication of new particles (see Chap. 1). Despite the lack of sophisticated biological machinery, viruses have found the way to efficiently infect the host, assemble, and egress the cell following, in many cases, a coordinated sequence of passive and spontaneous processes. This strongly suggests that, during their life cycle, viruses must rely on general physical and chemical mechanisms to succeed in their different tasks and to achieve the required resistance against possible extreme environmental conditions.

The assembly of viruses (Chaps. 10, 11, 12 and 13) provides a good example to illustrate the importance of physical mechanisms in the virus life cycle. In physiological conditions, viral capsids assemble directly from their building blocks, which involve the participation of the capsid proteins, the genetic material, and in some cases, auxiliary proteins. Indeed, it has been shown that many viruses can be reconstituted in a fully infective form by mixing these constitutive elements *in vitro* in the right conditions, as in the paradigmatic case of tobacco mosaic virus (TMV). This means that the interactions among these elements spontaneously drive the formation of viruses in a process that can be described by basic thermodynamic and kinetic principles. Moreover, the physical properties of the viral genome, which differ depending on the type of nucleic acid, have a crucial influence on the viral strategy of assembly and infection. In particular, single-stranded RNA

(ssRNA), which is a relatively flexible macromolecule, usually co-assembles with the viral capsid proteins to form infective viruses. Instead, double-stranded DNA (dsDNA) is a more rigid polymer that many viruses pack at high densities inside a spontaneously preformed shell, the procapsid, using the most powerful molecular motors known in nature. In this way, the elastic energy and high electrostatic self-repulsion of the packed dsDNA phosphate backbones generate an internal pressure that is crucial to initiate the subsequent infection of a new host (see Chaps. 9, 12 and 17).

The wealth of biological details described in previous chapters is essential to confer specificity to viruses and make possible their success infecting a particular host. However, even viruses that infect hosts from different species or kingdoms also share some common physical mechanisms, whose study would pave the way to understand key points in the viral life cycle and to a myriad of potentially useful applications. It is worth noting that viruses have been subjected to natural selection, some of them for billions of years, perhaps since the origin of life. Thus, they constitute a pool of optimal solutions at the nanoscale of incalculable value, and their efficiency in performing all kinds of functions make them a perfect model to guide the design of a new generation of nanostructures. In fact, the impressive properties of viral capsids have already grounded the development of promising biomedical and nanotechnological applications, as described in Chaps. 20, 21 and 22.

In this chapter, we will review the physics behind some of the major different steps of the virus infection cycle, and we will introduce the main theoretical tools developed in the emerging field of physical virology. In particular, we will discuss some of the underlying physical principles and likely mechanisms involved in virus architecture, viral particle assembly and mechanical stability, nucleic acid packaging, and virus exit from the infected cell and entry into another host cell; and we will see how simple theoretical models are able to give a very useful insight into many aspects of viruses, despite not incorporating most of the structural and biological elements introduced in other chapters of this book.

More specifically, first, we will see how the exquisite regular shape and architecture of viral particles is a natural consequence of simple geometrical principles and free energy minimization. Then, we will discuss how the favorable interactions among capsid proteins lead to the spontaneous formation of capsids in a self-assembly process that can be described using the same models that explain the phase transformations of well-known non-biological systems. The presence and the physical properties of the genome play a central role in this process, hence ssRNA and dsDNA viruses will be discussed separately. Once assembled, the physical stability of viruses is crucial to ensure their infectivity, and this stability is related to the mechanical properties of the capsid. Classical elasticity theory will be the starting framework to study capsid mechanical stability and, complemented with molecular modeling, will provide essential information to understand the physical resistance of viral shells. Finally, in the last section of this chapter, we will describe some basic physical mechanisms behind genome delivery and virus egress strategies during the infection cycle.

19.2 Architecture of Viral Shells

An astonishing aspect of viruses is the well-defined shape and symmetry of their capsids. As discussed in Chap. 2, viral shells essentially adopt four different architectures: quasi-spherical, rod-like, bacilliform, and, more rarely, conical. Thus, completely unrelated viruses can form similar structures, which is truly remarkable, specially taking into account that they generally infect different hosts and are made out of capsid proteins whose amino-acid sequences, sizes, or conformations can be very diverse. This suggests that, underlying the virus-specific structural details, there must be some common physical and geometrical principles dictating the shape and architecture of capsids.

The aim of this section is to provide a brief overview of the main theoretical ideas and models that explain the architecture of viral capsids. They have been grouped in three categories. First, the geometrical concepts that justify the overall shape and symmetry of viruses are discussed in Sect. 19.2.1. These geometrical principles provide a catalogue of the possible architectures adopted by viral capsids using very precise rules of construction. However, these ideas do not suffice to explain why a specific virus prefers a particular architecture. The selection of concrete structures obeys the physical principle of free energy minimization, and several models of increasing level of complexity have been proposed to understand the details of this process. The simplest models describe viral shells as continuum elastic materials, and their description in terms of thin shell theory (TST) is introduced in Sect. 19.2.2. Nevertheless, the fundamental understanding of architectural details of capsids requires the use of discrete models, and in Sect. 19.2.3 we discuss simulations of coarse-grained and discrete biomolecular models that allow exploring more complex and realistic scenarios. The combination of continuum and simulation approaches is fundamental to establish a solid theoretical framework that can rationalize our understanding of viral structures, helping us to guide further experiments and applications.

19.2.1 Geometrical Theories

One can justify the construction of the four main morphologies of viral capsids using basic geometrical ideas. As mentioned in Chap. 2, the small size of viral capsids, on the order of tens to hundreds of nanometers, restricts the amount of information that can be coded in the viral genome contained inside. Therefore, for the sake of genetic economy, capsids are typically built from multiple copies of a single or a few different small proteins [1]. These identical subunits interact with each other building a regular structure, in a process that shares some similarities with the crystallization of molecules that form regular lattices [1, 2]. However, in the case of viruses, the capsid proteins construct a hollow shell rather than a three dimensional solid. In a planar surface, the organization allowing the highest

packing of identical units, which maximizes the number of interactions, is the hexagonal lattice or, equivalently, the triangular lattice. Now, starting from this hexagonal network it is possible to rationalize the construction of all basic capsid shapes. In particular, the open helical tubes typical of rod-like viruses can be obtained by simply wrapping up the lattice. Alternatively, to obtain a closed shell, one needs to introduce 12 pentagonal defects (see Chap. 2). If those are evenly distributed, one gets an icosahedral shell, characteristic of quasispherical viruses. In a similar way, prolate or bacilliform capsids can be obtained by wrapping the lattice into a helical tube and closing each of its ends with hemispheres containing six defects. Finally, conical viruses are obtained by making a tube with a conical section closed with a different number of defects at the ends.

The geometrical characterization of viral capsids is particularly important because there is only a limited number of ways to introduce the pentameric defects that lead to perfectly closed regular shells. For spherical viruses, the different possibilities give rise to the triangulation T-number used in the structural classification of icosahedral capsids (see Chap. 2). For bacilliform viruses, the symmetry and arrangement of the pentameric defects in the caps determine a discrete set of compatible lengths, radii, and helicities that can be adopted. Finally, the different distribution possibilities of the 12 pentameric defects in both ends of a conical shell, *e.g.*, 7 and 5, or 9 and 3, explain the discretized values of cone angles found in viruses like human immunodeficiency virus (HIV). Indeed, these simple geometrical rules provide the basis for the structural classification of spherical viruses introduced back in the 60s by Caspar and Klug (CK), which was extended for some bacilliform viruses by Moody and has been generalized recently [3] (see Chap. 2).

Graphically, these structures can be considered as the result of covering a sphere, a spherocylinder, or a conical tube using equilateral triangular tiles that are associated to the capsid subunits (see Fig. 19.1a, b). Nevertheless, certain viruses do not fit strictly into this classification. For instance, polyomavirus has a structure similar to a $T = 7$ capsid where the expected hexagonal and pentagonal positions in the icosahedral lattice are occupied by clusters of five proteins (see also Chap. 2). To tackle this puzzling situation, an alternative tiling theory[1] has been developed recently, introducing tiles of other shapes as rhombs or kites rather than the triangular tessellation of the CK theory (see Fig. 19.1c). These ideas have been successful in characterizing the structure of some spherical and tubular all-pentamer viruses [5]. More recently, the same group showed that structural details on the tertiary structure of viral proteins, and even the organization of the genomic material inside the capsid could be reproduced using a sophisticated extension of the icosahedral symmetry, which works as a geometrical blueprint [15].

Therefore, all these geometrical models are very helpful characterizing the structure of viral capsids. Nonetheless, they do not explain *why* viruses adopt those particular architectures. In this sense, several simple mathematical models

[1] Tiling theory is a set of mathematical operations which can be used to cover an arbitrary surface with geometrical shapes (or tiles) with no overlaps and no gaps.

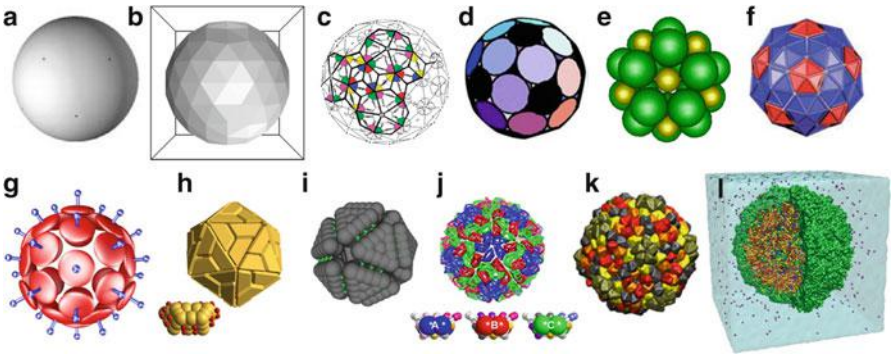

Fig. 19.1 Different levels of modelization of viral capsids. (**a**) Continuum elastic model of a spherical capsid in TST. (**b**) Representation of a shell as a triangulated network [4]. (**c**) Tiling representation of polyomavirus using kite and rhombic tiles [5]. (**d**–**g**) Coarse-grained models where capsomers and proteins are represented as (**d**) disks [6] or (**e**) balls [7] over the surface of a sphere, (**f**) hexagonal and pentagonal pyramids [8], and (**g**) ellipsoidal capsomers with additional attractive and repulsive sites [9]. (**h**–**j**) Coarse-grained models at the level of proteins described as: (**h**) multiple bead trapezoidal objects with many interaction sites (shown below) [10], (**i**) truncated pyramids [11], (**j**) trapezoidal multi-beads of three different types (shown below) [12]. (**k**) Shape-based coarse-grained model of phage ΦX174 [13]. (**l**) Full-atom model, including water and ions, of satellite tobacco mosaic virus (STMV) [14]

have tried to justify the prevalence of these structures in terms of their possible energetic or biological advantages. One of these desirable features is storage efficiency, which could be translated into maximizing the volume-to-surface ratio, *i.e.*, maximizing the storage capacity of the capsid with the minimum amount of proteins. The first ideas in this direction were based on exploring models that maximized the number of contacts, the packing, or the covering of a sphere by simple geometric objects like disks, which represented the capsomers or the structural subunits of the capsid [16]. These models were able to reproduce the symmetries of some all-pentamer viruses, but did not recover the sequence of T-numbers described by CK and adopted by most spherical viruses. Interestingly, the mathematical ideas of packing and covering are somehow connected to a more general principle in physics: the free energy minimization. This idea, already suggested by CK [2], has been explored using different continuum and discrete models as described in the next two subsections.

19.2.2 Continuum Theories

A first approximation to the energetic description of different capsid shapes can be made using the well-known continuum elasticity theory [17]. In this framework, the viral capsid is described as a continuum homogeneous material, and the focus is placed on the elastic energy required to deform the shell in order to obtain the

different capsid shapes. At first sight, this continuum description might seem very crude, since viruses are relatively small and made out of discrete elements (proteins). However, even the smallest virus contain hundreds of thousands of atoms that render this approximation not so unrealistic. Moreover, in most viruses the radius of their protein shell is substantially larger than its thickness, which means that the capsid can be approximated as a two-dimensional surface, and described using a simplified limit of the elastic theory known as thin shell theory (TST). This approach has been already used to describe the structural properties of several biological systems, like membranes and vesicles, and constitutes a very convenient starting point to understand the energetics of viral capsids.

Thin Shell Theory (TST)

In the context of TST, the elastic energy of a viral shell is decomposed into two main contributions [4]. First, one has a *stretching* term that is the energetic cost associated to elongating or compressing a planar piece of capsid material. This cost is commonly characterized by two elastic constants: the two-dimensional (2D) Young's modulus Y, which describes the stiffness of the material and is given by the ratio between the applied stress and the resulting elongation per unit length; and the Poisson ratio ν, which describes the relative contraction in the transverse directions when a material is stretched in the longitudinal direction (see Chap. 18). For many protein-like materials, this is taken to be around $\nu \sim 0.3$ or 0.4, which is similar to rubber. The second energetic contribution is the *bending* term, which accounts for the out-of-plane deformation. This is characterized by the bending constant κ, which describes the resistance to flexion, and a spontaneous curvature C_0, which is the inverse of the preferred radius of curvature of the surface. It is worth noting that C_0 is ultimately defined by the directionality of the interactions between capsid proteins and the steric effect of their surfaces in contact. Additionally, there is a contribution associated with the Gaussian rigidity, which for closed surfaces with fixed topology, like a sphere, a spherocylinder, or a cone, is a constant that is generally not taken into account.

Furthermore, as discussed in Chap. 2, spherical, bacilliform, and conical viruses require the insertion of 12 pentameric defects, or *disclinations*, in the original planar hexagonal network to form a closed shell. These disclinations are made of the same material as the rest of the capsid and generally accumulate an important degree of stress, which grows with the size of the shell and can promote the faceting of the structure, as we will see in Sect. 19.4.2. The contribution of these disclinations is implicitly captured in the TST description, although explicit expressions can be also derived. Interestingly, the total elastic energy of a shell obtained by adding up all these contributions depends only on a single non-dimensional parameter known as the Föppl-von Kármán (FvK) number, $\gamma = YR^2/\kappa$, where R is the radius of the shell. The FvK number is essentially the ratio of the stretching and bending energies and considerably simplifies the study of the capsid properties in the context of TST [4, 18, 19].

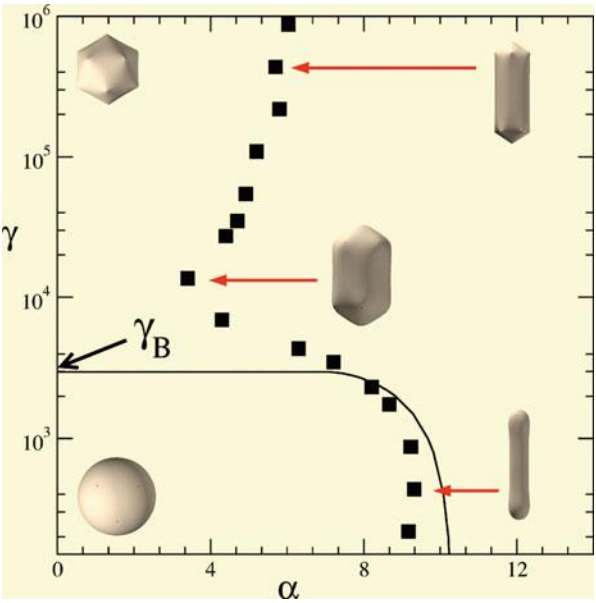

Fig. 19.2 Phase diagram of capsid shapes predicted by continuum theories (TST). Diagram showing the optimal shapes (*i.e.*, those that minimize the elastic energy in the context of TST) as a function of the FvK number, $\gamma = YS/\kappa$, and the dimensionless spontaneous curvature $\alpha = C_0 S^{1/2}$, where S is the area of the capsid. For low values of the FvK number, *i.e.* $\gamma < \gamma_B$, the optimal shape is spherical at low spontaneous curvatures and spherocylindrical, at high α. For $\gamma > \gamma_B$, a buckling transition is predicted and the optimal shapes are increasingly faceted spheres or spherocylinders. The continuous line is the TST analytical prediction, and the squares are the numerical results obtained from minimization of the elastic energy using a discretized triangular network. (Reproduced with permission from Ref. [18])

As has been discussed above, the particular shape and structure of viral capsids is the consequence of the free energy minimization principle. This means that, to obtain the possible architectures of viral capsids within TST, one has to minimize the energy functional containing the stretching and bending energies, including the contribution of the disclinations. This can be challenging in many cases, but, rather than solving the analytical expressions of TST, one can also minimize the elastic energy of a shell in a more convenient way: discretizing the structure into a triangular network and performing numerical simulations using the so-called *mass-and-spring model*, which approximates the interactions by a set of harmonic springs [18].

In this way, it has been possible to investigate the phase diagram of the optimal viral shapes as a function of the spontaneous radius of curvature and the FvK number [18] (see Fig. 19.2). The diagram shows that spherical capsids are more stable than spherocylindrical shapes at low values of the spontaneous curvature for a fixed surface, *i.e.*, number of capsid proteins. However, conical capsids were never found to be energetically better than spherical or spherocylindrical capsids. This suggests that in the context of TST some alternative physical constraints are

necessary to explain the existence of conical architectures, such as those adopted by the mature capsid of HIV-1 (see Chap. 2). In addition, for large FvK, capsids prefer to adopt a relatively faceted shape in order to release the stress accumulated in the disclinations, in a process known as the buckling transition [4], which has been well reported in the maturation of quasi-spherical capsids (see Sect. 19.4.2).

Thus, TST provides a very useful starting point to explain the overall shapes of viral capsids. However, TST does not generally capture the possible distribution of capsomers or capsid subunits in a shell. This means that in the context of continuum elastic theories there is not a simple way to distinguish between two quasispherical capsids with different T-numbers. To overcome this problem it is more convenient to use discrete models, as discussed below.

19.2.3 Coarse-Grained, Discrete Models, and Simulations

The particular structure adopted by a viral capsid is ultimately a consequence of the interactions between its fundamental building blocks. In the framework of TST, the molecular interactions between capsid proteins are effectively replaced by the stretching and bending elastic terms. However, to explore equilibrium and dynamical properties of capsids in a higher degree of detail, one cannot ignore the discrete nature of viral shells.

In this context, the discrete modeling of capsids has been performed at different levels of resolution, from full-atom approaches to simple coarse-grained physical models (see Fig. 19.1). The appropriate choice of the resolution depends on different factors, like the properties to be probed, the existence of experimental data, and the computational limitations. Here we have organized these modeling strategies in two groups. The first one contains *low-resolution* models that are usually grounded in basic physical ideas, aiming to understand the essential ingredients required to explain common aspects of viral structures; the second group includes *high-resolution* models, which generally use specific molecular reconstructions derived from experiments as a starting point, and then introduce force fields that can vary in complexity in order to properly characterize the finer structural details of particular virus capsids.

Low-Resolution Models

Low-resolution models might seem a rough approximation to describe complex molecular structures like capsids, but they constitute an excellent strategy to improve our understanding of viruses. The absence of molecular details allows the simulation of relatively large systems, and the simplicity of the models facilitates the qualitative characterization of the underlying mechanisms that explain many important properties of viral capsids. In general, these models replace the stretching and bending energies, discussed above in the context of continuum theory, by several types of effective interactions between the capsid building blocks

(CBBs). Three are essentially the minimal ingredients of these interactions. First, one needs to include a short-range repulsion force that mimics the steric effects between capsid proteins, which is intended to avoid the overlapping between the molecules. Then, a longer-range attraction is also introduced to drive the self-assembly of the capsid and to give stability to the final structure. The third essential ingredient is the existence of an imposed curvature, or preferential angle of interaction, between the building blocks, which is ultimately responsible for the selection of a particular size and curvature of the shell.

Remarkably, the seminal work in Ref. [6] proved that the icosahedral and other symmetrical structures adopted by spherical viruses are the natural consequence of the free energy minimization of a very generic interaction, which essentially contains the contributions discussed above. In their low-resolution model, the description was made in terms of capsomers, represented as disks or spheres of two different radii (see Fig. 19.1d), corresponding to pentamers and hexamers, which were confined to the surface of a sphere. Then, using a computational technique known as Monte Carlo simulation, it was possible to find the optimal arrangement of capsomers. For the case where hexamers and pentamers coexist in the capsid, the distribution that minimized the total energy corresponded precisely to the T-number architectures of the CK classification (see Fig. 19.3). Additionally, when only one type of morphological unit was allowed, the optimal free energy minima corresponded to the icosahedral and non-icosahedral architectures observed in all-pentamer viruses both *in vivo* and *in vitro*.

More recently, this model was extended to study the optimal structures of bacilliform capsids [7] (see Fig. 19.3c, d) showing that icosahedral symmetry is also predominant on the ending caps of these shells. Interestingly, the optimal structures followed the set of geometrical rules derived in Ref. [3] as extension of CK for elongated viruses. In addition, this study justified the existence of non-icosahedral-capped capsids for all-pentamer viruses (*e.g.*, polyomavirus), and demonstrated that the bacilliform geometrical construction also captures most of the properties of icosahedrally-capped all-pentamer viruses.

The important message of these simple coarse-grained models is that the impressive regular and symmetric architectures of viral capsids emerge naturally from free energy minimization of very generic interactions between their building blocks. In other words, the details of these interactions turn out to be not so crucial in determining the possible catalogue of viral shell architectures, although they can be essential to determine the physiological conditions of capsid formation. In fact, various models that use different types of interaction potentials and more realistic representations lead to similar conclusions [9, 21]. The exhaustive thermodynamic exploration of these models has shown that the set of basic parameters introduced in these simple physical approaches are able to spontaneously reproduce spheroidal shells, tubular and helical shapes, and even head-tail morphologies [9]. In this way, these low-resolution strategies have played a very important role in unveiling the key factors that determine the architecture of viral capsids. Furthermore, they are contributing to the determination of fundamental mechanisms controlling the mechanical properties, and the assembly and disassembly of viral capsids either in the presence or absence of genetic material, as will be discussed in Sects. 19.3 and 19.4.

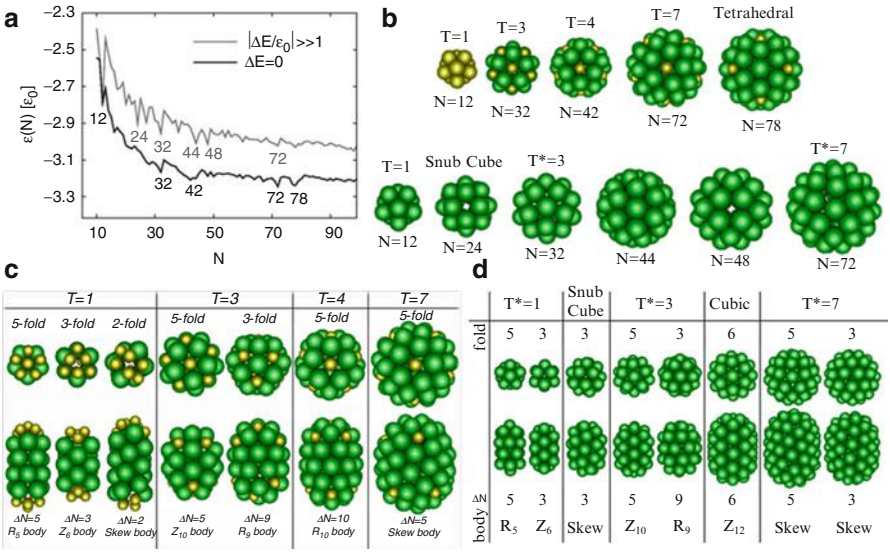

Fig. 19.3 Gallery of optimal structures obtained from coarse-grained simulations. (**a**) Energy landscape of the optimal structures obtained in a coarse-grained model of a quasi-spherical capsid with two types of capsomers (*solid line*), *i.e.*, pentamers and hexamers, or just one type (*grey line*). The figure plots the average energy per capsomer as a function of the total number of capsomers of the shell, and the deepest minima correspond to the optimal structures, which for the two capsomer-types coincide with the CK T-number structures. Notice that the apparently small energy differences *per capsomer* plotted correspond to quite large energy differences *per capsid*. (**b**) Gallery of optimal structures for quasispherical viruses made of two types of capsomers, *i.e.*, pentamers (*gold*) and hexamers (*green*) (*top*), or just of one type (*bottom*). (**c** and **d**) Gallery of optimal structures for spherocylindrical viruses with two (**c**) and just one (**d**) capsomer types. The top images show a zenithal view of the caps, and the lower images show a lateral view of the complete bacilliform virus. In each case, the T-number of the cap, the minimum number of capsomers in the body ΔN required to elongate the structure, and the arrangements of capsomers in the body (forming rings R_n or zig-zag patterns Z_n of n capsomers) are also indicated. (Figures adapted from Refs. [7, 20])

High-Resolution Models

High-resolution models are the most appropriate computational strategy to study in detail the properties of specific viruses. Contrary to the models discussed above, these are based on structures reconstructed from X-ray crystallography (see Chap. 4) and/or cryo-electron microscopy (cryo-EM; see Chap. 3) studies. The interactions between the atoms are modeled through effective force fields, and techniques like molecular dynamics or Monte Carlo are applied to sample the structural dynamics of the viruses.

Full-atom simulations are gaining popularity in the field of structural and chemical biology, and are of particular interest for biomedical applications. The constant improvement of computational power and sampling techniques has made this approach more feasible in the last years, though there is still ongoing research

to develop more accurate force fields and to reach longer time scales in simulations. The first full-atom simulation with explicit solvent of a complete virus was developed for the relatively small satellite tobacco mosaic virus (STMV), involving a million of atoms (135,960 protein atoms, 30,330 nucleic acid atoms, plus water and ions) and achieving timescales of nanoseconds. In that case it was shown that, at physiological conditions, the virus becomes unstable when the viral RNA is removed [14]. To study larger systems or processes that involve longer time scales it is required to use multiscale, or mesoscale models, and some degree of coarse-graining [22].

These simulations are starting to provide very useful insights into the detailed structural mechanisms of different viruses. For instance, in the case of HIV, it has been possible to study the interactions between the main capsid protein and the lipidic membrane in the immature virus [23], the role of the C-terminal binding interface between the capsid proteins controlling the curvature and shape of the capsid [23], and the phase diagram for the assembly of the capsid proteins [24]. Another well-studied case has been the plant virus cowpea chlorotic mottle virus (CCMV), where all-atom simulations have characterized its swelling process that is triggered by changes in the pH [25]. In particular, the combination of elastic network with coarse-grained models has allowed simulating the structure of this virus in the scale of microseconds. Besides, general multiscale models have been applied to simulate the *Nudaurelia capensis* ω virus and human papillomavirus (HPV). Alternatively, coarse-grained simulations of the bacteriophage MS2 indicates that the organization of the genetic material within the capsids of RNA viruses depends on the non-specific self-repulsion of the RNA in the inner layers, but the organization of the outer layers strongly depends on the interactions with the inner part of the capsid. Finally, simulations of the structure of Pariacoto virus suggest assembly mechanisms of RNA viruses in agreement with alternative physical models of assembly [26] (see Sect. 19.3).

19.3 Assembly of Viruses

The assembly of viral particles is a crucial step in the life cycle of viruses. Depending on the virus, the capsid can successfully assemble from the capsid proteins alone, or in combination with scaffolding proteins, or the genetic material, following specific assembly pathways reviewed in Chaps. 10, 11, 12 and 13. However, the assembly of viruses also shows some general trends that seem susceptible to be modeled using basic physical ideas. As discussed in Sect. 19.2, a very important fact is the spontaneity of the self-assembly process of viruses with single-stranded genomes, which is driven by free energy minimization, meaning that no external energy in form of ATP is required. This makes the proper assembly of some viral shells possible even in the absence of genetic material or in the presence of other types of cargoes. In fact, many viruses can be reconstituted *in vitro* from their basic components, yielding particles that are fully infective and indistinguishable from

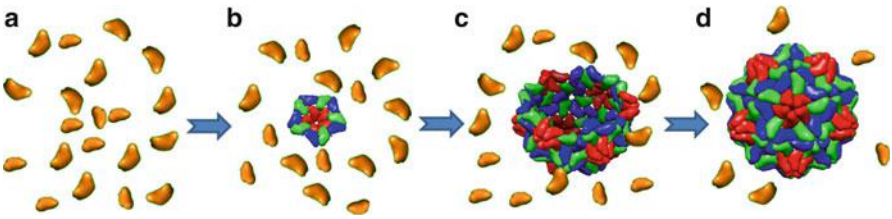

Fig. 19.4 Sketch of the different steps in capsid assembly. (**a**) The process is initiated by the presence of a sufficiently high concentration of free capsid building blocks (in *gold*). (**b**) These building blocks aggregate and disaggregate by thermal fluctuations until a sufficiently large cluster of subunits (in *color*) forms, which acts as the *nucleus* or embryo of the capsid; (**c**) the embryo grows by the addition of building blocks and forms a partial capsid or intermediate of *n* building blocks; (**d**) finally a complete capsid is formed

native virions (see Chaps. 10 and 21). The best-known example is TMV, which was the first virus to be fully reconstituted *in vitro* by mixing the capsid proteins and the viral RNA. In addition, by properly tuning the assembly conditions it has been possible to produce empty capsids of viruses that *in vivo* require the genetic material to assemble, as has been shown for plant viruses like CCMV and brome mosaic virus (BMV), or animal ones like human polyomavirus and hepatitis B virus (HBV). The modification of assembly conditions, like capsid protein concentration, pH, or ionic strength, can also lead viral capsid proteins to assemble into capsid structures different from the native one [27], and to encapsulate foreign genetic materials and even non-organic cargoes [28, 29]. Obviously, this has opened the possibility to engineer viral capsid proteins for different applications in nanoscience, *e.g.*, encapsulation of materials or templating of nanostructures (see Chap. 22).

Diverse viruses can follow different assembly pathways (see Chaps. 1, 10, 11, 12 and 13). In particular, their building blocks (CBBs) can be made of different number of capsid protein subunits and are typically defined as being the most stable oligomer in solution. For instance, the CBBs are single proteins in the case of *Penicillium chrysogenum* virus, dimers for CCMV, pentamers in HPV, or pentamers and hexamers in solution for bacteriophage Hong Kong 97 (HK97). Starting from these building blocks in solution, the assembly process for empty capsids proceeds first by the formation of some aggregates of CBBs, known as *nuclei*, which progressively grow incorporating more CBBs, forming what will be called *partial capsids* or *intermediates*, until a complete shell is assembled (see Fig. 19.4). It is worth noticing that understanding the steps during viral assembly is a key question to identify, for instance, possible targets for antivirals (see Chap. 20).

From a theoretical point of view, most modeling efforts have focused on the simplest case, *i.e.*, the *in vitro* self-assembly of *empty* viral capsids from capsid proteins that do not require any sort of auxiliary proteins. The abundance of *in vitro* experiments has facilitated the investigation of this scenario, and it is the first topic to be described in this section. Subsequently, we will discuss separately the peculiarities of the assembly of ssRNA and dsDNA viruses, since the mechanism followed in the assembly of nucleocapsids depends crucially on the type of nucleic acid of the viral genome.

19.3.1 Empty Capsids

The fact that many viruses can assemble in the absence of genetic material indicates that the interactions between viral capsid proteins generally contain all the information to form a capsid. The range and specificities of these interactions can vary among viruses, but they are usually dominated by two non-specific contributions: hydrophobic and electrostatic [30].

Hydrophobic and Electrostatic Interactions

Many viruses, whose structure has been solved close to atomic resolution, have interfaces between subunits with a strong component of hydrophobic amino acid side chains. In addition, structural and biochemical studies with some model viral particles have shown that the strength of viral capsid protein-protein interactions tends to increase when the temperature is raised, indicating that hydrophobic interactions are the main driving force in the assembly of those particles. The hydrophobic effect is ubiquitous in biological systems, although it is still a phenomenon under intense investigation. However, some effective characterizations have been already tested in non-biological soft matter systems, which conveniently serve as a starting point for the study of viruses.

In particular, the effective hydrophobic interaction potential between two capsid subunits can be approximated by that of two identical apolar surfaces, given by $V_H = -2\gamma A \exp(-D/\gamma)$. In our case γ denotes an effective surface tension between the hydrophobic protein surface and water, A the contact area between capsid proteins, D the separation of the hydrophobic surfaces, and λ a decay length, which is typically on the order of nanometers. The distance between interacting subunits in a capsid is on the atomic scale, *i.e.*, $D << \lambda$, and so the exponential term in the final capsid will be close to unity. Thus, for a complete capsid the total hydrophobic contribution to the free energy of formation will be $G_H = -A_H\gamma$, where A_H is the total hydrophobic area buried in the formation of the shell.

In addition, electrostatics plays an important role in many viral capsids. Capsid proteins are in many cases rich in basic residues leading to an excess of positive charge in the interior of viral shells. This allows the subunits to interact with the negatively charged phosphate backbone of the viral nucleic acids, promoting the assembly and stability of the capsid. In the absence of the genetic material this overcharge introduces an effective repulsion between the capsid proteins. As a consequence, the assembly of empty capsids may require higher salt concentrations (to screen the electrostatic repulsions) and/or lower pHs (to reduce the charge of protein residues), compared with the co-assembly with the genetic material. In addition, in several capsids, electrostatic repulsions between either positively or negatively charged residues that may occur under certain conditions *in vivo* are

used to regulate capsid conformational stability, disassembly or genome uncoating (*e.g.*, in TMV, CCMV and foot-and-mouth disease virus).

In a first approximation, the electrostatic contribution in a viral capsid can be described using Debye-Hückel (DH) theory [31]. This takes into account that the ions of the salt in the medium will screen the electrostatic interaction between the charged residues of a capsid. For two unit charges e separated by a distance r this description leads to the potential $V_c = k_B T(\lambda_B/r)\ \exp\ (-\kappa r)$, where k_B is the Boltzmann constant, T the absolute temperature, $\lambda_B = e^2/4\pi\varepsilon_0\varepsilon_r k_B T$ the Bjerrum length, and $\kappa^{-1} = \left(4\pi\lambda_{\mathrm{B}} \sum_i z_i^2 c_0^i\right)^{-1/2}$ the Debye screening length. Here $\varepsilon_0\varepsilon_r$ is the permittivity of the medium (water in this case), c_0^i is the number density of charged species i in the electrolyte (in units of m^{-3}), and z_i their corresponding valence. In the crudest approximation, a quasi-spherical capsid can be considered as a uniformly charged thin sphere of radius R and surface charge density σ, surrounded by an (ideal) ionic solution (1:1 electrolyte) of bulk concentration c_0. The electrostatic contribution to the free energy of the capsid is then obtained by integrating the DH potential over the surface of the sphere, yielding $G_{elec} = \frac{\pi\sigma^2 R^2}{\varepsilon_0\varepsilon_r\kappa}$, provided that the Debye length is short enough, *i.e.*, $\kappa R >> 1$, which is usually satisfied even at low salt concentrations, *e.g.*, on the order of mM (for a more detailed derivation of both terms, the reader is referred to Refs. [30, 32]). This simple approximation can be refined by incorporating the effects of a finite capsid thickness, a nonuniform distribution of charges in the capsid, lower ionic strengths, or even by solving the more rigorous description of this electrostatic problem given by the Poisson-Boltzmann equation. Interestingly, all these approaches lead to qualitatively similar results [32].

In this context, the total free energy of a complete spherical capsid is obtained by adding the hydrophobic and electrostatic contributions, yielding $G \approx -A_H\gamma + \frac{\pi\sigma^2 R^2}{\varepsilon_0\varepsilon_r\kappa}$. Rough estimates of the order of magnitude of both contributions can be made using HBV as an example. For HBV, the total hydrophobic buried surface is $A_H \sim 10^3\ \mathrm{nm}^2$, and the energy per unit area of exposed protein (accounting also for van der Waals interactions) is roughly $\gamma \sim 6\ \mathrm{mJ/m}^2$, yielding an estimate of $\sim 10^4\ k_B T$ as a typical attractive contribution. The order of magnitude of its surface charge is $\sigma \sim 0.7\ \mathrm{e/nm}^2$, and its radius is $R \sim 30$ nm, giving also an electrostatic contribution of $\sim 10^4\ k_B T$ at physiological conditions ($c_0 = 150$ mM) where $\kappa^{-1} \sim 1$ nm. Both contributions have thus similar strengths, illustrating the delicate balance between the two required for a successful assembly and its sensitivity to the environmental conditions.

The main parameter controlling the strength of the hydrophobic interaction is the temperature, whose increase favors a stronger attraction. In addition, the electrostatic contribution can be tuned by two accessible physicochemical conditions: the pH and the salt concentration of the solution. The pH controls the fraction of charged residues in the capsid protein, and the salt concentration varies the ionic strength, which screens the electrostatic repulsion between capsid proteins. These three factors have major roles in controlling the feasibility of capsid assembly both *in vivo* and *in vitro*.

Capsid Assembly as a Thermodynamic Process

The fourth important factor in the self-assembly of capsids is the concentration of subunits, and its impact on the assembly of empty capsids has been probed *in vitro* in the last years. In those assembly experiments, one starts from a given concentration of capsid proteins in solution and monitors the formation of capsids by scattering techniques, electron microscopy, or size-exclusion chromatography. These studies have revealed a steep dependence of the assembly yield on protein concentration as shown in Fig. 19.5a for the paradigmatic case of HBV. For low protein concentrations capsids are not formed; then the production starts above a certain threshold concentration, and grows steeply with concentration, until eventually reaching saturation. Another distinguishing feature observed experimentally is the lack of intermediates, *i.e.*, most subunits are found either free in solution or in fully formed capsids.

This behavior can be described theoretically by using standard ideas from equilibrium thermodynamics. The process of assembly can be thought as a phase transition between two states: the capsid building blocks in solution and in a fully formed capsid. The equilibrium state is reached when the chemical potentials of both states are equal, leading to the well-known law of mass action

$$K_{capsid}^{eq} = \frac{c_{capsid} c_s^{q-1}}{c_1^q} = e^{-\frac{G}{kT}}, \tag{19.1}$$

where q is total number of subunits in a complete capsid, c_{capsid} and c_1 are the equilibrium concentrations of capsids and subunits, respectively, c_s is the molarity of pure water taken as a reference concentration, and G is related to the free energy of capsid formation described above. The law of mass action can be conveniently rewritten in terms of the capsid mass fraction $f \equiv qc_q/c$ as

$$1 - f = f^{1/q} \left(\frac{c^*}{c} \right)^{1-1/q} \approx \begin{cases} 1, & c < c^* \\ c^*/c, & c \geq c^* \end{cases}, \tag{19.2}$$

where c is the total concentration of proteins in the solution, and $c^* = c_s e^{\frac{\Delta g}{kT}}$ is a threshold concentration related to the binding energy per protein $\Delta g \equiv G/q$. In terms of this threshold concentration, the final yield of capsid production can be described by a universal curve given by Eq. (19.2). Interestingly, since for viruses one typically has $q >> 1$, this universal curve can be simplified to $f = 0$ for $c < c^*$ and $f = 1 - c^*/c$ for $c \geq c^*$. This simple model successfully described the temperature and salt concentration dependence of the *in vitro* assembly of HBV [30], as shown in Fig. 19.5b.

Kinetic Aspects of Viral Assembly

In the past years, different studies using light scattering and turbidimetry techniques have been able to monitor also the kinetics of the assembly process for empty

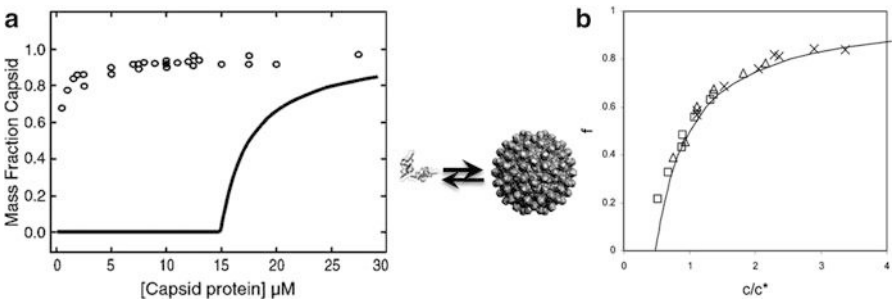

Fig. 19.5 *In vitro* assembly of HBV capsids. (**a**) Mass fraction of proteins in capsids as a function of the total capsid protein concentration in assembly (*solid line*) and disassembly (*open circles*) experiments. (**b**) Fraction of subunits in complete capsids, f, *versus* the total concentration of capsid protein scaled by the critical micelle concentration $c_{cmc} = 2c^*$. The symbols are data from Ref. [33] obtained from experiments of HBV assembly at 25 °C and different buffer concentrations of NaCl (crosses $c_{salt} = 0.7$ M; triangles $c_{salt} = 0.5$ M and squares $c_{salt} = 0.3$ M). The line is the universal aggregation curve predicted by Eq. (19.2). (Adapted from Ref. [30, 34])

capsids like HBV or HPV. These experiments suggest again the existence of common features in the kinetics of capsid assembly. It turns out that initially there is a lag time period where no capsids are formed, and then the production rapidly increases following a sigmoidal curve. Another characteristic is that the disassembly of the capsid takes place at different concentration conditions than the assembly, giving place to a hysteresis phenomenon, which is observed even in viruses that do not go through any apparent post-assembly maturation (see Fig. 19.5a). Indeed, once capsids are formed, they do not dissociate easily even if the concentration of free subunits is drastically reduced. This has a clear biological relevance, since viruses must resist a number of different environments to reach a new host, where no viral proteins are present.

Again, these features match the expected behavior of a standard first-order phase transition. Accordingly, different models have borrowed the physical concepts involved in the kinetics of phase transitions to quantify and characterize the assembly of viral capsids.

The first kinetic model for viral assembly was introduced by Zlotnick and collaborators [35]. There, the assembly of the capsid is conceived as a series of concurrent molecular reactions between individual subunits and intermediates of all possible sizes, which are schematically represented as subunits ⇔ nuclei + subunits ⇔ capsids (see Fig. 19.4). The population of intermediates containing n subunits is given by the set of equations

$$\frac{d[n]}{dt} = k_{elong,n-1}[n-1][subunit] + k_{dis,n+1}[n+1] - k_{elong,n}[n][subunit] - k_{dis,n}[n], \tag{19.3}$$

where $k_{elong,n}$ and $k_{dis,n}$ are the rates of association and dissociation of subunits to a capsid of size n, respectively. The kinetics is split into a nucleation step, which involves the formation of capsid embryos, and an elongation process, which corresponds to the growth of the nuclei by the sequential addition of subunits. In this context, nuclei are treated as a special kind of intermediates that cannot disassemble and form at a slower rate k_{nuc} than the elongation stage, in order to justify the initial lag time of viral assembly. Numerical simulations have shown that this set of reactions quickly reaches a steady-state concentration of intermediates that asymptotically approach an apparent equilibrium characterized by a law of mass action, Eq. (19.1), from which one can obtain the association free energy of the capsid.

Using this model, it has been possible to describe the kinetics of *in vitro* assembly experiments of empty capsids of viruses like HBV [33] and HPV [36]. Further refinements were proposed by Dragnea et al. [37] to describe the assembly of BMV capsids, which include certain simplifications in the rate equations and an additional activation step for the monomers to fit the short time initial takeoff of the assembly. Remarkably, these studies conclude that the interactions driving the assembly are on the order of a few k_BT per protein, typically 5–6 k_BT per contact, which is much weaker than a covalent bond [33]. This has two clear biological motivations: first, weak interactions facilitate the correction of possible mistakes during the assembly; second, disassembly has to be feasible for many viruses to release the genetic material and infect a new host. Interestingly, the capsid of viruses like bacteriophages that do not require disassembling undergo a maturation process that in many cases reinforces the shell with additional or alternative subunit-subunit interactions (see Sect. 19.4.2 and Chap. 13).

However, despite the utility of these kinetic models, they present several important drawbacks. In particular, they involve a large set of reaction equations that need to be solved numerically, which can be very demanding computationally; the reaction rates must be fitted to the experimental data, which limits the predictive power of the theory; and the size of the nucleus is usually guessed a priori, which means that different cases should be explored for each experiment. Moreover, these models assume that the assembly is an equilibrium aggregation; hence, to justify that disassembly takes place at much lower protein concentrations than expected, they have to assume that the binding energies for assembly and disassembly are different.

Classical Nucleation Theory of Capsids

Alternatively, a recent approach based on classical nucleation theory (CNT), has overcome some of these limitations, explaining in a simpler way the common trends found experimentally in the assembly of capsids, and providing also a more direct connection between theory and experiments [20, 38]. The key point in CNT is the development of a physical model for the free energy of formation of partial capsids that, contrarily to the equilibrium aggregation theory discussed

above, contains explicitly the dependences in both the interaction between subunits and the concentration of free subunits. As we show below, this energy can be related to the rate of capsid production, reducing considerably the number of parameters to be fit, if any. Let's describe the main elements of CNT applied to viral capsids.

As we discussed above, the aggregation of a capsid becomes thermodynamically favorable when a building block in a fully formed capsid has a lower chemical potential than in solution, *i.e.*, when the interaction energy of a building block in the capsid overcomes the entropic penalty of removing it from the solution. Nevertheless, if one takes a closer look to the assembly process, one realizes that, in the successive addition of building blocks to intermediate structures, those CBBs placed at the rim of a partial shell will be exposed and miss few energy contacts (see Fig. 19.4), *i.e.*, the formation of intermediates intrinsically entails an energetic cost that originates an energy barrier. It is precisely this barrier what justifies the existence of a lag time at the beginning of the assembly process and the scarcity of intermediates observed experimentally. This is a well-known scenario in physics and is analogous to a vapor-liquid phase transition, where the formation of a liquid drop in the vapor phase has an energetic penalty associated to the surface tension of the drop. Following this analogy, the simplest way to model the energetic cost of an intermediate shell is by introducing a rim energy penalty associated to a line tension, equivalent to the surface tension in a drop.

In the framework of CNT applied to quasispherical capsids, a partial shell can be approximated as a spherical cap made of n subunits, whose free energy of formation is given by $\Delta G(n) = n\Delta\mu + a\sqrt{n(q-n)}$. Here $\Delta\mu = -k_B T \ln\left(\frac{c}{c^*}\right)$ is the difference in chemical potentials between a subunit in a complete capsid and in solution, and a is a constant related to the line tension. Even at favorable assembly conditions, the competition between these two terms gives rise to an activation barrier (see Fig. 19.6a). The maximum of this barrier defines the critical nucleus size

$$n^* = \frac{q}{2}\left(1 + \frac{\Delta\mu}{\sqrt{\Delta\mu^2 + a^2}}\right), \tag{19.4}$$

and its height

$$\Delta G^*_{as} = \frac{q}{2}\left(\Delta\mu + \sqrt{\Delta\mu^2 + a^2}\right) \tag{19.5}$$

defines the activation (or nucleation) barrier. As we see in Fig. 19.6a, for $n < n^*$ the free energy of formation of an intermediate increases with the number of CBBs, so these intermediates will tend to disassemble back into free subunits. On the other hand, partial capsids of size $n > n^*$ reduce their free energy upon the addition of CBBs, and they will tend to grow spontaneously until complete a capsid. Thus n^* is the critical size that partial capsids have to reach to trigger the formation of complete shells and can be considered as the *nucleus* or the embryo of the assembly

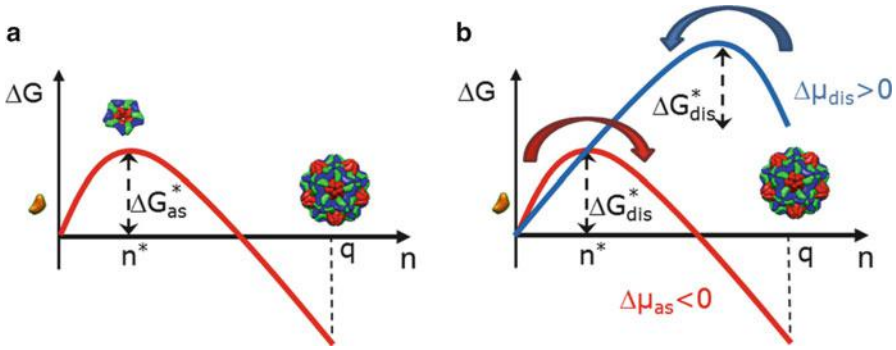

Fig. 19.6 Free-energy barriers of capsid formation in the context of CNT. (**a**) Total free energy of formation of a capsid as a function of the number of capsid building blocks n at favorable assembly conditions ($\Delta\mu_{as} < 0$). The complete capsid contains q building blocks. The nucleation barrier ΔG^*_{as} and the critical cluster size n^* are indicated. (**b**) Free energy of capsid formation for favorable assembly ($\Delta\mu_{as} < 0$, *red line*) and disassembly ($\Delta\mu_{dis} > 0$, *blue line*) conditions. For disassembly, the barrier has to be jumped in the opposite direction (*i.e.*, from *right* to *left*) than for the assembly. To have similar barrier heights for assembly and disassembly, different chemical potentials, and accordingly concentrations of capsid proteins, are required. That is ultimately the reason for the hysteresis between assembly and disassembly conditions found experimentally (see Fig. 19.5a)

process. It is worth emphasizing that, contrary to the assumption in the previous kinetic model, the critical nucleus is the most unstable intermediate because it corresponds to a maximum of the energy, and both its size and the height of the nucleation barrier depend on the concentration of subunits. In particular, for $c < c^*$ the formation of a capsid is not possible, and protein concentrations $c >> c^*$ are indeed needed to lower the nucleation barrier and the critical size enough to make the capsid formation feasible.

The rate of capsid formation J, *i.e.*, the number of capsids that form per unit volume and unit time, is controlled by the presence of this barrier. Like in other activation processes, it follows an Arrenhius behavior given by

$$J = c_s \beta^* Z e^{-\frac{\Delta G^*_{as}}{k_B T}}. \tag{19.6}$$

Hence it depends exponentially on the barrier height and linearly on β^*, which is the rate of attachment of CBBs to the critical cluster size. Here, Z is a correction factor associated with the local curvature at the top of the barrier and accounts for the possibility that nearly critical-sized clusters dissociate.

The activated mechanism described by CNT has several biological advantages. First, the requirement of a threshold concentration and the existence of a critical size that has to be reached to trigger capsid formation favors the production of enough proteins to guarantee the completion of capsids. In other words, if there were no barrier there would be multiple intermediates growing at the same time, which would rapidly deplete the protein by generating many incomplete shells, a

remarkably inefficient scenario. Additionally, the presence of a barrier is the origin of the hysteresis, which prevents the disassembly of capsids under conditions where the concentration of subunits is reduced or nonexistent, favoring the stability of the virus during the extracellular process before a new infection (see Sect. 19.5.2). A more detailed derivation of CNT for viral capsids can be found in Refs. [20, 38].

Simulations of the Assembly of Empty Capsids

Despite the advances in computational power, full-atom simulations of the assembly of even the smallest viruses are still out of reach. Accordingly, current simulations necessarily involve coarse-graining at different levels, which, apart from the computational advantage, represent also a great strategy to isolate the relevant details involved in the assembly process at a microscopic level.

At the coarser level, several models have explored the geometrical particularities required for the formation of T-number icosahedral shells. In particular, models based on local rules [39], which warrant the formation of the proper contacts between subunits, have shown that, due to the similarity between quasiequivalent positions (see Chap. 2), the set of rules to successfully assemble a T-number capsid can be less than T. For instance, four local rules suffices to produce $T = 7$ shells [40]. In a similar spirit, tiling approaches have studied the assembly of capsids focusing on the combinatorial aspects involved in the addition of subunits to intermediates of all possible sizes, concluding that only a small set of intermediates are relevant for the assembly pathway of empty shells [41, 42]. Various models have introduced a more physical implementation of these ideas by proposing effective interaction potentials with local constraints or strict orientational forces designed to produce a particular T-number target. This embraces patchy particle potentials [43] and bond-based potentials [44] (see Fig. 19.1).

The next level of complexity for coarse-grained strategies includes models that aspire to partially represent the shape of the assembly building blocks, like protein monomers, dimers, or pentamers and hexamers with a different degree of accuracy (see Fig. 19.1). In this way, subunits have been represented as pyramids [8], ellipsoids [9], triangular, and trapezoidal elements [10, 12, 45], which are agglomerates of relatively small spherical beads that help to place the interactions sites that are handcrafted to succeed in the assembly of a specific target structure.

Models based on a continuum elastic description of the assembly, based on growing a triangulated network, have also been successfully used to describe, for instance, the growth of spherical and polyhedral shells [46], and the (irregular) shapes of conical virus capsids like HIV [47, 48].

Finally, higher resolution coarse-grained strategies have been more challenging to investigate. However, some groups have studied, for instance, the assembly of HIV capsid proteins [24].

These simulation studies have confirmed many of the assumptions discussed in the aggregation and CNT theories, and have provided a better understanding of the difficulties and ingredients required for the successful assembly of empty viral

capsids. The combination of these theoretical and simulation studies is opening the door to a higher understanding of the assembly of viral capsids, and hopefully will facilitate the development of novel antiviral strategies.

19.3.2 RNA Co-assembly

Although *in vitro* experiments have demonstrated that, in certain conditions, the viral capsid proteins of ssRNA viruses alone are able to reconstruct the native capsid structure, *in vivo* their assembly generally requires the presence of the genetic material [49] (see Chap. 12). Consequently, the self-repulsion and entropy of the nucleic acid, and its interaction with the inner surface of the capsid are new ingredients that have to be considered to characterize the free energy of the complete system.

In physiological conditions ssRNA behaves like a flexible polymer with a persistence length on the range of the nucleotide size. In this scenario the main driving force in the co-assembly process is the electrostatic interaction between the ssRNA and the positively charged residues of the capsid, which often concentrate in disordered terminal segments of the capsid proteins. By virtue of electroneutrality one would expect a 1:1 correspondence between the length of the viral RNA and the charge of the inner part of the capsid in contact with the genetic material. However, experiments have determined that a variable part of the nucleic acid negative charge is neutralized by metal ions and/or polycations (*e.g.*, spermine or spermidine) in different viruses. Moreover, the distribution of charges in the capsid, the excluded volume effects and the Donnan potential (*i.e.* the difference in electrostatic potential between inside and outside the capsid due to the unbalance of charged macromolecules) have been shown to play a crucial role in the possible overcharging of viruses. Accordingly, there seems to be no universal genome to capsid charge ratio [50].

Several models have investigated the thermodynamics of co-assembly combining basic electrostatic and polymer physics theories, as discussed below. As a starting point, the complexation energy between ssRNA and the capsid can be evaluated using the DH mean field theory [32]. For spherical thin shells this yields $\Delta F_c = \frac{\pi\sigma^2 R^2}{\varepsilon_0 \varepsilon_r}(fd - \kappa^{-1})$, where d is the thickness of the ssRNA layer adsorbed on the inner capsid wall, f is a numerical factor, and κ^{-1} is the Debye screening length. The co-assembly is thermodynamically favored when this complexation energy becomes negative, which depends on the competition between the characteristic scales d and κ^{-1}. Contrary to the case of dsDNA bacteriophages discussed below, here the nucleocapsid reaches an energy minimum that can be evaluated from the previous energy of formation, indicating that ssRNA viruses may not be pressurized [32, 51]. More refined thermodynamic models that include the distribution of charges in the capsid, the possible presence of charged peptide arms, and the contribution of charged macromolecules, have been developed and can justify the

overcharging of some ssRNA viruses in nature and the favorable co-assembly of viruses with no net capsid charge [50].

Remarkably, many RNA viruses store several strands of genetic material rather than a long single sequence. This was thermodynamically characterized using a simple Flory mean-field theory [52], which accounts for the elastic compression of the chain, the self-repulsion of the genetic material, and the interaction of the polyelectrolyte with the inner wall of the capsid. This theory also justified the results from several *in vitro* experiments showing that the genome length can control the size of the final capsid [53]. Interestingly, the results of these thermodynamic models are valid also for other negatively-charged polymers different from RNA, as has been proved experimentally [28].

Regarding the organization of ssRNA, experimental studies have shown that it is distributed inhomogeneously inside the capsid, being very dense close to the inner surface and looser as we move towards the center, which in many cases is completely devoid of genetic material. The RNA in the central region does not adopt any particular structure and its organization is a consequence of the self-repulsion of the genetic material [54]; instead for some viruses the outer part sews the capsid proteins tracing a Hamiltonian path (*i.e.* a path that visits each vertex of the capsid only once) through the edges of a dodecahedron partially recovering the icosahedral symmetry of the shell, which could have an important role during the assembly of the capsid [26, 49].

Recently, several simulations have started to investigate the co-assembly of RNA viruses, but again this is a challenging task that requires coarse-graining and strong computational efforts. One of the first models explored the encapsulation of a polymer defined on a cubic lattice, but more realistic off-lattice simulations with flexible polymers encapsulated by coarse-grained capsomers or capsid proteins have been feasible lately [11, 55]. These simulations have assessed the role played by the genome length, the relative stoichiometry of genome *versus* capsid proteins, and the strength of the different interactions in the efficiency of the encapsulation process. In particular, it has been found that there is an optimal genome length that maximizes the encapsulation efficiency, which depends on the competition between capsid-capsid and capsid-genome interactions. Remarkably, depending on these interactions, two plausible assembly pathways have been identified (see Fig. 19.7). In the first one, the polymer is encapsulated in concert with the capsid formation, a process that is the natural extension of the assembly of empty capsids. Interestingly, in this case the capsid formation can be also described by classical nucleation theory, where the genome reduces significantly the nucleation barrier [55]. The other pathway takes place for strong subunit-RNA interactions. Initially many subunits (typically in excess to the total needed) attach to the genome and then they cooperatively rearrange to form the closed capsid. However, the process of nucleic acid-assisted assembly of ssRNA viruses is still far from well understood; further experimental and theoretical studies are clearly required, and the interested reader is referred to Chap. 12 for current views on the process based on experimental observations.

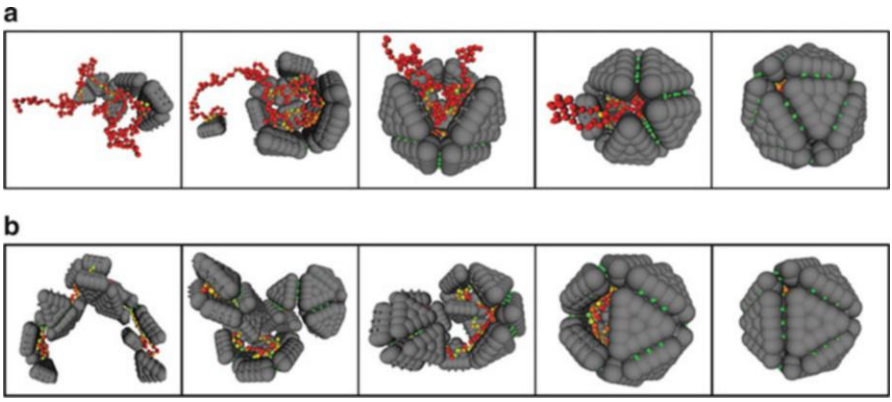

Fig. 19.7 Mechanisms of co-assembly for RNA virus. Different snapshots of a coarse-grained simulation showing two possible pathways in the formation of an RNA virus particle. (**a**) The RNA (*red chain*) is encapsulated in concert with the formation of the capsid by the addition of capsid proteins (modeled by multiple *grey* beads with *green* interaction sites). (**b**) For strong RNA-capsid protein interactions, first, many capsid subunits (in excess above the stoichiometric ratio) attach to the RNA in a disordered way and then cooperatively rearrange to form the complete capsid (Reproduced with permission from Ref. [11])

The encapsulation of non-genetic materials, such as gold nanoparticles, emulsions or negatively-charged polyelectrolytes by viral proteins, has also been explored in experiments and simulations, mainly driven by their potential technological applications (see Chap. 22).

19.3.3 dsDNA Packaging

There are two physical properties of dsDNA that play a crucial role in the assembly of many viruses: elasticity and charge. Contrary to ssRNA and ssDNA, dsDNA is a relatively stiff polymer. In a first approximation, its rigidity is characterized by its persistence length, *i.e.*, the length over which the DNA keeps straight under thermal fluctuations, which is about *50* nm for typical monovalent salt physiological conditions. This persistence length is similar to the characteristic capsid dimensions, and since dsDNA is usually stored at high densities, this implies that a significant elastic cost is paid during the packaging of the dsDNA genome inside the protective viral shell. Regarding the dsDNA charge, every base pair contributes with two negative charges. As for the viral ssRNA and ssDNA genomes, part of the dsDNA charge may be neutralized by metal ions and/or polycations. But in general, the charge causes the DNA chains to repel each other. The electrostatic term tends to dominate over the elastic one, although the latter seems to determine the conformation adopted by the confined dsDNA [56]. These two factors do not facilitate a spontaneous co-assembly of the capsid, except in the presence of condensing proteins like in

polyomavirus or adenovirus, and this is, most likely, one of the main reasons why many dsDNA viruses first assemble an empty capsid, the procapsid, and then package the genetic material using a molecular motor (see Chap. 12).

Along the years, several models have evaluated the energetics of the packed DNA [57–59]. Here we will describe the so-called inverse spool model that assumes that dsDNA is wrapped around with local hexagonal order in concentric hoops packed from the outside of the capsid towards the center (see Fig. 19.8). This theoretical approach assumes that the total free energy of confined DNA consists of the two contributions, elastic and electrostatic, discussed above. Then the total energy is given by $G_{tot}(L,d_s) = G_b + G_{int}$ and depends on both the total length of the genetic material L and the separation between strands d_s. The elastic contribution, G_b, is calculated assuming an inverse spool arrangement of the DNA, where the strands are packed in a hexagonal array with a spacing d_s. Then the elastic energy and the total DNA length become

$$G_b(L) = \frac{2\pi\xi_p k_B T}{\left(\sqrt{3}d_s\right)} \int_R^{R_{out}} (N(R')/R')dR'. \tag{19.7}$$

$$L = \frac{4\pi}{\left(\sqrt{3}d_s\right)} \int_R^{R_{out}} R'N(R')dR'. \tag{19.8}$$

where ξ_p is the persistence length, $N(R')$ is the number of hoops of radius R' in the capsid, and R and R_{out} are the inner and the outer radius of the inverse spool, the latter taken as the radius of the inner surface of the capsid (see Fig. 19.8).

An accurate first principle calculation of the DNA-DNA interactions is a difficult and unachieved task. Thereby, the electrostatic interaction G_{int} has been commonly characterized phenomenologically from osmotic stress experiments at the proper conditions [60]. These experiments have shown that the electrostatic interaction between DNA strands has a different behavior depending on the valence of the added salt. For monovalent and divalent salts, the contribution is purely repulsive and the dependence of the osmotic pressure $\Pi(d_s)$ with respect to the separation between strands is given by

$$\Pi(d_s) = F_0 e^{-\frac{d_s}{c}}, \tag{19.9}$$

where c and F_0 are parameters that characterize the decay length and strength of interactions, which depend on salt conditions. In the case of trivalent and tetravalent salts the effective electrostatic interaction has an optimal distance d_0 between strands, *i.e.*, at smaller separations leads to a repulsive interaction whereas at higher separations promotes attraction between the strands. A convenient empirical expression for the osmotic pressure in this situation is

$$\Pi(d_s) = F_0\left(e^{-\frac{(d_s-d_0)}{c}} - 1\right). \tag{19.10}$$

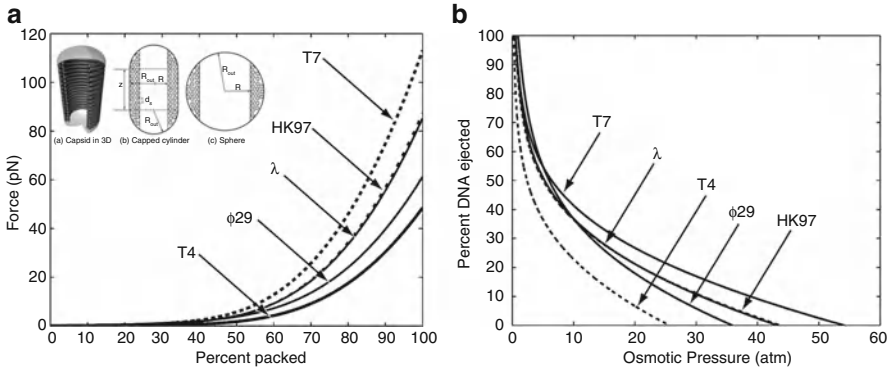

Fig. 19.8 Forces and pressures of confined dsDNA. (**a**) Forces resisting dsDNA packaging (and assisting dsDNA ejection) estimated using the inverse spool theory for different bacteriophages (T7, λ, T4, Φ29, HK97) in 0.5× PBS buffer. The inset shows the idealized organization of the dsDNA inside different capsids assumed in the model. (**b**) Osmotic pressure inside the capsid (in the horizontal axis) as a function of the percentage of DNA ejected estimated by the inverse spool theory, for different bacteriophages (T7, λ, T4, Φ29, HK97) in a buffer containing10 mM $MgSO_4$ (Figures reproduced with permission from Ref. [58])

From the osmotic pressure, the interaction energy can be obtained yielding

$$G_{\text{int}} = \sqrt{3}F_0 L\left(c^2 + cd_s\right)e^{-\frac{d_s}{c}} \tag{19.11}$$

in the purely repulsive regime and

$$G_{\text{int}} = \sqrt{3}F_0 L\left[\left(c^2 + cd_s\right)e^{-\frac{(d_s - d_0)}{c}} - \left(c^2 + cd_0\right) - \frac{1}{2}\left(d_0^2 - d_s^2\right)\right] \tag{19.12}$$

in the presence of tri- and tetra-valent counterions. It is important to emphasize that this interaction free energies G_{int} obtained from osmotic stress experiments not only account for the electrostatics of the DNA and counterions, but also for entropic and hydration contributions [58, 60].

Knowing the contributions of the bending and interaction terms, the optimal energy and DNA spacing are then obtained by minimizing the total energy $G_{tot}(L,d_s)$ with respect to d_s for a given total genome length L. The internal pressure P of the confined DNA or the force F at which it will be ejected can be finally obtained from it by their standard thermodynamic definition as $P = -\partial G_{tot}/\partial V$ or $F = -\partial G_{tot}/\partial L$. This model predicts the forces during the ejection process and the internal pressure in bacteriophages [58, 61], which are in the range of 20–60 atmospheres as determined by AFM and osmotic pressure experiments [62, 63] (see Chaps. 9 and 18 and Fig. 19.8). Remarkably, bacteriophages require the strongest molecular motors reported so far to store the dsDNA at such high densities. In many cases this stored energy is used to initially drive the injection of the dsDNA genome into a bacteria during the infection process.

Although the fact that the dsDNA is compacted in an inverse spool configuration is still controversial [56], the estimates of the pressures are not very sensitive to the configuration, since they are mostly controlled by the electrostatic contributions [64] and the packing density of the DNA. In fact, using in Eqs. 19.9 and 19.10 a simple estimate of the strand separation

$$d_s = \sqrt{\frac{2V}{\sqrt{3}}}\frac{1}{\sqrt{L}}, \tag{19.13}$$

which is obtained by equating the total volume available inside the capsid V with that of the closed packed DNA (assuming that it occupies the whole volume), one gets an accurate approximation of the capsid pressure due to the confined DNA.

19.4 Mechanical Stability of Capsids

Several studies indicate that virions can be subjected to substantial mechanical stress, for instance, due to osmotic shock, related to a sudden change of salt concentration in the environment, or to the packaging of the viral genome at high densities (see Chaps. 12 and 18). Interestingly, the last scenario concerns many dsDNA bacteriophages, which have to withstand up to tens of atmospheres arising from the confined viral genome, as we have just seen. Additionally, after the self-assembly of the capsid, to become infective many viruses undergo a maturation process that may properly tune their physical and mechanical properties. In bacteriophages, this maturation often involves the modification of the capsid shape through the so-called *buckling transition*, where an initial spherical procapsid undergoes a transformation into a polyhedral-like shell resembling an icosahedron (see Chap. 13). All these complex scenarios illustrate different sophisticated strategies that viruses have developed to protect their genetic material, remaining infective despite the hazards of the environment. Thus, their mechanical characterization becomes of particular interest in order to understand the stability of viruses.

Very little is known yet on the mechanical properties of viral capsids. A few studies relied in *bulk* experiments using biochemical and physical techniques, like inelastic Brillouin light scattering. However, the recent application of atomic force microscopy (AFM; Chap. 8) and optical tweezers (Chap. 9) have allowed the experimental determination of the mechanical properties and mechanochemical function of individual viral particles (see Chaps. 8, 9 and 18). This has also boosted the interest in developing and applying theoretical frameworks to characterize the mechanical response of viral capsids.

Thus, in this section we will discuss the main theoretical approaches applied for the mechanical characterization of viral capsids. The studies are grouped in three

blocks: first, we introduce some theoretical strategies to investigate the elastic response of viral capsids, which are essentially based on the TST framework and computational simulations; then some physical ideas behind the maturation of viruses are discussed; finally, we briefly review some techniques that, by analyzing thermal vibrations of the capsid, facilitate the extraction of important structural properties and general information about shell transformations.

19.4.1 Elastic Response of Viral Capsids

Elastic Properties from TST

TST constitutes the natural starting framework to study the mechanical properties of viral capsids. This model has been particularly useful to complement AFM indentation experiments (see Chap. 18), facilitating their interpretation and allowing the extraction of the elastic properties of the capsid [65]. To describe the local deformations of the capsids generated in AFM experiments, it is more convenient to rewrite the elastic energy in terms of a free energy functional of the indentation profile $\zeta(\mathbf{r})$ given by $\Delta F = \int dS \left\{ \frac{1}{2}\kappa(\Delta\zeta)^2 + \frac{1}{2}Y\left(\frac{2\zeta}{R}\right)^2 + \frac{1}{2}\tau(\Delta\zeta)^2 \right\}$. The first term describes the bending energy (with $\Delta\zeta$ being the curvature), the second term the stretching energy, and the third term the extra work required to deform the capsid, assuming that the shell is subjected to a tension τ [65]. The indentation profile can be obtained from the differential equation $\delta\Delta F/\delta\zeta(\mathbf{r}) = f(\mathbf{r})$, where $f(\mathbf{r})$ is the applied radial force per unit area. The previous equation is valid only for small deformations and can be solved analytically only in special situations, for instance, in the case of a point force applied on an empty spherical shell. Let's focus first on the scenario where the tension τ, which can be related, for example, to an internal pressure, vanishes.

For indentations much smaller than the shell thickness, TST predicts a linear relation between the applied force and the resulting indentation, as follows from the simple scaling arguments described in Chap. 18. The corresponding effective spring constants (as a measurement of mechanical stiffness, see Chap. 18) for spherical and cylindrical shells are $k_{sph} = \frac{2}{\sqrt{3(1-\nu^2)}}\frac{Eh^2}{R}$, and $k_{cyl} \propto Eh\left(\frac{h}{R}\right)^{3/2}$, respectively [17], where E is the (three-dimensional) Young's modulus, ν is the Poisson ratio, R is the radius of the capsid, and h its thickness. As shown in Chap. 18, using these expressions the Young's modulus of a virus can be derived from the slope of the force indentation curves obtained with the AFM, provided the size and the thickness of the capsid are known.

In addition, as described in Chaps. 9, 18 and Sect. 19.3.3, the presence of the confined dsDNA genome can build up a substantial pressure in the capsid, which may ultimately modify its mechanical response. In a first approximation, the classical Laplace law is able to estimate the tension generated in a shell by a

pressure difference Δp, which for a spherical shell is $\tau = \Delta pR/2$. Using this value in the previous free energy functional, and solving for a point force, one obtains the effective spring constant [66],

$$k = k_{empty} \frac{\pi}{2} \frac{\sqrt{\tilde{\tau}^2 - 1}}{\operatorname{arctan h} \sqrt{\left(1 - 1/\tilde{\tau}^2\right)}}, \tag{19.14}$$

which depends on the ratio $\tilde{\tau} = \Delta pR/k_{empty}$ between the pressure difference Δp multiplied by the radius R and the spring constant k_{empty} of the empty shell. This expression provides a simple way to estimate the internal pressure of spherical viruses by comparing the spring constants of the emptied mature capsid and the virion, provided that the cause of this difference is indeed the existence of internal pressure (see Chap. 18).

It is worth noting that experiments and calculations have shown that tension does not necessarily originate from an internal pressure. For instance, in the icosahedral minute virus of mice, viral ssDNA segments bind to equivalent sites at the capsid inner wall and act like molecular buttresses that reinforce anisotropically the viral particle (see Chap. 18); in the case of bacteriophage Φ29, it has been shown that the capsid is also mechanically anisotropic because of built-in stress [67]. The development of pre-stress is a common strategy in engineered structures and probably reinforces the capsid in order to better tolerate the pressure developed during the packaging of the viral genome.

An important mechanical property of a capsid, especially for dsDNA viruses, is its resistance to pressure before bursting. TST provides simple expressions that relate the largest pressure that a capsid can resist with the maximum relative radial expansion and the maximum tangential stresses. The largest stresses can be in turn evaluated in terms of the inflection point of the potential between subunits, *i.e.*, the maximum force that the interaction can hold [68, 69]. In addition, by simply using the Laplace law, one obtains that the maximum pressure Δp^* that a virus can tolerate, at equal interaction strengths, decays with the radius of the shell as $1/R$. In other words, at equivalent values of the molecular interactions, big capsids should be less efficient tolerating pressures. Notice that this decay also applies for the effective spring constant k, and it can be also shown that this follows for the bulk modulus [20, 70].

Finally, TST and scaling arguments can also be used to characterize large indentations of viral capsids using point-like axial loads. For large deformations the response of the viral capsid becomes strongly non-linear, which in many cases is associated to an inversion of the curvature (*buckle*) in a region of the shell around the indentation point [17, 71]. The elastic analysis shows that for spherical shells this inversion typically takes place when the indentation becomes comparable to the shell thickness h, or alternatively, when the load force exceeds a threshold value $F_{inv} \sim \kappa/R$. In the nonlinear deformation regime, the force-indentation curve is expected to show a squared root dependence on the indentation [17, 71]. However,

in these situations the characterization of the indentation using TST becomes very challenging, and other techniques, like finite element analysis (FEA) discussed in Chap. 18, are more suitable. FEA studies have demonstrated that capsids become stiffer and show higher non-linearities (associated to local bulging or buckling) for high FvK numbers [19, 65]. The relevance of the finite thickness and non-uniformity of viral shells has been also explored, for instance, in the softening of CCMV after its pH swelling transition, or in the anisotropic response of HBV when indented at the different axis of icosahedral symmetry [19, 65].

Discrete Models and Simulations

However, FEA is essentially a numerical solution of continuum elasticity equations, and accordingly it shares also certain limitations. For instance, the plasticity or the breaking of viral shells observed after repeated indentations cannot be captured. In addition, the discrete and possible anisotropic interactions between capsid building blocks cannot be easily incorporated in this description. To overcome these limitations, one has to resort to the use of discrete models.

Unfortunately, contrary to the case of the structure and assembly of viral capsids (described in Sects. 19.2 and 19.3), there have not been yet many models of this kind studying the deformation of viral capsids. The most popular simulation methods have been based on the discrete triangularization of TST, described in Sect. 19.1 [72], which have been able to reproduce the non-linearities in the indentation of viral capsids and estimate mechanical differences between some T-numbers. It is worth noticing that this should not be confused with FEA, which discretizes the capsid in tiny elements that do not correlate with the positioning of the capsid proteins in the shell.

Also at the low-resolution level, the model of Ref. [69] has highlighted the importance of the particular arrangement of the capsomers and the triangulation number on the mechanical properties of spherical capsids. At the high-resolution level, Schulten's group has recently developed a coarse-grained model of HBV complementing AFM experiments. The simulations were in agreement with the experiments without the necessity of any type of fitting [65, 73]. One of the most interesting features of this coarse-grained approach was the identification of the local rearrangements of capsid proteins and the irreversible conformational changes produced during the indentation of HBV [74], which was not captured by the FEA studies of HBV.

19.4.2 Maturation and Buckling

As described in Chap. 13, to become infective many viruses require a maturation step after the self-assembly of the procapsid. This process seems to tune different properties of the viral shell that are manifested in many cases by a noticeable

change in the capsid shape, generally associated to a *buckling transition*, which is of particular relevance among bacteriophages. In this transformation, the initial spherical procapsid with icosahedral symmetry undergoes a transition into a polyhedral shell with flatter triangular faces and an overall shape resembling an icosahedron.

TST provides again an excellent framework to establish the physical basis of the buckling transition [4]. In this context the asphericity of the capsid, *i.e.*, the degree of buckling, is exclusively determined by the FvK number, $\gamma = YR^2/\kappa$, which compares the stretching and bending contributions in the shell. When FvK exceeds a certain threshold value, $\gamma_B \sim 10^2$, the stretching energy of a spherical shell can be reduced if the 12 pentamers buckle out inducing the faceting of the shell into a polyhedral shape (see Fig. 19.2). Remarkably, it can be shown that within TST, the FvK number can be expressed as $\gamma = 12(1-\nu^2)(R/h)^2$, which means that it essentially depends on the square of the ratio between the capsid radius R and its thickness h. This suggests that big and thin shells should be more likely to adopt polyhedral rather than spherical shapes, in agreement with experimental observations of the faceted nature of large viruses [70].

Alternatively, a low-resolution discrete model has shown that the buckling transition depends on the triangulation number T and the icosahedral class[2] P of the virus structure [70]. In particular, P = 1 shells are most likely to produce polyhedral shapes, whereas viruses with P = 3 prefer to remain spherical (see Fig. 19.9). For big capsids the polyhedral shape becomes systematically the most stable, in agreement with TST. This study showed that in general the buckling transition can be also induced by a capsid expansion, in consonance with virus maturation. Furthermore, in comparison with the spherical shape, the resulting icosahedral shell is mechanically stiffer, tolerates larger expansions, and withstands higher internal pressures before failing. This suggests that the polyhedral shape could have a certain advantage for dsDNA phages that rely on the pressurization of their genetic material to be infective, justifying their buckling transition.

However, despite the very useful insights provided by the theoretical tools described above, one should not forget that the occurrence of the buckling transition during virus maturation is a complex process that often involves different molecular mechanisms including conformational changes, cleavage, and even the formation of covalent bonds. These mechanisms are virus-specific and cannot be easily characterized by most of the models described previously.

[2] The icosahedral class P was introduced by Caspar and Klug to classify triangulation numbers and it is related to them by $T = h^2 + k^2 + hk = Pf^2$, where f is the greatest common divisor of the integers h and k. It should not be confused with the pseudo-triangulation number P, used to label icosahedral viruses when the subunits that occupy quasiequivalent positions are chemically or structurally different (see Chap. 2).

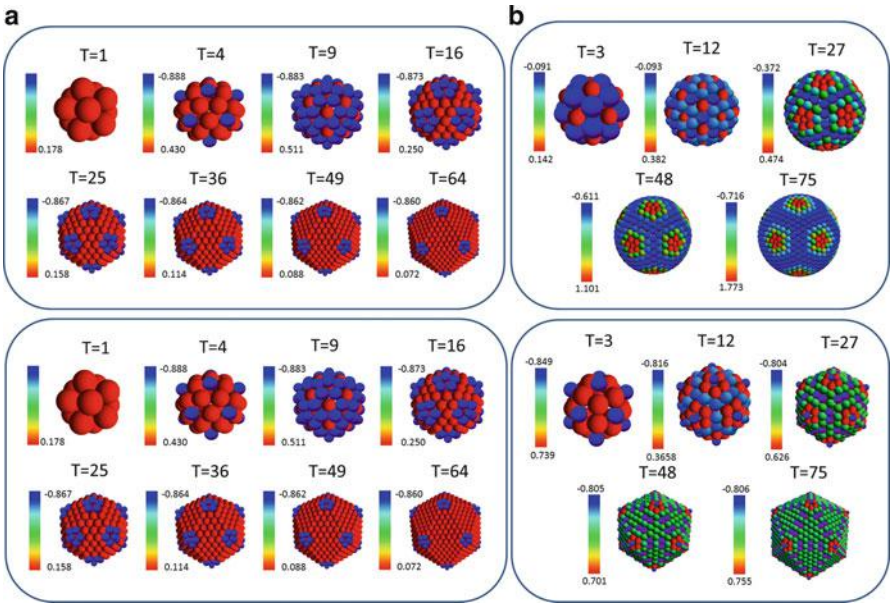

Fig. 19.9 Relevance of capsid structure in the buckling of spherical viruses. (**a**) The distribution of local stresses (indicated by the *color scales*) for spherical (*top panel*) and polyhedral (faceted, *bottom panel*) T-number shells belonging to the class P = 1. Capsomers in blue are stretched, whereas those in *red* are compressed. In the spherical capsids, positive stress concentrates on the lines connecting pentamers for T > 4, delimiting triangular areas with stretched hexamers. The absolute values of these stresses get larger as the T-number of the shell increases. The hexamer stretching is relieved if the triangular regions between pentamers are flattened; thus, these structures have a strong preference for buckling. (**b**) Same for spherical and polyhedral (faceted) T-number shells belonging to the class P = 3. In spherical P = 3 shells, compression concentrates on the pentamers, whereas the hexamers surrounding them are slightly stretched, leading to a dodecahedral pattern instead of the icosahedral one observed in P = 1. In this case, nothing indicates any particular relief of the stress on the hexamers upon adopting an icosahedral shape (Reproduced from Ref. [70])

19.4.3 Viral Capsid Vibrations

An interesting fundamental property of materials is the relation between the unavoidable thermal vibrations that they suffer at finite temperatures and their structural properties. In the case of capsids, this has been largely explored applying normal mode analysis (NMA) of elastic networks based on all-atom or cryo-EM structural models of specific viruses [75]. Essentially, in NMA the interaction between atoms or pseudo-atoms is approximated by a harmonic potential, and the individual vibrations of the subunits are characterized in terms of collective harmonic oscillations, called normal modes, which can be obtained by diagonalizing the second derivative of the potential energy (Hessian) [76]. In this description the

low frequency modes are associated with conformational changes in the capsid, and the high frequency modes describe localized motions.

In this way, NMA has allowed the study of the transition during the swelling of CCMV, and the maturation of *Nudaurelia capensis* ω virus and phage HK97 [76]. Recent investigations, combining high-resolution simulations and NMA, have shown that the maturation of HK97 follows a pathway dominated by icosahedral symmetry [77]. Remarkably, the combination of full-atom simulations and elastic network models constitutes a promising alternative method to obtain the Young's modulus of viruses that should complement the values obtained by the combination of TST and AFM [78]. Interestingly, these multiscale techniques have recently shown that cross-linking between capsomers during the maturation of HK97 is responsible for the increase of the shell resistance to breaking without noticeable changes in the Young's modulus [79].

19.5 Genome Delivery and Virus Egress

In this section we are going to discuss some of the physics behind particular mechanisms of genome delivery, and virus exit from and entry into cells. Once again, the relative simplicity and lack of an own metabolic machinery of viruses have led to biological solutions that strongly rely on basic physical processes. First, we describe, in physical terms, the injection of the genome of dsDNA bacteriophages into the host cell. Then, a mechanism of capsid disassembly is discussed in the context of CNT, which is important for many ssRNA viruses. Finally, we describe the modeling of the endocytosis and budding processes, which respectively mediate the entry and exit of many viral particles in the host.

19.5.1 Active and Passive Translocation

Bacteriophage infection relies on the translocation of the viral genome into the bacteria, leaving an intact empty capsid in the extracellular environment. In this infection pathway, the genetic material must rapidly get inside the cell, and the efficiency of the process is accomplished by a combination of different mechanisms, some of them not fully characterized yet (see Chap. 17).

In order to succeed in this translocation pathway, many bacteriophages rely on a proteinaceous tail that is attached to the capsid (head) and helps to initiate and direct the injection of their genomes through the bacterial wall and membrane (see Chap. 17). This process can involve complex conformational changes in the tail proteins requiring energy. A typical example is the contractile tail of bacteriophage T4. The contraction of the tail sheath drives the rigid inner tube through the bacterial wall and allows the virus to inject its DNA into the bacteria (see Chap. 17).

This contraction has been described in the context of physics as a *martensitic transition* [80, 81], a well-known phase transition occurring typically in solids. The idea is that the tail proteins can adopt two possible conformational arrangements: one extended and strained, and the other contracted and relaxed. A small activation barrier separates both states, where the initial one (the extended) has a higher energy. When the tail attaches to the membrane it triggers a rapid transition that releases the elastic energy stored by the proteins in the extended conformation adopted during the assembly of the virus. The contraction force derived from these studies is about 100 pN, roughly twice the stalling force of the motor packaging the DNA of Φ29 (see Chap. 9). This type of martensitic mechanism seems to be also behind other large conformational changes in the fusion proteins of different enveloped viruses, such as HIV and influenza virus [80] (see Chap. 16).

After a triggering event the dsDNA starts to enter the cell. As mentioned in Sect. 19.3.3, the capsid internal pressure drives the initial ejection of the genetic material. The initial forces and velocities of this entrance can be properly quantified by the inverse spool model, but after roughly the 50 % entrance of the DNA the pressure difference vanishes, and the remaining genetic material relies on other mechanisms to get in (see Chap. 12).

Different physical strategies can be conceived to explain the entrance of the remaining viral genome length inside the cell. The simplest one is by pure diffusion. In this case, thermal fluctuations shake the nucleic acid back and forth through the cell wall pore until the genome completely enters into the host. However, the translocation time of this mechanism scales as the square of the genetic material length L leading to $\tau_{dif} = L^2/D$, where D is the diffusion coefficient of the chain, and for typical viral parameters this time becomes prohibitively long. Alternatively, a second mechanism involving the presence of binding proteins has been studied to explain how the remaining viral genome is driven into the cell. Here, as soon as a portion of DNA gets inside the cell, certain proteins bind to it and impede its diffusion backwards, which rectifies the diffusion of the chain and speeds up the entrance by a factor proportional to the number of binding sites. Nevertheless, despite this rectification, the entrance time estimated for this mechanism is still too long. These two simple strategies cannot explain the entrance of the viral genome. However, it was recently shown that the proteins that bind to the DNA not only rectify its diffusion but simultaneously exert an effective translocating force that considerably increases the speed of the DNA entrance [82]. This additional force was characterized by a simple model evaluating the free energy of a stiff portion of DNA entering a spherical cell in the presence of nonspecific binding proteins (see Fig. 19.10). This simple model shows that the total free energy becomes more favorable the more genetic material gets inside the cell translating into a powerful effective force that drives the translocation. The effective force observed in coarse-grained simulations was also quantified using this free energy in a simple non-equilibrium kinetic model. Interestingly, this speed up effect associated to the binding energy of proteins has been verified in *in vitro* experiments that measure the kinetics of ejection (in that case in the presence of DNAse) [83]. On the other hand, although the model can also be refined by accounting for the entropy of the

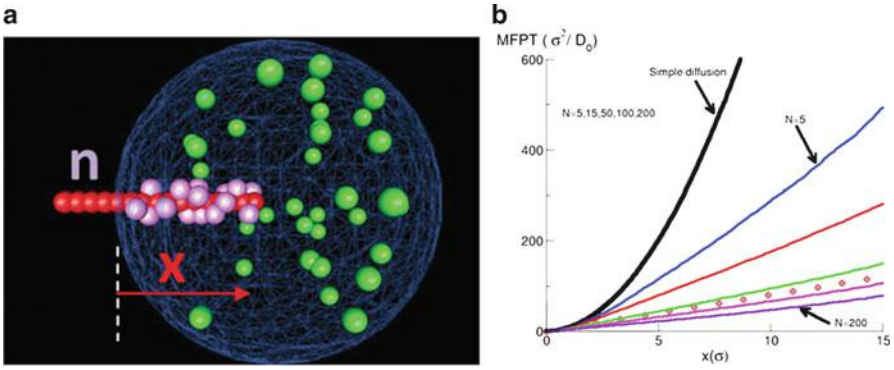

Fig. 19.10 Model for the translocation of a stiff portion of dsDNA in the presence of binding proteins. (**a**) Schematic illustration of the simulation setup, where a stiff portion of dsDNA (in *red*) enters a cell (in *blue*) in the presence of binding proteins (*green* (unbound) and *violet* (bound)). (**b**) Average time required for the entrance of a given distance x (measured in units of the bead size σ) by simple diffusion (*black line*) and in presence of different number N of binding proteins. For a sufficiently large concentration, the entrance becomes even faster than by ratcheted diffusion (*red dots*) (Reproduced from Ref. [82])

DNA chain [61], there are still many puzzling features of this translocation process of bacteriophages to be described, as has been shown by recent experimental studies [84].

DNA polymerases can also play an active role pulling in the DNA of some viruses, which has been studied experimentally with optical tweezers (see Chap. 9) and analyzed with stochastic models.

19.5.2 Capsid Disassembly

Eukaryotic viruses and some bacteriophages typically follow an infection pathway that requires virions to penetrate the host. Once inside, the capsid of these viruses have to change its conformation or, in many cases, disassemble to allow the release of the genetic material in specific regions of the cell, which will ultimately initiate the viral replication factory (see Chap. 14).

In many cases, viruses follow a post-assembly maturation stage that strengthens the initial weak interactions responsible for the assembly (see Chap. 13). Nevertheless, other viruses do not necessarily go through a maturation process, but they have still to avoid disassembly during their trip to infect a new host. If the assembly of viruses were an equilibrium process, viruses would disassemble as soon as the capsid protein concentration drops below the threshold value c^*. In this context the models introduced in Sect. 19.3.1 can be invoked. In particular, the CNT offers a good explanation of how viruses may accomplish this delicate balance. It is precisely the existence of an activation barrier that had to be overcome during the

assembly what prevents their disassembly. Then the process of viral disassembly can be though as the inverse of the assembly and can be mathematically described using the same CNT theory [20].

The disassembly of viruses like HBV has been probed by *in vitro* experiments (see Fig. 19.5a). This virus does not suffer any maturation or conformational change and the HBV capsid proteins can be reassembled repeatedly. In these experiments, one starts from a given total concentration of proteins in the form of complete capsids, and dilutes this concentration monitoring the disappearance of capsids. The results of these experiments have shown that the threshold concentration at which capsids disassemble, c_1^{dis}, is much lower than c_1^{as}, the value at which they start to assemble. This can be easily understood and quantified in the framework of CNT.

To initiate the disassembly, one has now to jump the barrier in the opposite direction respect the assembly (see Fig. 19.6b). The height of this disassembly barrier is given by $\Delta G^*_{dis} = \Delta G^*_{as} - \Delta G(q)$, where ΔG^*_{as} is given by Eq. (19.5). The disassembly rate depends also exponentially on ΔG^*_{dis} as in Eq. (19.6). At the conditions at which the assembly took place, this barrier is too high to be surmounted and capsids remain stable. In fact, positive $\Delta\mu = -k_B T \ln(c/c^*)$, implying c << c* are needed to lower the disassembly barrier up to a point that disaggregation is feasible. Since it is the height of the assembly and disassembly barriers what dominates the kinetics of both processes, a simple argument can be made to get a reasonable estimate of what would be the concentration required for disassembly. Assuming that the height of the barrier towards disassembly has to be the same than that of the assembly (to observe the processes at similar time scales) one gets the following simple expression

$$(c^*)^2 = c_1^{as} c_1^{dis} \tag{19.15}$$

that relates the threshold protein concentrations required for the assembly, c_1^{as}, and disassembly, c_1^{dis}, with the equilibrium concentration c^*, and through Eq. (19.1), with the binding energy per subunit. Thus, these simple ideas provide a way to extract useful information combining assembly and disassembly experiments [20].

19.5.3 Budding and Endocytosis

Many animal viruses require also crossing in and out cellular membranes to proceed with the infection. In enveloped viruses the lipid membrane and fusion proteins help to accomplish this task (see Chap. 16), and non-enveloped viruses are generally able to get inside the cell through endocytosis (see Chap. 15). Different physical factors control the feasibility of these processes: the membrane elasticity, the adhesion energy between the capsid and the membrane, the spontaneous thermal fluctuations, and the diffusion of spike proteins. Accordingly, several physical models have been proposed to address these interesting problems.

After the complete assembly of the nucleocapsid, most enveloped viruses acquire their envelope during the exit process from the cell by wrapping and pinching off of the host membrane. However, this deformation of the membrane entails an energetic cost. The first mechanism suggested to accomplish that was the spontaneous wrapping of the virus taking advantage of thermal fluctuations of the membrane [85]. However, the energy provided by thermal fluctuations alone is not enough to wrap completely the virus unless the membrane stiffness is low [86]. To overcome this cost, the virus counts on the adhesion energy between the membrane and its spike or capsid proteins. Several models accounting for these factors have been proposed to study the competence between these contributions and the optimal conditions for virus exit.

This process has been characterized by analyzing the wrapping of a spherical particle by an elastic membrane [87]. The elastic deformation of the membrane is described using a continuum elasticity model (see Sect. 19.2.2), where an additional term accounting for the adhesion energy between the virus and the membrane is introduced. This model has shown that a minimum value of the adhesion energy is required to have full wrapping, and that the process is very sensitive to the size of the viral particle. These results suggest the possibility of interfering with the entrance or budding of viruses by modifying the adhesion energy or the membrane tension. Recent simulations show that elasticity theory can account quantitatively for different mechanisms of budding, although some situations controlled by the kinetics of the process are not captured [88].

An extension of these ideas has been performed to account for the concurrent assembly and budding processes of HIV [89], ascertaining the conditions in terms of Gag-Gag interaction strength and membrane rigidity required to have complete budding.

During the budding and exit from an infected cell, many enveloped viruses embed in their lipid membrane spike-shaped proteins that play a crucial role to infect another cell. Statistical models of viral budding, which incorporate the entropy and distribution of spikes to the elastic and adhesion effective free energy terms, have also been proposed [90]. These studies analyzed the simultaneous formation of many buds and the competence of viruses to sequester the spike proteins. The results show that, for biologically reasonable values of the membrane elastic constant and viruses with relatively strong spike binding strength, all binding sites of the capsid get occupied by spikes upon successful budding.

Interestingly, with slight modifications, models developed to study budding can also be applied to study endocytosis. The endocytosis of the virus is mediated by the binding interactions between ligand molecules on the viral capsid and their receptor molecules on the cell membrane. These receptors are mobile and have to diffuse in the membrane to wrap around the virus and facilitate its entrance. In some case, this is the rate-limiting step for virus entry, and the corresponding kinetics can be modeled using a diffusion equation, as done in Ref. [91]. These studies show that particles in the size range of tens to hundreds of nanometers can enter (or exit) cells *via* wrapping without requiring the help of clathrin or caveolin. The optimal size to

get an optimal wrapping time was characterized, and turned out to be in the range of 25–30 nm in radius, which is quite similar to the size of many spherical viruses.

19.6 Perspectives and Conclusions

Theoretical models and simulations are starting to make a significant contribution to our understanding of how viruses function. As we have seen in this chapter, the regular architecture of viruses obeys general geometrical rules and basic physical principles. For instance, the icosahedral symmetry of spherical viruses and the structure of prolate viruses are a consequence of free energy minimization of generic interactions between the capsid proteins. In addition, the study of continuum elastic theories, and TST in particular, has established a reference framework to rationalize the distinguishing morphologies and the faceting of viral capsids, based on the minimization of the effective elastic and bending energies. Complementing TST, coarse-grained models have shown the dependence of buckling on the triangulation number, and high-resolution NMA have described basic aspects of capsid maturation for several viruses. The combination of TST, coarse-grained, FEA, and NMA simulations has been invaluable to characterize the remarkable elastic properties of capsids, facilitating the interpretation of AFM experiments.

In the context of viral assembly, CNT and kinetic models provide a solid background to explain the physical mechanisms involved in the assembly and disassembly of capsids. In CNT, the most important element is the existence of an energy barrier controlling the formation of capsids, which justifies the main features observed experimentally, *i.e.*, the scarcity of intermediates, the existence of a lag time at the beginning of the assembly, its sigmoidal kinetics, and the hysteresis in the disassembly. Coarse-grained simulations have started to complement CNT and other models by exploring the pathways of capsid assembly and the capsid-genome co-assembly of some ssRNA viruses, and they show that the genetic material is able to reduce the energy barrier of capsid formation. In addition, simulations have shown that the inhomogeneous distribution of RNA inside the capsid is a generic feature due to the self-repulsion of this highly flexible genetic material; several models have also justified the relation of charges between the genetic material and capsid proteins, the relation between the capsid size and the encapsulated cargo, the assembly competition of different ssRNA chains, or the ordered structure of the RNA layer in contact with the interior of the shell.

Alternatively, for some dsDNA viruses, it has been possible to estimate the pressures built up during the packaging of the genetic material and the corresponding ejection forces, which assist the initial stage of the dsDNA entry inside a new host cell. The translocation of the remaining material can be directed by either a nonspecific physical mechanism, like the binding of proteins to the viral genome, or by the pulling action of a polymerase. In the first case, the speed up due to the binding of proteins is associated to the lowering of the free energy with the

length inside and can be described by a simple adsorption model. Finally, statistical thermodynamical models are starting to describe the general physical mechanisms involved in the recruiting of the viral envelope and the membrane viral proteins, as well as the exit and entry of some viral particles by budding and endocytosis, respectively. The results indicate that viruses have the right size to efficiently favor the spontaneous occurrence of these processes.

Theoretical physical virology is a relatively young field, and most of the *smart* physical strategies developed by viruses to accomplish different stages in their life cycle are only starting to be unveiled. At the structural level, the understanding of irregular and conical capsids, and the mechanisms involved in the selection of a specific structure by a particular virus are major challenges to be faced. Regarding the mechanical stability of viruses, the main goal is to overcome the limitations of continuum techniques characterizing the elastic properties of viral shells. For this purpose, the development of coarse-grained discrete models, which incorporate the anisotropy of the interactions between the capsid building-blocks, is very timely and will facilitate a more accurate description of the protective role of the capsid, including the rich mechanical phenomenology obtained in AFM experiments. In the context of assembly, a clear direction is the further development of CNT to reach a quantitative and predictive understanding of assembly experiments on specific viruses, and its extension to explicitly incorporate the effects of the genetic material and auxiliary (scaffolding and/or condensing) proteins. Simulations will play a very important role to test these theories and to gain further insight into the molecular details of this process. Finally, regarding the delivery of the genetic material and the egress of viral particles, simulations and theory are needed to clarify specially the nonequilibrium and kinetic aspects of these processes.

Obviously, the answer to all these questions will require a close collaboration between physicists and molecular virologists. In this endeavor, the main challenge faced by theory is to properly adapt the main physical ideas into a precise biological context and relate them to accessible experimental magnitudes. This should allow us to progressively incorporate other important biological information, demonstrating the utility of theories in the prediction, guidance, and interpretation of experiments. With this effort, physics would provide not only a theoretical understanding of viruses but also a valuable tool in the development of novel antiviral strategies and useful applications.

Acknowledgements We acknowledge useful discussions with María Aznar, Robijn Bruinsma, Mauricio Comas-García, William Gelbart, William Klug, Charles Knobler, Joseph Rudnick, Jan Wedekind, and Roya Zandi. We also thank Charles Knobler for a careful reading of the manuscript and Alexander Bittner, and Carmen San Martín for fruitful suggestions. This work has been supported by the MICINN of the Spanish government through the I3 Program and grant No. FIS2008-01299 co-financed by European Union's FEDER funds. We also thank the support of the "Spanish Interdisciplinary Network on the Biophysics of Viruses" (BioFiVinet) (grants No. FIS2010-10552-E and FIS2011-16090-E).

References and Further Reading

1. Crick F, Watson J (1956) Structure of small viruses. Nature 177:473–475
2. Caspar D, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol 27:1–24
3. Luque A, Reguera D (2010) The structure of elongated viral capsids. Biophys J 98:2993–3003
4. Lidmar J, Mirny L, Nelson DR (2003) Virus shapes and buckling transitions in spherical shells. Phys Rev E 68:051910
5. Twarock R (2006) Mathematical virology: a novel approach to the structure and assembly of viruses. Philos Trans A Math Phys Eng Sci 364:3357–3373
6. Zandi R, Reguera D, Bruinsma RF et al (2004) Origin of icosahedral symmetry in viruses. Proc Natl Acad Sci USA 101:15556–15560
7. Luque A, Zandi R, Reguera D (2010) Optimal architectures of elongated viruses. Proc Natl Acad Sci USA 107:5323–5328
8. Fejer S, James T, Hernandez-Rojas J, Wales D (2009) Energy landscapes for shells assembled from pentagonal and hexagonal pyramids. Phys Chem Chem Phys 11:2098–2104
9. Fejer S, Chakrabarti D, Wales D (2010) Emergent complexity from simple anisotropic building blocks: shells, tubes, and spirals. ACS Nano 4:219–228
10. Rapaport DC (2010) Modeling capsid self-assembly: design and analysis. Phys Biol 7:045001
11. Elrad OM, Hagan MF (2010) Encapsulation of a polymer by an icosahedral virus. Phys Biol 7:045003
12. Nguyen HD, Reddy VS, Brooks CL (2009) Invariant polymorphism in virus capsid assembly. J Am Chem Soc 131:2606–2614
13. Arkhipov A, Freddolino PL, Schulten K (2006) Stability and dynamics of virus capsids described by coarse-grained modeling. Structure 14:1767–1777
14. Freddolino PL, Arkhipov AS, Larson SB et al (2006) Molecular dynamics simulations of the complete satellite tobacco mosaic virus. Structure 14:437–449
15. Keef T, Twarock R (2008) New insights into viral architecture *via* affine extended symmetry groups. Comput Math Meth Med 9:221–229
16. Bruinsma RF, Gelbart WM, Reguera D et al (2003) Viral self-assembly as a thermodynamic process. Phys Rev Lett 90:248101
17. Landau LD, Lifshitz EM (1975) Theory of Elasticity. Pergamon Press, London
18. Nguyen TT, Bruinsma RF, Gelbart WM (2005) Elasticity theory and shape transitions of viral shells. Phys Rev E 72:051923
19. Klug WS, Bruinsma RF, Michel J-P, Knobler CM (2006) Failure of viral shells. Phys Rev Lett 97:228101
20. Luque A (2011) Structure, mechanical properties, and self-assembly of viral capsids. Ph.D. Thesis, University of Barcelona
21. Chen T, Glotzer S (2007) Simulation studies of a phenomenological model for elongated virus capsid formation. Phys Rev E 75:051504
22. Saunders M, Voth G (2012) Coarse-graining of multiprotein assemblies. Curr Opin Struct Biol 20:144–150
23. Krishna V, Ayton GS, Voth GA (2010) Role of protein interactions in defining HIV-1 viral capsid shape and stability: a coarse-grained analysis. Biophys J 98:18–26
24. Chen B, Tycko R (2011) Simulated self-assembly of the HIV-1 capsid: protein shape and native contacts are sufficient for two-dimensional lattice formation. Biophys J 100:3035–3044
25. Miao Y, Johnson JE, Ortoleva PJ (2010) All-atom multiscale simulation of cowpea chlorotic mottle virus capsid swelling. J Phys Chem B 114:11181–11195
26. Devkota B, Petrov AS, Lemieux S et al (2009) Structural and electrostatic characterization of pariacoto virus: implications for viral assembly. Biopolymers 91:530–538
27. Lavelle L, Gingery M, Phillips M et al (2009) Phase diagram of self-assembled viral capsid protein polymorphs. J Phys Chem B 113:3813–3819
28. Hu Y, Zandi R, Anavitarte A et al (2008) Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size. Biophys J 94:1428–1436

29. Dixit SK, Goicochea NL, Daniel M-C et al (2006) Quantum dot encapsulation in viral capsids. Nano Lett 6:1993–1999
30. Kegel WK, van der Schoot P (2004) Competing hydrophobic and screened-Coulomb interactions in hepatitis B virus capsid assembly. Biophys J 86:3905–3913
31. Verwey EJW, Overbeek JTG (1999) Theory of stability of lyophobic colloids. Dover Press, Minneola
32. Siber A, Božič AL, Podgornik R (2012) Energies and pressures in viruses: contribution of nonspecific electrostatic interactions. Phys Chem Chem Phys 14:3746–3765
33. Ceres P, Zlotnick A (2002) Weak protein−protein interactions are sufficient to drive assembly of hepatitis B virus capsids. Biochemistry 41:11525–11531
34. Singh S, Zlotnick A (2003) Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly. J Biol Chem 278:18249–18255
35. Zlotnick A (2003) How does your virus grow? Understanding and interfering with virus assembly. Trends Biotechnol 21:536–542
36. Casini GL, Graham D, Heine D et al (2004) *In vitro* papillomavirus capsid assembly analyzed by light scattering. Virology 325:320–327
37. Chen C, Kao CC, Dragnea B (2008) Self-assembly of brome mosaic virus capsids: insights from shorter time-scale experiments. J Phys Chem A 112:9405–9412
38. Zandi R, van der Schoot P, Reguera D et al (2006) Classical nucleation theory of virus capsids. Biophys J 90:1939–1948
39. Schwartz R, Shor PW, Prevelige PE, Berger B (1998) Local rules simulation of the kinetics of virus capsid self-assembly. Biophys J 75:2626–6636
40. Berger B, Shor P, Tucker-Kellogg L, King J (1994) Local rule-based theory of virus shell assembly. Proc Natl Acad Sci USA 91:7732–7736
41. Keef T, Micheletti C, Twarock R (2006) Master equation approach to the assembly of viral capsids. J Theor Biol 242:713–721
42. Moisant P, Neeman H, Zlotnick A (2010) Exploring the paths of (virus) assembly. Biophys J 99:1350–1357
43. Wilber AW, Doye JPK, Louis AA et al (2007) Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. J Chem Phys 127:085106
44. Hagan MF, Chandler D (2006) Dynamic pathways for viral capsid assembly. Biophys J 91:42–54
45. Hagan MF, Elrad OM (2010) Understanding the concentration dependence of viral capsid assembly kinetics–the origin of the lag time and identifying the critical nucleus size. Biophys J 98:1065–1074
46. Morozov A, Rudnick J, Bruinsma R, Klug W (2010) Assembly and disassembly of deltahedral viral shells. In: Stockley P, Twarock R (eds) Emerging topics in physical virology. Imperial College Press, London, pp 159–183
47. Levandovsky A, Zandi R (2009) Nonequilibirum assembly, retroviruses, and conical structures. Phys Rev Lett 102:198102
48. Hicks SD, Henley CL (2006) Irreversible growth model for virus capsid assembly. Phys Rev E 74:031912
49. Bruinsma RF (2006) Physics of RNA and viral assembly. Eur Phys J E Soft Matter 19:303–310
50. Ting CL, Wu J, Wang Z-G (2011) Thermodynamic basis for the genome to capsid charge relationship in viral encapsidation. Proc Natl Acad Sci USA 108:16986–16991
51. van der Schoot P, Bruinsma R (2005) Electrostatics and the assembly of an RNA virus. Phys Rev E 71:061928
52. Zandi R, van der Schoot P (2009) Size regulation of ss-RNA viruses. Biophys J 96:9–20
53. Cadena-Nava RD, Comas-Garcia M, Garmann RF et al (2012) Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. J Virol 86:3318–3326
54. ElSawy KM, Caves LSD, Twarock R (2011) On the origin of order in the genome organization of ssRNA viruses. Biophys J 101:774–780
55. Mahalik JP, Muthukumar M (2012) Langevin dynamics simulation of polymer-assisted virus-like assembly. J Chem Phys 136:135101

56. Harvey S, Petrov A, Devkota B (2009) Viral assembly: a molecular modeling perspective. Phys Chem Chem Phys 11:10541–10542
57. Odijk T (1998) Hexagonally packed DNA within bacteriophage T7 stabilized by curvature stress. Biophys J 75:1223–1227
58. Purohit PK, Inamdar MM, Grayson PD et al (2005) Forces during bacteriophage DNA packaging and ejection. Biophys J 88:851–866
59. Kindt J, Ben-Shaul A, Tzlil S, Gelbart WM (2001) DNA packaging and ejection forces in bacteriophage. Proc Natl Acad Sci USA 98:13671–13674
60. Rau D, Lee B, Parsegian V (1984) Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: hydration forces between parallel DNA double helices. Proc Natl Acad Sci USA 81:2621–2625
61. Inamdar MM, Gelbart WM, Phillips R (2006) Dynamics of DNA ejection from bacteriophage. Biophys J 91:411–420
62. Smith DE, Tans SJJ, Smith SBB et al (2001) The bacteriophage phi29 portal motor can package DNA against a large internal force. Nature 413:748–752
63. Evilevitch A, Lavelle L, Knobler CM et al (2003) Osmotic pressure inhibition of DNA ejection from phage. Proc Natl Acad Sci USA 100:9292–9295
64. Cordova A, Deserno M, Gelbart WM, Ben-Shaul A (2003) Osmotic shock and the strength of viral capsids. Biophys J 85:70–74
65. Roos WH, Bruinsma R, Wuite GJL (2010) Physical virology. Nat Phys 6:733–743
66. Vella D, Ajdari A, Vaziri A, Boudaoud A (2012) The indentation of pressurized elastic shells: from polymeric capsules to yeast cells. J R Soc Interface 9:448–455
67. Carrasco C, Luque A, Hernando-Pérez M et al (2011) Built-in mechanical stress in viral shells. Biophys J 100:1100–1108
68. Purohit PK, Kondev J, Phillips R (2003) Mechanics of DNA packaging in viruses. Proc Natl Acad Sci USA 100:3173–3178
69. Zandi R, Reguera D (2005) Mechanical properties of viral capsids. Phys Rev E 72:021917
70. Aznar M, Luque A, Reguera D (2012) Relevance of capsid structure in the buckling and maturation of spherical viruses. Phys Biol 9:036003
71. Buenemann M, Lenz P (2008) Elastic properties and mechanical stability of chiral and filled viral capsids. Phys Rev E 78:051924
72. Vliegenthart GA, Gompper G (2006) Mechanical deformation of spherical viruses with icosahedral symmetry. Biophys J 91:834–841
73. Roos WH, Gibbons MM, Arkhipov A et al (2010) Squeezing protein shells: how continuum elastic models, molecular dynamics simulations, and experiments coalesce at the nanoscale. Biophys J 99:1175–1181
74. Arkhipov A, Roos WH, Wuite GJL, Schulten K (2009) Elucidating the mechanism behind irreversible deformation of viral capsids. Biophys J 97:2061–2069
75. Tama F, Wriggers W, Brooks CL (2002) Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. J Mol Biol 321:297–305
76. Gibbons MM, Klug WS (2007) Mechanical modeling of viral capsids. J Mater Sci 42:8995–9004
77. May ER, Feng J, Brooks CL (2012) Exploring the symmetry and mechanism of virus capsid maturation *via* an ensemble of pathways. Biophys J 102:606–612
78. May E, Brooks C (2011) Determination of viral capsid elastic properties from equilibrium thermal fluctuations. Phys Rev Lett 106:188101
79. Roos W, Gertsman I, May E et al (2012) Mechanics of bacteriophage maturation. Proc Natl Acad Sci USA 109:2342–2347
80. Olson G (1999) New directions in martensite theory. Mater Sci Eng A-Struct 273–275:11–20
81. Falk W, James R (2006) Elasticity theory for self-assembled protein lattices with application to the martensitic phase transition in bacteriophage T4 tail sheath. Phys Rev E 73:011917
82. Zandi R, Reguera D, Rudnick J, Gelbart WM (2003) What drives the translocation of stiff chains? Proc Natl Acad Sci USA 100:8649–8653

83. Evilevitch A (2006) Effects of condensing agent and nuclease on the extent of ejection from phage lambda. J Phys Chem B 110:22261–22265
84. Grayson P, Han L, Winther T, Phillips R (2007) Real-time observations of single bacteriophage lambda DNA ejections *in vitro*. Proc Natl Acad Sci USA 104:14652–14657
85. Simons K, Garoff H (1980) The budding mechanisms of enveloped animal viruses. J Gen Virol 50:1–21
86. Lerner DM, Deutsch JM, Oster GF (1993) How does a virus bud? Biophys J 65:73–79
87. Deserno M (2004) Elastic deformation of a fluid membrane upon colloid binding. Phys Rev E 69:031903
88. Ruiz-Herrero T, Velasco E, Hagan MF (2012) Mechanisms of budding of nanoscale particles through lipid bilayers. J Phys Chem B 116:9595–9603
89. Zhang R, Nguyen TT (2008) Model of human immunodeficiency virus budding and self-assembly: role of the cell membrane. Phys Rev E 78:051903
90. Tzlil S, Deserno M, Gelbart WM, Ben-Shaul A (2004) A statistical-thermodynamic model of viral budding. Biophys J 86:2037–2048
91. Gao H, Shi W, Freund LB (2005) Mechanics of receptor-mediated endocytosis. Proc Natl Acad Sci USA 102:9469–9474

Further Reading

Israelachvili J (2011) Intermolecular and surface forces, 3rd edn. Academic Press, London
Katen S, Zlotnick A (2009) The thermodynamics of virus capsid assembly. Methods Enzymol 455:395–417
Knobler CM, Gelbart WM (2009) Physical chemistry of DNA viruses. Annu Rev Phys Chem 60:367–383
Marenduzzo D, Micheletti C, Orlandini E (2010) Biopolymer organization upon confinement. J Phys Condens Matter 22:283102
Nurmemmedov E, Castelnovo M, Catalano CE, Evilevitch A (2007) Biophysics of viral infectivity: matching genome length with capsid size. Q Rev Biophys 40:327–356
Phillips R, Kondev J, Theriot J, Orme N (2010) Physical biology of the cell. Garland Science, New York
Stockley PG, Twarock R (2010) Emerging topics in physical virology. Imperial College Press, London

Also especially recommended for further reading are references [18, 20, 30, 32, 35, 58, 65, 71] *listed above*.