

CHAPTER 1

INTRODUCTION

1.1 SHOTGUN SEARCHING FOR DRUGS?

The search for new drugs is daunting, expensive, and risky.

If chemicals were confined to molecular weights of less than 600 Da and consisted of common atoms, the chemistry space is estimated to contain 10^{40} to 10^{100} molecules, an impossibly large space to search for potential drugs [1]. To address this limitation of vastness, “maximal chemical diversity” [2] was applied in constructing large experimental screening libraries. Such libraries have been directed at biological “targets” (proteins) to identify active molecules, with the hope that some of these “hits” may someday become drugs. The current target space is very small—less than 500 targets have been used to discover the known drugs [3]. This number may expand to several thousand in the near future as genomics-based technologies uncover new target opportunities [4]. For example, the human genome mapping has identified over 3000 transcription factors, 580 protein kinases, 560 G-protein coupled receptors, 200 proteases, 130 ion transporters, 120 phosphatases, over 80 cation channels, and 60 nuclear hormone receptors [5].

Although screening throughputs have massively increased since the early 1990s, lead discovery productivity has not necessarily increased accordingly [6–8]. Lipinski has concluded that maximal chemical diversity is an *inefficient* library design strategy, given the enormous size of the chemistry space, and especially that clinically useful drugs appear to exist as small tight clusters in chemistry space:

“one can make the argument that screening truly diverse libraries for drug activity is the fastest way for a company to go bankrupt because the screening yield will be so low” [1]. Hits *are* made in pharmaceutical companies, but this is because the most effective (not necessarily the largest) screening libraries are highly focused, to reflect the putative tight clustering. Looking for ways to reduce the number of tests, to make the screens “smarter,” has an enormous cost reduction implication.

The emergence of combinatorial methods in the 1990s has led to enormous numbers of new chemical entities (NCEs) [9]. These are the molecules of the newest screening libraries. A large pharmaceutical company may screen 3 million molecules for biological activity each year. Some 30,000 hits are made. Most of these molecules, however potent, do not have the right physical, metabolic, and safety properties. Large pharmaceutical companies can cope with about 30 molecules taken into development each year. A good year sees three molecules reach the product stage. Some years see none. These are just rough numbers, recited at various conferences.

A drug product may cost as much as \$880 M (million) to bring out. It has been estimated that about 30% of the molecules that reach development are eventually rejected due to ADME (absorption, distribution, metabolism, excretion) problems. Much more money is spent on compounds that fail than on those that succeed [10,11]. The industry has started to respond by attempting to screen out those molecules with inappropriate ADME properties during discovery, before the molecules reach development. However, that has led to another challenge: how to do the additional screening quickly enough, while keeping costs down [6,12].

1.2 SCREEN FOR THE TARGET OR ADME FIRST?

Most commercial combinatorial libraries, some of which are very large and may be diverse, have a very small proportion of drug-like molecules [1]. Should only the small drug-like fraction be used to test against the targets? The industry’s current answer is “No.” The existing practice is to screen for the receptor activity before “drug-likeness.” The reasoning is that structural features in molecules rejected for poor ADME properties may be critical to biological activity related to the target. It is believed that active molecules with liabilities can be modified later by medicinal chemists, with minimal compromise to potency. Lipinski [1] suggests that the order of testing may change in the near future, for economic reasons. When a truly new biological therapeutic target is examined, nothing may be known about the structural requirements for ligand binding to the target. Screening may start as more or less a random process. A library of compounds is tested for activity. Computational models are constructed on the basis of the results, and the process is repeated with newly synthesized molecules, perhaps many times, before satisfactory hits are revealed. With large numbers of molecules, the process can be very costly. If the company’s library is first screened for ADME properties, that screening is done only once. The same molecules may be recycled against existing or future targets many times, with knowledge of drug-likeness to fine-tune the

optimization process. If some of the molecules with very poor ADME properties are judiciously filtered out, the biological activity testing process would be less costly. But the order of testing (activity vs. ADME) is likely to continue to be the subject of future debates [1].

1.3 ADME AND MULTIMECHANISM SCREENS

In silico property prediction is needed more than ever to cope with the screening overload. Improved prediction technologies are continuing to emerge [13,14]. However, reliably measured physicochemical properties to use as “training sets” for new target applications have not kept pace with the in silico methodologies.

Prediction of ADME properties should be simple, since the number of descriptors underlying the properties is relatively small, compared to the number associated with effective drug–receptor binding space. In fact, prediction of ADME is difficult! The current ADME experimental data reflect a multiplicity of mechanisms, making prediction uncertain. Screening systems for biological activity are typically single mechanisms, where computational models are easier to develop [1].

For example, aqueous solubility is a multimechanism system. It is affected by lipophilicity, H bonding between solute and solvent, intramolecular H bonding, intermolecular hydrogen and electrostatic bonding (crystal lattice forces), and charge state of the molecule. When the molecule is charged, the counterions in solution may affect the measured solubility of the compound. Solution microequilibria occur in parallel, affecting the solubility. Few of these physicochemical factors are well understood by medicinal chemists, who are charged with making new molecules that overcome ADME liabilities without losing potency.

Another example of a multi-mechanistic probe is the Caco-2 permeability assay (a topic covered in various sections of the book). Molecules can be transported across the Caco-2 monolayer by several mechanisms operating simultaneously, but to varying degrees, such as transcellular passive diffusion, paracellular passive diffusion, lateral passive diffusion, active influx or efflux mediated by transporters, passive transport mediated by membrane-bound proteins, receptor-mediated endocytosis, pH gradient, and electrostatic-gradient driven mechanisms. The P-glycoprotein (P-gp) efflux transporter can be saturated if the solute concentration is high enough during the assay. If the substance concentration is very low (perhaps because not enough of the compound is available during discovery), the importance of efflux transporters in gastrointestinal tract (GIT) absorption can be overestimated, providing the medicinal chemist with an overly pessimistic prediction of intestinal permeability [8,15,16]. Metabolism by the Caco-2 system can further complicate the assay outcome.

Compounds from traditional drug space (“common drugs”—readily available from chemical suppliers), often chosen for studies by academic laboratories for assay validation and computational model-building purposes, can lead to misleading conclusions when the results of such models are applied to ‘real’ discovery compounds, which most often have extremely low solubilities [16].

Computational models for single mechanism assays (e.g., biological receptor affinity) improve as more data are accumulated [1]. In contrast, computational models for multimechanism assays (e.g., solubility, permeability, charge state) worsen as more measurements are accumulated [1]. Predictions of human oral absorption using Caco-2 permeabilities can look very impressive when only a small number of molecules is considered. However, good correlations deteriorate as more molecules are included in the plot, and predictivity soon becomes meaningless. Lipinski states that “The solution to this dilemma is to carry out single mechanism ADME experimental assays and to construct single mechanism ADME computational models. The ADME area is at least 5 or more years behind the biology therapeutic target area in this respect” [1].

The subject of this book is to examine the components of the multimechanistic processes related to solubility, permeability, and charge state, with the aim of advancing improved strategies for in vitro assays related to drug absorption.

1.4 ADME AND MEDICINAL CHEMISTS

Although ADME assays are usually performed by analytical chemists, medicinal chemists—the molecule makers—need to have some understanding of the physico-chemical processes in which the molecules participate. Peter Taylor [17] states:

It is now almost a century since Overton and Meyer first demonstrated the existence of a relationship between the biological activity of a series of compounds and some simple physical property common to its members. In the intervening years the germ of their discovery has grown into an understanding whose ramifications extend into medicinal chemistry, agrochemical and pesticide research, environmental pollution and even, by a curious re-invention of familiar territory, some areas basic to the science of chemistry itself. Yet its further exploitation was long delayed. It was 40 years later that Ferguson at ICI applied similar principles to a rationalization of the comparative activity of gaseous anaesthetics, and 20 more were to pass before the next crucial step was formulated in the mind of Hansch. . . . Without any doubt, one major factor [for delay] was compartmentalism. The various branches of science were much more separate then than now. It has become almost trite to claim that the major advances in science take place along the borders between its disciplines, but in truth this happened in the case of what we now call Hansch analysis, combining as it did aspects of pharmacy, pharmacology, statistics and physical organic chemistry. Yet there was another feature that is not so often remarked, and one with a much more direct contemporary implication. The physical and physical organic chemistry of equilibrium processes—solubility, partitioning, hydrogen bonding, etc.—is not a glamorous subject. It seems too simple. Even though the specialist may detect an enormous information content in an assemblage of such numbers, to synthetic chemists used to thinking in three-dimensional terms they appear structureless, with no immediate meaning that they can *visually* grasp. Fifty years ago it was the siren call of Ehrlich’s lock-and-key theory that deflected medicinal chemists from a physical understanding that might otherwise have been attained much earlier. Today it is glamour of the television screen. No matter that what is on display may sometimes possess all the profundity of a

five-finger exercise. It is visual and therefore more comfortable and easier to assimilate. Similarly, MO theory in its resurgent phase combines the exotic appeal of a mystery religion with a new-found instinct for three-dimensional colour projection which really can give the ingenu the impression that he understands what it is all about. There are great advances and great opportunities in all this, but nevertheless a concomitant danger that medicinal chemists may forget or pay insufficient attention to hurdles the drug molecule will face if it is actually to perform the clever docking routine they have just tried out: hurdles of solubilization, penetration, distribution, metabolism and finally of its non-specific interactions in the vicinity of the active site, all of them the result of physical principles on which computer graphics has nothing to say. Such a tendency has been sharply exacerbated by the recent trend, for reasons of cost as much as of humanity, to throw the emphasis upon *in vitro* testing. All too often, chemists are disconcerted to discover that the activity they are so pleased with *in vitro* entirely fails to translate to the *in vivo* situation. Very often, a simple appreciation of basic physical principles would have spared them this disappointment; better, could have suggested in advance how they might avoid it. We are still not so far down the path of this enlightenment as we ought to be. What is more, there seems a risk that some of it may fade if the balance between a burgeoning receptor science and these more down-to-earth physical principles is not properly kept.

Taylor [17] described physicochemical profiling in a comprehensive and compelling way, but enough has happened since 1990 to warrant a thorough reexamination. Then, combichem, high-throughput screening (HTS), Caco-2, IAM, CE were in a preingenu state; studies of drug-partitioning into liposomes were arcane; instrument companies took no visible interest in making pK_a , $\log P$, or solubility analyzers; there was no biopharmaceutics classification system (BCS); it did not occur to anyone to do PAMPA. With all that is new, it is a good time to take stock of what we can learn from the work since 1990. In this book, measurement of solubility, permeability, lipophilicity, and charge state of drug molecules will be critically reexamined (with considerable coverage given to permeability, the property least explored). Fick's law of diffusion [18] in predicting drug absorption will be reexplored.

1.5 THE "A" IN ADME

In this book we will focus on physicochemical profiling in support of improved prediction methods for *absorption*, the "A" in ADME. Metabolism and other components of ADME will be beyond the scope of this book. Furthermore, we will focus on properties related to *passive* absorption, and not directly consider active transport mechanisms. The most important physicochemical parameters associated with passive absorption are *acid-base* character (which determines the charge state of a molecule in a solution of a particular pH), *lipophilicity* (which determines distribution of a molecule between the aqueous and the lipid environments), *solubility* (which limits the concentration that a dosage form of a molecule can present to the solution and the rate at which the molecule dissolves from

the solid form), and membrane *permeability* (which determines how quickly molecules can cross membrane barriers). Current state of the art in measurement of these properties, as the ever important function of pH, will be surveyed, and in some cases (permeability), described in detail.

1.6 IT IS NOT JUST A NUMBER—IT IS A MULTIMECHANISM

Drugs exert their therapeutic effects through reactions with specific receptors. Drug–receptor binding depends on the concentration of the drug near the receptor. Its form and concentration near the receptor depend on its physical properties. Orally administered drugs need to be dissolved at the site of absorption in the gastrointestinal tract (GIT), and need to traverse several membrane barriers before receptor interactions can commence. As the drug distributes into the various compartments of the body, a certain (small) portion finds itself at the receptor site. Transport and distribution of most drugs are affected by passive diffusion, which depends on lipophilicity, since lipid barriers need to be crossed [19–24]. Passive transport is well described by the principles of physical chemistry [25–33].

The pK_a of a molecule, a charge-state-related parameter, is a descriptor of an acid–base equilibrium reaction [34,35]. Lipophilicity, often represented by the octanol–water partition coefficient K_p is a descriptor of a two-phase distribution equilibrium reaction [36]. So is solubility [37–39]. These three parameters are thermodynamic constants. On the other hand, permeability P_e is a rate coefficient, a kinetics parameter, most often posed in a first-order distribution reaction [40–42].

In high-throughput screening (HTS) these parameters are sometimes viewed simply as numbers, quickly and roughly determined, to be used to rank molecules into “good” and “bad” classes. An attempt will be made to examine this important aspect. In addition, how fundamental, molecular-level interpretations of the physical measurements can help to improve the design of the profiling assays will be examined, with the aim of promoting the data fodder of HTS to a higher level of quality, without compromising the need for high speed. Quality measurements in large quantities will lead to improved *in silico* methods. Simple rules (presented in visually appealing ways), in the spirit of Lipinski’s rule of fives, will be sought, of use not only to medicinal chemists but also to preformulators [12,43]. This book attempts to make easier the dialog between the medicinal chemists charged with modifying test compounds and the pharmaceutical scientists charged with physico-chemical profiling, who need to communicate the results of their assays in an optimally effective manner.