Part One
Physicochemical Aspects of Drug Dissolution and Solubility

2
Aqueous Solubility in Drug Discovery Chemistry, DMPK, and Biological Assays

Nicola Colclough, Linette Ruston, and Kin Tam

Abbreviations

AZ	AstraZeneca
BCS	Biopharmaceutics Classification System
BNN	Bayesian neural network
CAD	Charged aerosol detector
CD	Candidate drug
CLND	Chemiluminescent nitrogen detection
Clog *P*	Calculated log *P*
DMPK	Drug metabolism and pharmacokinetics
DMSO	Dimethyl sulfoxide
ELSD	Evaporative light scattering detector
HI	Hit identification
HPLC	High-performance liquid chromatography
HTS	High-throughput screening
HTSol	High-throughput solubility
LI	Lead identification
LO	Lead optimization
MLR	Multiple linear regression
PLM	Polarized light microscopy
PLS	Partial least squares
PXRD	Powder X-ray diffraction
QSPR	Quantitative structure–property relationship
SAR	Structure–activity relationship

Symbols

K_{sp}	Solubility product
MX	Salt of acid or base
M^+	Protonated base or cationic counterion to conjugate base

pK_a	Acid dissociation constant
S	Solubility of a compound at a particular pH
S_0	Solubility of a compound in its neutral form/intrinsic solubility
X^-	Conjugate base or anionic counterion to protonated base

2.1 Introduction

Aqueous solubility is one of the key physicochemical properties in drug discovery [1, 2]. High solubility in intestinal fluid provides the concentration gradient that drives the absorption of orally administered drugs and subsequent distribution to the site of action to elicit a pharmacological response. For intravenously administered agents, sufficiently high solubility in plasma is critical to minimize undesirable precipitation in the systematic circulation. Generally, poor aqueous solubility leads to formulation challenges in development, raising costs during this phase. Aqueous solubility data facilitate the interpretation of biological assay results. In particular, poorly soluble compounds can precipitate out of solution during a high-throughput screening (HTS) campaign, thereby giving undesirable false negatives and/or false positives, the latter via binding of the target to aggregates [3, 4]. In the absence of solubility information at the HTS stage, such false hits can go unnoticed and hamper structure–activity relationship (SAR) interpretation. Poor aqueous solubility can lead to problems *in vivo*, such as incomplete absorption following oral administration [5], variable bioavailability, fed/fasting effects [6, 7], and difficulties in establishing a sufficient safety margin following dose escalation studies. Moreover, poor solubility is relatively difficult to modulate in the later stage of a discovery project, where the core structure of the lead series is more or less defined. Enabling formulations, such as nanoparticle technology or polymer dispersion, may provide a solution by particle size reduction, offering an enhanced dissolution rate (Chapter 22). These approaches could show benefits in formulating BCS (Biopharmaceutics Classification System) class II compounds where solubility is low (dose limitation is likely) and permeability is high, and the limiting factor for absorption is the rate of dissolution, rather than the passage across the intestinal barrier [8].

The search for potent chemical series in drug discovery means that for certain biological target types there can be a tendency toward lipophilic and/or planar structures to maximize interactions at the active site. However, the solubilities of these compounds are generally low. On rare occasions, high potency may offset the issue of low solubility because a low dose is sufficient to show the clinical benefit. An example of this is montelukast (leukotriene D receptor antagonist, Clog $P = 8.47$), a very potent compound (0.1 nM) that due to low dose renders solubility no longer an issue [9]. To enable speedy progression of the chemical series of interest along the discovery pipeline, it is important to aim for a good balance of parameters, for example, solubility, exposure, and acceptable toxicity profile, while improving on potency. This translates into a need for physicochemical property assays with greater throughput or more reliable property prediction. The latter will be covered in

Chapter 4. In this chapter, we will focus on the experimental aspect of aqueous solubility and its interplay in discovery chemistry, DMPK, and bioscience assays. Particular emphasis will be placed on the latest technologies for determining aqueous solubility and the use of the solubility data in different phases of drug discovery.

2.1.1 Definition of Aqueous Solubility

Solubility is usually expressed as log S, where S is the saturated compound concentration in mol/l in equilibrium with its most stable crystalline form under certain defined conditions (e.g., physiological pH at room temperature over an extended period of time, typically 24–48 h). This is also known as the thermodynamic solubility. The typical log S values for discovery compounds vary from -3 (1 mM) down to -7 (0.1 μM). In contrast, kinetic solubility refers to the solubility value determined within a defined period of time, which is usually much shorter than 24 h. Since equilibrium conditions are not often achieved in this time frame, the compound is typically not in its most stable crystalline form. Therefore, the kinetic solubility value is normally higher than that obtained from the thermodynamic approach. Despite these caveats, kinetic solubility measurement can be set up in a high-throughput assay format and has been used by some pharmaceutical companies to identify poorly soluble compounds in the very early stage of drug discovery.

A search [10] of the World Drug Index revealed that 62.5% of marketed drugs are ionizable, which implies that these substances can exist in various charged states depending on the pH of the media. For ionizable drugs, solubility is pH dependent, and it is therefore important to understand the solubility in the context of pH. Ionization of a compound can be defined by the acid dissociation constant, pK_a. For the case of monoprotic compounds, the solubility at a given pH can be described by the following equations:

$$\text{base}: \quad S = S_0(1 + 10^{pK_a - pH}), \tag{2.1}$$

$$\text{acid}: \quad S = S_0(1 + 10^{pH - pK_a}), \tag{2.2}$$

where S_0 denotes the solubility of the compound in its neutral form, also referred to as intrinsic solubility. Figure 2.1 shows the pH–solubility profiles generated using Equations 2.1 and 2.2. It can be seen that the solubility of ionizable compounds is limited by the solubility of the neutral form of the compound. Depending on the charge types, solubility increases as pH decreases (base) or increases (acid) until a critical pH is reached, where the salt form and solubility product K_{sp} become solubility limiting (Equations 2.3 and 2.4) [11].

$$MX_{(s)} \rightleftharpoons M^+_{(aq)} + X^-_{(aq)}, \tag{2.3}$$

$$K_{sp} = [M^+_{(aq)}][X^-_{(aq)}]. \tag{2.4}$$

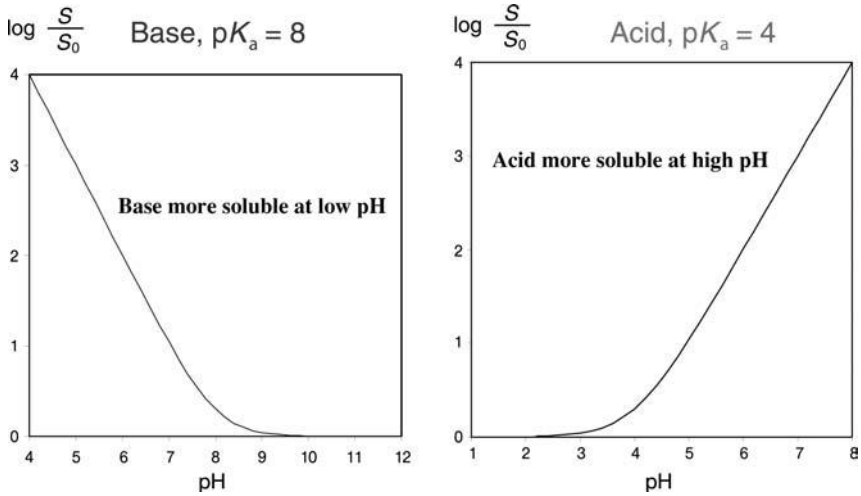


Figure 2.1 Theoretical solubility–pH profiles of a base with pK_a of 8 and an acid with pK_a of 4.

2.1.2
Aqueous Solubility in Different Phases of Drug Discovery

Figure 2.2 shows a schematic diagram of the drug discovery process. The target identification phase seeks to identify the biological target, signaling the start of a discovery project. Solubility evaluation typically begins in the hit identification (HI) phase enabling physicochemical characterization of hits and interpretation of biological and DMPK assay data. In lead identification (LI), solubility data facilitate the selection of "drug-like" lead series that will allow swift identification of a candidate drug (CD). Within the lead optimization (LO) phase, solubility data provide a formulation risk assessment for CDs entering development. Moreover, throughout the HI/LI/LO phases, solubility data are also used extensively as part of molecular design.

2.2
Aqueous Solubility in Hit Identification

In the HI phase of a discovery project, a variety of strategies are employed to identify potential hit series. HTS, focused subset library screening, and fragment library

Figure 2.2 Different phases of drug discovery.

screening represent a selection of the approaches undertaken in trying to find a good hit series. Besides an assessment of potency, it has become increasingly important to simultaneously understand the physicochemical properties of such potential leads to enable rapid identification of quality series most likely to progress quickly to CD delivery. Such front loading of physicochemical property assessment has required the introduction of solubility screens capable of handling the large number of compounds coming out of biological testing during this phase. Various solubility screens have been established in discovery to this end, but all share the common features of requiring compounds to be solubilized in DMSO as the starting point, since this facilitates rapid dispensing from company collections and the use of plate-based automation.

2.2.1
Aqueous Solubility from DMSO Solutions

It is the universally accepted practice within pharmaceutical companies to store compounds both as solids and as solutions in DMSO. Typically, as at AstraZeneca (AZ), a concentration of 10 mM or similar is used. From this, the stock samples are taken for a wide variety of tests including aqueous solubility screens. Using a DMSO solution as the starting point for a solubility assay presents a number of advantages. The automated nature of solution dispensing facilitates the study of a large number of compounds: sample consumption is usually significantly less than that in classic thermodynamic assays where solubility is measured from solid material and where DMSO solution in the aqueous sample is kept low, typically and at the 1% level or less, and there is minimal cosolvent effect [12] (see also Figure 2.3). The use of a DMSO stock solution does, however, mean that the

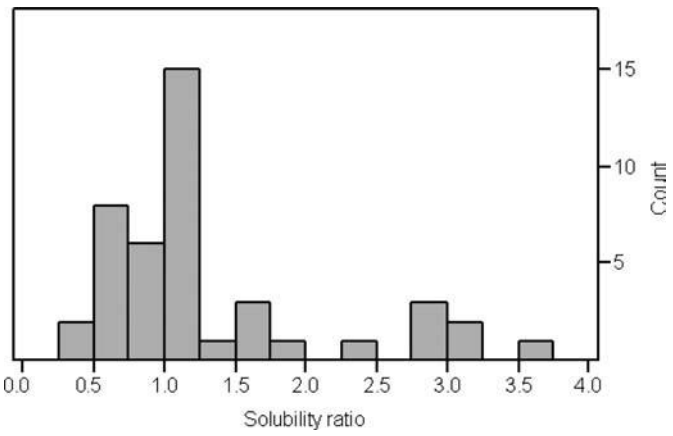


Figure 2.3 The solubility ratios of a set of 43 in-house compounds measured in phosphate buffer at pH 7.4 with and without 1% DMSO by the classic shake-flask method from solid. Compounds span a solubility range from 0.2 µM to 5 mM.

upper end of the dynamic range of the solubility assay is defined by the choice of the stock concentration in the company collection and the final DMSO percentage in the aqueous sample. In the case of AZ, a 10 mM DMSO stock combined with a 1% final cosolvent concentration in the high-throughput solubility (HTSol) assay means the assay upper limit is fixed at 100 μM. This is typically less than can be achieved from a solid-based assay.

Many of the HTSol assays reported in the literature using DMSO solution are "kinetic" HTSol assays. Such assays analyze the aqueous samples after only a short period of agitation with timescales ranging from immediately following precipitation to several minutes. In contrast, "thermodynamic" HTSol assays typically involve agitation for a minimum 24 h period and often longer. Kinetic solubility assays are consequently easier to run and as such are frequently used as a frontline assay to provide an initial ranking of lead series [13, 14]. In addition, kinetic solubility data have been applied to the interpretation of biological assay data where the measurement timescale and precipitation following dilution of a DMSO stock solution into aqueous media more closely mirror the process of biological assay testing [15].

Kinetic solubilities are by definition very time dependent and as such results can be less reproducible than thermodynamic solubility values. The short timescale also means that they are more dependent on the physical form of the initial precipitate. Consequentially, the correlation between kinetic and thermodynamic solubility is generally poor, with the kinetic measurement usually giving higher values [12, 16]. However, an advantage this can bring is that there will be few compounds excluded as false negatives in this phase.

HTSol assays employing DMSO stock solutions vary in the nature of the analysis of aqueous samples. Two approaches are commonly observed: (1) turbidimetric methods, where the formation or loss of precipitate is monitored against concentration, and (2) direct quantification of compound in solution by UV absorption spectroscopy following removal of precipitate by filtration.

2.2.1.1 **Turbidimetric Methods**

Turbidimetric methods rely on the measurement of light scattering from precipitate in solution to determine solubility. The initial approach, described by Lipinski, involves the stepwise addition of aliquots of DMSO stock solution at 1 min intervals to aqueous buffer in a UV cuvette until precipitation occurs when the kinetic solubility limit is achieved [1]. Precipitation is identified by an absorbance increase due to blockage of light by the particles in the range 600–800 nm using a diode array UV spectrometer. Using this approach, Lipinski is able to determine kinetic solubilities in the range 5–65 μg/ml with an upper limit of 0.67% DMSO cosolvent. Other turbidimetric methods have used fixed DMSO compositions (between 0.3 and 5% cosolvent) to avoid any potential cosolvent effects on solubility and have looked for precipitation following serial dilution [13, 14, 17, 18]. Alternative light scattering detection methods have also been used, including nephelometry [14, 17, 18] and flow cytometry [19]. Both make use of the perturbation of a laser beam passed through the sample. Nephelometric detection in a 96-well plate format is more amenable to automation and offers higher throughput than linear flow-through approaches, such

as cytometry, or incremental DMSO addition to a UV cuvette. However, nephelometer readings are very sensitive to plate quality, and the presence of scratches or dust can give rise to erroneously low solubility values.

2.2.1.2 UV Absorption Methods

The alternative method to turbidimetric detection used for measuring solubility in early discovery is to quantify the aqueous supernatant directly via UV absorbance [13, 20, 21]. Typically, DMSO stock solution is added to aqueous buffer such that the final DMSO composition is kept to a minimum (5% or less) and the resulting precipitate is removed by filtration. A UV plate reader is then used to determine the aqueous solubility by comparing the filtrate absorbance against that of a calibration solution prepared in an identical solvent. It is important to match the sample and calibration solutions to prevent solvochromic effects. Care must also be taken in the selection of the filter plate since nonspecific binding of compound can occur with some filter materials leading to erroneously low solubility values [22]. Like nephelometry, the plate-based UV detection approach is amenable to automation.

As with turbidimetric assays, many of the direct UV absorbance assays are set up to determine kinetic solubility. However, the UV absorbance method also lends itself well to thermodynamic solubility determination by extending the period of sample agitation prior to filtration to 24 h or more. This offers a number of advantages. The solubility data generated are less dependent on the physical form of the initial material precipitated from DMSO and are much closer to thermodynamic solubility values determined later in discovery and in early development. As such, it gives more consistent solubility data through the discovery phase and enables a better quality early assessment to be made of the likely difficulties or otherwise of progressing a lead series into development.

It is a version of this latter assay that has been established as our current frontline solubility measurement in the hit identification phase of discovery [23]. The assay uses 10 mM DMSO stock solution, which is diluted into aqueous buffer at pH 7.4 to give a final DMSO composition of 1%. Samples are agitated for 24 h using magnetic stirrer bars prior to plate filtration to remove precipitate. This sample is further diluted and compared against a calibration solution of known concentration also taken from the 10 mM DMSO stock and diluted to the same solvent composition. The assay is based on a 96-well plate format using a UV diode array plate reader. This has enabled full automation of the assay with over 600 compounds measured in each run.

One of the concerns raised with the direct UV absorbance approach is that without HPLC separation the presence of impurities may cause erroneous solubility values to be reported. This is also a concern for turbidimetric methods. However, with strict purity criteria for registration of compounds into the company collection, this has helped to lessen this concern. In addition, for this assay an algorithm has been written that checks the UV spectrum of the sample against that of the calibration. Any significant impurities or decomposition occurring during the 24 h agitation period are readily picked out as a spectral mismatch [23].

To determine how similar solubility values from our high-throughput DMSO-based thermodynamic assay are to classical thermodynamic solubilities measured from solid, a diverse test set of 200 predominantly in-house compounds were compared. In the case of the latter assay as well as starting from solid material, the method included HPLC analysis and separation of undissolved solid from the supernatant via double centrifugation. Figure 2.4 indicates that there is a good correlation between the two methods with most compounds giving solubilities within a factor of 3. Given the potentially different physical forms of the compound generated in the two assays, this seems reasonable. It is noted that the

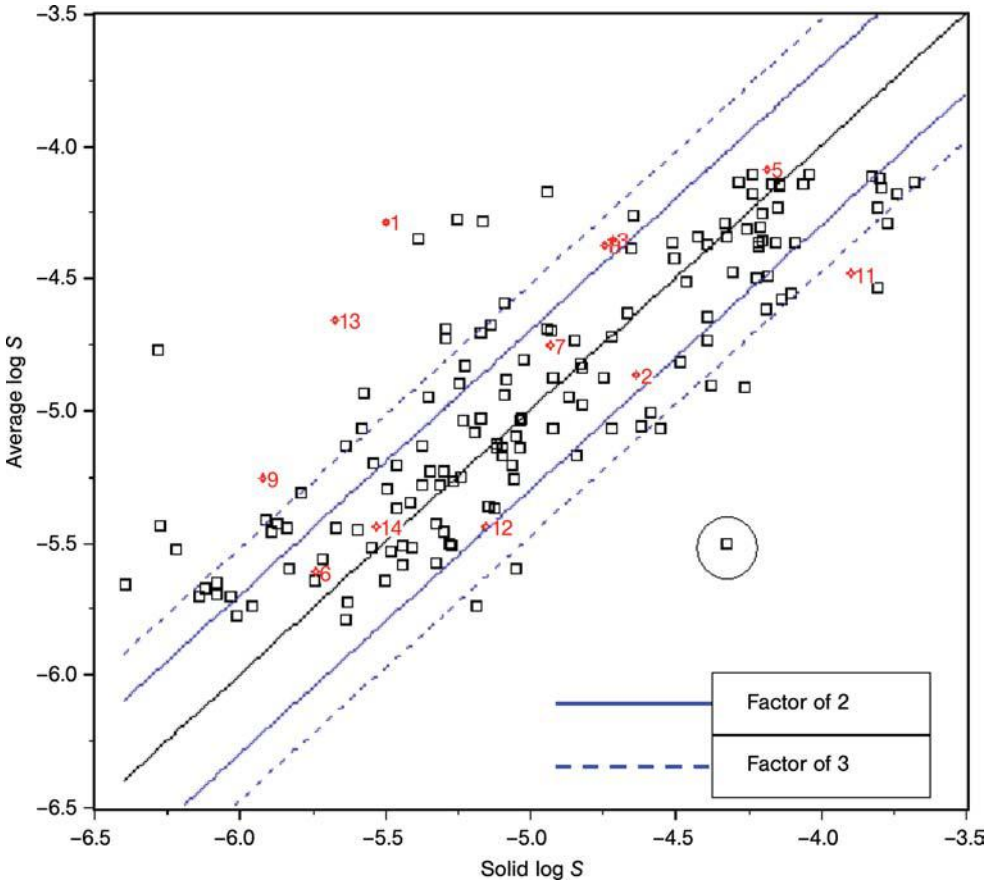


Figure 2.4 Correlation between the solubility of the test set obtained from the solid solubility assay (*x*-axis) and from the high-throughput solubility assay (*y*-axis; average of four experiments). The black line is the 1 : 1 line. Symbol ⊙ represents in-house compounds. Symbol ◇ represents commercial compounds: 1 – disulfiram, 2 – diethylstilbestrol, 3 – griseofulvin, 5 – haloperidol, 6 – mebendazole, 7 – glyburide, 8 – nifedipine, 9 – albendazole, 10 – bumetanide analogue, 11 – loperamide, 12 – astemizole, 13 – nimodipine, 14 – loratadine. Symbol in circle represents the negative outlier as discussed in the text. Reprinted with permission from [23].

effect of different crystalline polymorphs on solubility can typically be within this range [24, 25]. There are a number of outliers in the plot that predominantly fall on the side of greater solubility for the DMSO-based assay. Literature and our own studies suggest that this enhanced solubility effect is unlikely to be caused by the solubilizing effect of 1% DMSO [12]. Figure 2.3 shows the solubility ratios of a set of 43 compounds measured in buffer with and without 1% DMSO by the classic thermodynamic method. The compounds span a solubility range from 0.2 μM to 5 mM. The presence of 1% DMSO has minimal impact on the solubility, with a factor of 3.6 the largest enhancement observed. It is more likely that the enhanced solubility observed for the outliers reflects that the compounds have precipitated from DMSO solution as amorphous material and this has not yet reached true equilibrium with their crystalline form during the 24 h agitation period. Literature indicates that differences in solubility between amorphous and crystalline forms of a compound can be significant [26]. A similar explanation has also been reported for positive outliers observed in an extended agitation solubility study using HPLC–UV analysis [16]. There is one significant negative outlier in Figure 2.4. Further studies ruled out any retention to the filter membrane as a possible explanation. Powder X-ray diffraction on the postsolubility samples for this compound in a scaled-up experiment revealed that the DMSO method had generated a crystalline sample in 24 h whereas the sample from the experiment starting from solid material was predominantly amorphous after this period. This interesting observation highlights that for certain compounds DMSO precipitation conditions can facilitate formation of crystalline material.

2.2.1.3 **Alternative Detection Methodology**

A further recent approach taken to deliver higher throughput kinetic and thermodynamic aqueous solubility measurements from DMSO and solid, respectively, involved changing the assay detection method from UV to one that does not require compound-specific calibration, namely, chemiluminescent nitrogen detection (CLND) [27, 28]. The CLND detector is able to quantify the nitrogen content of the aqueous sample using a generic nitrogen calibration curve, and from the knowledge of the number of nitrogen atoms in the molecule the aqueous concentration and hence the solubility is determined. Eliminating the need for a compound-specific calibration solution reduces assay sample consumption relative to UV methods, and the technique is fast when coupled with direct flow injection onto the detector. However, the sensitivity of the CLND detector to nitrogen necessitates rigorous laboratory housekeeping to avoid contamination of the instrument from nonsample nitrogen sources. Regular recalibration of the CLND detector is required to retain accuracy, and a linear response is not observed for all nitrogen environments; for example, adjacent nitrogen atoms in a molecule are known to be a special class. Care must also be taken if HPLC is not used with the detector to ensure that no nitrogen-containing impurities are present in the sample.

Other universal detectors have been suggested as potential replacements for UV in solubility assays including ELSD and CAD, although at present no specific assay has been reported [29].

2.2.1.4 Application of DMSO-Based Solubility Assays

The advent of fully automated DMSO-based solubility assays has meant that aqueous solubilities can now be determined on HTS output in parallel with biological testing to enable rapid identification of quality hit series from a physicochemical perspective [2]. Figure 2.5 shows an AZ project example where simultaneous solubility and potency measurements highlighted 1 out of 19 hit series as having poor physicochemical properties. A comparison of solubility data with potency data also enables identification of potential false hits, which can be removed to facilitate SAR understanding. Automated DMSO-based solubility assays have also shown benefits in library screening. In relation to this, our thermodynamic HTSol assay described above has been successfully applied to the selection of compounds forming the AZ in-house generic fragment library [23]. Fragments typically show weak binding to molecular targets and so are normally screened at high concentrations. Consequently, good aqueous solubility is one of the key criteria to satisfy in establishing a fragment library. It has been demonstrated that by using a training set of 3234 neutral compounds, a clear relationship could be established between measured solubility and predicted log *P* (Clog *P*) (Figure 2.6) [23]. The use of binning and percentiles in Figure 2.6 enables maximum information to be extracted from data covering a small dynamic range. In particular, multiple percentiles (10, 20, 30, 40, and 50) capture the variation in log *S* and reveal the strength of the trend with Clog *P*. This analysis was applied to all the potential fragment library compounds. Neutral fragments with Clog *P* less than 2.19 were assumed to have acceptable solubility, while those with higher values were submitted for solubility measurement. Only those compounds with solubilities at or above the upper quantification limit for the solubility assay were accepted into the fragment library.

2.3 Aqueous Solubility in Lead Identification and Lead Optimization

In the lead identification and lead optimization phases of discovery, there is greater focus on thermodynamic solubility measurements. Thermodynamic solubility assays are designed to determine the solubility of the stable crystalline form of the compound, since this is the physical form that will be sought in the development phase for orally administered drugs. As such, thermodynamic solubilities provide discovery projects with a better risk assessment of likely formulation issues in development. Thermodynamic solubilities, unlike kinetic solubilities, are less dependent on the initial physical form of the compound and being less time critical also tend to be more reproducible. This is particularly important from a molecular design perspective where chemists are seeking to modify molecular structure to improve solubility.

Classically, thermodynamic solubility is measured using the shake-flask method [21, 30]. This method involves addition of an excess of solid to aqueous buffer at fixed pH. The solution is stirred for a minimum of 24 h prior to separation of the supernatant from undissolved material via centrifugation or filtration followed by

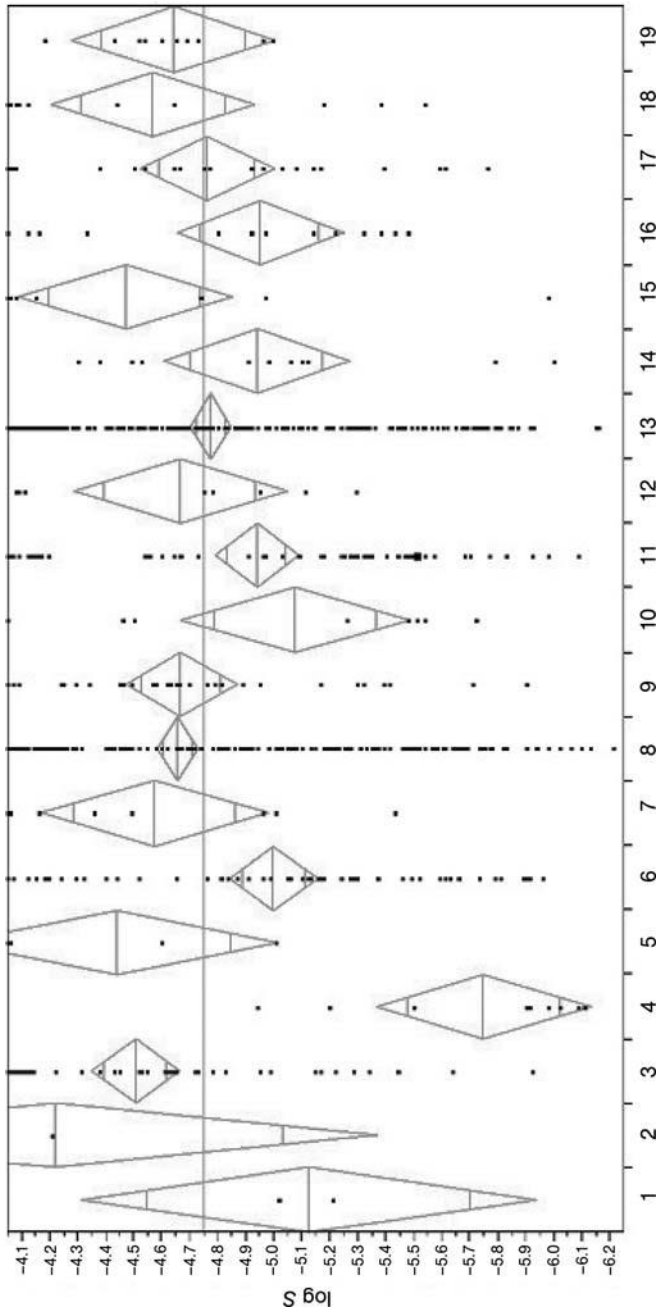


Figure 2.5 HT solubility of a discovery project HTS hit output clustered by structural series (1–19).

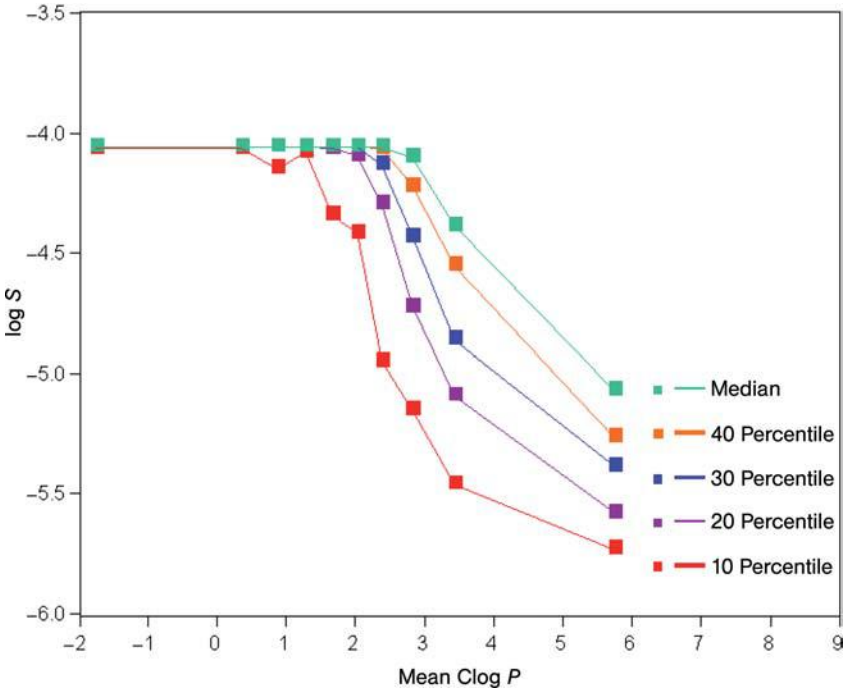


Figure 2.6 Percentiles for solubility for training set fragments as a function of mean Clog *P* for each bin. Data partitioned by Clog *P* into 10 bins each with 356 solubility measurements. Reprinted with permission from [23].

HPLC–UV analysis. The need for weighing out of solid and an analysis via HPLC mean that the method has generally low throughput. However, the increasing demand for thermodynamic solubilities in discovery has led to the development of automation-friendly methods on the basis of the shake-flask approach. Many of these methods make use of DMSO stock solutions as the starting point for the assay, thereby exploiting the automated compound dispensing facilities within pharmaceutical companies. The assays retain the minimum 24 h agitation of the shake-flask method with analysis of supernatant being either via HPLC–UV or direct UV.

2.3.1 Dried-Down Solution Methods

Another approach to increase the throughput of thermodynamic solubility measurement in discovery involves evaporating a DMSO stock solution to dryness at the start of the assay [12, 22, 31]. Aqueous buffer is added to the dried-down solid, which is then agitated for 24 h followed by HPLC–UV or UV analysis of the supernatant. Removal of all the DMSO solvent ensures that the solubility value is not enhanced by

any cosolvent effects. A further advantage of this method is that, as with the classic shake-flask approach, a larger solubility dynamic range can be achieved compared with direct DMSO-based methods. Appropriate selection of DMSO volume to be evaporated coupled with buffer volume selection allows the upper assay limit to be extended by an extra order of magnitude (~1 mM). It should be noted that care must be taken establishing evaporation conditions to avoid loss of volatile or thermally labile compounds. Solubility data using this approach have been reported to give a good correlation with thermodynamic solubility values determined by the classic shake-flask methods [12, 22, 31]. Incidences where differences have been observed are thought to reflect cases where equilibrium with the stable crystalline form has not been achieved by the dried solid method in the 24 h timescale [31].

2.3.2 Solubility from Solid

Many approaches to determine thermodynamic solubility in drug discovery have focused on miniaturizing the classic shake-flask method with solid samples manually weighed and dispensed into 96-well plates or 96-vial arrays [32–34]. The 96-well format enables the use of liquid handling robots to improve throughput. As with the standard shake-flask method, aqueous buffer is added and the solution is agitated for a minimum of 24 h prior to plate filtration or centrifugation to remove the supernatant, which is then analyzed by HPLC–UV or direct UV.

Given the large number of compounds evaluated in discovery for solubility and the small quantities of solid material available, the physical form of the starting solid and that at the end of the agitation period is rarely characterized. Often the starting solid material will be amorphous [35] reflecting the compound purification techniques used by chemists today, which are largely based on preparative HPLC. Recrystallization techniques are generally no longer used.

In all thermodynamic solubility methods, an assumption is made that in the 24 h of agitation the initial form of the compound is able to convert to the stable crystalline form. Evidence indicates that compounds can convert from amorphous to crystalline forms and can also change polymorphs in this period [24, 36]. However, not all compounds are able to equilibrate to the most stable polymorph during this time frame [37]. The question then arises what impact does the initial solid state have on the thermodynamic solubility result reported and how often does the compound convert to its most stable crystalline form in the solubility assay. Pudipeddi *et al.* [24] showed with a data set of 81 compounds that there was little difference in the solubilities of different crystalline polymorphs and that typically the solubility ratio was no greater than 2. Similar observations were made by Hancock *et al.* [26], who showed that the most significant solubility differences occurred between amorphous and crystalline materials. In an attempt to understand how often crystalline material is generated from amorphous samples in thermodynamic solubility assays, we have taken a diverse set of commercial compounds, measured the solubility starting from crystalline and amorphous solids, and compared the solid form by powder X-ray diffraction (PXRD) (Table 2.1). From Figure 2.7, it is apparent that most amorphous

Table 2.1 The aqueous solubility of 16 drug molecules at pH 7.4 measured from crystalline and amorphous solid materials using the shake-flask method (24 h agitation).

Drug name	Crystalline solubility (μM)	Amorphous solubility (μM)	Solubility ratio (amorphous/crystalline)
Disulfiram	2.2	3.2	1.5
Astemizole	3.5	4.3	1.2
Bicalutamide	4.6	6.1	1.3
Ketoconazole	5.2	5.6	1.1
Loperamide	6.6	444.3	67.5
Glyburide	9.5	56.8	6.0
Griseofulvin	15.3	18.0	1.2
Terfenadine	15.7	21.3	1.4
Nifedipine	41.9	882.4	21.1
Haloperidol	52.3	54.7	1.0
Testosterone	57.9	69.0	1.2
Flutamide	92.9	79.8	0.9
Bitolterol	103.0	77.2	0.7
Diazepam	130.0	132.3	1.0
Carbamazepine	428.6	456.2	1.1
Chlorzoxazone	1360.0	1626.2	1.2

compounds reach equilibration with crystalline material in the course of the assay (Figure 2.8). A few positive outliers were seen by PXRD to be amorphous at the end of the experiment (Figure 2.9). Similar observations have been made by Sugano, who examined the solubility of a series of compounds starting from DMSO solution following 20 h of agitation with characterization of the final solid state by polarized light microscopy (PLM) [16]. Indirect evidence also suggests that most compounds are able to generate a stable thermodynamic solubility value in 24 h. For example, subsequent batches of the same compound made in AZ discovery projects show small solubility differences ($\leq$0.44 log units, see Figure 2.10). Moreover, in general, it is observed that there are good correlations between data produced in thermodynamic solubility assays starting from solid, dried DMSO solution [12, 22, 31], and DMSO solution (Figure 2.4).

2.3.3
Thermodynamic Solubility Assays with Solid-State Characterization

More recently in discovery there has been a trend toward developing high-throughput thermodynamic solubility assays, which incorporate a solid-state assessment at the end of the period of agitation. This assessment aids interpretation of the solubility data and is an important consideration when relating the solubility data to molecular structure. Solid-state characterization methods include the use of PLM [16], microscopic analysis [34], PXRD [32, 33], and Raman microscopy [22]. With all these

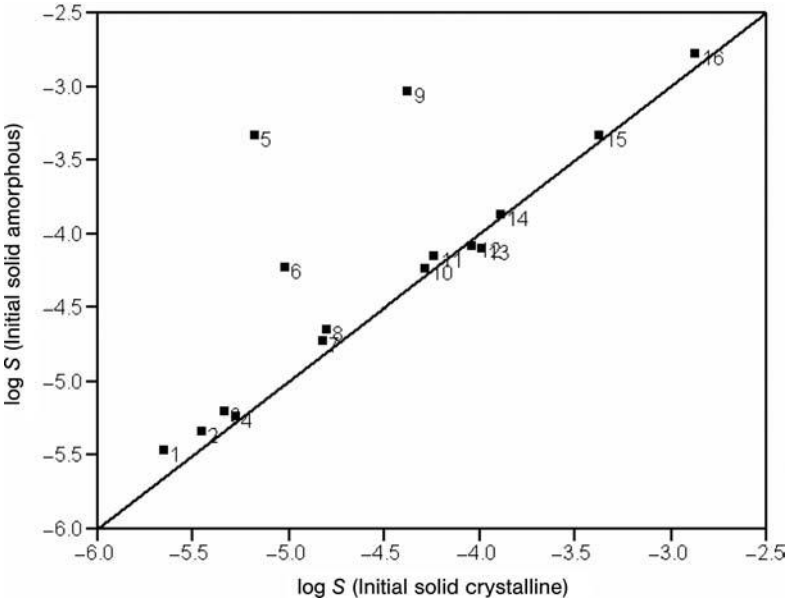


Figure 2.7 Plot of aqueous solubility using amorphous versus crystalline material at pH 7.4 following 24 h agitation. (1) Disulfiram, (2) astemizole, (3) bicalutamide, (4) ketoconazole, (5) loperamide, (6) glyburide, (7) griseofulvin, (8) terfenadine, (9) nifedipine, (10) haloperidol, (11) testosterone, (12) flutamide, (13) bitolterol, (14) diazepam, (15) carbamazepine, (16) chlorzoxazone.

methods having sufficient solid sample postincubation is key. It has been shown that PLM can be applied to a DMSO-based solubility assay with a sample size of 0.6 mg, enabling an amorphous/crystalline interpretation of remaining solid [16]. Direct analysis is made of the solid following centrifugation of the sample solution in a 96-well glass-bottomed plate. A similar interpretation has been reported using microscopic analysis from a solubility assay (PASS) using 0.5–4 mg of sample [34]. In this case, the supernatant is removed following centrifugation and the remaining solid resuspended in silicon oil prior to analysis. More recently, Raman spectroscopic analysis has been added to this assay to enable changes in solid form to be identified [22]. Those thermodynamic solubility assays that characterize the remaining solid by the gold standard method, PXRD, generally require larger amounts of solid than microscopy techniques [32, 33]. A high-throughput thermodynamic solubility assay has been reported including PXRD assessment using 3 mg of solid sample [32]. The sample solution is filtered following 17 h shaking using custom-built nickel filter plates. The nickel filter serves as an effective means of presenting the remaining solid to the PXRD instrument, since it does not give background diffraction in the analysis window.

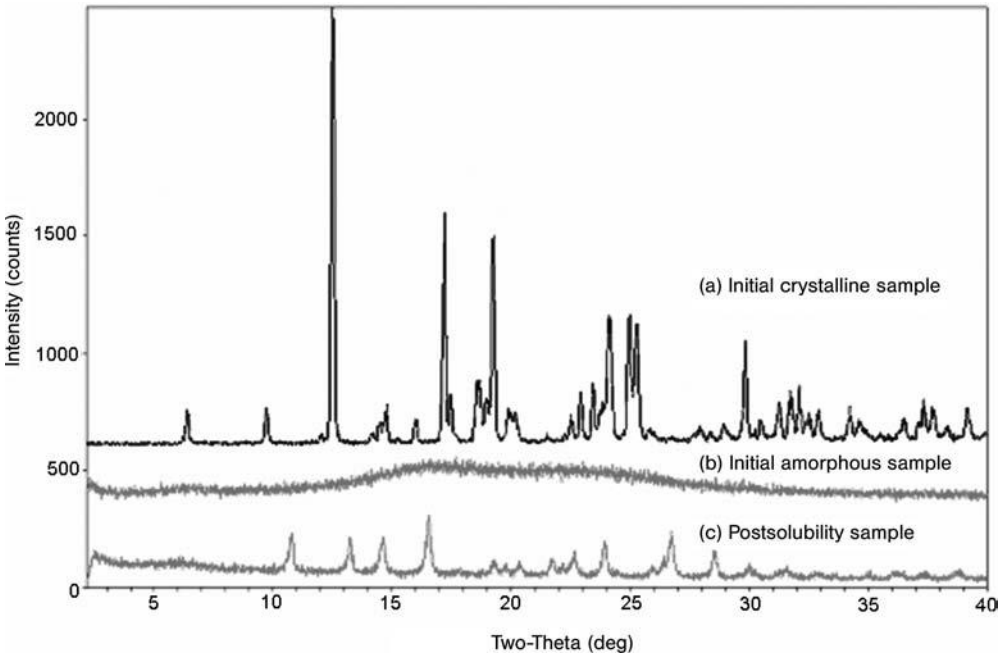


Figure 2.8 PXRD traces of bicalutamide: (a) the initial crystalline sample; (b) the initial amorphous sample; and (c) bicalutamide post-24 h agitation in phosphate buffer from the amorphous sample.

2.3.4
Solubility by Potentiometry

Potentiometry is a further technique used to measure aqueous solubility in discovery, although throughput limitations mean that this technique is used most often later in the LO stage. Potentiometric approaches specifically measure intrinsic solubility, which is the solubility of the neutral form of the molecule [38].

With the potentiometric approach, determination of intrinsic solubility is based upon the measurement of the pH shift caused by compound precipitation during acid–base titration of ionizable compounds. Two commercial potentiometric methods currently available are pSol [30, 39] and Cheqsol [40–42]. In the pSol method developed by Avdeef, a minimum of three titrations in the direction of dissolution are performed. Normal pH versus volume titration plots are reexpressed as Bjerrum plots, that is, average number of bound protons versus pH. The Bjerrum plots enable the shift in compound pK_a to be more readily observed and are used to determine intrinsic solubility (S_0) via Equation 2.5:

$$\log S_0 = \log\left(\frac{C}{2}\right) - (\mathrm{p}K_a^{\mathrm{app}} - \mathrm{p}K_a), \tag{2.5}$$

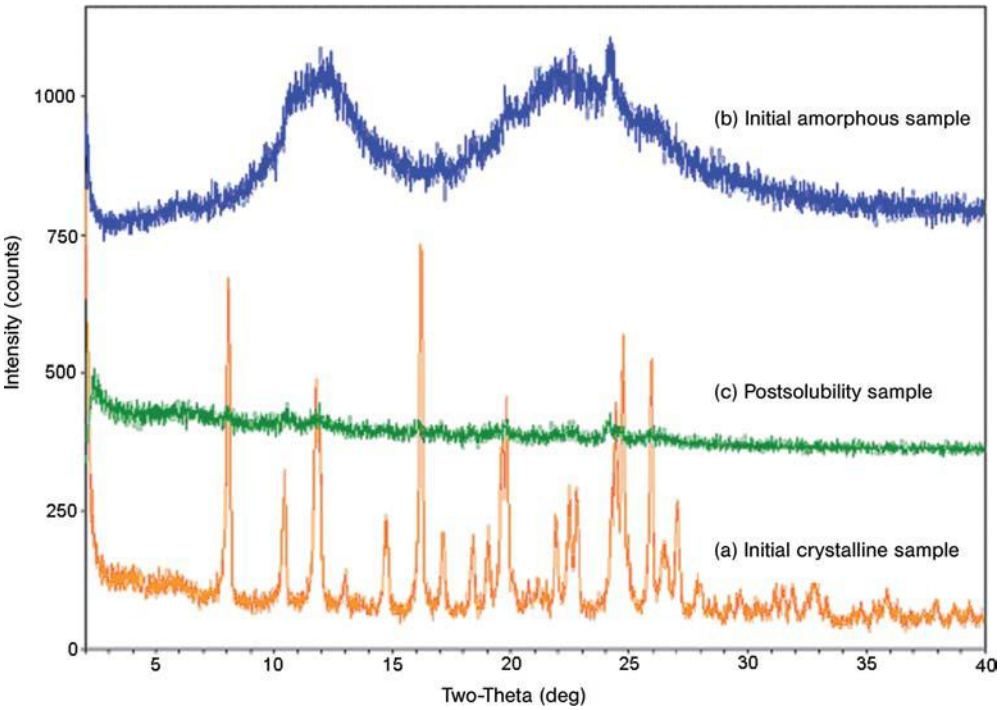


Figure 2.9 PXRD traces of nifedipine: (a) the initial crystalline sample; (b) the initial amorphous sample; and (c) nifedipine post-24 h agitation in phosphate buffer from the amorphous sample.

where S_0 is the intrinsic solubility, C is the total concentration of compound, pK_a^{app} is the measured pK_a in the presence of precipitation, and pK_a is the measured aqueous pK_a (no precipitation).

Intrinsic solubilities determined by pSol have been shown to agree well with values derived by shake-flask methods and have the advantage of requiring less compound since the technique does not require a sample calibration [39]. The time taken for each pSol solubility measurement is compound specific with 3–10 h quoted as typical [39]. Poorly soluble compounds can take longer than this and can be prone to reprecipitation causing outlying titration points. In these circumstances, the manufacturer's recommendation is to repeat the titrations in the presence of cosolvent spanning a range of percentage compositions and extrapolate back to the pure aqueous solubility value. Although this can improve the speed and accuracy of the titration, the need for further titrations generally adds to the overall experiment time. When undertaking the titrations, consideration should also be given to the compound physical form, which is usually not characterized. The initial physical form of the compound supplied may differ from that which is reprecipitated after each titration, and this can affect the solubility reported between the first and subsequent titrations. It should also be noted that since the time the solid spends in equilibrium with the aqueous solution at each pH is very short, there is less

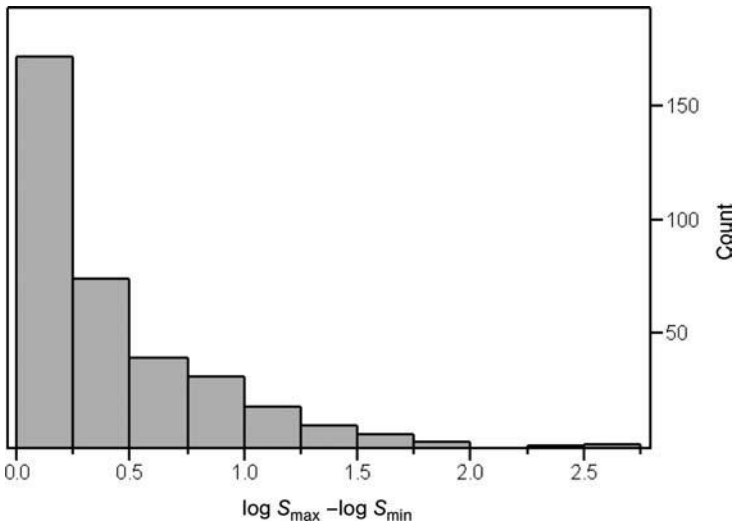


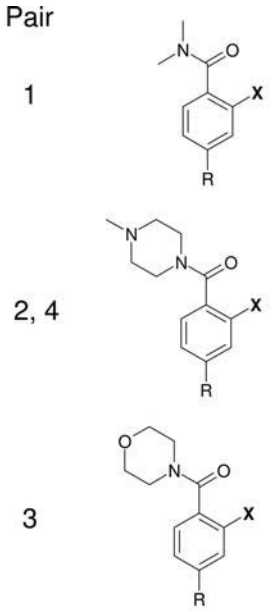
Figure 2.10 Difference between the maximum and minimum log solubility measured for different batches of each compound ($n = 360$). Values greater or less than assay detection limits (over-range values) have been excluded from the analysis. The mean $\log S_{max} - \log S_{min} = 0.44$.

opportunity for the thermodynamically stable form of the compound to be produced in the pSol method than in the 24 h of the classic shake-flask solubility method.

The Cheqsol method uses a similar experimental setup to pSol. However, in contrast to pSol, the Cheqsol method begins with the compound in solution and titration is followed until precipitation is detected by light scattering via a UV spectroscopic probe. At this point, the solution is switched repeatedly between a state of subsaturation and that of supersaturation by the addition of small amounts of acid and base to rapidly generate the point of equilibrium, from which the intrinsic solubility is calculated. As with the pSol technique, determination of the intrinsic solubility requires an accurate measure of the compound aqueous pK_a, and for poorly soluble compounds, titration in cosolvent with extrapolation to the aqueous state is used. The Cheqsol technique has the advantage of being relatively quick to achieve a state of equilibrium with experiment times in the region of 1 h for compounds that do not require cosolvent extrapolation. It has been shown with two compounds, diclofenac and sulindac, that changes in the crystalline polymorphs can occur as a result of the potentiometric cycling in Cheqsol [41, 43]. How the physical forms of the solid generated via Cheqsol compare to those observed in classic shake-flask studies is yet to be established.

2.3.5
Application of Thermodynamic Solubility Data in LI and LO

In LI and LO, thermodynamic solubility data are used to aid the understanding of DMPK data and guide formulation. When solubility is low, DMPK issues arise, such



Pair	X	S (μM)	log S	log D	Δlog S	Δlog D
1	H	112	-3.95	2.8	1.36	0.1
	Me	2550	-2.59	2.9		
2	H	42	-4.37	2.5	1.67	0.1
	Me	1950	-2.71	2.6		
3	H	943	-3.02	2.2	>0.54	0.1
	Me	>3300	>-2.48	2.3		
4	H	39	-4.41	2.8	0.7	0.1
	Me	197	-3.7	2.9		

Figure 2.11 Crystalline solubility match pairs showing effects of the *ortho*-methyl substituent.

as poor and variable bioavailability, fed/fasting effects [6], and lack of dose linearity. Poorly soluble molecules can also present significant and time-consuming formulation challenges during development. In trying to address these issues and challenges, solubility data are frequently used as a part of molecular design in discovery. In this respect, thermodynamic solubility data have been successfully applied to local and generic QSPR models using techniques such as BNN, MLR, and PLS [44] and to matched molecular pair analysis to determine substituent effects [45, 46] (Chapter 4). The latter approach has proved particularly useful for identifying novel structural effects on solubility. For example, Figure 2.11 shows a project that was able to improve solubility in a chemical series through incorporation of a single methyl group, not obvious based solely on log *P*, but which is suggestive of an effect on crystal packing. Interpretation of such observations requires that quality solubility data are used, coupled with the knowledge of the solid state to remove any concern about physical form effects. It is to be hoped that the trend toward increased characterization of the solid state earlier in discovery will lead to greater exploitation of solubility data as part of molecular design.

2.4 Conclusions

Aqueous solubility is an important property in discovery that has an impact across chemistry, DMPK, and formulation, and in the interpretation of biological assay results. The difficulties faced in the accurate prediction of solubility mean that measurement of aqueous solubility is essential from early HI onward. Automated approaches based on kinetic solubility and turbidimetric readouts have been developed in response to the high numbers of compounds requiring characterization in HI and also as a result of the desire to better understand biological assay data. However, more recently this demand has switched to high-quality thermodynamic solubilities to enable an early risk assessment of formulation issues in development and to identify quality hit series. Such frontloading of thermodynamic solubility has necessitated modification of the classic shake-flask approach to automation-friendly formats, which offers higher throughputs and can exploit the ease of dispensing provided by DMSO solution. In addition to this, there has also been a growing trend toward increased solid-state characterization of the sample in the solubility experiment during the LO stage of discovery. Such characterization means that there is a better understanding of the discovery solubility data and consequently greater confidence in its use to assess formulation risk when entering development and enhanced application of solubility data in molecular design.

Acknowledgments

We would like to thank our colleagues in the Alderley Park Physical Chemistry group for the AZ physical properties measurements captured in this chapter. In particular, thanks to Mark Timms and June Kenworthy who undertook most of the measurement work in Figures 2.3–2.7. We would also like to thank our computational chemistry colleague, Peter Kenny, for his contribution and valuable discussions on the use of solubility in fragment library design. Thanks also to our pharmaceutical colleagues Rhea Brent, Rebecca Booth, Rod Kittlety, and George Kirk for support around PXRD studies and valuable discussion on solid state and solubility. Thanks also to Trevor Johnstone and Jamie Scott for the preparation of amorphous material, enabling the study in Figure 2.7 to be undertaken.

References

1 Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, **46**, 3–26.

2 Kerns, E. (2001) High throughput physicochemical profiling for drug discovery. *Journal of Pharmaceutical Sciences*, **90** (11), 1838–1858.

3 Di, L. and Kerns, E.H. (2006) Biological assay challenges from compound solubility: strategies for bioassay

optimisation. *Drug Discovery Today*, **11** (9–10), 446–451.

4 McGovern, S.L., Caselli, E., Grigorieff, N. and Shoichet, B.K. (2002) A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *Journal of Medicinal Chemistry*, **45**, 1712–1722.

5 van de Waterbeemd, H., Smith, D.A., Beaumont, K. and Walker, D.K. (2001) Property-based design: optimization of drug absorption and pharmacokinetics. *Journal of Medicinal Chemistry*, **44**, 1–21.

6 Sunesen, V.H., Vedelsdal, R., Kristensen, H.G., Christrup, L. and Mullertz, A. (2005) Effect of liquid volume and food intake on the absolute bioavailability of danazol, a poorly soluble drug. *European Journal of Pharmaceutical Sciences*, **24**, 297–303.

7 Gu, C.H., Li, H., Levons, J., Lentz, K., Gandhi, R.B., Raghavan, K. and Smith, R.L. (2007) Predicting effect of food on extent of drug absorption based on physicochemical properties. *Pharmaceutical Research*, **24** (6), 1118–1130.

8 Mass, J., Kamm, W. and Hauck, G. (2007) An integrated early formulation strategy – from hit evaluation to preclinical candidate profiling. *European Journal of Pharmaceutics and Biopharmaceutics*, **66**, 1–10.

9 Schmid, E.F. and Smith, D.A. (2007) Pharmaceutical R&D in the spotlight: why is there still unmet medical need? *Drug Discovery Today*, **12**, 998–1006.

10 Comer, J.E.A. (2007) High-throughput measurement of log *D* and pK_a, in *Drug Bioavailability. Methods and Principles in Medicinal Chemistry*, Vol. 8 (eds H. van de Waterbeemd, H. Lennernas and P. Artursson), Wiley-VCH Verlag GmbH, Weinheim, pp. 21–43.

11 Serajuddin, A.T.M. (2007) Salt formation to improve drug solubility. *Advanced Drug Delivery Reviews*, **59**, 603–616.

12 Zhou, L., Yang, L., Tilton, S. and Wang, J. (2007) Development of a high throughput equilibrium solubility assay using miniaturized shake-flask method in early drug discovery. *Journal of Pharmaceutical Sciences*, **96** (11), 3052–3071.

13 Pan, L., Ho, Q., Tsutsui, K. and Takahashi, L. (2001) Comparison of chromatographic and spectroscopic methods used to rank compounds for aqueous solubility. *Journal of Pharmaceutical Sciences*, **90** (4), 521–529.

14 Fligge, T.A. and Schuler, A. (2006) Integration of a rapid automated solubility classification into early validation of hits obtained by high throughput screening. *Journal of Pharmaceutical and Biomedical Analysis*, **42**, 449–454.

15 Popa-Burke, I.G., Issakova, O., Arroway, J.D., Bernasconi, P., Chen, M., Coudurier, L., Galasinski, S., Jadhav, A.P., Janzen, W.P., Lagasca, D., Liu, D., Lewis, R.S., Mohney, R.P., Sepetov, N., Sparkman, D.A. and Hodge, C.N. (2004) Streamlined system for purifying and quantifying a diverse library of compounds and the effect of compound concentration measurements on the accurate interpretation of biological assay results. *Analytical Chemistry*, **76**, 7278–7287.

16 Sugano, K., Kato, T., Suzuki, K., Keiko, K., Sujaku, T. and Mano, T. (2006) High throughput solubility measurement with automated polarized light microscopy analysis. *Journal of Pharmaceutical Sciences*, **95** (10), 2115–2122.

17 Bevan, C. and Lloyd, R.S. (2000) A high-throughput screening method for the determination of aqueous drug solubility using laser nephelometry in micro-titer plates. *Analytical Chemistry*, **72**, 1781–1787.

18 Dehring, K.A., Workman, H.L., Miller, K.D., Mandagere, A. and Poole, S.K. (2004) Automated robotic liquid handling/laser-based nephelometry system for high throughput measurement of kinetic aqueous solubility. *Journal of Pharmaceutical and Biomedical Analysis*, **36**, 447–456.

19 Lilley, M., Shanter, M., Ardiff, M., Ciolkosz, T., Kashdan, M., Aubin, C. and Goodwin, J. A fully automated workstation for testing flow based kinetic solubility of compounds. Poster Presentation, BD

Biosciences, Two Oak Park, Bedford, MA, USA.

20 Chen, T., Shen, H. and Zhu, C. (2002) Evaluation of a method for high throughput solubility determination using a multi-wavelength UV plate reader. *Combinatorial Chemistry & High Throughput Screening*, **5**, 575–581.

21 Avdeef, A. (2001) Pharmacokinetic optimization in drug research, in *Biological, Physicochemical, and Computational Strategies* (eds B. Testa, H. van de Waterbeemd, G. Folkers and R. Guy), Verlag Helvetica Chimica Acta, Zurich, pp. 305–326.

22 Alsenz, J. and Kansy, M. (2007) High throughput solubility measurement in drug discovery and development. *Advanced Drug Delivery Reviews*, **59**, 546–567.

23 Colclough, N., Hunter, A., Kenny, P.W., Kittlety, R.S., Lobedan, L., Tam, K.Y. and Timms, M.A. (2008) High throughput solubility determination with application to selection of compounds for fragment screening. *Bioorganic & Medicinal Chemistry*, **16**, 6611–6616.

24 Pudipeddi, M. and Serajuddin, A.T.M. (2005) Trends in solubility of polymorphs. *Journal of Pharmaceutical Sciences*, **94** (5), 929–939.

25 Yalkowsky, S.H. (1999) *Solubility and Solubilisation in Aqueous Media*, Oxford University Press, Cambridge, pp. 81–115, Table 4.8.

26 Hancock, B. and Parks, M. (2000) What is the true solubility advantage for amorphous pharmaceuticals. *Pharmaceutical Research*, **17** (4), 397–404.

27 Bhattachar, S.N., Welsey, J.A. and Seadeek, S. (2006) Evaluation of the chemiluminescent nitrogen detector for solubility determinations to support drug discovery. *Journal of Pharmaceutical and Biomedical Analysis*, **41**, 152–157.

28 Hill, A. (2007) High-throughput solubility. PhysChem Forum 3, June 2007.

29 Reilly, J. (2007) CAD detector validation and potential for physicochemical measurements. PhysChem Forum 3, June 2007.

30 Glomme, A., Marz, J. and Dressman, J.B. (2005) Comparison of a minaturized shake-flask solubility method with automated potentiometric acid/base titrations and calculated solubilities. *Journal of Pharmaceutical Sciences*, **94** (1), 1–16.

31 Alelyunas, Y. (2006) Exploring the solubility limit of CNS discovery candidates. CHI's Molecular Medicine Tri-Conference, February 21–25, San Francisco, CA.

32 Seadeek, C., Ando, H., Bhattachar, S.N., Heimbach, T., Sonnenberg, J.L. and Blackburn, A.C. (2007) Automated approach to couple solubility with final pH and crystallinity for pharmaceutical discovery compounds. *Journal of Pharmaceutical and Biomedical Analysis*, **43**, 1660–1666.

33 Wyttenbach, N., Alsenz, J. and Grassmann, O. (2007) Miniaturized assay for solubility and residual solid screening (SORESOS) in early drug development. *Pharmaceutical Research*, **24** (5), 888–898.

34 Alsenz, J., Meister, E. and Haenel, E. (2007) Development of a partially automated solubility screening (PASS) assay for early drug development. *Journal of Pharmaceutical Sciences*, **96** (7), 1748–1762.

35 Lipinski, C.A. (2004) Solubility in water and DMSO: issues and potential solutions. *Pharmaceutical Profiling in Drug Discovery*, **1**, 93–125.

36 Buckton, G. (2002) Solid state properties, in *Pharmaceutics: The Science of Dosage Form Design* (ed. M.E. Aulton), Churchill Livingstone, Edinburgh, pp. 141–151.

37 Huang, L.F. and Tong, W.Q. (2004) Impact of solid state properties on developability assessment of drug candidates. *Advanced Drug Delivery Reviews*, **56**, 321–334.

38 Avdeef, A. (2007) Solubility of sparingly-soluble ionizable drugs. *Advanced Drug Delivery Reviews*, **59**, 568–590.

39 Avdeef, A., Berger, C.M. and Brownell, C. (2000) pH-metric solubility. 2. Correlation between the acid–base titration and the

saturation shake-flask solubility–pH methods. *Pharmaceutical Research*, **17** (1), 85–89.

40 Stuart, M. and Box, K. (2005) Chasing equilibrium: measuring the intrinsic solubility of weak acids and bases. *Analytical Chemistry*, **77**, 983–990.

41 Llinas, A., Burley, J.C., Box, K.J., Glen, R.C. and Goodman, J.M. (2007) Diclofenac solubility: independent determination of the intrinsic solubility of three crystal forms. *Journal of Medicinal Chemistry*, **50**, 979–983.

42 Box, K.J., Volgyi, G., Baka, E., Stuart, M., Takacs-Novak, K. and Comer, J.E.A. (2006) Equilibrium versus kinetic measurements of aqueous solubility, and the ability of compounds to supersaturate in solution: a validation study. *Journal of Pharmaceutical Sciences*, **95** (6), 1298–1307.

43 Llinas, A., Box, K.J., Burley, J.C., Glen, R.C. and Goodman, J.M. (2007) A new method for the reproducible generation of polymorphs: two forms of sulindac with very different solubilities. *Journal of Applied Crystallography*, **40**, 379–381.

44 Bruneau, P. (2001) Search for predictive generic model of aqueous solubility using Bayesian neural nets. *Journal of Chemical Information and Computer Sciences*, **41**, 1605–1616.

45 Leach, A.G., Jones, H.D., Cosgrove, D.A., Kenny, P.W., Ruston, L., Macfaul, P., Wood, J.M., Colclough, N. and Law, B. (2006) Matched molecular pairs as a guide in the optimization of pharmaceutical properties: a study of aqueous solubility, plasma protein binding and oral exposure. *Journal of Medicinal Chemistry*, **49** (23), 6672–6682.

46 Haubertin, D.Y. and Bruneau, P. (2007) A database of historically observed chemical replacements. *Journal of Chemical Information and Modeling*, **47**, 1294–1302.

3
Gastrointestinal Dissolution and Absorption of Class II Drugs

Arik S. Dahan and Gordon L. Amidon

Abbreviations

BA/BE	Bioavailability/bioequivalence
BCS	Biopharmaceutics Classification System
CMC	Critical micelle concentration
FDA	Food and Drug Administration
GI	Gastrointestinal
IR	Immediate release
IVIVC	*In vitro–in vivo* correlation
NSAIDs	Nonsteroidal anti-inflammatory drugs

Symbols

A_n	Absorption number
C_S	Equilibrium solubility
D	Diffusion coefficient
D_0	Dose number
D_n	Dissolution number
P_{eff}	Effective permeability
t_{abs}^{-1}	Effective absorption rate constant
t_{Diss}	Dissolution time
t_{res}	Residence time

3.1 Introduction

Modern drug discovery techniques (i.e., advances in *in vitro* high-throughput screening methods, the introduction of combinatorial chemistry) have resulted in

an increase in the number of low water solubility drug substances being selected as drug candidates. According to some estimates, more than 40% of new drug candidates are lipophilic and exhibit poor water solubility [1–3]. With very few exceptions, dissolution of the drug substances in the aqueous gastrointestinal (GI) milieu is a prerequisite for its absorption following oral administration. Hence, low-solubility compounds often suffer from limited oral bioavailability. A great challenge facing the pharmaceutical scientist is making these molecules into orally administered medications with sufficient bioavailability. This chapter reviews the fundamentals of low-solubility, high-permeability drug substances and the intestinal absorption process, including introduction to the Biopharmaceutics Classification System (BCS) focusing on Class II drugs (see Chapter 19). We will discuss the relevant variables affecting the absorption process of these compounds. In addition, this chapter provides a perspective on regulatory issues concerning low-solubility, high-permeability drug substances.

3.2 Drug Absorption and the BCS

The absorption of drugs following oral administration is a cascade of complex events, and the rate and extent of the drug absorption are affected by many factors. These include physicochemical factors (e.g., pK_a, solubility, stability, diffusivity, lipophilicity, surface area, particle size, and crystal form), physiological factors (e.g., GI pH, GI blood flow, gastric emptying, transit time through the different GI segments, and absorption mechanisms), and factors related to the dosage form (e.g., tablet, capsule, solution, suspension, emulsion, and gel) [4, 5].

When Fick's first law is applied to a membrane, the absorption of a drug across the GI mucosal surface under sink conditions can be written as

$$J_W = P_W \times C_W = \frac{dM}{dt} \times \frac{1}{A}, \tag{3.1}$$

where J_W is the mass transport across the GI wall (mass/area/time), P_W is the effective permeability, C_W is the concentration of the drug at the membrane, and A is the surface area. As developed by Amidon *et al.* [6–8], the analysis of this equation reveals that the fundamental events controlling oral drug absorption are the permeability of the drug through the GI membrane, the dissolution of the drug in the GI milieu, and the dose. These key parameters are characterized in the BCS by three dimensionless numbers [7]: absorption number (A_n), dissolution number (D_n), and dose number (D_0). These numbers take into account both physicochemical and physiological parameters and are fundamental to the oral absorption process.

The absorption number (A_n) is the ratio of permeability (P_{eff}) and the intestinal radius (R) multiplied by the residence time (t_{res}), which can be interpreted as the effective absorption rate constant (t_{abs}^{-1}) times the residence time:

$$A_n = \text{absorption number} = \frac{P_{eff}}{R} t_{res} = t_{abs}^{-1} t_{res}. \tag{3.2}$$

The dissolution number (D_n) is the ratio of the residence time and the dissolution time (t_{Diss}), which comprises the equilibrium solubility (C_S), diffusivity (D), density (ρ), the initial particle radius (r_0), and the intestinal residence time (t_{res}):

$$D_n = \text{dissolution number} = \frac{DC_S}{r_0} = \frac{4\pi r_0^2}{(4/3)\pi r_0^3 \rho} t_{res} = \frac{3t_{res} DC_S}{\rho r_0^2} = \frac{t_{res}}{t_{Diss}}. \tag{3.3}$$

Finally, the dose number (D_0) is the ratio of dose to dissolved drug:

$$D_0 = \text{dose number} = \frac{M_0/V_0}{C_S}, \tag{3.4}$$

where C_S is the equilibrium solubility, M_0 is the dose, and V_0 is the volume of water taken with the dose, which is generally set to be 250 ml. This volume was selected based on a typical bioequivalence study that administered an 8 oz (240 ml) glass of water with the oral dosage form. Thus, 250 ml, allowing a small GI residual volume, represents the initial gastrointestinal volume to which an oral dosage form is exposed in the fasting state. This number may be viewed as the number of glasses of water required to dissolve the drug dose.

Based on their solubility and intestinal permeability characteristics, drug substances have been classified into one of the four categories according to the BCS proposed by Amidon *et al.* [6] (Figure 3.1) (see Chapter 19):

- *BCS Class I:* High-solubility, high-permeability drugs. BCS Class I drugs are generally very well absorbed. An immediate release (IR) product of this class is

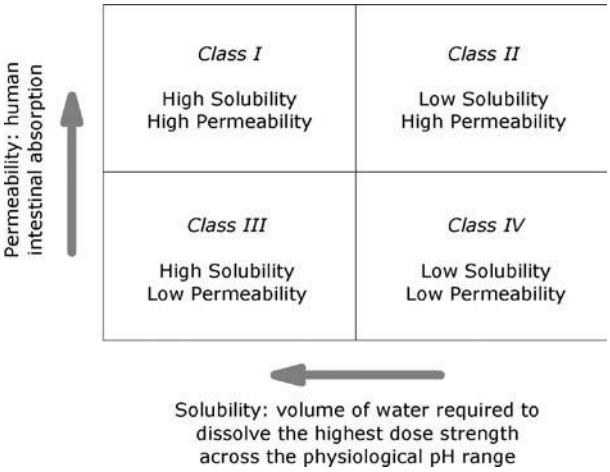


Figure 3.1 The Biopharmaceutics Classification System as defined by Amidon *et al.* [6]. The BCS is a classification of drug substances according to their solubility and permeability properties, in order to stand for the most fundamental view of the drug intestinal absorption process following oral administration.

expected to yield 100% intestinal absorption if at least 85% of the drug is dissolved within 30 min across the physiological pH. Hence, a waiver is granted for bioavailability/bioequivalence (BA/BE) studies of BCS Class I IR drug products.

- *BCS Class II:* Low-solubility, high-permeability drugs. These drugs are the main scope of this chapter. In general, BCS Class II drug products are likely to be limited by the dissolution/solubility rate.
- *BCS Class III:* High-solubility, low-permeability drugs. The intestinal absorption of this class of drugs will be limited by the permeability rate, as the dissolution is likely to occur rapidly. Hence, it has been suggested that as long as the drug product does not contain permeability modifying agents, a waiver for BA/BE studies for this class of drugs should be considered [9–12].
- *BCS Class IV:* Low-solubility, low-permeability drugs. These drugs are characterized by very poor oral bioavailability and tend to exhibit very large inter- and intrasubject variability. Hence, unless the dose is very low, they are generally poor oral drug candidates [1].

This BCS is one of the most significant prognostic tools created to facilitate product development in recent years and has been adopted by the US Food and Drug Administration (FDA) for setting BA/BE standards for drug product approval. The validity and applicability of the BCS have been the subject of extensive research and discussion, and classification of many drugs by the BCS is available in the literature [13–16]. Of particular interest are BCS Class II low-solubility, high-permeability drug substances, which account for the majority of new chemical entities. This chapter will focus on the different aspects of the intestinal absorption process of this class of drugs.

3.3 Class II Drugs

Being low-solubility, high-permeability compounds, Class II drug substances are characterized by high absorption number (A_n) and typically a high dose number (D_0). In these cases, dissolution might play a major role in the rate and extent of the oral absorption. In general, the dissolution of low-solubility drugs is low, that is, $D_n < 1$, while A_n and D_0 are high. In cases where both A_n and D_n are low, the compound will be classified as a Class IV drug [17, 18].

The intestinal absorption of Class II drug substances can be broadly viewed as dissolution limited or solubility limited. The concentration of the drug in the GI tract milieu will be determined by dissolution rate, while the upper limit will be the solubility. The classical example, still relevant today, which illustrates the effect of dissolution number and dose number on the fraction of dose absorbed of highly permeable drugs, is the case of digoxin and griseofulvin. A typical profile of the fraction of dose absorbed as a function of the dissolution number and the dose number for a large absorption number (i.e., highly permeable drug) is shown in

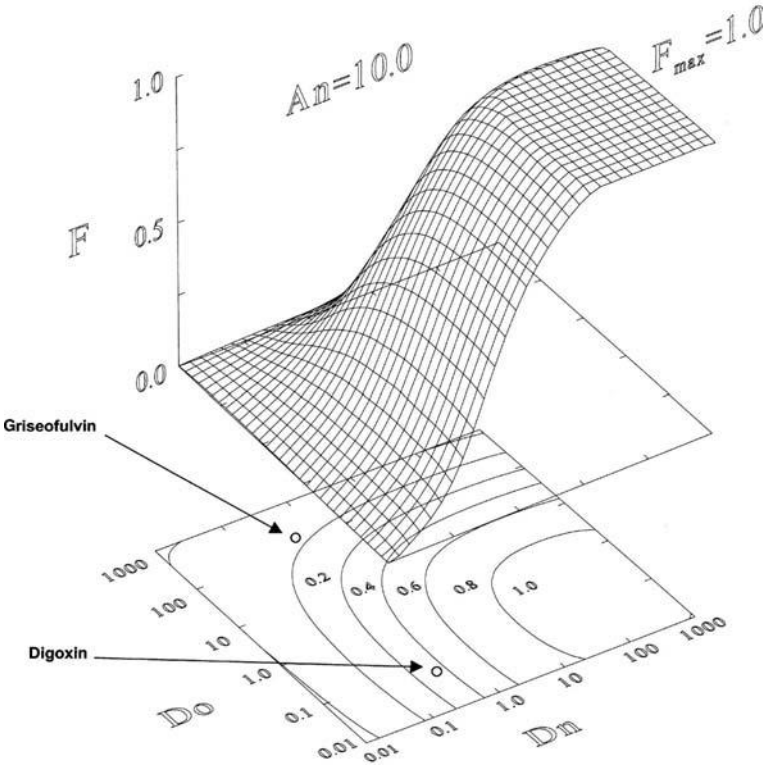


Figure 3.2 Graph of estimated fraction dose absorbed (F) vs dissolution number (D_n) and dose number (D_0) for a high permeability drug [6]. The dose number and the dissolution number of digoxin and griseofulvin are marked in the figure.

Figure 3.2. It can be seen that for high A_n, the critical range of the dose number and the dissolution number is around 1, where sharp changes in the fraction of drug absorbed are obtained due to small changes in D_0 and D_n [6].

Digoxin and griseofulvin have about the same solubility (~20 µg/ml), but a very different dose (0.5 and 500 mg, respectively). Consequently, while digoxin has a very low dose number (0.08), griseofulvin has a dose number, D_0, of 133, which indicates that over 33 l of water is required to dissolve a single griseofulvin dose.

The dose number and the dissolution number of digoxin and griseofulvin are marked in Figure 3.2. It can be seen that for digoxin fraction of drug absorbed is highly dependent upon the dissolution number. A complete intestinal absorption can be expected for digoxin if the drug particle size is small enough (i.e., high D_n); however, digoxin might be a dissolution rate limited drug if the drug particle size is too large (small D_n) [6, 19, 20]. Hence, micronized digoxin powder will lead to a faster dissolution rate, and the intestinal residence time would be sufficient for complete absorption. However, changes in griseofulvin D_n alone would not be sufficient to

influence the fraction of the drug absorbed. Micronization of the drug powder is not expected to improve griseofulvin absorption unless accompanied by reduced dose number, for example, by formulation that enables efficient solubilization in the GI milieu [21, 22]. Without that, griseofulvin absorption will be solubility limited, and complete absorption of the drug cannot be achieved.

3.4 GI Physiological Variables Affecting Class II Drug Dissolution

With very few exceptions, dissolution of the drug substance in the GI tract milieu is a prerequisite for drug absorption following oral administration. For Class II compounds, the rate-limiting factor in their intestinal absorption is dissolution /solubility [23–25]. Hence, in-depth understanding of this process is essential in the oral delivery of low-solubility compounds. Factors governing the dissolution process can be directly identified from the following equation, based on the Nernst–Brunner and Levich modifications of the Noyes–Whitney model [26–28]:

$$\frac{dX_d}{dt} = \frac{A \times D}{h}\left(C_S - \frac{X_d}{V}\right). \tag{3.5}$$

The dissolution rate is a function of the surface area of the solid drug (A), the diffusion coefficient of the drug (D), the effective diffusion boundary layer thickness adjacent to the dissolving surface (h), the saturation solubility of the drug (C_S), the volume of the fluid available for dissolution (V), and the amount of drug already dissolved (X_d). Thus, the rate of dissolution is highly affected by the physicochemical properties of the drug and by many GI physiological factors that will be discussed in this section.

3.4.1 Bile Salts

Bile acids affect both solubility and dissolution by micellization and wetting effects [29, 30]. Hence, they play a significant role in Class II drugs' intestinal absorption, which is a dissolution/solubility rate limited process. The bile fluid is secreted from hepatocytes in the liver and stored and concentrated in the gall bladder before release into the small intestine. The major organic solutes of the bile are bile acids, phospholipids (particularly lecithin), and cholesterol. The bile acids are derivatives of cholesterol in which hydroxyl and carboxylic acid are attached to the steroid moiety, converting it into a powerful natural surfactant. Average typical intestinal concentrations for bile acids and phospholipids are 5 and 0.2 mM, respectively, in the fasting state [31, 32] and 15 and 4 mM, respectively, in the fed state [32]. Above their critical micelle concentration (CMC), these biliary secretions aid in drug dissolution by forming submicron mixed micelles in which the low water-soluble molecule is solubilized and gets to the absorptive membrane of the enterocyte [33–35].

Enhanced solubility of the low water-soluble drug may be obtained by bile salts through the wetting mechanism [36, 37]. This is the main mechanism when the bile salts are present at a level below their CMC [35, 38].

3.4.2 GI pH

BCS II class includes nonionizable substances (e.g., carbamazepine, griseofulvin) and ionizable compounds, either acids (e.g., ibuprofen) or bases (e.g., diazepam). For nonionizable drugs, pH changes along the GI tract would not have an impact on drug solubilization/dissolution. However, the intestinal absorption of BCS Class II ionizable drugs' is highly dependent on pH [30, 39]. Generally, aqueous solubility is directly proportional to the number of hydrogen bonds that can be formed with water, and hence the ionized form exhibits greater aqueous solubility than do the unionized species. The nonionized form has generally better membrane permeability than the ionized species; however, since permeability is not the rate-limiting step for Class II drugs absorption process, this effect is less significant. Hence, for these drugs, an alteration in the degree of conversion of the unionized drug to its ionized form upon dissolution as a function of the pH may dictate the rate of absorption.

Throughout the passage along the GI tract, a drug product experiences a wide pH range. Gastric pH highly depends upon food intake and values from 1 to 8 are reported, while the fasting-state stomach pH is 1.4–2.1 [40–42]. In general, the pH values in the small intestine are higher than those in the stomach, much less dependent upon food intake, and show an upward gradient from the proximal to the distal segments, covering a range of 4.4–7.4 [23, 40, 43].

For BCS Class II weak base drugs (e.g., dipyridamole, ketoconazole), ionization will occur in the gastric acidic environment, leading to a rapid dissolution in the stomach. As the drug is emptying from the stomach to the duodenum, the degree of ionization is significantly reduced due to the elevated pH, with possible precipitation of the drug [44, 45]. This leads to a complicated intestinal absorption pattern controlled by many factors including the extent of supersaturation and solid form of the weak base, pH, fluid volume, viscosity, and bile salts' concentration [30, 46, 47].

For BCS Class II weak acid drugs (e.g., ibuprofen, ketoprofen) with pK_a in the GI physiological range, extensive ionization is expected in the small intestine. As the intestinal pH is on average higher than the pK_a in more than one unit, the apparent solubility of the weak acid increases by 10–100-fold. Thus, the *in vivo* solubility and dissolution of these drugs would be high, presumably behaving more likely as Class I compounds, as discussed in Section 7.1 [38, 48, 49].

3.4.3 GI Transit

The two major components of the GI transit are the gastric residence time, dictated by the gastric emptying, and the small intestinal transit time. In general, the rate of gastric emptying is of significance in cases where dissolution is relatively fast [50]. For

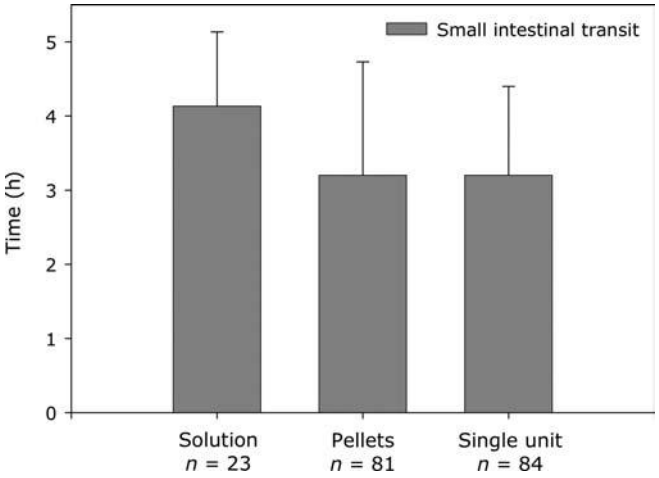


Figure 3.3 Human small intestinal transit time of different pharmaceutical dosage forms measured by γ-scintigraphy [40]. The intestinal transit time is fairly fixed and largely independent of the physical properties of the system or food intake.

BCS Class II drug substances, dissolution rate is expected to be slower than the gastric emptying, and hence this factor is not thought to be significant in their intestinal absorption.

The transit time through the small intestine has the potential to affect Class II intestinal absorption, as increased time in the main absorption site is expected to yield higher absorption. However, the intestinal transit time is fairly fixed and largely independent of food intake (i.e., fasting versus fed states) or the physical properties of the system (Figure 3.3) [40]. On average, the small intestinal transit time is around 3–4 h [51, 52].

3.4.4 Drug Particle Size

The particle size of the drug powder might be an important physical parameter in the dissolution rate, as shown in the case of digoxin (Section 3.3). Smaller particle size will lead to higher surface area available for dissolution, resulting in a faster dissolution rate [53, 54]. Hence, particularly for Class II dissolution rate limited drug substances, the dissolution rate is expected to increase proportionally with an increase in surface area, and the particle size will be a critical formulation variable. The density of particles might also affect the dissolution rate through alteration of the *in vivo* particle dispersion, as greater dispersion leads to improved dissolution [19, 55]. The effect of reduced particle size seems to depend upon food intake. It has been shown that under the fasting state, reduced particle size had a major effect on *in vitro* dissolution and *in vivo* oral absorption of DPC961, a BCS Class II reverse

transcriptase inhibitor, while no effect was observed under the fed state [56]. This may be attributed to the differences in the solubilization capacities in the fed and fasting states. Improved *in vivo* dissolution rate in humans was observed for decreased spironolactone particle size, however, with no influence on the relative bioavailability [57]. The authors suggested that the lack of difference between the bioavailability of the two particle sizes may be related to insufficient washout of particles after perfusion ends, reabsorption of surface active ingredients along the GI tract, relatively small difference in particle size, and the large inter- and intraindividual differences in pharmacokinetic variables [57].

In regard to the drug powder particle size, it is worth noting that it is the effective surface area that is important, that is, the surface area that is available to the dissolution fluid (the wetted surface), and not the actual particle size. This is important in the case of a highly hydrophobic drug in a dissolution medium that has poor wetting properties and in the case of a manufacturing process that changes the particle size during manufacture. In this case, decreased particle size will slow the dissolution rate [58, 59].

3.4.5
Volume Available for Dissolution

The volume of fluids available in the GI tract for drug dissolution depends on the volume of fluids coadministered with the drug, secretions into the GI lumen, and water flux across the gut wall. This factor is of high importance particularly for BCS Class II compounds, as higher volume of available fluids will enable the dissolution of higher amounts of drug. Average fluid volumes approximately representing the usual physiological range are 300–500 ml for the fasting-state stomach (although the volume may be as low as 20–30 ml) and 800–900 ml for the fed-state stomach; 500 ml for the fasting-state small intestine and 900–1000 ml for the fed-state small intestine [23, 60, 61] (although volumes in the upper small intestine in the fed state can reach as much as 1.5 l [62]).

3.5
In Vitro Dissolution Tests for Class II Drugs

As denoted above, the rate-limiting step in the oral absorption of Class II drug substances is often the *in vivo* dissolution [23–25]. Hence, a well-designed dissolution test should be capable of providing adequate *in vitro–in vivo* correlation (IVIVC).

3.5.1
Biorelevant Media

The choice of medium is expected to play a very important role in the dissolution of BCS Class II drug substances. The media used need to closely represent the *in vivo* conditions in the upper GI tract to achieve a meaningful IVIVC. As discussed in this

chapter, Class II drug dissolution depends on a wide variety of parameters, such as surfactants, pH, buffer capacity, ionic strength, and the volume available for dissolution. Class II dissolution test media often fail to produce adequate IVIVC due to composition that does not take into account these essential factors.

Examples of suitable media for simulating the composition of proximal GI tract include simulated gastric fluid (SGF) and surfactant, used to simulate the stomach conditions under the fasting state; long-life milk 3.5% fat, used to simulate the stomach conditions under fed state; fasting-state simulating intestinal fluid (FaSSIF), the medium used to represent fasting state in the proximal small intestine; and fed-state simulating intestinal fluid (FeSSIF), the fed-state proximal small intestine medium [23, 44, 61, 63]. Very good IVIVC was obtained with these media for the low-solubility drugs albendazole, danazol, ketoconazol, atovaquone, and troglitazone [44, 61, 64–66]. For these drugs, fed-state versus fasting-state effects, as well as formulation effects, could be predicted by appropriate dissolution tests, showing the potential of biorelevant dissolution tests to adequately simulate the *in vivo* GI tract milieu composition.

The relationship between the hydrodynamics in the GI tract and that in the current available dissolution tests is another factor that must be considered. It has been reported that, provided an appropriate composition is chosen for the dissolution test, the United States Pharmacopeia (USP) paddle apparatus can be used to reflect variations in hydrodynamic conditions in the upper GI tract [67–69]. However, more data are warranted, as this might insert uncertainty into the interpretation of dissolution tests, even when the composition of the GI milieu is well simulated.

3.5.2 Dynamic Lipolysis Model

One of the approaches to improve the oral bioavailability of Class II drug substances is the utilization of lipid-based drug delivery systems [1, 70, 71]. Enhanced dissolution/solubilization of the coadministered lipophilic drug by stimulation of biliary and pancreatic secretions is a major factor in this phenomenon [72]. Additional mechanisms by which lipid-based formulations might enhance the absorption of lipophilic drugs include presentation of the low water-soluble drug as a solution and hence avoiding the complexities associated with solid state; mild prolongation of GI tract residence time; possible reduced metabolism and efflux activity [73, 74]; and changing the intraenterocyte transport route by stimulation of the lymphatic transport pathway [75, 76].

Following oral administration, the lipidic component of the lipid-based formulation is subjected to enzymatic hydrolysis. Pancreatic lipase, upon complexation with colipase, acts at the surface of the emulsified triglyceride droplets to produce the corresponding 2-monoglyceride and two fatty acids. These amphiphilic lipid digestion products interact with the endogenous bile salts and phospholipids, forming colloidal structures holding different levels of surface activity, which enables the solubilization of the coadministered poorly water-soluble compound and prevents their precipitation in the aqueous GI tract milieu. In most cases, this process, which

maintains the poorly water-soluble drug in solution and prevents its precipitation, is thought to be the primary mechanism by which lipid-based drug delivery systems augment the oral absorption of lipophilic drugs [32, 77]. Owing to the dependence of lipid-based delivery systems' performance on GI processing, a meaningful release test of such formulations will require the presence of lipolytic enzymes that catalyze GI lipid digestion *in vivo* [78].

A dynamic *in vitro* lipolysis model, which provides simulation of the *in vivo* lipid digestion process, has been developed over the past few years [79–81]. This model has been used for assessing different lipid-based delivery systems, and a correlation between *in vitro* dissolution and *in vivo* absorption in the lipolysis model was reported for a number of low water-soluble drugs, including halofantrine [82], griseofulvin [83], and progesterone [84]. Overall, this model looks promising; however, the number of studies evaluating the IVIVC provided by this method is still limited, and a more thorough characterization of the model in different physiological conditions is warranted.

3.6
BCS-Based FDA Guidelines: Implications for Class II Drugs

The current FDA guidelines on waiver of *in vivo* BA/BE studies for BCS Class I drugs in rapid dissolution IR solid oral dosage form are generally considered highly conservative, especially with respect to the class boundaries of solubility, permeability, and dissolution. Thus, the possibility of modifying these boundaries and criteria to allow biowaivers for additional drug products has received increasing attention [24, 85–88].

3.6.1
Potential of Redefining BCS Solubility Class Boundary

Currently, drug substances are classified as high-solubility compounds if the highest strength is soluble in 250 ml or less of aqueous media throughout the pH range of 1.0–7.5. Three factors in this requirement are considered highly conservative and may be reevaluated: (1) the required pH range; (2) the nature of the media; and (3) the volume of the media [86, 88].

Under fasting state, the GI pH varies from 1.4 to 2.1 in the stomach, 4.9–6.4 in the duodenum, 4.4–6.6 in the jejunum, and 6.5–7.4 in the ileum [42, 89]. Hence, it seems reasonable to redefine the BCS class boundary pH range from 1.0–7.5 to 1.4–6.8. Moreover, if a drug product meets the dissolution criterion, that is, not less that 85% dissolved within 30 min, its dissolution process is probably completed during the jejunum, as it generally takes 85 min for a drug to reach the ileum [5, 90]. Thus, it might be reasonable to narrow the pH range requirement even more. This would somewhat ease the requirement for Class II basic drug substances. Many ionizable Class II acidic drugs have low solubility at low pH, but they are highly soluble at higher pH values. For example, most nonsteroidal anti-inflammatory drugs

(NSAIDs) are poorly soluble in the stomach but are highly soluble in the distal intestine, which is the main absorption region of most compounds. These drugs are classified as Class II drugs according to the current solubility definitions. However, their absolute human bioavailability is 90% or higher, thus exhibiting BCS Class I behavior [91]. In the same manner, the NSAID ketoprofen is classified as Class II drug due to low solubility at low pH values; however, it has been demonstrated that this drug behaves in a manner essentially equivalent to Class I drugs and could be considered for a waiver of BA/BE studies [38, 48]. Yazdanian *et al.* [92] evaluated the possible impact of changing the current high-solubility definition on the BCS classification of 18 acidic NSAIDs. While 15 of the 18 drugs were classified as Class II based on the current high-solubility definition, 8 of the 18 could be classified as Class I by considering pH 5 and above. Overall, consideration for an intermediate solubility classification for such compounds, which will take into account only the intestinal relevant pH range, seems warranted.

While the solubility classification is based on the dissolution of the drug in aqueous buffers, the *in vivo* conditions for drug dissolution contain bile salts and phospholipids, even under fasting state. As denoted in Section 3.4.1, these are powerful natural surfactants that aid in the dissolution/solubilization of the drug substances. A medium that more adequately reflects physiological conditions may be more relevant in assessing *in vivo* solubility and dissolution, and potentially, drugs that are classified as Class II according to the current solubility definitions could be classified as Class I under these conditions [66, 86, 88].

Under the fasting state, the physiological volume of the small intestine varies from 50 to 1100 ml with an average of 500 ml [17, 93]. Upon administration, the drug is usually ingested with 250 ml of water that is immediately available to dissolve the solid particles in the stomach. If the drug is not in solution on gastric emptying, it is then exposed to additional fluids in the small intestine. Hence, the dose volume of 250 ml is a conservative estimate of the actual *in vivo* volume available for solubilization and dissolution. However, the wide variability of the small intestinal fluid volume makes an appropriate volume definition difficult to set.

3.6.2
Biowaiver Extension Potential for Class II Drugs

As discussed above, the rate-limiting step in the oral absorption of Class II drug substances is likely to be the *in vivo* dissolution [23–25]. For Class II dissolution rate limited drugs, hence, if *in vivo* dissolution can be estimated *in vitro*, an *in vitro–in vivo* correlation may be established. As discussed in Section 3.5, such media have been developed, and an adequate IVIVC was shown for number of Class II drugs. However, due to the numerous *in vivo* parameters involved, it appears that more research is needed to develop uniform dissolution media reflecting *in vivo* dissolution conditions, to establish an adequate IVIVC, and to asses the risk of bioinequivalence [86, 88]. In addition, the relationship between the hydrodynamics in the currently available dissolution tests and the actual *in vivo* situation is not adequately characterized and might interfere to obtain the correlation.

As discussed in detail above, the intestinal absorption of Class II drug substances may be limited by dissolution rate or solubility rate. In the latter case, when the absorption is limited by the drug equilibrium solubility, an IVIVC is not likely to be obtained. The GI tract drug concentrations in this case will be close to the saturation concentration, and since standard dissolution tests are carried out under sink conditions, they can predict only processes occurring well below the saturation concentration [85]. Hence, at this point, Class II solubility rate limited drugs are probably poor candidates for BA/BE waiver.

3.7 Conclusions

Looking into the future, more BCS Class II drug candidates are likely to be produced, and the delivery of these molecules through the oral route is expected to be a continuing challenge.

In this chapter, we have reviewed the rate and extent of oral absorption of this class of drugs and discussed the numerous factors, physicochemical, physiological, and dosage form, that must be considered in effectively delivering these drug candidates. In-depth comprehension of these factors and their influence on the intestinal absorption process is essential in the effective oral delivery of BCS Class II drug substances.

References

1 Dahan, A. and Hoffman, A. (2006) *Enhancement in Drug Delivery* (eds E. Touitou and B.W. Barry), CRC Press, pp. 111–127.

2 Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, **46** (1–3), 3.

3 van de Waterbeemd, H., Smith, D.A., Beaumont, K. and Walker, D.K. (2001) Property-based design: optimization of drug absorption and pharmacokinetics. *Journal of Medicinal Chemistry*, **44** (9), 1313–1333.

4 Lennernas, H. (1998) Human intestinal permeability. *Journal of Pharmaceutical Sciences*, **87** (4), 403–410.

5 Yu, L.X., Lipka, E., Crison, J.R. and Amidon, G.L. (1996) Transport approaches to the biopharmaceutical design of oral drug delivery systems: prediction of intestinal absorption. *Advanced Drug Delivery Reviews*, **19** (3), 359.

6 Amidon, G.L., Lennernas, H., Shah, V.P. and Crison, J.R. (1995) A theoretical basis for a biopharmaceutic drug classification: the correlation of *in vitro* drug product dissolution and *in vivo* bioavailability. *Pharmaceutical Research*, **12** (3), 413.

7 Oh, D.M., Curl, R.L. and Amidon, G.L. (1993) Estimating the fraction dose absorbed from suspensions of poorly soluble compounds in humans: a mathematical model. *Pharmaceutical Research*, **10** (2), 264.

8 Sinko, P.J., Leesman, G.D. and Amidon, G.L. (1991) Predicting fraction dose

absorbed in humans using a macroscopic mass balance approach. *Pharmaceutical Research*, **8** (8), 979.

9 Becker, C., Dressman, J.B., Amidon, G.L., Junginger, H.E., Kopp, S., Midha, K.K., Shah, V.P., Stavchansky, S. and Barends, D.M. (2008) Biowaiver monographs for immediate release solid oral dosage forms: pyrazinamide. *Journal of Pharmaceutical Sciences*, **97** (9), 3709–3720.

10 Blume, H.H. and Schug, B.S. (1999) The biopharmaceutics classification system (BCS): Class III drugs – better candidates for BA/BE waiver? *European Journal of Pharmaceutical Sciences*, **9** (2), 117.

11 Cheng, C.L., Yu, L.X., Lee, H.L., Yang, C.Y., Lue, C.S. and Chou, C.H. (2004) Biowaiver extension potential to BCS Class III high solubility-low permeability drugs: bridging evidence for metformin immediate-release tablet. *European Journal of Pharmaceutical Sciences*, **22** (4), 297.

12 Jantratid, E., Prakongpan, S., Amidon, G.L. and Dressman, J. (2006) Feasibility of biowaiver extension to biopharmaceutics classification system class III drug products: cimetidine. *Clinical Pharmacokinetics*, **45** (4), 385–399.

13 Kasim, N.A., Whitehouse, M., Ramachandran, C., Bermejo, M., Lennernäs, H., Hussain, A.S., Junginger, H.E., Stavchansky, S.A., Midha, K.K., Shah, V.P. and Amidon, G.L. (2004) Molecular properties of WHO essential drugs and provisional biopharmaceutical classification. *Molecular Pharmaceutics*, **1** (1), 85–96.

14 Lindenberg, M., Kopp, S. and Dressman, J.B. (2004) Classification of orally administered drugs on the World Health Organization Model list of Essential Medicines according to the biopharmaceutics classification system. *European Journal of Pharmaceutical Sciences*, **58** (2), 265.

15 Takagi, T., Ramachandran, C., Bermejo, M., Yamashita, S., Yu, L.X. and Amidon, G.L. (2006) A provisional biopharmaceutical classification of the top 200 oral drug products in the United States, Great Britain, Spain, and Japan. *Molecular Pharmaceutics*, **3** (6), 631–643.

16 Yang, Y., Faustino, P.J., Volpe, D.A., Ellison, C.D., Lyon, R.C. and Yu, L.X. (2007) Biopharmaceutics classification of selected beta-blockers: solubility and permeability class membership. *Molecular Pharmaceutics*, **4** (4), 608–614.

17 Lobenberg, R. and Amidon, G.L. (2000) Modern bioavailability, bioequivalence and biopharmaceutics classification system. New scientific approaches to international regulatory standards. *European Journal of Pharmaceutics and Biopharmaceutics*, **50** (1), 3–12.

18 Martinez, M.N. and Amidon, G.L. (2002) A mechanistic approach to understanding the factors affecting drug absorption: a review of fundamentals. *Journal of Clinical Pharmacology*, **42** (6), 620–643.

19 Dressman, J. and Fleisher, D. (1986) Mixing-tank model for predicting dissolution rate control or oral absorption. *Journal of Pharmaceutical Sciences*, **75** (2), 109–116.

20 Jounela, A., Pentikainen, P. and Sothmann, A. (1975) Effect of particle size on the bioavailability of digoxin. *European Journal of Clinical Pharmacology*, **8** (5), 365–370.

21 Crison, J.R., Weiner, N.D. and Amidon, G.L. (1997) Dissolution media for *in vitro* testing of water-insoluble drugs: effect of surfactant purity and electrolyte on *in vitro* dissolution of carbamazepine in aqueous solutions of sodium lauryl sulfate. *Journal of Pharmaceutical Sciences*, **86** (3), 384–388.

22 Shah, V.P., Konecny, J.J., Everett, R.L., McCullough, B., Noorizadeh, A.C. and Skelly, J.P. (1989) *In vitro* dissolution profile of water-insoluble drug dosage forms in the presence of surfactants. *Pharmaceutical Research*, **6** (7), 612.

23 Dressman, J.B., Amidon, G.L., Reppas, C. and Shah, V.P. (1998) Dissolution testing as a prognostic tool for oral drug absorption: immediate release dosage forms. *Pharmaceutical Research*, **15** (1), 11.

24 Fagerholm, U. (2007) Evaluation and suggested improvements of the Biopharmaceutics Classification System (BCS). *The Journal of Pharmacy and Pharmacology*, **59**, 751.

25 Wilding, I. (1999) Evolution of the biopharmaceutics classification system (BCS) to oral modified release (MR) formulations; what do we need to consider? *European Journal of Pharmaceutical Sciences*, **8** (3), 157–159.

26 Levich, V. (1962) *Physicochemical Hydrodynamics*, Prentice-Hall, Englewood Cliffs, NJ.

27 Nernst, W. and Brunner, E. (1904) Theorie der reaktionsgeschwindigkeit in heterogenen systemen. *Zeitschrift für Physikalische Chemie*, **47**, 52–110.

28 Noyes, A. and Whitney, W. (1897) The rate of solution of solid substances in their own solutions. *Journal of the American Chemical Society*, **19**, 930–934.

29 Bakatselou, V., Oppenheim, R.C. and Dressman, J.B. (1991) Solubilization and wetting effects of bile salts on the dissolution of steroids. *Pharmaceutical Research*, **8** (12), 1461.

30 Sheng, J.J. (2007) Toward an *in vitro* bioequivalence test. PhD Thesis, University of Michigan, Ann Arbor.

31 Persson, E., Gustafsson, A.-S., Carlsson, A., Nilsson, R., Knutson, L., Forsell, P., Hanisch, G., Lennernas, H. and Abrahamsson, B. (2005) The effects of food on the dissolution of poorly soluble drugs in human and in model small intestinal fluids. *Pharmaceutical Research*, **22** (12), 2141.

32 Porter, C.J.H., Trevaskis, N.L. and Charman, W.N. (2007) Lipids and lipid-based formulations: optimizing the oral delivery of lipophilic drugs. *Nature Reviews. Drug Discovery*, **6** (3), 231.

33 Amidon, G.E., Higuchi, W. and Ho, N. (1982) Theoretical and experimental studies of transport of micelle-solubilized solutes. *Journal of Pharmaceutical Sciences*, **71** (1), 77–84.

34 Granero, G., Ramachandran, C. and Amidon, G.L. (2005) Dissolution and solubility behavior of fenofibrate in sodium lauryl sulfate solutions. *Drug Development and Industrial Pharmacy*, **31** (9), 917–922.

35 Horter, D. and Dressman, J.B. (2001) Influence of physicochemical properties on dissolution of drugs in the gastrointestinal tract. *Advanced Drug Delivery Reviews*, **46** (1–3), 75.

36 Luner, P.E. (2000) Wetting properties of bile salt solutions and dissolution media. *Journal of Pharmaceutical Sciences*, **89** (3), 382–395.

37 Luner, P.E. and Kamp, D.V. (2001) Wetting behavior of bile salt–lipid dispersions and dissolution media patterned after intestinal fluids. *Journal of Pharmaceutical Sciences*, **90** (3), 348–359.

38 Sheng, J.J., Kasim, N.A., Chandrasekharan, R. and Amidon, G.L. (2006) Solubilization and dissolution of insoluble weak acid, ketoprofen: effects of pH combined with surfactant. *European Journal of Pharmaceutical Sciences*, **29** (3–4), 306.

39 Balakrishnan, A., Rege, B.D., Amidon, G.L. and Polli, J.E. (2004) Surfactant-mediated dissolution: contributions of solubility enhancement and relatively low micelle diffusivity. *Journal of Pharmaceutical Sciences*, **93** (8), 2064–2075.

40 Davis, S.S., Hardy, J.G. and Fara, J.W. (1986) Transit of pharmaceutical dosage forms through the small intestine. *Gut*, **27** (8), 886–892.

41 Lindahl, A., Ungell, A.L., Knutson, L. and Lennernäs, H. (1997) Characterization of fluids from the stomach and proximal jejunum in men and women. *Pharmaceutical Research*, **14** (4), 497.

42 Oberle, R. and Amidon, G.L. (1987) The influence of variable gastric emptying and intestinal transit rates on the plasma level curve of cimetidine: an explanation for the double peak phenomenon. *Journal of Pharmacokinetics and Biopharmaceutics*, **15** (5), 529–544.

43 Kalantzi, L., Goumas, K., Kalioras, V., Abrahamsson, B., Dressman, J. and Reppas, C. (2006) Characterization of the human upper gastrointestinal contents under conditions simulating bioavailability/bioequivalence studies. *Pharmaceutical Research*, **23** (1), 165.

44 Galia, E., Nicolaides, E., Hörter, D., Löbenberg, R., Reppas, C. and Dressman, J.B. (1998) Evaluation of various dissolution media for predicting *in vivo* performance of class I and II drugs. *Pharmaceutical Research*, **15** (5), 698.

45 Kalantzi, L., Persson, E., Polentarutti, B., Abrahamsson, B., Goumas, K., Dressman, J. and Reppas, C. (2006) Canine intestinal contents vs. simulated media for the assessment of solubility of two weak bases in the human small intestinal contents. *Pharmaceutical Research*, **23** (6), 1373.

46 Russell, T.L., Berardi, R.R., Barnett, J.L., O'Sullivan, T.L., Wagner, J.G. and Dressman, J.B. (1994) pH-related changes in the absorption of dipyridamole in the elderly. *Pharmaceutical Research*, **11** (1), 136.

47 Zhou, R., Moench, P., Heran, C., Lu, X., Mathias, N., Faria, T.N., Wall, D.A., Hussain, M.A., Smith, R.L. and Sun, D. (2005) pH-dependent dissolution *in vitro* and absorption *in vivo* of weakly basic drugs: development of a canine model. *Pharmaceutical Research*, **22** (2), 188.

48 Granero, G., Ramachandran, C. and Amidon, G.L. (2006) Rapid *in vivo* dissolution of ketoprofen: implications on the biopharmaceutics classification system. *Pharmazie*, **61** (8), 673–676.

49 Jinno, J., Oh, D.M., Crison, J.R. and Amidon, G.L. (2000) Dissolution of ionizable water-insoluble drugs: the combined effect of pH and surfactant. *Journal of Pharmaceutical Sciences*, **89** (2), 268–274.

50 Langguth, P., Lee, K., Spahn-Langguth, H. and Amidon, G.L. (1994) Variable gastric emptying and discontinuities in drug absorption profiles: dependence of rates and extent of cimetidine absorption on motility phase and pH. *Biopharmaceutics & Drug Disposition*, **15** (9), 719–746.

51 Yu, L.X. and Amidon, G.L. (1999) A compartmental absorption and transit model for estimating oral drug absorption. *International Journal of Pharmaceutics*, **186** (2), 119.

52 Yu, L.X., Crison, J.R. and Amidon, G.L. (1996) Compartmental transit and dispersion model analysis of small intestinal transit flow in humans. *International Journal of Pharmaceutics*, **140** (1), 111.

53 Dressman, J.B., Vertzoni, M., Goumas, K. and Reppas, C. (2007) Estimating drug solubility in the gastrointestinal tract. *Advanced Drug Delivery Reviews*, **59** (7), 591.

54 Lobenberg, R., Kramer, J., Shah, V.P., Amidon, G.L. and Dressman, J.B. (2000) Dissolution testing as a prognostic tool for oral drug absorption: dissolution behavior of glibenclamide. *Pharmaceutical Research*, **17** (4), 439.

55 Sinko, C.M., Yee, A.F. and Amidon, G.L. (1990) The effect of physical aging on the dissolution rate of anionic polyelectrolytes. *Pharmaceutical Research*, **7** (6), 648.

56 Aungst, B.J., Nguyen, N.H., Taylor, N.J. and Bindra, D.S. (2002) Formulation and food effects on the oral absorption of a poorly water soluble, highly permeable antiretroviral agent. *Journal of Pharmaceutical Sciences*, **91** (6), 1390–1395.

57 Bonlokke, L., Hovgaard, L., Kristensen, H.G., Knutson, L. and Lennernäs, H. (2001) Direct estimation of the *in vivo* dissolution of spironolactone, in two particle size ranges, using the single-pass perfusion technique (Loc-I-Gut(R)) in humans. *European Journal of Pharmaceutical Sciences*, **12** (3), 239.

58 Kostewicz, E.S., Brauns, U., Becker, R. and Dressman, J.B. (2002) Forecasting the oral absorption behavior of poorly soluble weak bases using solubility and dissolution studies in biorelevant media. *Pharmaceutical Research*, **19** (3), 345.

59 Takano, R., Sugano, K., Higashida, A., Hayashi, Y., Machida, M., Aso, Y. and

Yamashita, S. (2006) Oral absorption of poorly water-soluble drugs: computer simulation of fraction absorbed in humans from a miniscale dissolution test. *Pharmaceutical Research*, **23** (6), 1144.

60 Custodio, J.M., Wu, C.Y. and Benet, L.Z. (2008) Predicting drug disposition, absorption/elimination/transporter interplay and the role of food on drug absorption. *Advanced Drug Delivery Reviews*, **60** (6), 717.

61 Dressman, J.B. and Reppas, C. (2000) *In vitro–in vivo* correlations for lipophilic, poorly water-soluble drugs. *European Journal of Pharmaceutical Sciences*, **11** (Suppl. 2), S73.

62 Fordtran, J. and Locklear, T. (1966) Ionic constituents and osmolality of gastric and small-intestinal fluids after eating. *The American Journal of Digestive Diseases*, **11** (7), 503–521.

63 Macheras, P., Koupparis, M. and Tsaprounis, C. (1986) Drug dissolution studies in milk using the automated flow injection serial dynamic dialysis technique. *International Journal of Pharmaceutics*, **33** (1–3), 125.

64 Galia, E., Horton, J. and Dressman, J.B. (1999) Albendazole generics – a comparative *in vitro* study. *Pharmaceutical Research*, **16** (12), 1871.

65 Nicolaides, E., Symillides, M., Dressman, J.B. and Reppas, C. (2001) Biorelevant dissolution testing to predict the plasma profile of lipophilic drugs after oral administration. *Pharmaceutical Research*, **18** (3), 380.

66 Vertzoni, M., Dressman, J., Butler, J., Hempenstall, J. and Reppas, C. (2005) Simulation of fasting gastric conditions and its importance for the *in vivo* dissolution of lipophilic compounds. *European Journal of Pharmaceutics and Biopharmaceutics*, **60** (3), 413.

67 Scholz, A., Abrahamsson, B., Diebold, S.M., Kostewicz, E., Polentarutti, B.I., Ungell, A.-L. and Dressman, J.B. (2002) Influence of hydrodynamics and particle size on the absorption of felodipine in labradors. *Pharmaceutical Research*, **19** (1), 42.

68 Scholz, A., Kostewicz, E., Abrahamsson, B. and Dressman, J.B. (2003) Can the USP paddle method be used to represent *in-vivo* hydrodynamics? *The Journal of Pharmacy and Pharmacology*, **55**, 443.

69 Sheng, J.J., Sirois, P.J., Dressman, J.B. and Amidon, G.L. (2008) Particle diffusional layer thickness in a USP dissolution apparatus II: a combined function of particle size and paddle speed. *Journal of Pharmaceutical Sciences*, in press.

70 Charman, W.N. (2000) Lipids, lipophilic drugs, and oral drug delivery – some emerging concepts. *Journal of Pharmaceutical Sciences*, **89** (8), 967–978.

71 Gershanik, T. and Benita, S. (2000) Self-dispersing lipid formulations for improving oral absorption of lipophilic drugs. *European Journal of Pharmaceutics and Biopharmaceutics*, **50** (1), 179–188.

72 Fleisher, D., Li, C., Zhou, Y., Pao, L.H. and Karim, A. (1999) Drug, meal and formulation interactions influencing drug absorption after oral administration. Clinical implications. *Clinical Pharmacokinetics*, **36** (3), 233–254.

73 Dintaman, J.M. and Silverman, J.A. (1999) Inhibition of P-glycoprotein by D-α-tocopheryl polyethylene glycol 1000 succinate (TPGS). *Pharmaceutical Research*, **16** (10), 1550.

74 Nerurkar, M.M., Burton, P.S. and Borchardt, R.T. (1996) The use of surfactants to enhance the permeability of peptides through Caco-2 cells by inhibition of an apically polarized efflux system. *Pharmaceutical Research*, **13** (4), 528.

75 Dahan, A. and Hoffman, A. (2005) Evaluation of a chylomicron flow blocking approach to investigate the intestinal lymphatic transport of lipophilic drugs. *European Journal of Pharmaceutical Sciences*, **24** (4), 381–388.

76 Gershkovich, P. and Hoffman, A. (2007) Effect of a high-fat meal on absorption and disposition of lipophilic compounds: the importance of degree of association with

triglyceride-rich lipoproteins. *European Journal of Pharmaceutical Sciences*, **32** (1), 24.

77 Tso, P. (1985) Gastrointestinal digestion and absorption of lipid. *Advances in Lipid Research*, **21**, 143–186.

78 Dahan, A. and Hoffman, A. (2008) Rationalizing the selection of oral lipid based drug delivery systems by an *in vitro* dynamic lipolysis model for improved oral bioavailability of poorly water soluble drugs. *Journal of Controlled Release*, **129** (1), 1–10.

79 Porter, C.J.H. and Charman, W.N. (2001) *In vitro* assessment of oral lipid based formulations. *Advanced Drug Delivery Reviews*, **50** (Suppl. 1), S127.

80 Sek, L., Porter, C.J.H., Kaukonen, A.M. and Charman, W.N. (2002) Evaluation of the *in-vitro* digestion profiles of long and medium chain glycerides and the phase behaviour of their lipolytic products. *The Journal of Pharmacy and Pharmacology*, **54**, 29.

81 Zangenberg, N.H., Mullertz, A., Gjelstrup Kristensen, H. and Hovgaard, L. (2001) A dynamic *in vitro* lipolysis model. II. Evaluation of the model. *European Journal of Pharmaceutical Sciences*, **14** (3), 237.

82 Porter, C.J.H., Kaukonen, A.M., Taillardat-Bertschinger, A., Boyd, B.J., O'Connor, J.M., Edwards, G.A. and Charman, W.N. (2004) Use of *in vitro* lipid digestion data to explain the *in vivo* performance of triglyceride-based oral lipid formulations of poorly water-soluble drugs: studies with halofantrine. *Journal of Pharmaceutical Sciences*, **93** (5), 1110–1121.

83 Dahan, A. and Hoffman, A. (2007) The effect of different lipid based formulations on the oral absorption of lipophilic drugs: the ability of *in vitro* lipolysis and consecutive *ex vivo* intestinal permeability data to predict *in vivo* bioavailability in rats. *European Journal of Pharmaceutics and Biopharmaceutics*, **67** (1), 96.

84 Dahan, A. and Hoffman, A. (2006) Use of a dynamic *in vitro* lipolysis model to rationalize oral formulation development for poor water soluble drugs: correlation with *in vivo* data and the relationship to intra-enterocyte processes in rats. *Pharmaceutical Research*, **23** (9), 2165–2174.

85 Lennernäs, H. and Abrahamsson, B. (2005) The use of biopharmaceutic classification of drugs in drug discovery and development: current status and future extension. *The Journal of Pharmacy and Pharmacology*, **57**, 273.

86 Polli, J.E., Yu, L.X., Cook, J.A., Amidon, G.L., Borchardt, R.T., Burnside, B.A., Burton, P.S., Chen, M.-L., Conner, D.P., Faustino, P.J., Hawi, A.A., Hussain, A.S., Joshi, H.N., Kwei, G., Lee, V.H.L., Lesko, L.J., Lipper, R.A., Loper, A.E., Nerurkar, S.G., Polli, J.W., Sanvordeker, D.R., Taneja, R., Uppoor, R.S., Vattikonda, C.S., Wilding, I. and Zhang, G. (2004) Summary workshop report: biopharmaceutics classification system – implementation challenges and extension opportunities. *Journal of Pharmaceutical Sciences*, **936** 1375–1381.

87 Rinaki, E., Dokoumetzidis, A., Valsami, G. and Macheras, P. (2004) Identification of biowaivers among class II drugs: theoretical justification and practical examples. *Pharmaceutical Research*, **21** (9), 1567.

88 Yu, L.X., Amidon, G.L., Polli, J.E., Zhao, H., Mehta, M.U., Conner, D.P., Shah, V.P., Lesko, L.J., Chen, M.L., Lee, V.H.L. and Hussain, A.S. (2002) Biopharmaceutics classification system: the scientific basis for biowaiver extensions. *Pharmaceutical Research*, **19** (7), 921.

89 Youngberg, C., Berardi, R., Howatt, W., Hyneck, M., Amidon, G., Meyer, J. and Dressman, J. (1987) Comparison of gastrointestinal pH in cystic fibrosis and healthy subjects. *Digestive Diseases and Sciences*, **32** (5), 472–480.

90 Kaus, L.C., Gillespie, W.R., Hussain, A.S. and Amidon, G.L. (1999) The effect of *in vivo* dissolution, gastric emptying rate, and intestinal transit time on the peak concentration and area-under-the-curve of drugs with different gastrointestinal

permeabilities. *Pharmaceutical Research*, **16** (2), 272.

91 Dressman, J. (1997) *Scientific Foundations for Regulating Drug Product Quality* (eds G.L. Amidon, J. Robinson and R. Williams), AAPS Press, pp. 155–168.

92 Yazdanian, M., Briggs, K., Jankovsky, C. and Hawi, A. (2004) The "high solubility" definition of the current FDA guidance on biopharmaceutical classification system may be too strict for acidic drugs. *Pharmaceutical Research*, **21** (2), 293.

93 Lipka, E. and Amidon, G.L. (1999) Setting bioequivalence requirements for drug development based on preclinical data: optimizing oral drug delivery systems. *Journal of Controlled Release*, **62** (1–2), 41.

4
In Silico Prediction of Solubility

Andrew M. Davis and Pierre Bruneau

Abbreviations

ASNN	Associative neutral network
BRNN	Bayesian regularised neural networks
DMSO	Dimethyl sulfoxide
GI	Gastrointestinal
GPs	Gaussian processes
HTS	High-throughput screening
KNNs	*k*-nearest neighbors
LFERs	Linear free-energy relationships
log $D_{7.4}$	logarithm of the distribution coefficient between *n*-octanol and water at pH 7.4
log *P*	logarithm of the partition coefficient of the neutral form of the compound between *n*-octanal and water
LSER	Linear solvation energy relationship
MAD	Maximum absorbable dose
MLR	Multiple linear regression
n	Number of compounds within the training/test set
PCB	Polychlorinated biphenyl
pH	Acidity of solution measured represented as $-\text{logarithm}_{10}$ of the activity of hydronium ions in solution
PLS	Partial least squares
PSA	Polar surface area (dynamic or static – calculated from 2D or 3D structure as defined)
QSAR	Quantitative structure–activity relationship
QSPR	Quantitative structure–property relationship
RMSE	Root mean square error of predictions – average deviation of the predictions from the measured value for a test set
s	Standard error
SD	Sample standard deviation
SVMs	Support vector machines

Symbols

2D	Two-dimensional
3D	Three-dimensional
log P	$\log_{10}$(*n*-octanol–water partition coefficient)
log Sol	$\log_{10}$(solubility measured according to the defined protocol) – usually, on molar scale
MP	Melting point
pK_a	Ionization constant
r	Correlation coefficient
R^2	Coefficient of variation of the fitted multivariate model

4.1 Introduction

Although the pharmaceutical industry has over 100 years of experience in drug discovery, attrition through clinical development remains refractory to our efforts to use experience and modern tools to successfully design drugs. Compound toxicology remains a major cause for uncontrolled attrition. While we seek a detailed understanding of the linkage between molecular mechanisms, *in vivo* toxicity, and adverse drug reactions in humans, increasing focus is now being placed upon exercising control of the simplest of molecular properties, molecular weight, lipophilicity, and drug solubility [1]. Ideally, we would wish to exercise control through the accurate prediction of these properties at the point of design of chemical targets, prior to synthesis; hence, there is an increasing focus upon the development of accurate computational models of these properties. In this chapter, we will focus upon the prediction of one of the most important properties, solubility.

Solubility is widely regarded as one of the most difficult physical properties to predict. But when building predictive solubility models, or in fact any model, one needs to answer a number of questions: what solubility measures are required to be modeled? Do we have a suitable data set on which to build a computational model? What descriptors and what modeling methods should be used? How accurate are the models required to be? What is the influence of the domain of applicability? Do we know when good predictions have been made? In this chapter, we will highlight recent research, in an attempt to answer these key questions in solubility modeling.

4.2 What Solubility Measures to Model?

Solubility is a fundamental compound quality indicator and plays a critical role in many aspects of drug research. Although most research has focused upon modeling solubility in water or aqueous buffered solutions, solubility in other milieu may be equally important.

Oral solid dosage forms need to dissolve with high enough solubility in gastrointestinal (GI) fluids and rapidly enough relative to GI transit time to provide sufficient systemic exposure. The amount of material ultimately absorbed after an oral dose will depend, among other factors, upon the dose, the solid-state form, the rate and extent of dissolution in the gastrointestinal milieu, and the GI transit time, as well as the dissolved drug's inherent permeability.

A number of mathematical models have been developed to describe the interplay of solubility and these physiological parameters to model drug absorption. The most simplistic model is the maximal absorbable dose (MAD) calculation. The MAD calculation combines the amount of drug that can dissolve to form a saturated solution in water equal in volume to the small intestinal volume, with an estimate of the absorption rate and the small intestinal transit time. The maximal absorbable dose is then related to the dose required to achieve the desired therapeutic effect [2]. If the estimated MAD is much greater than the predicted dose to achieve a therapeutic effect, this can give confidence enough to take the drug toward clinical use. Predictions of aqueous solubility may then be useful in predicting the extent of absorption in man.

The MAD calculation penalizes low-solubility compounds, as the predicted maximum absorbable dose, limited by solubility, may be less than the predicted human dose. Low-soluble compounds may result in an acceptable prediction for poorly soluble drugs if the predicted dose is also low. A recent validation study has reported that the MAD calculation underestimates the known human dose range for low aqueous solubility clinical drugs. This is because poorly aqueous soluble drugs often show enhanced solubility in gastric fluids. Measurement of solubility in either aspirated human intestinal fluids or simulated intestinal fluids is often more relevant to poorly soluble lipophilic drugs, as the bile salts, for example, have a significant enhancing effect on solubility [3]. Food may also have a significant solubility-enhancing effect by altering gastric emptying time and affecting the solubility of drugs. Computational models predicting solubility in intestinal fluids may be a promising future area for research.

More sophisticated approaches to predict the extent of oral absorption of drugs use mathematical models based on the known physiology and utilizing simple physicochemical measurements as input. The GastroPlus [4] program is a commercial tool that utilizes an advanced compartmental and transit model, based on the work of Amidon and Yu [5], and allows what–if questions to be posed to enable pharmaceutical optimization (see Chapter 17). For instance, the impact of morphology, formulation, and/or particle size changes, and sensitivity analysis to include errors in parameters on the prediction, can be considered.

Once absorbed, the drug needs to stay in solution as it equilibrates with all the body compartments. For extensively renally cleared drugs, precipitation or crystallization in the kidneys is a particular concern, as it leads to crystalluria [6, 7]. Changes in pH and salt concentrations in the kidneys, the therapeutic dose required, and the rate of renal clearance will affect the risk of crystalluria. Computational models predicting solubility at differing pH values may be useful in the context of renally cleared drugs.

Controlled dissolution from a formulation may be critical for the control of duration of drug action. Extended release formulations are useful for most drug delivery routes.

For instance, fluticasone propionate is dosed topically to the lung via metered dose inhalers or dry-powder inhalers, and its high lipophilicity and low aqueous solubility are important for the drug's pharmacokinetic and pharmacodynamic profile [8, 9]. Models predicting solubility in specific formulations, or models predicting dissolution rate, may be useful in formulation development and drug delivery.

Solubility in nonaqueous solvents is also important in drug discovery. For instance, solubility in DMSO underpins most *in vitro* screening, and HTS depends upon the long-term stability of solutions of compounds in DMSO in screening collections. Although solubility is seldom a limiting factor in chemical synthesis in the drug discovery phase, as drug development progresses, solubility can be a limiting factor in large-scale synthetic processes. Tetko has published one of the first studies describing DMSO solubility models [10]. Published data sets with 20–70 000 cases exist, but they are only categorical in nature. Using common modeling methods such as random forests, recursive partitioning, and linear discriminant analysis produced classification models with 70–80% successful classification of random test sets. Computational models for predicting solubility in DMSO and other nonaqueous solvents are an area for further research throughout the discovery process.

It is apparent from the preceding discussion that there are many choices over solubility measures on which to base a computational model, and data sets and models are sparse beyond solubility in water or aqueous buffered solutions.

4.3 Is the Data Set Suitable for Modeling?

Solubility is a complex property, and this complexity confounds our ability to develop computational models to predict it. Most computational solubility models are empirical QSPR models, trained on solubility data sets either sourced from the literature and corporate databases or generated specifically for the purposes of modeling. Hence, it is not surprising that the quality of the computational model depends on the quality of the data set of experimental measurements used to train the model.

The solubility measure describes the concentration reached in solution, when a pure phase of the material is allowed to dissolve in the solvent for a defined period of time, at a defined temperature (and pressure). Most often for pharmaceutical purposes, the pure phase is a solid, ideally a crystalline solid, and the liquid is water or a buffered aqueous solution, at a controlled temperature (often 25 or 37 °C) and ambient pressure. The purity of the solid can have a large effect on measured solubility. Solubility can be measured in water or in pH-controlled buffers. In water, the extent of solubility for ionizable compounds will depend upon the pK_a values and the nature of the counterion. In pH-controlled aqueous buffered solution, at equilibrium, the solubility will depend upon the compound's intrinsic solubility, its pK_a, and the ionic strength. It may also depend upon the relative solubility of the initial added compound and the solubility of the salt formed by the compound with the buffer salts, with which the solid may equilibrate. In any buffer or solvent system, the measured solubility may depend on the time of sampling, as solubility kinetics

may be important. Sampling at different time points may prove useful not only to define the thermodynamic equilibrium solubility but also sometimes to describe the kinetics of dissolution. The "undissolved" solid may also change its morphology during the experiment. As one solid-state form dissolves and another crystallizes, the solubility measured at the end of an experiment may therefore be the solubility of a form different from the one the experiment began with.

In generating data sets of solubility measurements for modeling purposes, the degree of control of the parameters discussed above will undoubtedly contribute to, or compromise, the accuracy and precision of any predictive model built upon them. Taskinen and Norinder in their recent comprehensive survey [11] of data sources for solubility models commented that inadequate documentation made it difficult to assess whether many data sets represent suitably consistent values regarding thermodynamics or ionization for modeling purposes. For these reasons, an ideal data set for *in silico* modeling would be one from a single laboratory, measured under a consistent and controlled experimental protocol with an understood precision. Ideally, the conditions will be relevant to the end point being optimized and the solid-state form of the compounds being studied should be known before and after the experiment. Most data sets used as the basis for solubility models do not conform to this idealized description. This may contribute to the perceived lack of accuracy and precision of solubility models in general.

Bergström [12] made an interesting discussion about the advantages of the various methods of measuring solubility and permeability. Bergström showed that some methods, such as kinetic measurement of solubility by precipitation of a DMSO solution, by adding increasing amounts of aqueous buffer, can only lead to a qualitative classification in terms of low, intermediate, and high solubility. As a consequence, these results can only be used in classification predictions. On the contrary, if quantitative models are looked for, more labor-intensive and maybe less-automated methods are needed to collect more precise data of solubility and permeability. Bergström emphasizes the need of a database of good, accurate, and reproducible measurements to produce good predictive models. Even if we have a good database, descriptors, and mathematical tools to make a good model, there will remain the problem of applicability of the model to various chemical spaces.

The quality of measured data is often a problem in the research environment where the various batches of a compound show different measured solubility. In an in-house study at AstraZeneca, 75% of repeated measures on the same batch were found to have a standard deviation (SD) of less than 0.29 log unit, whereas the corresponding figure for the interbatch average measurements reached 0.49 log unit and 10% of these interbatch measurements had a variability of more than 0.81 log unit. This discrepancy was tentatively explained by the differences in solid state of the samples issued from different batches. Although an attractive hypothesis, and supported by the data, this suggested explanation is not supported by Delaney [13], who holds that the difference of physical state between compounds is not important to the accuracy of the solubility prediction, and by the recent study of Bergström and coworkers [14], who show that the solubility of poorly soluble compounds is limited more by their weak solvation ability than by their solid state.

4.4
Descriptors and Modeling Methods for Developing Solubility Models

The choice of descriptors is not always clear-cut. The time required to calculate elaborate descriptors by quantum methods is not always justified compared to the results obtained with simpler and more rapidly calculated descriptors. For example, Bergström [12] compared 2D polar surface area (PSA) with 3D PSA and static, instead of dynamic, calculations. No definitive gain was obtained by using the most sophisticated method(s) of calculating PSA descriptors.

There is no such clear-cut judgment about the statistical methods of modeling solubility. There are models as simple as the relationship between log *P* and melting point (MP), established some time ago by Yalkowsky and coworkers, and the very complex linear solvation energy relationships (LSERs). The limitation of the simple Yalkowsky relationship is that it uses two variables, obtained with accuracy only by measurement, and thus the simple relationship turns out to be very complicated when calculated log *P* and MP are used.

Among the recent reviews, the most comprehensive one is by Dearden [15]. In his review, he discusses the fundamentals of aqueous solubility, which lead to the Yalkowsky equation. Dearden also reviews the oldest approaches of predicting aqueous solubility, from a very simple fragment-based approach of 1924 to the numerous approaches post-1990, for which he made an extensive tabular comparison.

Delaney [13] describes the solubilization mechanism as controlled by a double phenomenon: the affinity of the compound for itself and the affinity of the compound for the solvent. The latter effect is simply described either by the log *P* property or by very sophisticated methods such as statistical thermodynamic or quantum mechanical techniques. These very intensive calculation methods have not yet proved their superiority over the simpler and faster methods that tend to mimic the successful log *P* fragment calculator.

The effects of solid-state structure on solubility are even more complicated and so far less successful to calculate, if we exclude some very crude methods of entropy of melting estimation by evaluating the number of rotatable bonds and the symmetry of the molecule. But Delaney estimates that the error on calculation of the compound–compound interaction is small (about half a log unit) compared to the potential error due to the compound–solvent interaction that can be estimated as 2 log units. It is even possible to neglect the variation of the melting points of compounds by using an average value of 125 °C for all compounds without influencing the accuracy of the prediction. This latter argument justifies the use of empirical approaches that neglect the fundamental mechanisms of solubilization but try to correlate the measured values with various calculated descriptors more or less related to the solubility-like parameters accounting for hydrogen bonding or solvent cavity formation. This is reassuring, as the use of solid-state descriptors would first require the *ab initio* prediction of the solid-state structure of compounds, which even for simple compounds is still in its infancy. In this part of his review, Delaney emphasizes the fact that linear methods are suitable for restricted homogeneous series, whereas large

data sets of diverse chemistry should be treated by nonlinear methods. Delaney surveys the various mathematical tools that have been successfully used in the data regression without pointing out a simple method, or a group of methods, which may be better than others. Delaney, similar to many other authors, concludes that the quality of the data is of primary importance and that there are some problems not yet fully solved, such the prediction of charged compounds at various pH values and DMSO solubilities. The influence of the crystal stability on solubility and the success of the Yalkowsky equation have initiated many attempts at predicting the melting point.

A recent publication relates the work of Varnek *et al*. [16] on this subject. The authors have gathered a database of 717 compounds with measured melting points. These compounds were all bromides of nitrogen-containing organic cations. They used associative neural network (ASNN), *k*-nearest neighbors (KNNs), support vector machines (SVMs), and multiple linear regression (MLR). Their validation method was a leave-one-fifth-out method, that is, they built five models on four-fifths of the data using one-fifth of the data reserved for the validation set. The five subvalidation sets were combined to form the internal validation set. Similar to all leave-*n*-out validation methods, these internal validation sets will contain close analogs of the training sets, but it does not seem that there is another good method to validate such models. The results show that there is a slight advantage of using the nonlinear method and that an RMSE of around 40 °C can be achieved.

Nigsch *et al*. [17, 18] made a similar attempt; by using a *k*-nearest neighbors algorithm with a genetic parameter optimization on a training set of 4110 diverse organic molecules and 277 drugs, they obtained an RMSE of 42.2 °C. They explained that the remaining error is due to "the lack of information about the interactions in the liquid state."

4.5 Comparing Literature Solubility Models

The reviews published by Bergström [12] and Delaney [13] in 2005 and Dearden [15] in 2006 give good overviews of the predictive solubility literature.

But the main problem with the multitude of solubility models is how to compare them? The only way to obtain an objective judgment of their predictive power is to test each model using a common external test set. For this purpose, Dearden made a compilation of results from 21 models, whose authors published results of a common test set comprising 20 drugs and pesticides and one PCB, as it was initially used by Yalkowsky and Banerjee to evaluate their own model. A summary of the most homogeneous results from these models is reproduced in Table 4.1.

This comparison is very useful, but one must keep in mind that 20 molecules in a test set does not represent a large chemical space, and although it constitutes a comparative test, it does not prove that the models are predictive for *any* drug-like molecule, as we shall discuss later. Dearden widens his comparative study to 17 commercially available programs able to predict solubility. The comparison was done

Table 4.1 Aqueous solubility prediction studies using the Yalkowsky and Banerjee test set [15].

Author	Year of publication	Compounds in training set	Compounds in test set	Std of training set	Std of test set
Yalkowsky	1992	41	21	0.50	0.79
Klopman	1992	483	21	0.53	1.25
Yalkowsky	1993	41	19	0.39	1.34
Kuhner	1995	694	21	0.38	1.05
Huuskonen	1998	160	21	0.46	1.25
Huuskonen	2000	884	21	0.47	0.63
Huuskonen	2000	675	21	0.52	0.75
Huuskonen	2001	674	21	0.58	0.84
Livingstone	2001	552	21	0.52	0.77
Liu	2001	1033	21	0.70	0.93
Yan	2003	797	21	0.50	0.77
Wegner	2003	1016	21	0.52	0.79
Butina	2003	2688	11	0.61	0.94
Yan	2003	741	21	0.51	0.80
Hou	2004	878	21	0.59	0.64
Delaney	2004	2874	21	0.97	0.78

with a common test set of 122 compounds with accurately measured intrinsic solubility in pure water. In judging the results, one must be careful of the fact that some or even most of the 122 test compounds may have been used on the training set of the software concerned. Keeping in mind this possibility, the standard error of the log Sol of commercial software has a range of 0.47–1.96 with a median at 0.95 log unit. Dearden points out that the best scores, around 0.50, reach the limits of the experimental errors of solubility measurements [18, 19]. A good method to compare the commercially available software is to evaluate the percentage of compounds that are well predicted with an error less than 0.5 log unit and the percentage of compounds that have an acceptable error less than 1 log unit, as shown in Table 4.2.

Schwaighofer *et al*. [19] have done a remarkable work by gathering solubility data on 23 516 compounds from literature sources that after data cleaning led to reliable solubility measurements on 3307 neutral compounds combined with data from about 1100 compounds, mainly electrolytes, which have been measured in-house. They also used data on 704 compounds, which were used by Huuskonen and numerous other researchers. The complete database was used either to predict pure solubility of neutral compounds or to predict solubility at a given pH for neutral compounds and electrolytes. Apart from using a comprehensive database, the main originality of this work is the use of Gaussian process (GP) as the mathematical tool to build the models. GPs are techniques from the field of Gaussian statistics, and similar to the Bayesian regularized neural networks (BRNNs) [20], they have the great advantage to allow an evaluation of the prediction error for the individual predicted solubilities. Their model achieves a performance, as measured by the RMSE, comparable to commercial packages at around 0.6 log unit on neutral compounds,

Table 4.2 Predictive ability of some commercially available software for aqueous solubility, based on a 122-compound test set of drugs [15].

Software	Percentage of compounds predicted within	
	±0.5 log unit	±1.0 log unit
SimulationsPlus	65	91
Admensa	72	87
Pharma Algorithms ADME Boxes	59	87
ChemSilico	60	86
ACDLabs	59	85
ALogS	52	81
PredictionBase	47	81
ESOL	55	79
MOLPRO	62	78
Absolv 2	44	75
QikProp	48	79
SPARC	43	73
Cerius ADME	38	73
WSKOWWIN	41	67
ADMEWORKS Predictor	34	66
AlogP98	38	62
CHEMICALC	23	46

while it is much better on electrolytes with an RMSE of 0.77, compared to well above 1 log unit for commercial packages on the same validation set.

Du-Cuny *et al*. [21] have used a large data set of 2473 compounds likely to be more "drug-like" as they are from their own employer's pharmaceutical company collection. The originalities of their work are

- modeling of intrinsic solubility ($\log 1/S_0$);
- use of 170 structural fragment descriptors and four fragment-based correction factors;
- use of 81 congeneric series indices.

The obtained PLS model gave $R^2 = 0.81$ on an independent test set of 958 compounds. This measure of predictivity is optimistic, since the test set was selected in a manner favorable to good prediction. The initial data set of 2473 compounds was clustered and the singletons were eliminated. Then the remaining data set was randomized and the training set and test set were selected. Therefore, the members of the test set were very likely not too far from members of the training set, thus leading to better prediction statistics. The authors submitted their test set to prediction with a commercial software that led to a poor $R^2 = 0.10$. Although the figure of 0.81 obtained with their model is probably optimistic, the improvement over $R^2 = 0.10$ obtained with WSKOWWIN is large enough to conclude the superiority of the models developed using in-house data over the commercial models.

As Du-Cuny *et al.* [21] have described, it is tempting to derive the intrinsic solubility from a measurement of the solubility at a given pH and measured pK_a and using the Henderson–Hasselbalch equation. Hansen *et al.* [22] followed a reverse path. They built an intrinsic solubility model from intrinsic solubility data and, using a calculated pK_a from a commercial tool, predicted the pH-dependent solubility of a validation set of 27 drugs with experimentally determined pH-dependent solubility profile, to obtain an RMSE of 0.79 log unit. Bergström *et al.* [23] have shown that one must be very cautious in doing this kind of calculation. They measured a set of 25 monoamine compounds. The Henderson–Hasselbalch equation shows a slope of −1 for the relationships of solubility log unit with pH. The Bergström *et al.* measurements reveal that −1 is very rarely observed since the experimental slopes range from −0.5 to −8.6. This is explained by various experimental conditions leading to dimers, aggregates, and micelles and also to the presence of counterions with multi-ionizable sites. The consequences of these observations are that solubility measurements are extremely difficult to obtain with a reproducible and reliable quality.

In his review, Dearden [15] has tabulated the statistical characteristics of nearly 90 models published between 1990 and 2005 in a chronological order. If we exclude two studies showing very low standard error of estimate ($s = 0.08$ and 0.19), which according to Dearden are probably heavily overfitted models, and a model having a very high $s = 2.4$, the remaining reported data of s and RMSE have a mean value of 0.57 log unit, which fits quite well with the generally accepted figure of 0.6 log unit for the experimental error of the solubility measurements. But the data show a remarkable increase with time. Dearden points out that the first studies to use drugs in the training set did not appear until 1998; shortly after, Lipinski *et al.* published their study on the rule-of-5 [24]. If we separate the s and RMSE published before 1998 from the ones published after that year, we observe a clear increase from 0.41 log unit for the pre-1998 data to 0.62 for the post-1998 data. Even after 1998, the increasing trend is shown with an average of 0.49 in 2000 to an average of 0.67 for the models published in 2004. The difference between these two figures is statistically significant at the 0.05 level of a Student t-test. The results for the other years show the trend, although they do not show a statistical difference from each other (Table 4.3).

The reasons for the degradation of modeling performance are unclear. Dearden proposed that there are more and more drugs involved in the modeling process. It is probably true, but there is no fundamental difference between a chemical entity

Table 4.3 Variation of the mean RMSE of the models published between 2000 and 2004.

Year of publication	Mean RMSE	Number of publications
2000	0.49	7
2001	0.62	9
2002	0.60	11
2003	0.65	15
2004	0.67	11

being "drug" and another a "nondrug." Since Dearden's data consider statistics obtained on training sets, the results should at least be stable or should rather improve with the improvement of modeling processes and of the descriptors sets. Any difference explainable by the usual more complex and higher molecular weight of drug-like compounds should be seen only on validation results.

4.6 What Is the Influence of the Domain of Applicability?

As highlighted by Bergström [12], the problem of the applicability domain of a model has not been definitively solved. Stanforth *et al*. [25, 26] have recently addressed this problem by clustering the training set by a *k*-means clustering algorithm in the descriptors space. Then, the distance to domain was evaluated and correlated to the error of the model. Their method compares well with other methods of evaluating the distance to model to estimate the likely error of the prediction. In a similar manner, Bruneau and McElroy [20] set up a method that used the average of the Mahalanobis distances to the three nearest neighbors, calculated in the descriptors space to describe the distance to the models. These authors show a very good dependency of the RMSE of a $\log D_{7.4}$ model on the calculated distance from the model. They also show that the standard deviation of the prediction obtained by sampling a few hundreds of BRNNs is also a very good indication of the likely error of the model.

Bruneau has demonstrated the importance of domain of applicability of solubility models using three training set/test set pairs described by the same set of descriptors and using the same modeling method, Bayesian neural networks with automatic relevance determination variable selection [26]. He defined internal proprietary training and test sets and "public" training and test sets from the literature. A combined data set representing both "public" and "in-house" proprietary data was also constructed. The "public" compounds were on average smaller, less lipophilic, and more soluble than the proprietary drug-like compounds. He showed that similar results were obtained from the "public" model applied to a "public" validation set with an RMSE $= 0.84$ log unit and the so-called "in-house" model applied to "in-house" validation set with an RMSE $= 0.78$ log unit. This would be judged as two equivalent models with similar performance from a publication point of view. But when the validation sets were crossed over, the "public" model gave an RMSE $= 1.0$ on the "in-house" validation set, and the "in-house" model' gave an RMSE $= 1.88$ on the "public" validation set. It is clear that even models with apparently similar performances may give very different results when applied to different data sets. It must be noted that the "in-house" model had much more difficulties in predicting "public" data than the reverse. This was explained by the higher diversity of the "public" database compared to the "in-house" database, which is more series dependent. To solve this, Bruneau combined the training and test set databases and obtained a "mixed" model that when applied to the "public" validation set gave an RMSE $= 0.82$ and when applied to the "in-house" validation set gave an RMSE $= 0.79$. The

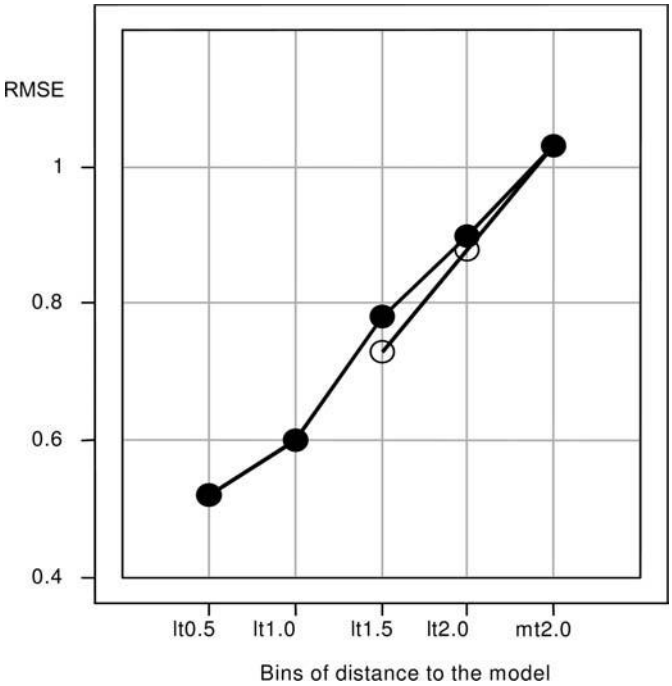


Figure 4.1 Distance-to-model versus RMSE. Solid circles represent "public model" applied to "public" validation set. Open circles represent "public" model applied to "in-house" validation set. The missing open circles correspond to the lack of examples in the corresponding bin. lt0.5, lt1.0, lt1.5, and lt2.0 stand for bins of the Mahalanobis distances of compounds to be tested to the nearest compounds in the training set of less than 0.5, 1.0, 1.5, and 2.0 distance arbitrary units, respectively. mt2.0 stands for bin of the Mahalanobis distances of compounds to be tested to the nearest compounds in the training set of more than 2.0 distance arbitrary units [26].

predictions from the literature model for each test set were then binned by the distance to the training space of the "public" model, and the test set's RMSE values were analyzed per distance bin. It was concluded that once the distance to the training set model space was considered, the literature model now performed similarly on all three test sets, when bins of a similar distance were compared as shown in Figure 4.1.

This analysis is very revealing. It suggests that comparing different computational models on common test sets may not be a very useful exercise. It does tell you which model predicts that particular test set well, but this gives little information on the relative generalizability of the models. The dependence of predictive performance on distance-to-model must be taken into account. As Bruneau states, "It is utopian to expect that a model can predict any structure." This view may undermine many reports in the literature, but the most important consideration is how well does a particular model predict "my test set of interest" or, more importantly, how well does the model predict the next compound I am going to make.

4.7
Can We Tell when Good Predictions Are Made?

Whatever the results of training sets, and test sets, the users ask what is the best model to use? The question is difficult, and even impossible, to answer from the published data on training and validation sets. This is because the performance of any model on any particular test set will heavily depend on the distance of the test set compounds to the training set model space, as well as the inherent predictive poser of the model itself, as has been described in the previous section. The only way to answer this question is for the users to try the models on their own data sets.

The modelers of physical properties face a difficult situation: they are satisfied when their models have an RMSE on validation set of less than the fatidic one order of magnitude, but at the same time the medicinal chemists are frustrated when they use the model in their own projects. The discrepancy of judgments is explained on pure statistical elements. An RMSE of 0.8 log unit on a global model of solubility validated on a validation set spreading from −2 to 8 log unit with an SD of say 5 log units leads to a very satisfactory r^2 of nearly 1, which classified the model as very good. When the medicinal chemist uses the model on his own data, the spread of which is often narrowed to maybe two orders of magnitude with an SD of say 1 log unit, he observes a deceptive r^2 of 0.36. The situation is usually even worse since the series on which the chemist is working might be poorly represented in the training set, and the RMSE of the model on this new series may well be higher than the SD of the measured data leading to an even more disappointing negative r^2. The usual answer to this problem is often a request for "local" models. This approach may sometime be successful, especially when linear regression methods are used, since these methods are able to pick up "local" specificities.

So why cannot we improve the models? Johnson [27] attempts to answer this disturbing question. Although his discussion is centered on QSAR, it has many common points relevant to QSPR, which is used in the field of solubility modeling. He concludes that we should not be led by mere statistics that too often dictates what is good or wrong, but rather by scientific rationale, experimental design, and personal observation.

4.8
Conclusions

The literature and commercial companies abound with computational solubility models. Many data sets have been studied, with many different descriptor sets, and using a multitude of statistical methods. It appears that diverse drug-like data sets are often predicted by our best methods with an RMSE of 0.8–1 log unit. This compares with an error in replicate measurements of approximately 0.5 log unit. A common view is that there is still room for improvement in the computational modeling of solubility. There are a number of suggestions that the quality control of the ideal data set is still lacking. This may be true for some literature data set compilations, but it is

less likely to be the case for proprietary data sets described by large pharmaceutical companies, where a large number of compounds have been measured through single well-controlled assays. There may be further work that could be done to describe the solute–solute interactions in the solid state, but this may be very challenging and may be a minor controlling factor compared to the importance of solvation, which is well described by current QSAR descriptors. Maybe an RMSE of 0.7–0.8, as observed on diverse test sets predicted by a number of models, is the best one should expect to be able to achieve. When distance to the training set model space of the test set is considered, compounds closest to the training sets may be predicted close to the replicate error in measurement. The problem then reduces to keeping the training set space of your computational model close to the compounds you wish to predict and building your computational model to predict to the most relevant end point.

One should always keep in mind that a correlation established on a training set between descriptors and experimental data may not be an indication of causation. Without the causation factor, that is, the descriptors explain the phenomenon that induces the variation in the data, it is unlikely that the model can predict any so far unknown molecular feature. Of course, this is the heart of molecular modeling and this is why the models are so dependent on the distance to the model space. All you can do to gain confidence in your model is to assess the predictivity of your model on your compounds. You can do this by testing the model on compounds synthesized and tested subsequent to the model derivation, with the test set similar to the compound(s) required to be predicted, and consider the effect of distance to model in the prediction errors. Once one has built confidence that the model has some predictive power on similar compounds of interest, a judgment has to be made whether the predictivity achieved has the resolution required for the purpose in hand. If it does have, then all that is left is to hold your breath and make a real prediction. As Niels Bohr said, "Predictions are difficult, especially those of the future." In the end, you have to take your chances that your model really does have enough control of solubility to guide to you to your goal, and this chance is often worth taking than leaving solubility to random chance.

References

1 Leeson, P.D. and Springthorpe, B. (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Reviews. Drug Discovery*, **6** (11), 881–890.

2 Johnson, K.C. and Swindell, A.C. (1996) Guidance in the setting of drug particle size specifications to minimize variability in absorption. *Pharmaceutical Research*, **13** (12), 1795–1798.

3 Dressman, J.B., Vertzoni, M., Goumas, K. and Reppas, C. (2007) Estimating drug solubility in the gastrointestinal tract. *Advanced Drug Delivery Reviews*, **59** (7), 591–602.

4 Agoram, B., Gilman, T., Woltosz, W. and Bolger, M. (2001) Application of the ACAT model for simulating the Cp-time profiles of controlled release dosage forms of metoprolol. Proceedings of the 28th

International Symposium on Controlled Release of Bioactive Materials and 4th Consumer & Diversified Products Conference, June 23–27, 2001, San Diego, CA, Vol. 1, pp. 840–841.

5 Yu, L.X., Crison, J.R. and Amidon, G.L. (1996) Compartmental transit and dispersion model analysis of small intestinal transit flow in humans. *International Journal of Pharmaceutics*, **140** (1), 111–118.

6 Stormann, H. and Nigitz, H.P. (1980) Experimental crystalluria induced in rats by sulfamides. *Investigacion Medica Internacional*, **7** (3), 245–250.

7 Merschman, S.A., Rose, M.J., Pearce, G.E.S., Woolf, E.J., Schaefer, B.H., Huber, A.C., Musson, D.G., Petty, K.J., Rush, D.J., Varsolona, R.J. and Matuszewski, B.K. (2005) Characterization of the solubility of a poorly soluble hydroxylated metabolite in human urine and its implications for potential renal toxicity. *Pharmazie*, **60** (5), 359–363.

8 Meibohm, B., Mollmann, H., Wagner, M., Hochhaus, G., Mollmann, A. and Derendorf, H. (1998) The clinical pharmacology of fluticasone propionate. *Reviews in Contemporary Pharmacotherapy*, **9** (8), 535–549.

9 Derendorf, H., Hochhaus, G., Meibohm, B., Mollmann, H. and Barth, J. (1998) Pharmacokinetics and pharmacodynamics of inhaled corticosteroids. *Journal of Allergy and Clinical Immunology*, **101** (4 Pt 2), S440–S446.

10 Balakin, K.V., Savchuk, N.P. and Tetko, I.V. (2006) *In silico* approaches to prediction of aqueous and DMSO solubility of drug-like compounds: trends, problems and solutions. *Current Medicinal Chemistry*, **13** (2), 223–241; Meltzer, P.S., Kallioniemi, A. and Trent, J.M. (2002) Chromosome alterations in human solid tumors, in *The Genetic Basis of Human Cancer* (eds B. Vogelstein and K.W. Kinzler), McGraw-Hill, New York, pp. 93–113.

11 Taskinen, J. and Norinder, U. (2007) *ADME-Tox Approaches*, Vol. 5 (eds B. Testa and H. van de Waterbeemd), in *Comprehensive Medicinal Chemistry II* (series eds D. Triggle and J. Taylor), Elsevier, Oxford, pp. 627–648.

12 Bergström, C.A.S. (2005) Computational models to predict aqueous drug solubility, permeability and intestinal absorption. *Expert Opinion on Drug Metabolism & Toxicology*, **1** (4), 613–627.

13 Delaney, J.S. (2005) Predicting aqueous solubility from structure. *Drug Discovery Today*, **10** (4), 289–295.

14 Bergström, C.A.S., Wassvik, C.M., Johansson, K. and Hubatsch, I. (2007) Poorly soluble marketed drugs display solvation limited solubility. *Journal of Medicinal Chemistry*, **50** (23), 5858–5862.

15 Dearden, J.C. (2006) *In silico* prediction of aqueous solubility. *Expert Opinion on Drug Discovery*, **1** (1), 31–52.

16 Varnek, A., Kireeva, N., Tetko, I.V, Baskin, I.I. and Solov'ev, V.P. (2007) Exhaustive QSPR studies of a large diverse set of ionic liquids: how accurately can we predict melting points? *Journal of Chemical Information and Modeling*, **47** (3), 1111–1122.

17 Nigsch, F., Bender, A., Van Buuren, B., Tissen, J., Nigsch, E. and Mitchell, J.B.O. (2006) Melting point prediction employing *k*-nearest neighbor algorithms and genetic parameter optimization. *Journal of Chemical Information and Modeling*, **46** (6), 2412–2422.

18 Katritzky, A.R., Wang, Y., Sild, S., Tamm, T. and Karelson, M. (1998) QSPR studies on vapor pressure, aqueous solubility, and the prediction of water–air partition coefficients. *Journal of Chemical Information and Computer Sciences*, **38** (4), 720–725.

19 Schwaighofer, A., Schroeter, T., Mika, S., Laub, J., Ter Laak, A., Suelzle, D., Ganzer, U., Heinrich, N. and Mueller, K.-R. (2007) Accurate solubility prediction with error bars for electrolytes: a machine learning approach. *Journal of Chemical Information and Modeling*, **47** (2), 407–424.

20 Bruneau, P. and McElroy, N.R. (2006) Log $D_{7.4}$ modeling using Bayesian regularized neural networks. Assessment and correction of the errors of prediction. *Journal of Chemical Information and Modeling*, **46** (3), 1379–1387.

21 Du-Cuny, L., Huwyler, J., Wiese, M. and Kansy, M. (2008) Computational aqueous solubility prediction for drug-like compounds in congeneric series. *European Journal of Medicinal Chemistry*, **43** (3), 501–512.

22 Hansen, N.T., Kouskoumvekaki, I., Jorgensen, F., Steen, B., Soren, J. and Svava, O. (2006) Prediction of pH-dependent aqueous solubility of druglike molecules. *Journal of Chemical Information and Modeling*, **46** (6), 2601–2609.

23 Bergström, C.A.S., Luthman, K. and Artursson, P. (2004) Accuracy of calculated pH-dependent aqueous drug solubility. *European Journal of Pharmaceutical Sciences*, **22** (5), 387–398.

24 Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, **23** (1–3), 3–25.

25 Stanforth, R.W., Kolossov, E. and Mirkim, B. (2007) A measure of domain of applicability for QSAR modelling based on intelligent *K*-means clustering. *QSAR & Combinatorial Science*, **26** (7), 837–844.

26 Bruneau, P. (2001) Search for predictive generic model of aqueous solubility using Bayesian neural nets. *Journal of Chemical Information and Computational Sciences*, **41**, 1605–1616.

27 Johnson, S.R. (2008) The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *Journal of Chemical Information and Modeling*, **48** (1), 25–26.