

Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Pharmacy 295



Computational and Experimental Models for the Prediction of Intestinal Drug Solubility and Absorption

BY

CHRISTEL A. S. BERGSTRÖM



ACTA UNIVERSITATIS UPSALIENSIS
UPPSALA 2003

Dissertation for the Degree of Doctor of Philosophy (Faculty of Pharmacy) in Pharmaceutics presented at Uppsala University in 2003

ABSTRACT

Bergström, C. A. S., 2003. Computational and Experimental Models for the Prediction of Intestinal Drug Solubility and Absorption. Acta Universitatis Upsaliensis. *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy* 295. 66 pp. Uppsala. ISBN 91-554-5747-9.

New effective experimental techniques in medicinal chemistry and pharmacology have resulted in a vast increase in the number of pharmacologically interesting compounds. However, the number of new drugs undergoing clinical trial has not augmented at the same pace, which in part has been attributed to poor absorption of the compounds.

The main objective of this thesis was to investigate whether computer-based models devised from calculated molecular descriptors can be used to predict aqueous drug solubility, an important property influencing the absorption process. For this purpose, both experimental and computational studies were performed. A new small-scale shake flask method for experimental solubility determination of crystalline compounds was devised. This method was used to experimentally determine solubility values used for the computational model development and to investigate the pH-dependent solubility of drugs. In the computer-based studies, rapidly calculated molecular descriptors were used to predict aqueous solubility and the melting point, a solid state characteristic of importance for the solubility. To predict the absorption process, drug permeability across the intestinal epithelium was also modeled.

The results show that high quality solubility data of crystalline compounds can be obtained by the small-scale shake flask method in a microtiter plate format. The experimentally determined pH-dependent solubility profiles deviated largely from the profiles predicted by a traditionally used relationship, highlighting the risk of data extrapolation. The *in silico* solubility models identified the non-polar surface area and partitioned total surface areas as potential new molecular descriptors for solubility. General solubility models of high accuracy were obtained when combining the surface area descriptors with descriptors for electron distribution, connectivity, flexibility and polarity. The used descriptors proved to be related to the solvation of the molecule rather than to solid state properties. The surface area descriptors were also valid for permeability predictions, and the use of the solubility and permeability models in concert resulted in an excellent theoretical absorption classification. To summarize, the experimental and computational models devised in this thesis are improved absorption screening tools applicable to the lead optimization in the drug discovery process.

Christel A. S. Bergström, Department of Pharmacy, Uppsala Biomedical Centre, Box 580, SE-751 23 Uppsala, Sweden

© Christel A. S. Bergström 2003

ISSN 0282-7484

ISBN 91-554-5747-9

Printed in Sweden by Universitetsstryckeriet, Uppsala 2003

To my family

The whole is simpler than the sum of its parts.

W. J. Gibbs

Gå upp och pröva dina vingar,
Och känn hur underbart det är
Där ovan molnen du dig svingar
Och fröjdas åt att vingarna bär
Se på fåglarna som svävar i det blå
Det är deras väg vi gå
Gå upp och pröva dina vingar
Och snart är hela jorden din!

L. Dahlquist

CONTENTS

1. PAPERS DISCUSSED	7
2. ABBREVIATIONS AND SYMBOLS	8
3. INTRODUCTION	9
3.1. The Drug Discovery Setting: From Lead Structure to Candidate Drug	9
3.2. Intestinal Drug Absorption	11
3.2.1. Mechanisms of Intestinal Solubility	11
3.2.2. Mechanisms of Intestinal Membrane Permeation	13
3.3. <i>In Vitro</i> Screening for Drug Absorption	14
3.3.1. Solubility Measurements	14
3.3.2. Permeability Measurements	15
3.4. <i>In Silico</i> Screening for Drug Absorption	16
3.4.1. Computational Calculation of Molecular Descriptors	17
3.4.2. Model Development	19
3.4.3. Solubility Models	21
3.4.4. Permeability Models	22
4. AIMS OF THE THESIS	24
5. METHODS	25
5.1. Investigated Drugs	25
5.2. Differential Scanning Calorimetry (DSC)	25
5.3. Solubility Determinations	25
5.4. Cell Culture	27
5.5. Transport Studies	27
5.6. Analytical Methods	28
5.7. Biopharmaceutical Classification	28
5.8. Molecular Descriptors	29
5.9. Statistics	30
5.9.1. Solubility and Permeability Experiments	30
5.9.2. Model Development	31
6. RESULTS AND DISCUSSION	32
6.1. Datasets	32
6.2. Solubility Measurements <i>In Vitro</i>	34
6.2.1. The Small-Scale Shake Flask Method (SSF)	34
6.2.2. Temperature and Buffer Effects	36

6.2.3. pH-dependent Solubility	38
6.3. Permeability Measurements <i>In Vitro</i>	39
6.4. Solubility Predictions <i>In Silico</i>	40
6.4.1. Global Solubility Models	40
6.4.2. Subset Specific Solubility Models	43
6.4.3. Computational Protocol for Rapid Calculation of Molecular Surface Areas	44
6.4.4. Molecular Surface Areas as Descriptors of the Solvation Process	45
6.5. Computational Solid State Characterization	46
6.6. Permeability Prediction <i>In Silico</i>	47
6.7. Biopharmaceutical Classification	48
7. CONCLUSIONS	51
8. PERSPECTIVES	53
9. POPULÄRVETENSKAPLIG SAMMANFATTNING	54
10. ACKNOWLEDGEMENTS	56
11. REFERENCES	58

1. PAPERS DISCUSSED

This thesis is based on the following papers, which will be referred to by the Roman numerals assigned below:

- I. Bergström C.A.S., Norinder U., Luthman K. and Artursson P.
Experimental and computational screening models for prediction of aqueous drug solubility.
Pharmaceutical Research, 19:2, 182-188, 2002.
- II. Bergström C.A.S., Strafford M., Lazorova L., Avdeef A., Luthman K. and Artursson P.
Absorption classification of oral drugs based on molecular surface properties.
Journal of Medicinal Chemistry, 46:4, 558-570, 2003.
- III. Bergström C.A.S., Wassvik, C.M., Norinder U., Luthman K, Artursson P.
Global and local computational models for aqueous solubility prediction of drug-like molecules. Submitted.
- IV. Bergström C.A.S., Norinder U., Luthman K. and Artursson P.
Molecular descriptors influencing melting point and their role in classification of solid drugs.
Journal of Chemical Information and Computer Sciences, 43, 1177-1185, 2003.
- V. Bergström C.A.S., Luthman K and Artursson P.
Accuracy of calculated pH-dependent aqueous drug solubility. Submitted.

Reprints were made with the permission of the journals.

2. ABBREVIATIONS AND SYMBOLS

ADMET	absorption, distribution, metabolism, elimination/excretion, toxicity
BCS	biopharmaceutics classification system
Caco-2	adenocarcinoma cell line derived from human colon
CC	combinatorial chemistry
ClogP	calculated partition coefficient between octanol and water
CD	candidate drug
DMSO	dimethylsulphoxide
DSC	differential scanning calorimetry
FA	fraction of the dose absorbed
FDA	Food and Drug Administration
GI	gastrointestinal tract
HH	Henderson-Hasselbalch equation
HTS	high throughput screening
$\log P_{\text{oct}}$	partition coefficient between octanol and water
LSER	linear solvation energy relationship
MLR	multiple linear regression
MTS	medium throughput screening
NN	neural networks
NPSA	non-polar surface area
P_{app}	apparent permeability coefficient
PCA	principal component analysis
PK/PD	pharmacokinetics/pharmacodynamics
PLS	partial least square projection to latent structures
PSA	polar surface area
PTSA	partitioned total surface area
Q^2	cross-validated coefficient of determination
QSAR	quantitative structure-activity relationship
QSPR	quantitative structure-property relationship
R^2	coefficient of determination
RMSE_{tr}	root-mean square error of the training set
RMSE_{te}	root-mean square error of the test set
S_0	intrinsic solubility
SA	surface area
SSF	small-scale shake flask

3. INTRODUCTION

Throughout the last decade, new rapid experimental techniques have resulted in the production of a large number of pharmacologically interesting compounds. The tremendous amount of data generated makes rationalization and prioritization more important in order to identify compounds with favorable developability characteristics. In this thesis, models for the prediction of intestinal drug absorption, one of the major factors influencing drug developability, will be discussed. The work has been focused on the development of computational (*in silico*) and experimental (*in vitro*) models for the prediction of aqueous drug solubility, which is considered to be one of the rate limiting steps to absorption of orally administered drugs.

3.1. The Drug Discovery Setting; From Lead Structure to Candidate Drug

The ability to identify and validate target proteins for drug treatment has recently been improved by the use of genomics, proteomics and bioinformatics.¹⁻³ When the target has been identified, the search for a lead structure starts, i.e., for a compound that binds to the target and exerts an acceptable therapeutic effect. After finding such a structure, the lead optimization process is initiated. Combinatorial chemistry (CC) and high throughput screening (HTS) are used to synthesize and test new compounds and to optimize them with regard to increased potency.⁴⁻⁹ The lead optimization is performed in cycles and, in the end, the resulting compounds with the highest potency might be structurally quite different from the starting structure. Until recently, the lead optimization and the screening for developability were performed in serial (Figure 1).

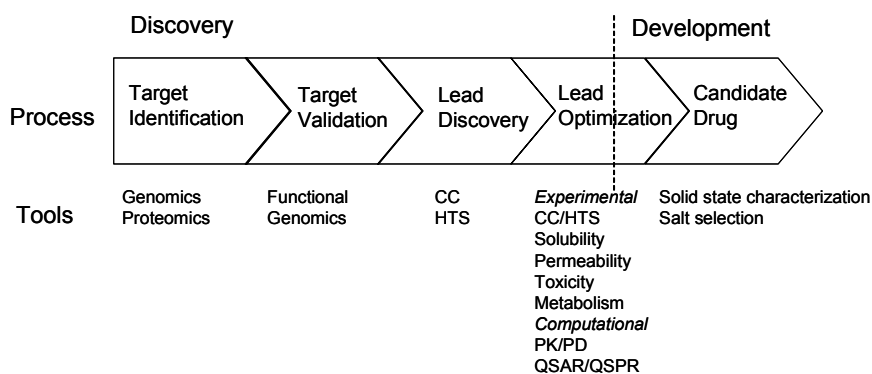


Figure 1. From lead structure to candidate drug. Drug discovery and development have traditionally been performed in serial rather than in parallel, resulting in the evaluation of developability late in the drug discovery process.

Hence, the lead structures were primarily optimized for the pharmacological effect, and not until the end of the lead optimization process important developability properties, e.g. solubility, permeability and toxicity, were evaluated. After these determinations, a few candidate drugs (CDs) were selected for further development.

Contrary to expectation, the increased number of new structures generated each year has not resulted in a corresponding increase of drugs undergoing clinical trial. In part this has been attributed to poor pharmacokinetic (PK) properties of the CDs, and as much as 40% of the attrition rate of CDs has been related to poor PK profiles.¹⁰ Thus, reliable screening filters for factors such as absorption, distribution, metabolism, elimination/excretion and toxicity (ADMET) are highly desired.¹¹⁻¹³ Ideally, these screens should be computer-based to allow ADMET analysis of computationally designed drug-like molecules prior to the chemical synthesis. In this mode, only structures predicted to have acceptable potency and developability are selected for synthesis (Figure 2). This results in knowledge-based synthesis of fewer compounds with improved PK properties.¹⁴ After the synthesis of such prioritized compound libraries, the potency and the developability of the compounds are determined experimentally in parallel. Thus, methods for rapid *and* reliable experimental methods for screening of these important properties are warranted.

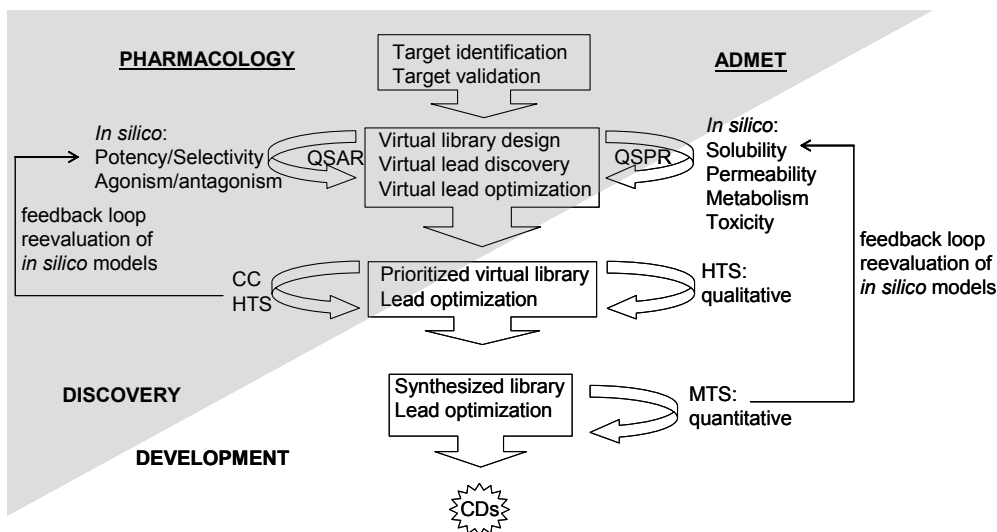


Figure 2. Knowledge-based, parallel drug discovery setting. Computational prioritization of a virtual library is followed by chemical synthesis. *In silico* and *in vitro* pharmacological and ADMET screening are performed simultaneously.

3.2. Intestinal Drug Absorption

Drug administration via the oral route is the most convenient for patients. Thus, intestinal absorption is one of the first molecular properties to be studied for new chemical entities to estimate the developability of an oral dosage form. The extent to which a drug will be absorbed, i.e., transported from the intestinal fluid across the mucosal membrane,¹⁵ is dependent on the physicochemical properties of the compound, the pharmaceutical dosage form and physiological factors.¹⁶ A prerequisite for drug absorption is that the drug dissolves in the intestinal fluid (Figure 3). However, the dissolved compound can be subjected to processes that lower the fraction of the dose absorbed (FA) by, e.g., enzymatic and/or chemical degradation in the intestinal fluid and formation of complexes or micelles with proteins, ions and/or food residues. In fact, only the unbound molecules will diffuse to the intestinal wall, permeate the enterocytes and eventually reach the systemic circulation.

3.2.1. Mechanisms of Intestinal Solubility

The intestinal solubility of a compound is dependent on the physicochemical properties of the molecule, the location in the gastrointestinal (GI) tract, the GI physiology and the dosage form. The close relationship between drug dissolution and drug solubility can be illustrated by the Noyes-Whitney equation (adjusted for sink condition)¹⁷:

$$\frac{dm}{dt} = \frac{DA(C_s)}{h} \quad \text{Eq. 1}$$

where, dm/dt is the dissolution rate, C_s is the maximum amount of drug that can be dissolved in the fluid, i.e. the maximum solubility of the compound in the dissolution medium, A is the surface area of the undissolved compact, D is the diffusion coefficient in the intestinal fluid and h is the height of the diffusion layer adjacent to the solid compact.

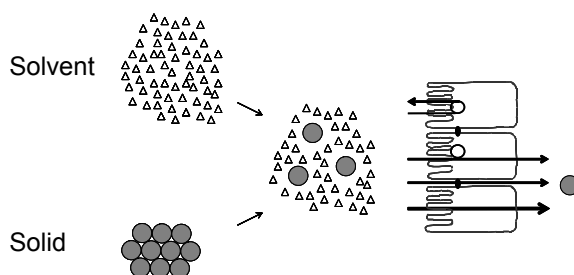


Figure 3. Drug dissolution followed by intestinal wall permeation. The dissolution of the drug is dependent on the solubility of the drug molecules in the intestinal fluid, and is a prerequisite for permeation of the enterocytes. The compound can passively diffuse para- or transcellularly, or be subjected to active transport/efflux.

Physicochemical properties such as size, lipophilicity and charge will influence the aqueous solubility and thereby the dissolution rate. Firstly, in order to incorporate the drug molecule, the tight water structure has to open up and form a large enough cavity for the solute (Figure 3). Thus, the larger the cavity that has to be formed, the more the energy required.¹⁸ Secondly, large molecules often are more lipophilic than smaller ones. In agreement with the “like-dissolve-like” theory, the solubility of lipophilic compounds will generally be poorer in the water-based intestinal fluid than the solubility of hydrophilic compounds.¹⁹⁻²¹ Thirdly, for proteolytic compounds, the solubility increases with increased ionization, as described by the Hendersson-Hasselbalch (HH) equation (exemplified for a weak base):²²

$$pK_a = pH + \log \frac{S_{tot} - S_0}{S_0} \quad \text{Eq. 2}$$

where pK_a is the dissociation constant of the solute, S_{tot} is the solubility at the specific pH used for the calculation and S_0 is the intrinsic solubility. Hence, the GI pH-gradient, which varies from pH 1 in the stomach up to pH 8 in the distal ileum,^{16,23} will result in charged compounds with increased solubility. An exception to this rule is zwitterionic compounds, which can display a positive and a negative charge within this pH-gradient. At the pH-value corresponding to the isoelectric point of the compound, the net charge of the compound is zero resulting in the lowest, i.e. the intrinsic, solubility.

The ionic strength of the intestinal fluid is dependent on the food and fluid intake as well as the absorption and secretion of fluid within the intestine.¹⁶ In general, the solubility decreases with increased ionic strength due to the salting-out effect and/or the common ion effect.²³⁻²⁸ The salting-out effect occurs when electrolytes in the solution compete with the drug molecules for interactions with water; with higher concentrations of electrolytes present in the water, more water molecules will be “occupied”. The common ion effect occurs when ionic complexes with no net charge are formed between electrolytes and the ionized drug molecules, which can result in precipitation. However, additives such as electrolytes may also improve drug solubility by salting-in effects. These can arise from specific interactions between the compound and the electrolytes²⁹⁻³¹ or from the formation of solvent cavities with the capacity to incorporate drug molecules.³² The intake of food may cause a salting-out effect, but the dissolution rate and the solubility may also decrease because of an increased viscosity.³³ Furthermore, food induces the secretion of bile salts, which are surfactants secreted by the gall bladder. These surfactants often improve the solubility of poorly soluble compounds and wetting has been attributed as the most important mechanism of bile salt solubilization.³³⁻³⁵ However, drug molecules can also be incorporated within micelles formed at higher bile salt concentrations, which further improves the solubility of the drug.^{33,36-39}

Solubility problems can be pharmaceutically treated by optimizing the compound and/or the dosage form. For instance, the solubility of the compound can be improved by salt formation⁴⁰⁻⁴² or by synthesis of a more soluble prodrug.⁴³⁻⁴⁶ The apparent solubility of the dosage form can be improved by micronization of the material in order to increase the surface area that will be in contact with the intestinal fluid.^{47,48} This processing of material can also result in solid state disorder of amorphous character, which has higher solubility than the crystalline compound.⁴⁹ Excipients such as cyclodextrines^{50,51} and disintegrating agents can further increase the solubility.^{52,53}

3.2.2. Mechanisms of Intestinal Membrane Permeation

As with solubility, the rate and extent of intestinal membrane permeation is dependent on both physicochemical properties of the compound and physiological factors. Drugs are mainly absorbed in the small intestine owing to its much larger surface area and because the epithelium is less tight than in the colon.¹⁶ The intestine is lined with enterocytes. These are polarized cells, with the apical membrane facing the intestinal lumen being separated from the basolateral membrane facing the sub-epithelial tissues by tight junctions. The apical and basolateral membranes have different phospholipid and protein compositions and therefore also different permeability properties.¹⁶

The drug molecules can either passively diffuse through the intestinal wall, or utilize active transport mechanisms (Figure 3). Passively absorbed compounds will diffuse through the cells (transcellularly) or in between the cells (paracellularly); which pathway is used is dependent on the physicochemical properties of the drug. The pH-partition theory suggests that only the un-ionized form of drugs permeate the intestinal epithelium,⁵⁴ but highly ionized compounds have been reported as exceptions to this rule.⁵⁵ Hydrophilic and/or charged compounds, which cannot easily permeate the lipophilic cell membrane, may diffuse through the aqueous pores. However, the limited surface area of the pores together with the size restriction by the tight junctions,^{16,56,57} limits the contribution of this pathway significantly. To diffuse transcellularly, a reasonable balance between hydrophobicity and hydrophilicity of the compound is important, since the compound diffuses both through lipophilic membranes and the aqueous cytoplasm.⁵⁸ Although the transport by the transcellular route can be regarded as a rather complex process, the majority of drug-like compounds utilizes this pathway.

Actively transported compounds permeate the membrane by binding to a membrane protein. This transport is energy dependent, site-specific, substrate-specific and saturable.⁵⁹ Thus, concentration dependent absorption can occur *in vivo* after administration of actively transported compounds, resulting in non-linear dose-response relationships. The carrier-mediated route can be useful for compounds with structural

features restricting transcellular absorption and, lately, compounds with poor passive permeability have been designed to target e.g. the PepT1 transporter,^{60,61} and nucleoside carriers⁶² in order to increase the FA. However, membrane proteins can also actively secrete, i.e. efflux drugs, a process resulting in a reduced FA. It has been proposed that efflux proteins cooperate to a large extent with metabolizing enzymes present in the cytoplasm, which could further limit the uptake of drug molecules.⁶³⁻⁶⁷ However, the clinical importance of this cooperation on FA has been questioned.⁶⁸

3.3. *In Vitro* Screening for Drug Absorption

The drug permeability over the intestinal lumen has previously been regarded as the major factor influencing the total amount of drug absorbed. However, hits identified by CC and HTS are usually large and lipophilic compounds, since an increased hydrophobicity generally results in an increased potency by non-specific binding to the target protein. Such physicochemical properties result in poor aqueous solubility, which stresses that not only drug permeability, but also the aqueous drug solubility needs to be analyzed to estimate the FA.⁶⁹ Experimental *in vitro* analysis of these properties allows rapid and qualitative estimations of the drug absorption *in vivo*, but *in vitro* methods can also be applied to mechanistically study the solubility and transport processes.

3.3.1. Solubility Measurements

Lead compounds are often delivered from the medicinal chemists as dimethylsulphoxide (DMSO) solutions and therefore high throughput ADMET screening based on DMSO solutions is requested. The turbidimetric method measures solubility by aqueous titration of a DMSO solution of the compound.⁷⁰⁻⁷² The solubility is determined as the value when precipitation occurs, suggesting that the aqueous solution has become oversaturated and the maximum solubility has been reached. The method results in qualitative solubility values, i.e. the classification of substances as “poorly” and “highly” soluble rather than measurements of absolute values. This together with the possibility to automate the method makes it applicable for solubility estimations of a large number of compounds.

The potentiometric technique determines the solubility of proteolytes from a pH titration of a drug suspension into a clear solution.^{23,73-76} Excess material is present during the titration and an apparent pK_a (pK_a^{app}) is determined under condition of precipitation. The obtained difference between the pK_a values in solution and precipitation (ΔpK_a) is used to calculate the intrinsic solubility (S_0), i.e. the solubility of the uncharged species, according to Eq. 3

$$\log S_0 = \log(C/2) - |\Delta \text{pK}_a| \quad \text{Eq. 3}$$

where C is the concentration. The method results in quantitative solubility values, i.e. solubility values of high accuracy. The titration is rather time consuming and therefore suitable in the interface of drug discovery and drug development, when fewer compounds are analyzed.

The method traditionally used for quantitative solubility determinations is the shake flask method,⁷⁷ in which the intrinsic solubility is determined after the equilibrium between the dissolved and undissolved compound has been reached. This method allows the determination of solubility values of highest possible quality. However, there are many factors influencing the measured solubility. The influence of ionic strength and pH has already been discussed (Section 3.2.1.). Solubility is affected by the temperature and, therefore, the solubility of a series of compounds should be performed at a specific temperature. Moreover, the solubility is determined when equilibrium between the dissolved compound and the suspension is reached, making the time-scale important. The separation of solid from the solution by either filtration or centrifugation may further influence the final solubility value. Spectrometry and HPLC are commonly used for analysis, but use of electrical stream sensing has also been reported in the literature.^{78,79} Therefore, to allow a correct comparison of solubility values there is a need for standardization of the experimental setting. Solubility determinations by shake flask are time consuming and labor intensive, making them suitable for application in the drug development setting when an exact solubility value for the CD is required.

To summarize, solubility determinations are performed either in a screening mode or in experimentally more demanding settings, making the methods applicable at different stages of the drug discovery and development process. Moreover, the solubility data obtained are of different levels of accuracy, which should be carefully considered when solubility data are selected for *in silico* model development.

3.3.2. Permeability Measurements

The complexity of *in vitro* models used for the estimation of drug permeability varies greatly, from simple partition systems allowing automatized screening to more advanced and labor demanding physiological models. Physicochemical characterization used for estimation of passive diffusion through cell membranes, such as the lipophilicity coefficient obtained from octanol-water partitioning ($\log P_{\text{oct}}$),⁸⁰⁻⁸² $\Delta \log P$ ($\log P_{\text{oct}} - \log P_{\text{cyclohexane/water}}$)^{83,84} and artificial membranes⁸⁵⁻⁸⁷ are qualitative methods that rather describe distribution into a lipophilic environment than permeability through a membrane. In order to obtain quantitative permeability values, epithelial cell monolayers can be used. The most commonly used epithelial cell line for intestinal

permeability screening is Caco-2.^{88,89} This cell line originates from the human colon and forms enterocyte-like monolayers under standardized conditions. Caco-2 cell monolayers display active transporter proteins and under certain conditions metabolizing enzymes such as CYP3A4,⁹⁰ making Caco-2 cells useful in the screening for active transport/efflux⁹¹ and enterocytic metabolism.⁹⁰ The Caco-2 cell monolayer is tighter than the small intestine, and hence the paracellular transport is underestimated by the use of this system. The leakier 2/4/A1 cells have been suggested for screening for paracellular transport.⁹²⁻⁹⁴

As for solubility the *in vitro* methods for permeability determinations are of different complexity making them useful at different stages of the discovery process. Physicochemically based methods producing qualitative data are useful in early drug discovery when rough estimations of permeability for a large number of compounds are wanted. The more sophisticated cell lines which generate quantitative data are applicable later in the drug discovery process, when highly accurate data for a smaller number of compounds are required.

3.4. *In Silico* Screening for Drug Absorption

In the last few years the number of publications on *in silico* prediction of absorption has increased significantly (Figure 4). The potential of computational models is obvious since the use of virtual tools for prediction of drug absorption would allow new compounds to be evaluated prior to synthesis.

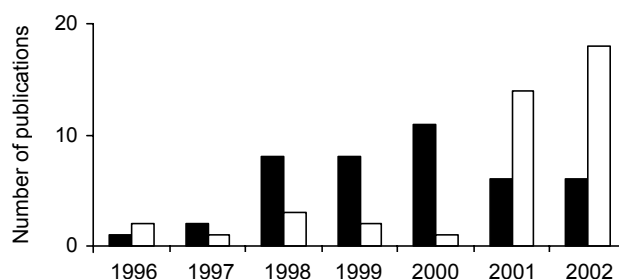


Figure 4. Number of publications on *in silico* absorption predictions.

Black and white bars show the number of publications on permeability and solubility predictions, respectively, from the year of 1996 and onwards. The PubMed search was performed in July 2003.

The experimental input data must be of high quality to build computational models of high accuracy. Moreover, the models should be based on drug-like molecules, since drug molecules often contain a larger variety of functional groups. Virtual filters for drug-likeness can provide a guide for the selection of the dataset for the model development.^{70,95-99} One example of a drug-likeness filter is the ChemGPS methodology.^{96,98,100} This method has defined the chemical space of drugs through multivariate data analysis of physicochemical properties of drugs and non-drugs. The Lipinski rule-of-five is an example of a non-complex, easy-to-use filter for drug-likeness in terms of their developability characteristics.⁷⁰ The rule-of-five states that compounds with a molecular weight of less than 500, a $\log P_{\text{oct}}$ value of less than five, fewer than five hydrogen bond donors and ten hydrogen bond acceptors will probably be absorbed after oral administration. It is likely that the intestinal absorption will be poor if two or more of these cut-off values are violated. Pickett and coworkers have shown that drug-like libraries with good absorption characteristics can be designed by use of molecular weight, the calculated $\log P_{\text{oct}}$ (ClogP) and the polar surface area (PSA).¹⁰¹ To conclude, tools for drug-likeness should be considered in the selection of the datasets used in the generation of ADMET models applicable in the drug discovery process.

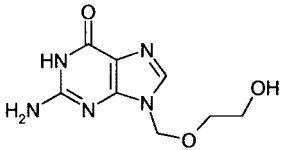
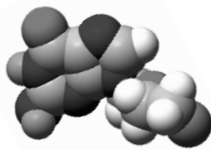
3.4.1. Computational Calculation of Molecular Descriptors

Ideally, the descriptors used for model development should be rapid to calculate and easy to interpret. Descriptors can be classified as one-, two- or three-dimensional (1D, 2D and 3D, respectively), depending on the representation needed for the calculation (Table 1). Simplistically, the time needed for the calculation of properties increases with the increase in dimension, with quantum mechanics calculations based on the wave function obtained from the 3D structure being the more time consuming (see below). However, the speed of calculation of descriptors based on the 3D representation has increased through the marketing of software for 2D to 3D conversion.^{102,103}

The simplest descriptors are calculated from a 1D representation of the compound. Typical 1D properties are atom counts and molecular weight. Computational languages describing the bond order of atoms can be used for the calculation of 2D descriptors.¹⁰⁴ Typically, the 2D descriptors are related to the size, flexibility/rigidity, electron distribution, hydrophilicity and lipophilicity.¹⁰⁵⁻¹⁰⁷ Several of the properties are calculated from the group contribution approach, which is based on data from large sets of compounds that have been experimentally determined for the response parameter of interest.¹⁰⁸⁻¹¹⁰ The calculation of such 2D properties is largely dependent on the size of the experimental database, and calculations of compounds with fragments missing in the database can give erroneous results. While lipophilicity and hydrogen bond strength are descriptors that can be rather easily interpreted, electrotopological and

electrogeometrical descriptors might be more difficult to understand. However, they contain information on the electron distribution of the structural features comprising the drug molecule.

Table 1. Examples of descriptors calculated from different representations of the molecule.

	Typical Representation	Typical Descriptors
1D	$C_8H_{10}N_5O_3$	Molecular weight Atom counts
2D		Fragment counts Topological indices Connectivity Flexibility
3D		Molecular surface areas Molecular volume Interaction energies
	Wave function	Valence properties

The molecule has to be converted to its 3D structure if information associated with the conformation of the molecule is needed. Molecular mechanics and/or molecular dynamics calculations are used to investigate the conformational space of the molecule and to identify low energy conformers. Typical descriptors calculated from the 3D structure are properties related to the molecular surface area, i.e. the PSA,¹¹¹ non-polar surface area (NPSA)¹¹² and partitioned total surface areas (PTSAs),¹¹³ and the volume. The PSA is commonly defined as the surface area occupied by oxygen atoms, nitrogen atoms and hydrogen atoms bound to these heteroatoms,^{111,114} but also sulfur atoms and phosphorus atoms have been defined as polar.^{115,116} The NPSA is defined as the total surface area minus PSA.¹¹² PTSA is the calculated surface area occupied by each type of atom.¹¹³ Other descriptors obtained from the 3D representation of the molecule are hydrophobicity/hydrophilicity balance, amphiphilic moments and critical packing parameters, which can be calculated on the basis of molecular interaction fields.¹¹⁷ More advanced descriptors related to the distribution of valence electrons can be calculated after conversion of the 3D structure into its wave function by the use of quantum mechanics calculations.^{118,119} Unfortunately, the quantum mechanical calculation for a single structure can take hours, making such descriptors unsuitable in the screening of large chemical libraries.

3.4.2. Model Development

The establishment of absorption models can be considered to have two aims: firstly, there is a need for tools facilitating the rapid estimation of the developability of lead compounds.¹²⁰ Secondly, information on the impact of structural features on developability is wanted to guide medicinal chemists in the drug design process (Figure 2). The estimations can be qualitative with the resulting classification being “high/intermediate/poor” and “yes-no” answers, or quantitative, resulting in predictions of higher accuracy. The lack of large drug-like datasets with solubility and permeability values of high quality has resulted in models developed from large series of non drug-like molecules¹²¹⁻¹²⁶ or from small series on drug-like molecules.^{113,127-129} The large datasets are compiled of data generated in different laboratories using different techniques, reducing the quality of the data.¹³⁰⁻¹³²

The simplest models for prediction are based on the correlations between two properties. Correlations to solubility and permeability have been obtained from linear and non-linear regression; for instance $\log P_{\text{oct}}$ has been linearly correlated to solubility,¹³³ while PSA has been correlated to permeability by sigmoidal regression.¹³⁴ These models are transparent for the user, but fail to predict drug-like datasets with broad structural diversity.

Predictive solubility and permeability models have been devised by use of multivariate statistics. Multivariate statistics are defined as methods that examine multiple variables simultaneously, and hence, models built from several descriptors are regarded as multivariate.¹³⁵ These models can be obtained from multiple linear or non-linear regression (MLR and MNLR, respectively) of the variables. MLR and MNLR require stepwise regression, mathematical independence of the x-variables and a larger number of observations than the number of variables. More advanced treatment is provided by techniques using the projection of latent variables, such as principal component analysis (PCA)¹³⁶ and partial least square projection to latent structures (PLS).¹³⁷ These methods are suitable when handling datasets with few observations and many variables. Moreover, PCA and PLS methods can use correlated variables and data matrices with missing values in the model development. PCA summarizes the variation in the x-space, gives an overview of the data, and reveals groups of observations, trends and outliers.¹³⁶ PLS, in contrast, is used for prediction of response parameters and relates two data matrices to each other by a linear multivariate model using latent structures.¹³⁷ Recently, non-linear PLS was introduced for quantitative structure-property relationships (QSPR),¹³⁸ but its usefulness in the drug discovery process remains to be shown. All of the above mentioned techniques are rather transparent for the user, revealing the influence and importance of each variable included in the prediction.

Less transparent models are obtained from neural networks (NN). The NN mimic the way neurons are connected to each other within the brain. In each layer, information obtained from several neurons is compressed and further transmitted by a new neuron into the next layer.¹³⁹ Thus, the neural network models are generally viewed as black boxes, since the influence of each input variable cannot be revealed. A significant pitfall of NN is the ease by which the system is over-trained. This results in models which are specific for the datasets used in the development, with little or no capacity to accurately predict new data.¹³⁹ However, NN have gained much attention as prediction tools for PK properties lately, and models with high accuracy have been developed for the prediction of solubility¹⁴⁰ and absorption.^{141,142}

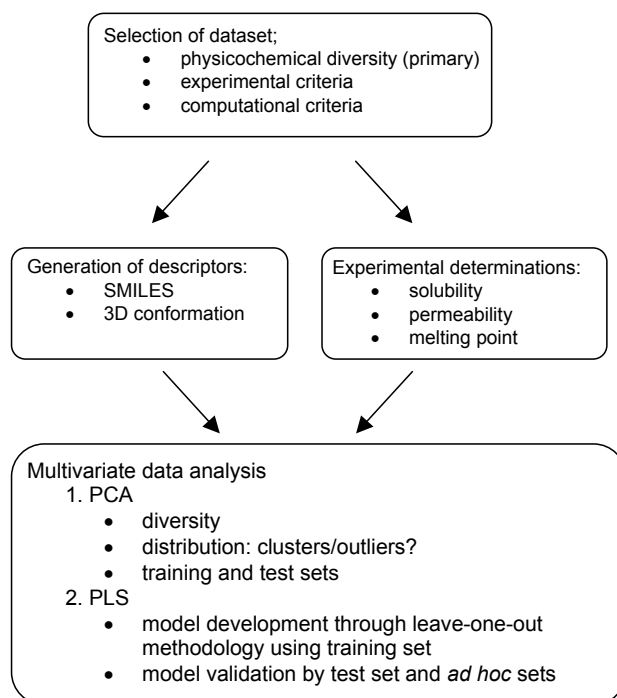


Figure 5. Flow chart of model development. The example given is applied in Papers I-IV.

Irrespective of the statistical tool used for model development, the validity of the model will be dependent on the dataset used, i.e. the training set (Figure 5). The requirement of a large and structurally diverse database for the development of global models with general applicability may initiate the generation of models applicable to a smaller volume of the drug-like space. Cross-validation of the model, i.e. iteratively keeping a portion of the training set out of the model development, is one way to avoid over-fitted

models.¹⁴³ Moreover, the external predictivity of models can be assessed by prediction of datasets that have not been involved in the model development, i.e. test sets. The use of test sets indicates to what extent new structures are correctly predicted.¹⁴³

3.4.3. Solubility Models

Solubility is a thermodynamic process dependent on the enthalpy and entropy of mixing. Studies have suggested that the entropy effect on solubility is constant for rigid organic molecules.¹⁴⁴ However, aqueous solubility has been predicted from both enthalpy related¹⁴⁵ and entropy related descriptors.^{21,144,146,147} The linear solvation energy relationship (LSER) is an extension of Hildebrand's and Scatchard's work on the enthalpy related solubility parameters^{19,20,145} and considers solubility as a function of volume, dipolarity and hydrogen bonding capacity.^{18,148} LSERs have been used in the prediction of blood solubility and distribution to tissues, such as the brain, lung, muscle, kidney and fat,¹⁴⁹ and has been widely accepted for aqueous solubility prediction.^{150,151} Recently, an amended LSER for solubility prediction was presented by Abraham and coworkers.^{152,153} The non drug-like training set was predicted with good accuracy from descriptors of solute hydrogen bond acidity/basicity, molar refraction describing dispersion forces, solute polarizability and solute volume.

In the late 60's Hansch and coworkers reported a linear correlation between $\log P_{\text{oct}}$ and solubility ($R^2=0.87$) for liquids.¹³³ Yalkowsky and coworkers combined $\log P_{\text{oct}}$ with a solid state characteristic, i.e. the melting point, which successfully predicted the solubility of solids.^{130,144,146,154-156} Unfortunately, the melting point has to be experimentally determined, and hence, chemical synthesis is required. However, Meylan and collaborators investigated several approaches to predict the solubility of a diverse, but non drug-like, dataset of 1450 substances.¹²² This study suggested that lipophilicity and a size descriptor alone can predict solubility with the same accuracy as if the melting point was to be included. McFarland and coworkers used the lipophilicity and hydrogen bond charges to predict a series of 22 crystalline drugs with high accuracy ($R^2=0.88$).¹²⁹ The hydrogen bond charges have also been used to predict the solubility from a similarity index approach.^{157,158}

Molecular surface area descriptors have proven important in solubility predictions of liquids in liquids.¹⁵⁹ Recently, Jorgensen and Duffy presented a solubility model applicable to solids ($R^2=0.88$) based on descriptors of surface area and hydrogen bonds obtained from molecular modeling of molecules in an aqueous environment.^{160,161} Solubility models of equal accuracy have been obtained from descriptors of electrotopology/geometry and flexibility when treated with MLR¹⁶², PLS¹⁶³ and/or NN.^{121,123,125-127,162,164-168} However, the majority of these datasets are non drug-like and

several of the models are based on chemicals such as pesticides, alcohols and aliphatic hydrocarbons, which are located in a different part of the chemical space than that of drugs. Hence, the applicability of these models in the drug discovery process is unclear.

3.4.4. Permeability Models

Permeability models are generally models of transcellular passive transport, and descriptors of lipophilicity, hydrophilicity and molecular size have proven to be important. The $\log P_{\text{oct}}$ descriptor is an important predictor of membrane permeability (Section 3.2.2.), and hence, the ClogP descriptor is incorporated into a large number of the models developed. For less complex datasets, ClogP, PSA and hydrogen bond counts have each been used as a single predictor of permeability.^{84,111,128,134,169-172}

However, $\log P_{\text{oct}}$ can be regarded as a composed property, largely dependent on both the size and the hydrophilicity of the compound.¹⁷³ Indeed, the use of molecular weight and hydrogen bond descriptors have been shown to predict permeability.^{101,174}

The introduction of datasets with large structural diversity in model development has highlighted the need for several descriptors and multivariate data analysis to obtain good models. For instance, the introduction of larger and more flexible structures showed that PTSAs and descriptors related to the flexibility of the molecule are also useful in permeability predictions.^{113,175}

Electrotopological indices have resulted in permeability models of good accuracy when treated with PLS.^{176,177} Other descriptors applicable for permeability predictions are the solubility descriptors, the amended LSER descriptors and hydrogen bond charges (see Section 3.4.3.).^{153,158,161} Descriptors such as ClogP, polarizability, polarity, the strength of the Lewis base and the Lewis acid, and the number and strength of hydrogen bond donors or acceptors obtained from quantum mechanics have been correlated to permeability.^{113,118,119,178} These descriptors gave good results ($R^2 > 0.79$), even though less complex and more rapidly calculated descriptors of PTSAs were more accurate ($R^2 = 0.85$). Thus, since quantum mechanics descriptors are not outperforming more rapidly calculated descriptors with respect to accuracy of the permeability prediction, they are of limited use in the drug discovery setting until the calculations become faster.

To conclude, models using different descriptors and statistical tools for the prediction of solubility and permeability have been developed. Unfortunately, the majority of the models are based on datasets compiled of non drug-like molecules, which may restrict their usability in drug discovery. Moreover, the limited number of experimental data available and the interlaboratorial variability of such data further confine the possibility to devise good models of drug absorption. Thus, in order to improve the prediction of intestinal absorption, there is a need to further expand the experimental database

available for model training.¹¹ Furthermore, several models of intestinal absorption are based on FA,^{141,179} which is a composed measure affected by for example solubility and permeability. Simultaneous prediction of solubility and permeability would give information on the relative importance of each property on absorption and result in amended absorption models.

4. AIMS OF THE THESIS

The general aim of the thesis was to develop new protocols for prediction of intestinal drug solubility and absorption. In the first part of the thesis, screening approaches for the prediction of solubility and permeability were studied (Papers I-III). In the second part an analysis of the melting point, a commonly used solid state characteristic included in computational solubility predictions, and the pH-dependent solubility were performed (Papers IV and V, respectively). The specific aims were the following;

- ✓ to devise a small-scale experimental method for generation of high quality solubility data
- ✓ to develop *in silico* models for aqueous drug solubility based on calculated molecular descriptors
- ✓ to devise computational protocols applicable to the prediction of aqueous drug solubility and intestinal drug permeability in an effort to predict the absorption of orally administered drugs
- ✓ to investigate molecular descriptors influencing the solid state and to evaluate to what extent the solid state needs to be incorporated in *in silico* solubility models
- ✓ to analyze the accuracy of calculated pH-dependent solubility of drug molecules and to evaluate the effect of extrapolation of solubility data from one pH-value to another.

5. METHODS

5.1. Investigated Drugs

The selection criteria for the compounds investigated in the different studies were that: a) the drugs should be structurally, physicochemically and therapeutically diverse; b) the compounds should be stable at the pH used for the solubility and permeability determinations; c) it should be possible to analyze the conformational preferences of the molecules using molecular mechanics calculations; d) the compounds should not display polymorphism or pseudopolymorphism, e) the compounds included in the permeability study should mainly be passively transported through the Caco-2 cell monolayers or display a concentration-independent absorption *in vivo*, and f) the melting point data used for computational prediction of the solid state should be determined for the pure compound, i.e. no salt forms were included in this study.

5.2. Differential Scanning Calorimetry (DSC)

The solid state characterization described in Paper I and V was performed with a Mettler DSC 20 TC10A/15 (Switzerland) and a DSC 220C (Seiko, Japan), respectively. The samples were kept in aluminium pans and heated at a rate of 10°C/minute in the interval 25-350°C. The equipment utilized in Paper V kept the samples in an atmosphere of nitrogen to avoid oxidation taking place during the experiment.

5.3. Solubility Determinations

The experimental solubility values were obtained by the small-scale shake flask (SSF) method (Papers I and V) or by potentiometric titration (Paper II). The SSF method used volumes of 50-1,000 μL solvent to determine the solubility value of the drugs. Each drug was added in excess and the test tubes were placed on a plate shaker (300 rpm) at room temperature ($22.5 \pm 1^\circ\text{C}$). The pH of each drug suspension was adjusted to a value of at least 1 pH unit below or above the pK_a for acids and bases, respectively. This allowed the solubilities of uncharged compounds to be determined. The pH of ampholyte suspensions was adjusted to the isoelectric point of the compound in order to determine the solubilities of the zwitterionic species. In the pH-dependent solubility study (Paper V), additional solubility determinations within the pH-range 2-12 were performed to obtain the complete pH-solubility curve. Several end-points were investigated for the solubility determinations (Paper I) in order to study the time-range needed for the solubility experiment. The samples taken from the suspensions were centrifuged in an Eppendorf centrifuge at 23,000 g for 15 minutes to separate the solid

material from the solution and the supernatant was analyzed with HPLC (Figure 6). Poorly soluble compounds showing no detectable solubility using the SSF method and HPLC analysis, were studied using methanol as cosolvent. The solubility was measured at different concentrations of methanol in water and the aqueous solubility (0% w/w methanol) was extrapolated by linear regression.^{156,180}

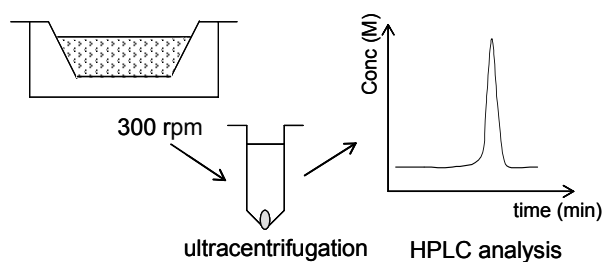


Figure 6. Method setting for the SSF solubility measurements.

Suspensions were shaken for 24 -72 h at room temperature, thereafter they were ultracentrifuged and the supernatant was analyzed for solute concentration using HPLC.

In Paper V, pH-dependent solubility plots were drawn using the mean values with standard deviations. The range of the solubility was obtained from the solubility_{max} and solubility_{min} calculated using the following sigmoidal function:

$$\log S_{\text{tot}} = \frac{\log S_{\text{max}} - \log S_{\text{min}}}{1 + \left(\frac{\text{pH}}{\text{pH}_{50\%}} \right)^{\gamma}} + \log S_{\text{min}} \quad \text{Eq. 4}$$

where S_{tot} is the solubility at a specific pH, S_{max} is the solubility for the completely ionized compound, S_{min} is the intrinsic solubility, $\text{pH}_{50\%}$ is the pH value at 50% of the solubility range and γ is the slope factor. The equation was fitted to the solubility values by minimizing the sum of squared residuals. Experimentally determined values within 20-80% of the range of S_{min} and S_{max} were used to obtain the slope of the linear part of the pH-dependent solubility curve. A prediction of the pH-solubility profiles was obtained from the intrinsic solubility value of the compounds using the Henderson-Hasselbalch equation (see Eq. 2).

Solubility determinations at 25 and 37°C (Paper II) were performed with the potentiometric technique as implemented in the pSOL apparatus (pION inc., Boston, MA).⁷³⁻⁷⁵ Prior to the solubility experiment the pK_a was measured, which was used to calculate the intrinsic solubility (see Eq. 3). The intrinsic solubility was determined by a

pH titration of a suspension of the drug into a clear solution. The compounds were titrated in volumes of 1.7–17.0 ml and stirred with a magnetic stirrer.

5.4. Cell Culture

Caco-2 cells obtained from American Tissue Collection, Rockville, MD, USA, were maintained in an atmosphere of 90% air and 10% CO₂, as described previously.⁸⁹ For transport experiments, 5×10⁵ cells of passage number 94–100 were seeded on polycarbonate filter inserts (12 mm diameter; pore size 0.4 µm; Costar) and allowed to grow and differentiate for 21–35 days before the cell culture monolayers were used for transport experiments.

5.5. Transport Studies

The intestinal permeability of the compounds was determined from transport rates across Caco-2 cell monolayers.^{82,89} In general, the drugs were dissolved in Hank's Balanced Salt Solution (HBSS) containing 25 mM HEPES at pH 7.4. The amount of compound added to the transport buffer depended on the solubility of the compound, its expected permeability, the presence of saturable active transport mechanisms, and the HPLC detection limit for the compound. Transport studies were initiated by incubating the monolayers in HBSS, pH 7.4, at 37°C for 20 minutes in a humidified atmosphere. Filter inserts with Caco-2 cells were stirred at 500 rpm during the transport experiments to obtain data that were unbiased by the aqueous boundary layer.¹¹³ Permeability coefficients were determined both in the apical to basolateral direction and in the basolateral to apical direction (pH 7.4 in both chambers) in order to determine the possible involvement of active transport mechanisms or efflux. Monolayer permeability to the paracellular marker [¹⁴C]-mannitol was routinely used to investigate the integrity of the monolayers under the experimental conditions.

In general, the transport studies were performed under sink conditions and the apparent permeability coefficients (P_{app}) were calculated from

$$P_{app} = \frac{\Delta Q}{\Delta t} \times \frac{1}{A \times C_0} \quad \text{Eq. 5}$$

where $\Delta Q/\Delta t$ is the steady-state flux (mol/s), C_0 is the initial concentration in the donor chamber at each time interval (mol/mL), and A is the surface area of the filter (cm²). For rapidly transported compounds where sink conditions could not be maintained for the full duration of the experiments, P_{app} was calculated as described previously⁵⁵ from

$$C_R(t) = \frac{M}{V_D + V_R} + \left(C_{R,0} - \frac{M}{V_D + V_R} \right) \times e^{-P_{app} \times A \times \left(\frac{1}{V_D} + \frac{1}{V_R} \right) \times t} \quad \text{Eq. 6}$$

where $C_R(t)$ is the time-dependent drug concentration in the receiver compartment, M is the amount of drug in the system, V_D and V_R are the volumes of the donor and receiver compartment, respectively, and t is the time from the start of the interval. P_{app} was obtained from nonlinear regression, minimizing the sum of squared residuals ($\Sigma(C_{R,i,obs} - C_{R,i,calc})^2$), where $C_{R,i,obs}$ is the observed receiver concentration at the end of the interval and $C_{R,i,calc}$ is the corresponding concentration calculated according to Eq. 6.

5.6. Analytical Methods

Reversed-phase HPLC was used to determine the drug concentration of the solubility samples in Papers I and V, and the amount of drug transported through the Caco-2 cell monolayers in Paper II. In Paper I an isocratic HPLC system was used, and methods were developed for each specific drug that was analyzed. The sample analysis performed in Papers II and V used an HPLC gradient; the same method was applied for all compounds.

Radioactive samples (Paper II) were analyzed with a liquid scintillation counter (Packard Instruments 1900CA TRI-CARB; Canberra Instruments, Downers Grove, IL).

5.7. Biopharmaceutical Classification

The drugs were classified into six different biopharmaceutical classes according to their permeability¹¹⁵ and solubility¹⁸¹: I. high solubility - high permeability; II. low solubility - high permeability; III. high solubility - low permeability; IV. low solubility - low permeability; V. high solubility - intermediate permeability; and VI. low solubility - intermediate permeability. A drug was regarded as a highly soluble compound if the maximum dose given orally was soluble in 250 mL fluid in the pH interval 1-7.5. The maximum dose found in the Physicians' Desk Reference¹⁸² and/or in FASS¹⁸³ was compared with the minimum solubility value at a pH between 1 and 7.5. The permeability was defined as "low" if <20% and as "high" if >80% of the given dose is absorbed in humans. Drugs with FA data in between these values were defined as having intermediate permeability.¹¹⁵ The P_{app} values discriminating between the three classes of permeability were obtained from the correlation between drug permeability in Caco-2 cells and the FA established in our laboratory. The sigmoidal function used was

$$FA = \frac{100}{1 - \left(\frac{P_{app}}{P_{app50\%}} \right)^\gamma} \quad \text{Eq. 7}$$

where $P_{app50\%}$ is the apparent permeability corresponding to 50% of the dose absorbed and γ is the slope factor. This curve was used to calculate the permeability values corresponding to 20% and 80% of the dose absorbed.

The theoretical biopharmaceutical classification performed in Paper II was based on a combination of PLS models for solubility and permeability (see section 5.9.2). The predicted solubility value was compared to the maximum dose given, and thereafter sorted as a low or high solubility. The predicted permeability value was compared to the experimentally determined permeability cut-offs, and thereafter sorted as a low, intermediate or high permeability.

5.8. Molecular Descriptors

The lipophilicity was calculated using the ClogP program (version 2.0) from BioByte Corp. (Claremont, CA). The 2D descriptors used in Papers III-IV were calculated with Molconn-Z¹⁸⁴ and the AstraZeneca in-house program SELMA.¹⁸⁵ Molconn-Z was used to calculate electrotopological state indices. Briefly, the electrotopological state indices for a particular atom result from the topological and electronic environment. The indices will encode the electronegativity and the local topology of each atom by considering perturbation effects from the neighboring atoms. The program SELMA generates descriptors related to size, ring structure, flexibility, hydrogen bonds, polarity, connectivity,^{106,107,186} electronic environment, partial atom charge and lipophilicity. In total, Molconn-Z and SELMA generated 566 different descriptors.

The 3D descriptors used in Paper I-IV were obtained after a 500 (Papers III-IV) to a 250,000 (Paper II)-step Monte Carlo conformational analysis with the BatchMin program as implemented in MacroModel version 6.5. The MM2 force field was used for the smaller datasets (Papers I and II), while MMFF was applied on the larger datasets (Papers III-IV). Water was used as environment for the conformational studies in the papers investigating solubility alone (Papers I and III), whereas vacuum was used as environment in the studies predicting permeability values (Paper II) and the melting point (Paper IV).

Both the dynamic molecular surface areas obtained by Boltzmann averaging of all the low energy conformers (Paper I) and the static molecular surface areas for the global

minimum conformation identified in the conformational analysis (Papers I-IV) were calculated by MAREA.¹⁸⁷ Composite properties, such as NPSA and PSA, were calculated as were PTSA descriptors. The PSA was defined as the surface area occupied by oxygen and nitrogen, and hydrogen atoms bound to these heteroatoms, whereas the NPSA was defined as the SA minus the PSA. The PTSA descriptors correspond to the surface area of a certain type of atom (Figure 7). For example, the NPSA originating from carbon atoms can be partitioned into the surface areas of sp -, sp^2 -, and sp^3 -hybridized carbon atoms and the hydrogen atoms bound to these carbon atoms. In a similar way, the PSA originating from oxygen atoms can be partitioned into the surface areas of single-bonded oxygen, double-bonded oxygen, and hydrogen atoms bound to single-bonded oxygen atoms. The absolute surface areas and the surface areas relative to the SA were calculated.

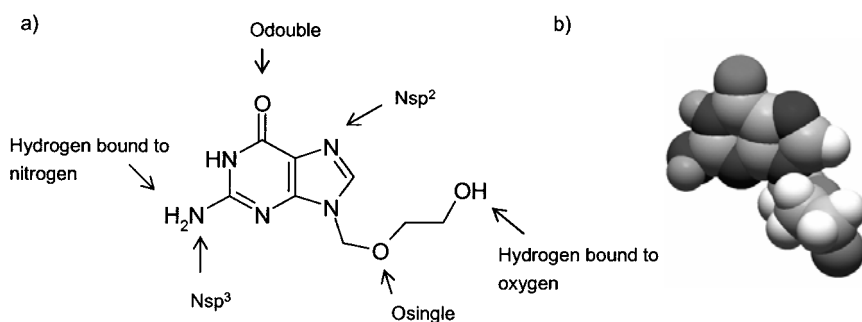


Figure 7. Molecular surface areas calculated for acyclovir. a) The PTSAs represent the surface areas of each type of atom in the molecule. PSA is comprised of the PTSAs of oxygen atoms, nitrogen atoms and hydrogen atoms bound to these heteroatoms. All other atom types are included in the NPSA. b) 3D conformation used for PTSA calculations. Oxygen and nitrogen atoms are shown in dark gray.

5.9. Statistics

5.9.1. Solubility and Permeability Experiments

The experimentally determined data included in this thesis were measured in at least triplicate. ANOVA was used to test whether the difference between two mean values in Paper I was statistically significant ($p < 0.05$). The coefficient of determination (R^2) assesses the goodness of fit of linear and sigmoidal regressions.

5.9.2. Model Development

The diversity of the descriptor space used for the prediction of melting point, solubility and permeability values was analyzed by PCA¹³⁶ (Papers I-IV) and by the ChemGPS methodology⁹⁸ (Paper III). Skewed descriptors were transformed prior to the multivariate data analysis to avoid overweighting in the models and descriptors with a skewness that exceeded ± 1.5 were excluded from the model development. The PCA of the input matrix was used to divide the compound datasets into training sets and test sets. In general, the training set was selected to cover a maximum range in descriptor space and included approximately two thirds of the dataset investigated. Qualitatively determined values obtained in Paper II were included in the test set. In the large dataset used for melting point prediction, the PCA approach was too complicated to use, since the number of compounds made the PCA plot less transparent. Therefore, every third compound when listed in ascending melting point order was included in the test set and, thereafter, the diversity of the selected training and test sets was checked with PCA.

The models were obtained by linear regression and MLR in Paper I, by PLS in Papers I-IV and by consensus modeling of PLS models in Papers III-IV. The number of PLS components computed was assessed by the cross-validated coefficient of determination (Q^2). The training set was divided into four (Paper III) or seven groups (Papers I, II, and IV), and the Q^2 was obtained by leaving out one group at time from the R^2 calculation. Only PLS components resulting in a positive Q^2 were computed and the number of principal components was never allowed to exceed one-third of the number of observations used in the model. The models were refined through step-wise selection of the descriptors. Initially, all the non-skewed descriptors were included in the PLS model. After the first round, the descriptor with the least influence on the prediction was deleted and the PLS repeated. If this exclusion resulted in a more predictive model (as assessed by a higher Q^2), the descriptor was permanently excluded from the model. This procedure was repeated until no further improvement of the model could be achieved. The predictivity of the models was assessed by root-mean square error of the test sets ($RMSE_{te}$) in Papers I-IV and the external test sets ($RMSE_{ext}$) in Papers II and III. When applying consensus modeling (in Papers III and IV), the results of several models were averaged to predict the solubility and the melting point, respectively.

6. RESULTS AND DISCUSSION

6.1. Datasets

In the first part of the thesis, screening approaches for the prediction of solubility and permeability were studied (Papers I-III). An experimental method, which measures solubility using small amounts of crystalline compounds, was devised in Paper I. The solubility data obtained were used to test the applicability of surface area descriptors in solubility predictions. In Paper II, solubility and permeability models based on molecular surface areas were generated, with the goal to allow theoretical prediction of the intestinal absorption of drug-like compounds. The general usefulness of the molecular surface areas and descriptors of electrotopology, bond energies, connectivity indices, flexibility, hydrophobicity and hydrophilicity for solubility prediction was tested in Paper III. In the second part of the thesis, the impact of the solid state and the ionization of the compound on the solubility were analyzed (Papers IV and V, respectively). Since the five studies had such different purposes, the selection criteria varied for the datasets investigated. However, all datasets were selected to be drug-like and structurally and physicochemically diverse.

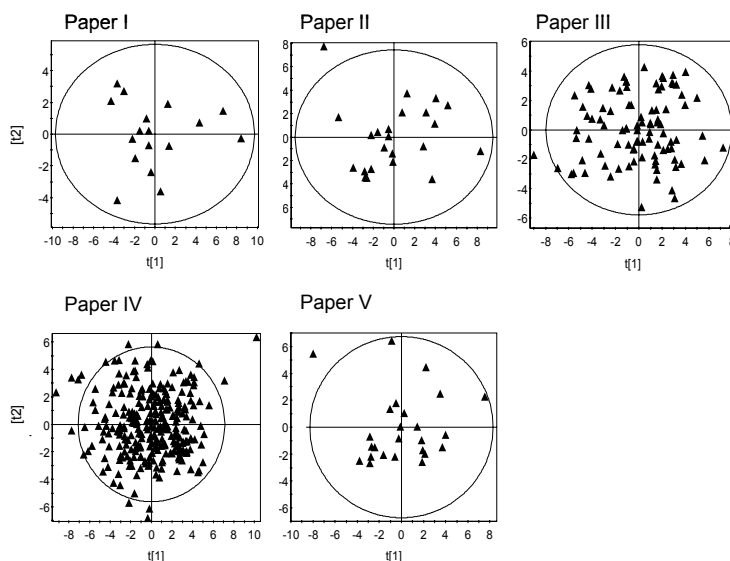


Figure 8. The physicochemical diversity identified by use of PCA and 2D and 3D descriptors. The two first principal components of each data analysis are shown, representing the following structural diversity of the datasets; 67% in paper I, 55% in paper II, 54% in paper III, 52% in paper IV and 64% in paper V.

The physicochemical and structural diversity was identified with PCA for all datasets studied. Indeed, most of the datasets were diverse (Figure 8) and, thus, they were considered to be suitable training sets for the model development. The ChemGPS analysis of the 85 compounds investigated in Paper III showed that the compounds were scattered within the drug-like space. As a result, this dataset was considered to be challenging to predict *in silico*, since it was not restricted to certain volumes of the drug-like space. The dataset most structurally restricted was the series of amines studied in Paper V. However, also this dataset proved to be diverse (Figure 8), mainly due to the large variation in physicochemical properties, e.g. size, lipophilicity and hydrophilicity, but also because of the inclusion of compounds with primary, secondary and tertiary amines.

A large range in data of the response parameters (solubility, permeability and melting point) was desired to avoid obtaining models with limited applicability. The aqueous drug solubility of the datasets studied ranged over almost seven log units, from 0.7 ng/mL of SKF105657 in Paper I to more than 20 mg/mL of ergonovine and zidovudine in Paper II. The permeability coefficients of the drugs ranged from 3×10^{-8} cm/s of folic acid and methotrexate to 4×10^{-4} cm/s of ethinyl estradiol (Paper II). In Paper IV, the dataset had melting points from 40°C up to 345°C, with the majority of the compounds displaying melting points between 140°C and 160°C.

The solubility and permeability data were determined in-house with standardized methods to maximize the reliability in the data used for the model development (Papers I and II). In Paper III, the number of compounds investigated was increased by compiling solubility data of Papers I and II, and by including solubility data obtained from the pharmaceutical industry. The experimental quality was prioritized over the number of compounds studied, so only compounds for which the intrinsic solubility at room temperature had been determined were included. The advantage of this compilation was that the larger number of compounds resulted in the investigation of a larger portion of the drug-like space (Figure 8).

The external test set of 207 compounds applied in Paper III was compiled from data in the literature.^{160,166} The aqueous drug solubility data found in these publications have been used repeatedly for model development,^{160,163,166,188-190} and hence provide a suitable means of comparing new models with existing ones. The PCA of the compounds showed that the external test set occupied a limited drug-like space in comparison to the 85 compounds in the training and test sets (Figure 9a). Furthermore, large homologous series were found within the dataset and, for instance, analogues of barbituric acids and steroids constituted 17% and 14% of the 207 compounds, respectively. The size of the subsets of acids, ampholytes, bases and non-proteolytes also differed largely from the 85 compounds in the training and test sets (Figure 9b).

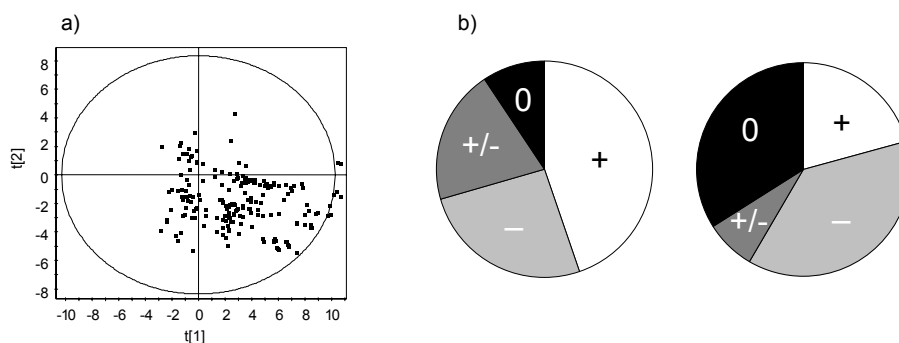


Figure 9. Distribution of external test set applied in Paper III. a) The external test set (207 compounds) occupied a limited volume of the drug-like space defined by the training and test set (85 compounds). b) The distribution of the proteolytic subsets found within the training and test sets (left hand side) and the external test set (right hand side). Bases are shown in white, acids in light gray, ampholytes in gray and non-proteolytes in black.

The two largest subsets of the external test set were non-proteolytes and acids. In contrast, the majority of the dataset used in the model development in Paper III were bases and only a minority were non-proteolytes, a distribution in accordance with that of registered drugs.¹⁹¹ The external test set was therefore regarded as biased to certain therapeutics groups which probably would have been better predicted by a training set including similar structures. However, the literature dataset allows comparison of the general applicability of our solubility models to previously published solubility models, and was therefore used.

6.2. Solubility Measurements *In Vitro*

6.2.1. The Small-Scale Shake Flask Method (SSF)

A new experimental method for high quality measurements of solubility in a medium throughput mode was developed. The method should be applicable early in the drug discovery setting when only small quantities of drugs are available. The commonly used shake flask method was therefore adjusted for these demands (Paper I). The results obtained with the SSF were of equal accuracy to those obtained with the traditionally used large-scale shake flask (Figure 10a) and state-of-the-art potentiometric determinations (Figure 10b). The SSF used ultracentrifugation to separate the remaining solid from the solution after equilibrium had been reached. This method was successful as assessed by light scattering measurements of the supernatant: no colloidal particles could be identified within the supernatant. Solubility measurements of high quality could be performed in 50 μL of solvent, allowing thermodynamic solubility determinations using microgram quantities of the drug (Figure 10c). The study of the

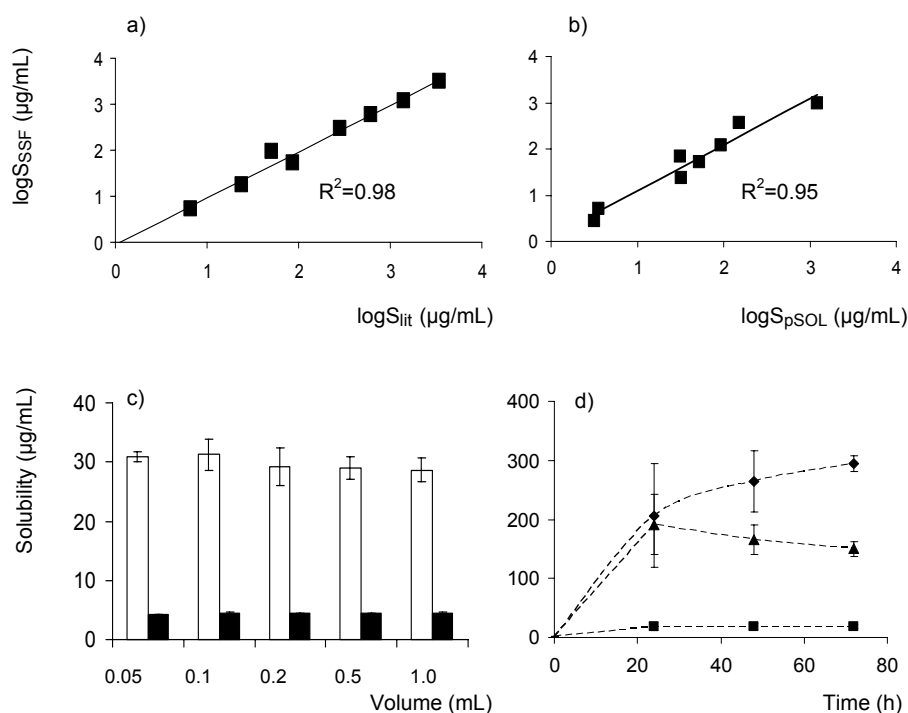


Figure 10. Results of SSF development. a) Correlation between solubility data available in the literature obtained from large-scale shake flask and SSF solubility data. b) Correlation between solubility data obtained by potentiometric titration (pSOL) and SSF solubility data. c) Scaling down the sample volumes showed that the thermodynamic solubility of solids (pindolol in white bars, probenecid in black) can be determined in suspension volumes of 50 μL . d) The time-dependent solubility is illustrated with three examples taken from the study performed in Paper I. The majority of the compounds behaved as exemplified by testosterone (■), and reached their solubility value within 24 h. However, hydrocortisone (◆) approached the equilibrium solubility slowly, as did cimetidine (data not shown). In contrast, the solution of amiloride (▲) first became oversaturated and thereafter reached its thermodynamic solubility value.

time-scale needed for the determination of thermodynamic solubility revealed that the majority of the model drugs had reached their solubility equilibrium within 24 h (Figure 10d). The compounds that had not attained equilibrium within this time differed by a factor of less than 1.5 from the solubility value at equilibrium. These results suggest that when screening for the intrinsic solubility of solid drugs, the time-scale can be set to 24 h if smaller deviations from the thermodynamic solubility value can be tolerated. In conclusion, the method devised allows thermodynamic solubility determinations to be performed for solids in a microtiter plate format. By using this experimental setting, the rate limiting process will be the analysis of sample concentration rather than the solubility experiment *per se*.

6.2.2. Temperature and Buffer Effects

The temperature dependence was analyzed in two steps. First, the effect of ambient temperature on the obtained solubility value was studied in a comparison with a water-bath controlled temperature. The comparison between solubility data obtained with the SSF ($22.5^{\circ}\text{C} \pm 1^{\circ}\text{C}$) and the potentiometric technique (25°C) showed that this minor difference in temperature did not influence the solubility value to a large extent. This is in agreement with the finding that solubility values determined at temperatures of $\pm 2^{\circ}\text{C}$ only differed marginally.⁷⁷ Hence, the simpler temperature setting of the SSF method, i.e. the determination of solubility at room temperature, can replace the more sophisticated water-bath experiments. Second, in Paper II the solubility values obtained at 25°C were compared to solubilities at 37°C (Figure 11). In contrast to what is generally expected, 60% of the compounds determined at both temperatures showed a somewhat lower solubility at 37°C than at room temperature. Indomethacin and verapamil were most strongly affected by the temperature increase, resulting in a 6-fold higher and 9-fold lower solubility at 37°C , respectively. All other compounds had 37°C solubility values that were ± 3 times the solubility value determined at 25°C . Hence, these results suggest that solubility values at 37°C can be approximated from 25°C determinations in early stages of drug development.

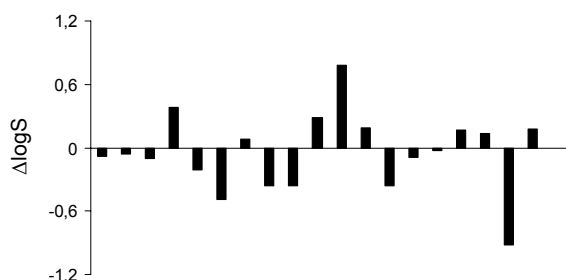


Figure 11. Temperature differences for the compounds in paper II measured at 25°C and 37°C . $\Delta \log S = \log S_{37^{\circ}\text{C}} - \log S_{25^{\circ}\text{C}}$. Solubility values at 37°C were determined for 19 of the 23 compounds investigated.

The effect of the solvent used was studied by performing solubility determinations in water and buffer systems. The intrinsic solubility value obtained in different solvents, i.e., MQ-water (Paper I), 0.15 M KCl buffer (Paper II) and phosphate buffers (Paper V), resulted in slightly different solubilities (Table 2). The majority of the compounds showed a lower intrinsic solubility in the buffer systems than in the pH-adjusted MQ-water. Hence, the higher ionic strength of the buffers in comparison to the MQ-water resulted in salting-out effects of the compounds.²⁵⁻²⁸ The two buffers used affected the

compounds differently, exemplifying the difficulties associated with obtaining a general prediction of salting-in and/or salting-out effects by specific buffers.¹⁹²

Table 2. Intrinsic solubility values given in µg/mL determined in MQ-water, 0.15 M KCl and/or 0.15 M phosphate buffer at room temperature.

Compound	S _{MQ-water} (µg/mL)	S _{0.15 M KCl} (µg/mL)	S _{0.15M Ph buffer} (µg/mL)
SKF105657	0.0007	0.005	
Prazosin	3.2	2.8	
Probenecid	3.6	5.2	
Propranolol	31	70	
Pindolol	33	23	
Ciprofloxacin	54	61	
Ketoprofen	94	120	
Amiloride	150	125	
Acyclovir	1213	1200	
Promethazine	1.6	12	0.6
Verapamil	8.2	9.2	11
Desipramine	96	47	54
Chlorprothixene	0.5		0.5
Dipyridamole	1.0		2.0
Mifepristone	1.3		0.6
Procyclidine	9.3		8.1
Propafenone	14		2.1
Orphenadrine	55		23
Bupivacaine	83		44
Pramoxine	178		97
Disopyramide	373		298
Hydralazine	749		367
Terazosin	2743		6580
Lidocaine	3800		2398
Trimethoprim	3916		3324
Celiprolol	4354		5688

To conclude, the experimental studies performed in Papers I, II and V show that the time-scale, the temperature and the selected solvent will influence the solubility value obtained. A majority of the compounds investigated reached their solubility equilibrium within 24 h. No simple rules-of-thumb were revealed for the contribution of temperature, counter-ions and ionic strengths. The results indicate that in early drug development an approximation of the drug solubility in the intestinal fluid (in the fasted state) can be obtained from aqueous solubility determinations performed at room temperature, provided that the solvent has a physiologically relevant pH. However, if

the purpose of the generation of solubility data is to establish an experimental database for model development, it is important to standardize the experimental setting to reduce experimental variation.

6.2.3. pH-dependent Solubility

In Paper V, the HH relationship and its application to solubility predictions was investigated. The pH-dependent solubility was found to be substance specific, with variations in the solubility range of 1.1 to 6.3 log units between the uncharged and the completely charged species (Figure 12a). Thus, this study did not support the generalization that cationic drugs display a solubility range of three log units.²³ On the contrary, as many as 76% of the compounds showed a solubility range outside 3 ± 0.5 log units, consistent with a substance specific response to the counter-ion of the buffer. The counter-ion mediated effect on the pH-dependent solubility has also been reported by others.^{193,194} Moreover, the considerable variation in the slope of the linear part of the pH versus solubility curve (-0.5 to -8.6 ; Figure 12b) revealed a specific response of each of the compounds to the pH variation and the buffer system. Slopes of less than -1 can occur due to self-association of the drug molecules into larger aggregates such as micelles.¹⁹⁵⁻¹⁹⁷ However, no colloidal particles were identified in the supernatants and, hence, values of less than -1 were not caused by formation of micelles. Instead the negative deviations in slope were probably related to the formation of low molecular weight aggregates such as dimers or oligomers. The same result would be obtained by salting-in effects caused by the electrolytes in the buffer.^{30,31} This scenario is less likely, since the buffer was shown to predominantly exert salting-out effects (Table 2). Slopes larger than -1 can be attributed to the formation of salts with electrolytes available in the buffer.¹⁹⁸ Such salts would have a stoichiometry different from 1:1, e.g., of 2:1 drug molecules to counter-ions. However, the solid state characterization of the remaining solid after the experiment did not confirm such formation at pH-values corresponding to the linear part of the pH-dependent solubility curve. Hence, the positive deviations in slope might be the result of salting-out effects of the buffer.^{27,28,30}

In a drug discovery setting it is common to determine the aqueous solubility at a single pH-value, e.g. pH 7.4 or at the pH-value giving the intrinsic solubility. Thereafter, the solubility at other pH-values is calculated using the measured solubility value and the HH equation (Eq. 2). The results obtained in Paper V clearly show that such calculated solubility values can differ largely from the experimental observations. This is in agreement with a recent study performed by Hendriksen and coworkers, who showed that at least two solubility measurements have to be performed to predict the pH-dependent solubility.¹⁹⁹ Moreover, it is important to consider the impact of the inclusion of HH calculated solubilities in computer-based model development. If such values are included in the training set, models based on incorrect solubility values are obtained.

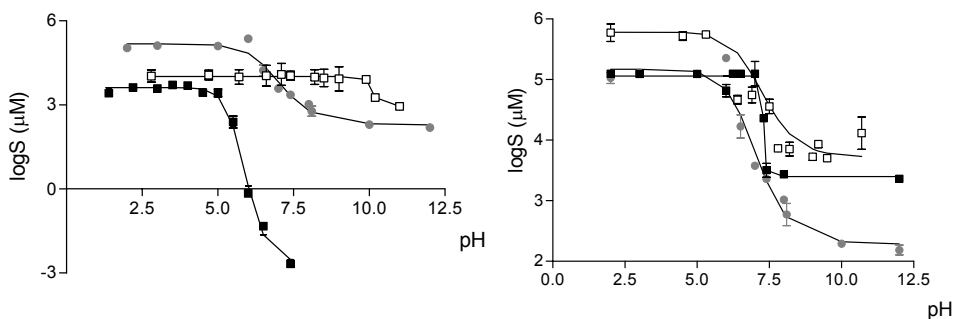


Figure 12. pH-dependent aqueous drug solubility. a) Variation in solubility range shown by disopyramide (□), which displayed the smallest range (1.1 log units) and amiodarone (■), which displayed the largest range (6.3 log units) in comparison to bupivacaine (●), a compound with a solubility range of 3 log units. b) The variation in slope in the solubility profiles observed for celiprolol (□), which had the smallest slope (-0.5 log units) and hydralazine (■), which displayed the largest slope (-8.6 log units) in comparison to bupivacaine (●), one of the compounds with a slope of -1.

6.3. Permeability Measurements *In Vitro*

The passive, transcellular transport of a series of 23 compounds through Caco-2 cell monolayers was determined (Paper II). It has been suggested that four of the compounds, i.e. erythromycin,^{200,201} verapamil,²⁰² folic acid and methotrexate,²⁰³ are actively transported. However, to enable studies on a structurally diverse dataset, also compounds representing structural features more often associated with active rather than passive transport, e.g., larger size and polarity,^{204,205} were included.

Most compounds had a less than twofold difference in permeability values determined in the apical to basolateral (a–b) direction compared to the basolateral to apical direction (b–a) and were therefore considered to be mainly passively transported. Ciprofloxacin had a 3.7-fold difference in permeability in the b–a direction compared to the a–b direction. However, a previous study in our laboratory showed that the transport rate of ciprofloxacin in the absorptive (a–b) direction is concentration-independent.¹¹³ Moreover, erythromycin was found to be actively secreted at the applied concentration (1.2 mM), probably by a mechanism mediated by an ABC transporter such as P-glycoprotein.^{200,201} The passive permeability coefficient of this compound was obtained after inhibition of the secretion by addition of verapamil.

6.4. Solubility Predictions *In Silico*

6.4.1. Global Solubility Models

The experimental solubility data obtained in Paper I were used to investigate the accuracy of commonly used models based on easily comprehended physicochemical descriptors for the prediction of aqueous solubility.^{122,133,144} These models have been devised from non drug-like datasets and it was therefore not surprising that they resulted in only poor solubility predictions for the series of drugs studied. The best prediction of the four models tested was obtained from ClogP and the melting point,¹⁴⁴ which resulted in an R^2 of 0.67 and RMSE of 1.27 log units for the tested compounds. The statistics improved when the drug-like dataset was used to train the model, resulting in an R^2 of 0.85 and RMSE_{te} of 0.66 log units. These results show that non drug-like compounds poorly predict aqueous drug solubilities. Moreover, the test of a software available at the time of the investigation clearly indicated that further development of computational models for the prediction of aqueous drug solubility was required.²⁰⁶

Table 3. Final *in silico* models obtained.

Paper	Response	n _{tr}	n _{te}	n _{ext te}	R ² , RMSE _{tr}	RMSE _{te}	RMSE _{ext te}	ClogP	2D	3D
I	logS	12	5		0.91, 0.61	0.90		x		x
II	logS	14	6	31	0.93, 0.37	0.76	1.05			x
II	logP _{app}	13	9	26	0.93, 0.35	0.99	0.85			x
III	logS	56	29	207	0.80, 0.90	0.83	0.82	x	x	x
IV	mp	185	92		0.63, 35.1	44.6		x	x	x

The response parameters studied were the solubility (logS), permeability through Caco-2 cell monolayers (logP_{app}) and the melting point (mp). n denotes the number of compounds, tr, te and ext te denote the training, test and external test sets, respectively. The crossed descriptor boxes show the type of descriptors included in the final model. In Paper III subset specific models and models for homologous series were also devised. RMSE-values are given in log M (solubility), log cm/s (permeability) and °C (melting point).

Molecular surface descriptors obtained by Monte Carlo simulations have recently been found to be good predictors of drug solubility.¹⁶⁰ In order to evaluate the use of PSA, NPSA and PTSAs in solubility predictions of drug-like molecules, PLS models based on these descriptors were generated by the use of four different datasets (Table 3). In Paper I, the approach of adopting surface area descriptors as predictors of solubility was tested for the first time. The initial input matrix used in the PLS model development contained PTSAs and composite surface areas such as PSA and NPSA. In addition, ClogP and descriptors for hydrogen bond donors and acceptors, which previously have proven to be important for predictions of solubility,^{70,133} were included. The best

computational model was obtained from ClogP, PTSAs and composite surface areas. The PLS analysis showed that properties that are negatively correlated with solubility, such as size, lipophilicity and surface areas of non-polar atoms were the most important for the solubility predictions of this dataset. Only one hydrogen bond descriptor (surface area of hydrogen bound to nitrogen atoms) remained after the descriptor selection. The predictivity of the surface areas was tested with another dataset in Paper II. The results from this study confirmed the results from Paper I, since a highly accurate theoretical solubility model was developed ($R^2=0.93$, $RMSE_{tr}=0.37$ log units). For this dataset too, the predominant descriptors selected by the PLS analysis were those restricting solubility, i.e., NPSA, SA and PTSAs related to hydrophobic atoms. Only one descriptor for hydrogen bonding was selected, the surface area of double-bonded oxygen, which correlated positively with aqueous drug solubility. The findings in these two papers show that PTSAs, NPSA, and SA descriptors are new potential predictors of solubility (Figure 13). Moreover, the results indicate that the surface areas of non-polar atoms are general molecular descriptors for aqueous drug solubility, i.e., the selection of these variables does not seem to be dataset dependent.

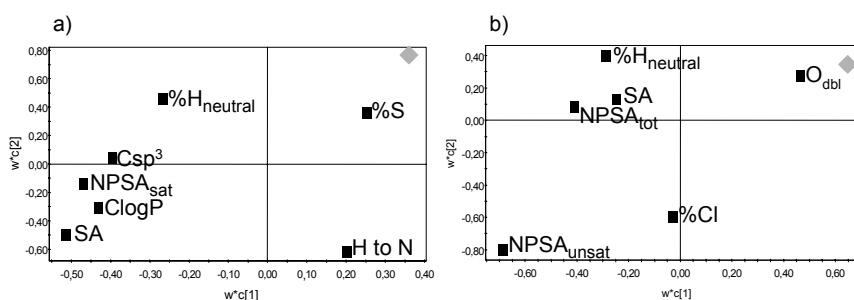


Figure 13. Loading plots for the PLS solubility models. The loadings of the first ($w*c[1]$) and second ($w*c[2]$) principal components extracted in the PLS analysis of solubility (\blacklozenge), identifying the most important descriptors for solubility of the datasets studied in a) Paper I and b) Paper II.

To assess the general applicability of the surface area descriptors, a larger dataset was investigated in Paper III. Here, a training set of 56 compounds was used, and the model obtained was validated with two test sets. The model had an intermediate predictive power, with RMSEs ranging from 1.03 to 1.06 log units for the training, test and external test sets. Also for this dataset, ClogP, size and hydrophobic surface area descriptors proved to be important. Moreover, PTSAs of chloride and sulfur atoms, which were important predictors of solubility in Papers I and II, were excluded prior to the PLS analysis since these descriptors showed a high skewness. Hence, in this study it became evident that the previous selection of descriptors obtained for specific heteroatoms was largely dataset dependent.

As several publications show that Molconn-Z¹⁸⁴ generated descriptors are successful in the prediction of solubility,^{121,123,125-127,162-168} the accuracy of the surface area descriptors was compared to the accuracy of models based on these descriptors (Paper III). In addition, the program SELMA¹⁸⁵ was used to generate 2D related descriptors. Indeed, the 2D descriptors obtained from Molconn-Z and SELMA proved to be strong predictors of aqueous drug solubility. The resulting model had a good predictive power, giving values of R^2 and Q^2 of 0.75 and 0.68, respectively, in comparison to the values of 0.57 and 0.53, respectively, obtained with the surface area descriptors. Lipophilicity was identified as an important descriptor also in this model. Furthermore, descriptors for flexibility, connectivity indices, electron energy and ionization were important variables in the model. The RMSEs obtained for the training set and the two test sets were of the same order of magnitude as the model derived from molecular surface areas ($RMSE_{tr}$ and $RMSE_{ext\ te}$ of 0.92 to 1.01 log units, respectively).

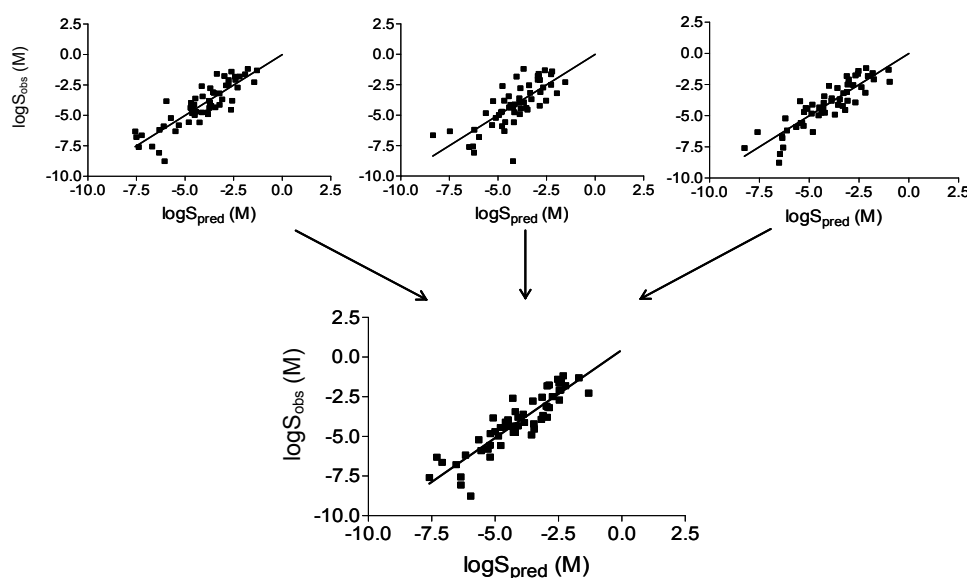


Figure 14. Averaged consensus model for prediction of aqueous drug solubility. The global consensus model (bottom panel) trained on 56 compounds was based on three models (from the left hand side): a solubility model built from 2D descriptors; a solubility model built from surface areas obtained from the 3D representation; and a solubility model derived from a matrix of both 2D and 3D descriptors.

The result of the model development in Paper III indicated that the different descriptor spaces (2D and 3D) described different molecular properties. Thus, the use of both types of descriptors in the model development could be expected to result in improved models. This was studied in two ways: first, both 2D and 3D descriptors were used

together in the variable selection of the model development, a strategy that marginally improved the model in comparison to the model established from 2D descriptors alone. Second, averaged consensus modeling was applied, allowing the three derived models to influence the prediction. This approach resulted in the best global model devised. It was more robust with predictions of higher accuracy and a reduced number of outliers in comparison to each of the individual models (Figure 14).

The assessment of the general application of the global model was performed using a drug-like test set compiled from the literature.^{160,166} The model predicted the dataset well (RMSE 0.82), although it was revealed that it failed to predict isomeric compounds (Figure 15). The model obtained in this work unfortunately predicted these compounds to display equal solubilities. To obtain solubility models that can handle the prediction of closely related structures, two approaches need to be taken. Firstly, the model has to be trained on isomeric structures, so that electrotopological variables describing the change in the electronic environment for these analogues can influence the multivariate data analysis. This is supported by the impressive results of models obtained from datasets containing isomeric structures in the training set.^{160,163,166,189,190} Secondly, since the influence of the position of substituents on the solubility could be linked to the solid state properties rather than to the solvation process,^{130,144} inclusion of solid state descriptors may result in better solubility models.

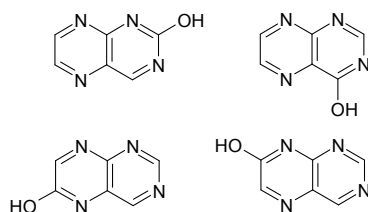


Figure 15. Examples of isomeric compounds included in the external test set in Paper III. 2-, 4-, 6-, and 7- hydroxypteridines are shown.

6.4.2. Subset Specific Solubility Models

The aim of the subset specific models presented in Paper III was to investigate whether the improved solubility predictions obtained with the consensus model were a result of the 2D and 3D descriptors having different preferences for the proteolytic groups of the molecules. If so, solubility of a heterologous dataset would be best predicted when both types of descriptors are included. To investigate this, the dataset was divided into subsets based on the proteolytic function of the compound. The subset of ampholytes was predicted by the 2D descriptors with good accuracy ($R^2=0.80$). This was probably a

result of the selection of descriptors of positive and negative charge reflecting the proteolytic functions of the molecule. Both 2D and 3D descriptors were successful in the prediction of bases (R^2 of 0.80 and 0.83, respectively). The ease of which this subset was predicted from either of the two matrices could be attributed to the low variation in the proteolytic function of the compounds. None of the matrices succeeded in the prediction of acids. This discouraging result may be an effect of the definition of acids applied in the study: functional groups which have the ability to lose a proton within the pH-interval of 2-12 were regarded as acidic functions. Therefore, weak acids such as phenols were included in this subset.

To conclude, both global and subset specific solubility models have been generated. The present studies clearly showed that both 2D and 3D descriptors are valid predictors of solubility. By studying models for proteolytic subsets, it was demonstrated that ampholytes and bases were well predicted by the 2D and 3D descriptors investigated. Consensus modeling resulted in a global model with good predictive accuracy and few outliers. This model was based on rather easily comprehended descriptors related to hydrophobicity, hydrophilicity, flexibility, size, connectivity and electron energy. Moreover, the consensus model derived is relatively transparent, giving the user feedback on the impact structural features have on solubility.

6.4.3. Computational Protocol for Rapid Calculation of Molecular Surface Areas

The computer-based conformational analysis can either be thorough, in order to find all low energy conformers, or rapid, to obtain one conformer that represents the low energy population fairly well. Furthermore, the conformational search can be performed using different solvent models. If the permeability is to be modeled, vacuum usually is chosen as the “solvent”. This is a hydrophobic environment and, hence, is assumed to mimic the cell membrane. For aqueous solubility models, the descriptors are calculated from conformers generated in water. However, the results presented in Paper II showed that surface areas calculated from conformers obtained in vacuum and water were highly correlated. This is in agreement with the recent findings by Stenberg and coworkers.¹¹³ They found that the PSA generated from water and vacuum conformers were highly correlated, suggesting that the conformational search in water can be replaced with the faster search performed in vacuum. Compounds with high flexibility are exceptions to this rule.¹³⁴

The surface area descriptors can be calculated either as dynamic or static surface areas. The use of a static surface area, calculated on one conformer would result in faster calculations than the generation of Boltzmann weighted dynamic surface areas

calculated from all low energy conformers. Indeed, analysis of the accuracy of carefully calculated dynamic molecular surface areas and the more rapidly calculated surface areas showed that these descriptors produced models of equal accuracy (Papers I and II). The static PSA has previously been used as a descriptor of permeability with good results.^{172,207,208} Hence, it could be concluded that fast protocols for the generation of surface areas can be applied both for permeability and solubility predictions, making surface area-based absorption models more attractive for drug discovery settings.

6.4.4. Molecular Surface Areas as Descriptors of the Solvation Process

In none of the solubility models obtained in Papers I-III was the PSA an important descriptor for solubility. Thus, by combining the results from the solubility studies with the results from the theoretical analysis of the solid state characteristic melting point (Paper IV), the following generalization can be forwarded: the solubility of non-polar compounds is not restricted by the formation of strong crystal structures, but rather, by the energy required for the solvation process. This can be described from the solvation theory:^{19,20} if the non-polar surface area of a molecule, and therefore its hydrophobicity, increases, the dissolution capacity of the compound in an aqueous environment will decrease. The influence of hydrogen bond descriptors on solubility seems to be more dataset dependent in comparison to hydrophobic descriptors. This might be due to the fact that hydrogen bond descriptors can affect drug solubility in two different ways: some polar atoms form energetically favorable hydrogen bonds with water, while others participate in strong intermolecular interactions within the crystal lattice. The same type of polar atom may either promote or prevent aqueous drug solubility, depending on its position in the 3D molecular structure of the compound. Poor solubility of compounds with large PSAs will mainly be an effect of the formation of strong crystal structures. If the stabilization provided by the crystal bonds is beneficial to the molecule the melting point will increase. As a result, the aqueous solubility will decrease, an effect described by the van't Hoff equation:

$$\ln a = -\frac{\Delta S_{f,T_m}}{R} \left(\frac{T_m}{T} - 1 \right) \quad \text{Eq. 8}$$

where a is the activity, which corresponds to the concentration in an ideal solution, $\Delta S_{f,T_m}$ is the entropy of fusion, R is the gas constant, T_m the melting point and T the absolute temperature. Finally, the selection of the size descriptor reflects the energy required to form cavities in the water with a sufficiently large volume to accommodate the molecules.^{18,20,145}

6.5 Computational Solid State Characterization

Before the work on the solid state prediction started, a search was performed on the Internet for publications on similar investigations. Not a single publication could be found on melting point prediction of drugs, even though crystal lattice energies have recently been studied by use of molecular modeling.^{209,210} This highlights the need for development of theoretical tools for solid state prediction. Thus, a theoretical analysis of factors influencing the melting point was performed with the intention of obtaining calculated solid state characteristics which could be included in computational predictions of solubility (Paper IV).^{144,146,155,211,212} The derived model was based on 2D descriptors representing molecular flexibility, polarity and electron distribution (Figure 15). The model allowed a classification of the compounds to be made, with excellent separation being found between the “low” and “high” melting point classes. Unfortunately, the model obtained could not quantitatively predict the melting point.

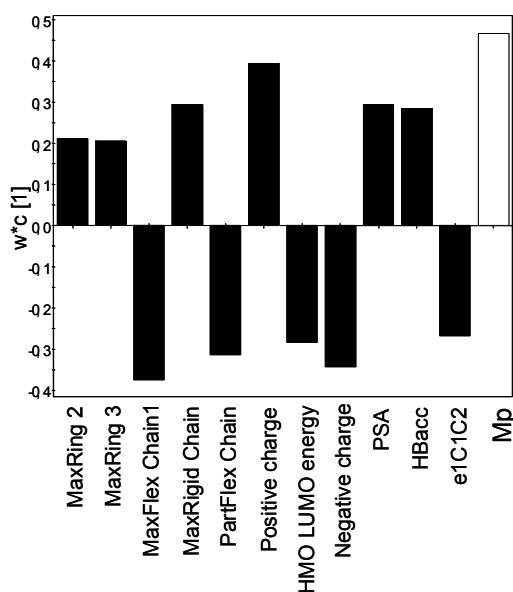


Figure 15. Loading plot for the melting point model. The relationship between the descriptors (black) and the response parameter (white) is shown. The most important 2D descriptors were: the size of the second and third largest ring (MaxRing 2 and MaxRing 3); the length of the longest chain with only rotatable bonds (MaxFlex Chain1); the lengths of the three longest chains containing only rigid bonds (MaxRigid Chain); the length of the longest partially flexible chain (PartFlex Chain); the averaged positive and negative partial atom charge calculated from all charged atoms within the molecule; the energy of the lowest unoccupied molecular orbital (HMO LUMO energy); the PSA; the number of hydrogen bond acceptors (HBacc); and the electron accessibility between sp^2 - and sp^3 -hybridized carbon atoms (e1C1C2).

6.6. Permeability Prediction *In Silico*

In Paper II, a theoretical permeability model based on composite surface areas and PTSAs was generated from the experimentally determined P_{app} values in Caco-2 cell monolayers (Figure 16). Polar surface area descriptors have previously shown a negative correlation to permeability,^{111,113,128,172} and this was confirmed by the model obtained in this study. The negative correlation between the polar descriptors and permeability was interpreted as a result of the increase in desolvation energy required for molecules to enter the hydrophobic interior of the membrane from the aqueous surroundings.^{213,214} For the first time, the SA descriptor was included in a permeability model based on molecular surface areas. This might be a result of the large range in size (180-734 in molecular weight) of the dataset, which give size as a descriptor enough weight to influence the model. The SA was found to be negatively correlated to the permeability, which was interpreted as an effect of the steric hindrance to diffusion across the cell membrane caused by the ordered membrane structure.^{213,214} Therefore, the larger the molecule, the greater the steric hindrance to diffusion through the cell membrane. The accuracy of the permeability model obtained ($R^2=0.93$, $RMSE_{tr}=0.35$ log units) was comparable to that of a previous in-house model, generated from a different dataset.¹¹³

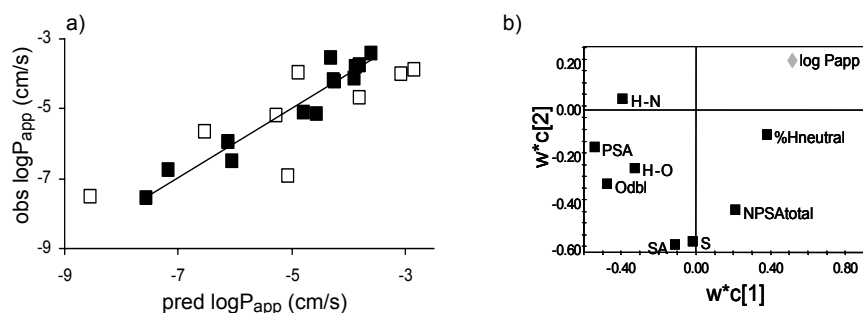


Figure 16. The permeability model based on molecular surface area descriptors. a) The model was derived from a training set (■) of 14 compounds and validated with a test set (□) of eight compounds. b) Loading plot of the permeability model showing the interrelationship between the PTSAs and Caco-2 permeability. The following surface areas were important descriptors for the permeability (in order of importance): the PSA, double-bonded oxygen (O_{dbl}), hydrogen atoms bound to nitrogen (H-N), the fraction of surface areas occupied by neutral hydrogen atoms (% $H_{neutral}$), hydrogen atoms bound to oxygen (H-O), $NPSA_{total}$, sulfur atoms (S) and the SA.

6.7. Biopharmaceutical Classification

The biopharmaceutics classification system (BCS) was originally introduced by the Food and Drug Administration (FDA) to simplify the production of generic drugs.¹⁸¹ The FDA suggests that compounds with a high solubility in comparison to the given dose and a high permeability (Class I compounds, see Section 5.7) do not need to be studied for bioequivalence after minor changes in the formulation. Various BCSs have previously been applied as qualitative screening tools for drug absorption in drug discovery and development.^{115,181,215} In Paper II, a BCS containing six categories was used, according to which solubility was classified as “low” or “high”¹⁸¹ and permeability was classified as “low”, “intermediate”, or “high”.¹¹⁵ This classification was chosen because it provides a better tool for absorption ranking of compounds in drug discovery than does the stricter classification provided by the FDA.¹⁸¹ A computer-based BCS that predict absorption characteristic with high accuracy would sort compounds in accordance with their developability (absorption-wise) prior to synthesis. Such virtual tools applied in early drug discovery would result in a decreased number of CDs with formulation problems. Only one approach to theoretical biopharmaceutical classification has previously been published, defining the classes in terms of molecular weight and PSA.²¹⁶

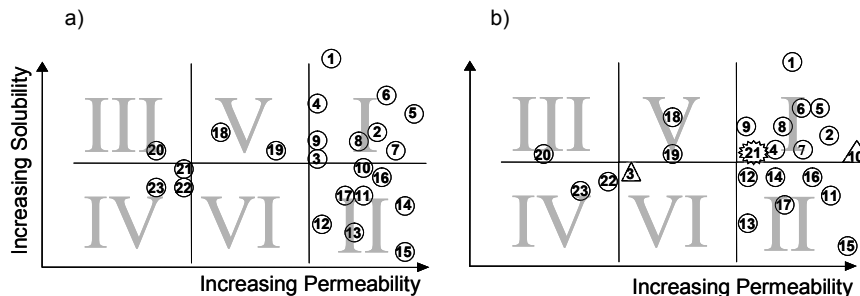


Figure 17. Comparison of experimental and theoretical biopharmaceutical classification of 23 drugs. The six classes are marked in light grey and the compounds are numbered as in Table 4. Relative scales are used since solubility is dose-dependent and, therefore, compound-specific. a) Experimental determination of BCS class. The compounds were mainly distributed in classes I (39%) and II (35%). b) Theoretical prediction of the biopharmaceutical classes. The BCS classes were predicted with a success rate of 87%. Deviations into an adjacent class are shown by triangles. Acyclovir (compound 21) was falsely predicted from Class IV to Class I.

The experimentally determined, dose-adjusted solubility values showed that 44% of the compounds had poor solubility characteristics, indicating that the maximum oral dose would not be completely soluble. The compounds were found to be distributed between

all three permeability classes, but only 17% of the compounds displayed low permeability. Thus, the compounds studied were better optimized for permeability than for solubility. This is in agreement with recent publications, which conclude that the selection of candidate drugs is biased towards molecular properties giving good permeability as opposed to good solubility characteristics.^{69,70} The combination of experimentally determined solubility and permeability data showed that the compounds were distributed into five out of the six biopharmaceutical classes (Figure 17a). The experimental study identified nine new BCS Class I compounds (Table 4).

The experimental classification identified 14 compounds with poor solubility and/or permeability (BCS Classes II–VI), 12 of which were correctly predicted using the *in silico* models. In comparison, the Lipinski rule-of-five⁷⁰ predicted that only four of the 14 compounds would exhibit poor solubility and/or permeability. The surface area based models resulted in 87% of the compounds being correctly sorted into their respective class, i.e. only three compounds (acyclovir, amitriptyline and doxycycline) were falsely predicted (Figure 17b).

Table 4. Experimental determination of BCS classes according to the FDA guidelines.¹⁸¹

BCS Class I	BCS Class II	BCS Class III	BCS Class IV
1. Atropine	10. Amitriptyline	18. Amiloride	21. Acyclovir
2. Desipramine	11. Chlorpromazine	19. Erythromycin	22. Amoxicillin
3. Doxycycline	12. Ciprofloxacin	20. Folinic acid	23. Methotrexate
4. Ergonovine	13. Indomethacin		
5. Ethinyl estradiol	14. Phenazopyridine		
6. Primaquine	15. Tamoxifen		
7. Promethazine	16. Verapamil		
8. Theophylline	17. Warfarin		
9. Zidovudine			

The guidelines given by FDA sort the compounds into the following classes: I. high solubility - high permeability; II. low solubility - high permeability; III. high solubility - low permeability; IV. low solubility - low permeability. Solubility is defined as high if the maximum dose given orally is dissolved in 250 mL of fluid in the pH-interval 1-7.5. Permeability is defined as high if 90% or more of the dose is absorbed, otherwise it is low.¹⁸¹

To further evaluate the usefulness of the combinations of *in silico* models in biopharmaceutical classification, the recommended set of reference drugs listed in the FDA guidelines was used as an external test set for the model.¹⁸¹ The theoretical classification correctly predicted 77% of the external test set. None of the compounds in the external test set was falsely predicted with regard to both the solubility and the permeability characteristics. To summarize, the biopharmaceutical model developed in

Paper II has so far been tested on 36 compounds and the results indicate that more sophisticated *in silico* models combining computational analysis of solubility and permeability can successfully predict the absorption process. If these and other models of drug absorption¹²⁰ are used in the drug discovery process, the resulting CDs will probably show improved developability.

7. CONCLUSIONS

This thesis presents experimental and computational models of solubility and permeability. The devised experimental protocol for solubility determinations is a medium throughput method and does not demand the involvement of expensive or specialized equipment. Moreover, the system is flexible, allowing the determination of solubility in accordance with the demands specified by the user. The investigations of *in silico* solubility predictions showed that molecular surface area descriptors are potential new descriptors for solubility estimations. The descriptors are rapidly generated in the computer and easy to interpret, leading to relatively transparent models for the user. In studies on the prediction of the absorption class of the compounds, a computational protocol was devised that is suitable for the drug discovery process. The combination of computational models for the two main properties influencing drug absorption, i.e., the solubility and permeability, successfully sorted the drug molecules into the correct absorption class. These results suggest that accurate and rapid *in silico* screening of drug absorption will be further implemented in the drug discovery process within the near future.

The specific conclusions of this thesis are:

- ✓ The shake flask method has been modified to produce reliable solubility measurements of crystalline compounds in a microtiter plate format and a medium throughput mode. This format makes it possible to perform high quality solubility determinations when only a limited amount of crystalline compound is available.
- ✓ Evaluation of existing theoretical solubility models showed that these approaches failed to predict the solubility of registered drugs. In order to avoid incorrect predictions of aqueous drug solubility, experimental data of high quality and drug-like training sets should be used in the model development.
- ✓ For the first time, surface area descriptors such as NPSA, PSA and PTSA were used to generate predictive solubility models. In general, the surface areas related to non-polar atoms and size, i.e. descriptors restricting solubility, were important solubility predictors. Furthermore, the studies presented showed that the combination of surface area descriptors and descriptors for electron distribution, connectivity indices and flexibility results in generally applicable solubility models.
- ✓ An analysis of the melting point, a solid state descriptor commonly used in solubility predictions, showed that the descriptors restricting solubility in the devised *in silico* solubility models were related to the solvation process of the compounds rather than to the solid state.

- ✓ The determination of pH-dependent solubility profiles of a series of amines showed that calculations based on the HH equation can deviate from the experimentally determined values by a factor of 100. Large variations in the slope of the linear part of the pH versus solubility curve and in the solubility range were observed within the series studied. The results demonstrate the limitations of extrapolating experimentally measured solubility values obtained at one pH-value to the solubility at another.

- ✓ The validity of PTSAs in the prediction of intestinal drug permeability was confirmed. Moreover, when the solubility and permeability models were used in concert, the absorption class of the compounds could be predicted with high accuracy. This result indicates that *in silico* biopharmaceutical classification can be performed prior to the synthesis of compounds, which may increase the number of high quality leads that result in marketed drugs.

8. PERSPECTIVES

The theoretical models obtained in this thesis and models published by others have an accuracy of 0.5-1 log units for the prediction of solubility and permeability. The models can be used as qualitative computational filters, i.e. providing yes and no answers, or sorting compounds into classes denoted “poor”, “intermediate” or “good”. Some models may also be regarded as quantitative filters, providing that the deviations mentioned are acceptable. However, many different approaches need to be adopted to improve the success of the predictions of these response parameters. Firstly, the computational models have to be developed using drug-like compounds if they are to predict such molecules. Moreover, the experimental databases used for model development need to be increased, so that a larger volume and higher density of the drug-like space is represented by the training sets. The setting for the experiments should be standardized to not restrict the model development by variability in the data. Secondly, the incorporation of computationally derived descriptors reflecting the solid state would result in better solubility predictions e.g. for isomeric structures. Thirdly, the models should be formatted so that constructive feedback can be given to the medicinal chemists on how to design better molecules. Since several properties are modeled (solubility, permeability, metabolism, toxicity, etc.), all of which are important for the developability and success of the compound, it is difficult to draw straightforward conclusions on how to improve the drug design process based on the predictions. However, the adaptation of chemo- and bioinformatics to pharmaceutical screening may result in computational models of drug developability that are highly accurate and user friendly. In conclusion, the generation of extended experimental drug-like databases can be regarded as the rate-limiting step in future model development, since the predictions will not improve further until larger volumes of high quality experimental data are treated. If such databases are produced, the prerequisites for quantitative rather than qualitative ADMET screening are given.

9. POPULÄRVETENSKAPLIG SAMMANFATTNING

Under det senaste årtiondet har nya tekniker gjort det möjligt att identifiera och producera hundratusentals nya presumtiva läkemedelssubstanser varje år. Gemensamt för dessa potentiella läkemedel är att de har en farmakologisk effekt, d v s de kan hämma, lindra eller bota sjukdomar och/eller sjukdomssymptom. Tyvärr visar studier att många substanser tas upp dåligt från tarmen till blodet. Sådana läkemedel kan inte tillverkas som tabletter eller kapslar, vilket annars är de beredningsformer som vanligen uppskattas mest av patienten. För att kunna förutsäga huruvida en potentiell läkemedelssubstans kommer att absorberas eller inte krävs nya beräkningsmetoder, samt kapacitet att behandla stora mängder data.

I avhandlingsarbetet presenteras teoretiska modeller för att studera både läkemedels löslighet i tarmsaft och läkemedels transport över tarmväggen, egenskaper som är avgörande för hur mycket av ett läkemedel som tas upp av kroppen. Syftet med dessa studier har varit att utveckla nya, snabba och billiga metoder för att sälla ut de substanser som absorberas bra i tarmen. Därmed väljs bara de bäst lämpade läkemedelskandidaterna ut för vidare studier. Inom ramen för avhandlingsarbetet har datorbaserade modeller för att beräkna lösligheten av läkemedel och dess transport över tarmväggen utvecklats, liksom en ny metod för att experimentellt kunna bestämma lösligheten av läkemedel. De datorbaserade modellerna är tänkta att användas för att rangordna substanserna efter deras absorptionsegenskaper då läkemedlet fortfarande bara finns som idé och inte som färdig substans. Efter denna rangordning kan de substanser som uppvisar lovande absorptionsegenskaper framställas kemiskt. Den utvecklade experimentella löslighetsmetoden kan därefter användas för att avgöra hur löslig substansen är i verkligheten. För att uppskatta transporten av substansen över tarmväggen kan en metod baserad på tarmceller användas.

De experimentellt bestämda löslighetsdata och resultaten från tarmcellsexperimenten har använts som utgångspunkt för de teoretiska löslighetsmodellerna och transportmodellerna. De teoretiska modellerna baseras på hur läkemedelsmolekylen ser ut. Beräkningsprogram av olika komplexitet har använts för att beräkna egenskaper som hur stor och hur fettlöslig substansen är. Dessa egenskaper har använts i modellbyggandet för att förutsäga löslighet och upptag, och ett urval av de viktigaste egenskaperna har gjort m h a statistiska metoder.

De erhållna resultaten visar att det är mycket viktigt att basera sina modeller på experimentella data av hög kvalitet, eftersom de teoretiska modellerna aldrig blir bättre än de experimentella data som de grundar sig på. Den nya experimentella metoden för löslighetsbestämning kan användas för att skapa en sådan experimentell databas, då den ger mycket pålitliga löslighetsvärden. Metoden är lämplig att användas tidigt i

utvecklingsfasen av nya läkemedel eftersom den möjliggör experimentell löslighetsbestämning även för en substans som endast tillverkats i några få milligram. Vidare är det viktigt att många olika typer av substanser ingår vid utvecklingen av teoretiska modeller så att dessa blir så generella som möjligt, d v s kan användas för att förutsäga absorptionsegenskaperna hos vilken ny substans som helst. Slutligen visar studierna att teoretiskt beräknade egenskaper som är länkade till ytan på molekylerna kan användas för att förutsäga både lösligheten och upptaget. Datormodeller baserade på dessa ytegenskaper kunde användas för att identifiera substanser med absorptionsproblem. Dessa resultat visar alltså att två av de viktigaste egenskaperna som styr upptaget av läkemedel från tarmen kan förutsägas med enkla datorbaserade modeller. De modeller som presenteras i avhandlingen kan resultera i att nya läkemedel utvecklas effektivare och till lägre kostnad, vilket leder till billigare läkemedel för patienten.

10. ACKNOWLEDGEMENTS

The studies in this thesis were carried out at the Department of Pharmacy, Faculty of Pharmacy, Uppsala University.

The financial support from the Swedish Foundation for Strategic Research, the Knut and Alice Wallenberg Foundation, Grant No. 9478 from the Swedish Research Council, the Swedish Fund for Research without Animal Experiments, GlaxoSmithKline, CD Carlssons Stiftelse and IF Stiftelse is gratefully acknowledged.

I wish to express my sincere gratitude to

My supervisors; **Professor Per Artursson**, for being brave enough to enter the solubility field and for always aiming at research and papers of highest possible quality. Also a big thanks for your positive attitude during my pregnancy and maternity leave! **Professor Kristina Luthman**, for your eagerness to teach me medicinal chemistry, your ability to see the (sometimes small...) positive things and your commitment during the intensive writing periods and the finalization of my thesis.

Professor Göran Alderborn and **Professor Christer Nyström** for providing good working facilities.

My co-authors; **Dr Ulf Norinder** for your magic treatment of the datasets, for always being so humble and for your peculiar sense of humor; **Dr Lucia Lazorova** for performing the best Caco-2 studies in the world (no problem to build good computational models with such great data!), your fun jokes and big heart; **Melissa Strafford** for the help with pK_a and solubility determinations and for your rapid answers to all potentiometric related questions; **Dr Alex Avdeef**, "Mr potentiometry", for being the only one in the world (except for me) who understands how great fun solubility is, for being such a happy person and for showing great hospitality during my stay in Boston; **Carola Wassvik** for being the bravest person I know – combining solubility *and* the solid state (!), and for understanding the importance with laughing loudly. Together we have learnt a lot that does *not* work...

Johan Gråsjö; for being such an enthusiastic mathematician, statistician, skater and hiker. For your support during my time in the "computer group" and for all fun times when we have been "galeniker" in the tablet lab, preparing samples in the last minute...

Past and present members of the "prediction group"; **Katrin Palm** and **Patric Stenberg** for setting a (too) high standard of the surface area based models. **Carola Wassvik** for solubility discussions and for arranging a longed-for UNIX course; **Pär Matsson** for finally bringing some computational knowledge into the prediction group! Also thanks to Carola and Pär for the insightful comments on the thesis.

My students, **Andreas Lundquist, Olle Lagerlund, Håkan Berndt, Emma Boström, Per Wessman, Helena Hansson, Gustav Ahlin** and **Sara Lindberg** for contributing to my studies and for fun times (at least I thought so...) in the lab.

Eva Nises Ahlgren, Christin Magnusson, Ulla Wästberg-Galik, Eva Lide, Harriet Pettersson, Aysegul Magnusson, Göran Ocklind and **Leif Dahlberg** for your help with practical matters.

My room-mates; **Anders Lindahl** for always having a new saying or a good fishing story to tell; **Helene Hägerström**, my short-term room-mate who made me understand that it is possible to generate PLS models with Q^2 of 0.98; **Björn Jansson**, my *rumbo* for putting up with me and my laugh for almost five years! Sometimes I wonder if it is you or I who has a speciality in computers?! Thanks for being a true friend and for not (always) laughing at all my private projects...

All **past and present colleagues** at the department for your support, for being such great friends and for all fun times; borrel, wc, parties, travels, skiing and competitions of all kinds.

Kulturföreningen Dragskåpet, Östgöta Nationskapell and **all my friends** for contributing to that I enjoyed my stay in Uppsala as an undergraduate so much that I wanted to continue my education here.

DO and its members for all times of joy, happiness and glamor!

Bergströms, my new family, for always showing interest in what I do at work. Special thanks to **Sara & Gunnar** for all your help with the house in Dalarna and baby-sitting during this summer and fall. Also thanks to **Jakob** for being our "byggare Bob", knowing every nail in the roof, the floors and the walls of our house. I guess we will notice when it is pay-back time!

Annelie & Kjell, Linda & Johan, and **Linus** for always being there for me. Tack!

My brother **Tony** and sister-in-law **Petra** for your support during my time in Uppsala and for making me practice my discussion skills in terms of football, hockey and golf. **Oscar** and **Clara** for being such great kids questioning for instance how long time a person can spend at school. OK – I am done now!

Ann-Marie & Eskil, my fantastic parents, for never doubting my capacity, for being such believers in that nothing is impossible and for always having time to help although we live so far away from each other. Tack av hela mitt hjärta!

Selma, our little miracle and sunshine, who has opened a new world to me. Your happy face has kept me going!

Daniel, my love, for your endless support and encouragement during these years, for making me see new things in life and for being a great father to Selma!

11. REFERENCES

- 1 Chanda, S. K.; Caldwell, J. S. Fulfilling the promise: drug discovery in the post-genomic era. *Drug Discov Today* 2003, 8, 168-174.
- 2 Jeffery, D. A.; Bogyo, M. Chemical proteomics and its application to drug discovery. *Curr Opin Biotech* 2003, 14, 87-95.
- 3 Whittaker, P. A. What is the relevance of bioinformatics to pharmacology. *TRENDS Pharmacol Sci* 2003, 24, 434-439.
- 4 Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies, and future directions. *J Med Chem* 1994, 37, 1385-1401.
- 5 Houghten, R. A. Combinatorial libraries. Finding the needle in the haystack. *Curr Biol* 1994, 4, 564-567.
- 6 Janda, K. D. Tagged versus untagged libraries: methods for the generation and screening of combinatorial chemical libraries. *Proc Natl Acad Sci U S A* 1994, 91, 10779-10785.
- 7 Gayo, L. M. Solution-phase library generation: methods and applications in drug discovery. *Biotechnol Bioeng* 1998, 61, 95-106.
- 8 Spencer, R. W. High-throughput screening of historic collections: observations on file size, biological targets, and file diversity. *Biotechnol Bioeng* 1998, 61, 61-67.
- 9 Dolle, R. E.; Nelson, K. H., Jr. Comprehensive survey of combinatorial library synthesis: 1998. *J Comb Chem* 1999, 1, 235-282.
- 10 Kennedy, T. Managing the drug discovery /development interface. *Drug Discov Today* 1997, 2, 436-444.
- 11 Clark, D. E.; Grootenhuys, P. D. Progress in computational methods for the prediction of ADMET properties. *Curr Opin Drug Discov Devel* 2002, 5, 382-390.
- 12 Modi, S. Computational approaches to the understanding of ADMET properties and problems. *Drug Discov Today* 2003, 8, 621-623.
- 13 van de Waterbeemd, H.; Gifford, E. ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov* 2003, 2, 192-204.
- 14 Stenberg, P.; Bergström, C. A. S.; Luthman, K.; Artursson, P. Theoretical predictions of drug absorption in drug discovery and development. *Clin Pharmacokinet* 2002, 41, 877-899.
- 15 Chiou, W. L. The rate and extent of oral bioavailability versus the rate and extent of oral absorption: clarification and recommendation of terminology. *J Pharmacokinet Biopharm* 2001, 28, 3-6.
- 16 Kararli, T. T. Comparison of the gastrointestinal anatomy, physiology, and biochemistry of humans and commonly used laboratory animals. *Biopharm Drug Dispos* 1995, 16, 351-380.
- 17 Noyes, A. A.; Whitney, W. R. The rate of solution of solid substances in their own solutions. *J Am Chem Soc* 1897, 19, 930-934.
- 18 Kamlet, M. J.; Doherty, R. M.; Abboud, J.-L. M.; Abraham, M. H.; Taft, R. W. Linear solvation energy relationships: 36. Molecular properties governing solubilities of organic nonelectrolytes in water. *J Pharm Sci* 1986, 75, 338-349.
- 19 Hildebrand, J. Solubility. XII. Regular solutions. *J Am Chem Soc* 1929, 51, 66-80.
- 20 Scatchard, G. Equilibria in non-electrolyte solutions in relation to the vapor pressures and densities of the components. *Chem Rev* 1931, 8, 321-333.
- 21 Ruelle, P.; Rey-Mermet, C.; Buchmann, M.; Nam-Tran, H.; Kesselring, U. W.; Huyskens, P. L. A new predictive equation for the solubility of drugs based on the thermodynamics of mobile disorder. *Pharm Res* 1991, 8, 840-850.
- 22 Hasselbalch, K. A. Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebunden Kohlensäure desselben, und die Sauerstoffbindung des Blutes als Funktion der Wasserstoffzahl. *Biochem Z* 1916, 78, 112-144.
- 23 Avdeef, A. Physicochemical profiling (solubility, permeability and charge state). *Curr Top Med Chem* 2001, 1, 277-351.
- 24 Ito, K.; Sekiguchi, K. Studies on the molecular compounds of organic medicinals. II. Application of the solubility product principle and consideration by the phase rule to the solubility phenomena of the molecular compound of sulfanilamide and sulfathiazole. *Chem Pharm Bull (Tokyo)* 1966, 14, 255-262.

- 25 Bogardus, J. B. Common ion equilibria of hydrochloride salts and the Setschenow equation. *J Pharm Sci* 1982, 71, 588-590.
- 26 Serajuddin, A. T.; Sheen, P. C.; Augustine, M. A. Common ion effect on solubility and dissolution rate of the sodium salt of an organic acid. *J Pharm Pharmacol* 1987, 39, 587-591.
- 27 Khalil, E.; Najjar, S.; Sallam, A. Aqueous solubility of diclofenac diethylamine in the presence of pharmaceutical additives: a comparative study with diclofenac sodium. *Drug Dev Ind Pharm* 2000, 26, 375-381.
- 28 Ni, N.; Yalkowsky, S. H. Prediction of Setschenow constants. *Int J Pharm* 2003, 254, 167-172.
- 29 Arakawa, T.; Timasheff, S. N. Preferential interactions of proteins with salts in concentrated solutions. *Biochemistry* 1982, 21, 6545-6552.
- 30 Arakawa, T.; Timasheff, S. N. Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding. *Biochemistry* 1984, 23, 5912-5923.
- 31 Arakawa, T.; Timasheff, S. N. Abnormal solubility behavior of beta-lactoglobulin: salting-in by glycine and NaCl. *Biochemistry* 1987, 26, 5147-5153.
- 32 Breslow, R.; Guo, T. Surface tension measurements show that chaotropic salting-in denaturants are not just water-structure breakers. *Proc Natl Acad Sci U S A* 1990, 87, 167-169.
- 33 Charman, W. N.; Porter, C. J.; Mithani, S.; Dressman, J. B. Physicochemical and physiological mechanisms for the effects of food on drug absorption: the role of lipids and pH. *J Pharm Sci* 1997, 86, 269-282.
- 34 Bakatselou, V.; Oppenheim, R. C.; Dressman, J. B. Solubilization and wetting effects of bile salts on the dissolution of steroids. *Pharm Res* 1991, 8, 1461-1469.
- 35 Naylor, L. J.; Bakatselou, V.; Dressman, J. B. Comparison of the mechanism of dissolution of hydrocortisone in simple and mixed micelle systems. *Pharm Res* 1993, 10, 865-870.
- 36 Poelma, F. G.; Breas, R.; Tukker, J. J.; Crommelin, D. J. Intestinal absorption of drugs. The influence of mixed micelles on the disappearance kinetics of drugs from the small intestine of the rat. *J Pharm Pharmacol* 1991, 43, 317-324.
- 37 Humberstone, A. J.; Porter, C. J.; Charman, W. N. A physicochemical basis for the effect of food on the absolute oral bioavailability of halofantrine. *J Pharm Sci* 1996, 85, 525-529.
- 38 Hammad, M. A.; Muller, B. W. Increasing drug solubility by means of bile salt-phosphatidylcholine-based mixed micelles. *Eur J Pharm Biopharm* 1998, 46, 361-367.
- 39 Hammad, M. A.; Muller, B. W. Solubility and stability of lorazepam in bile salt/soya phosphatidylcholine-mixed micelles. *Drug Dev Ind Pharm* 1999, 25, 409-417.
- 40 Gu, L.; Strickley, R. G. Preformulation salt selection. Physical property comparisons of the tris(hydroxymethyl)aminomethane (THAM) salts of four analgesic/antiinflammatory agents with the sodium salts and the free acids. *Pharm Res* 1987, 4, 255-257.
- 41 Tong, W. Q.; Whitesell, G. In situ salt screening--a useful technique for discovery support and preformulation studies. *Pharm Dev Technol* 1998, 3, 215-223.
- 42 Engel, G. L.; Farid, N. A.; Faul, M. M.; Richardson, L. A.; Winneroski, L. L. Salt form selection and characterization of LY333531 mesylate monohydrate. *Int J Pharm* 2000, 198, 239-247.
- 43 Varia, S. A.; Stella, V. J. Phenytoin prodrugs V: In vivo evaluation of some water-soluble phenytoin prodrugs in dogs. *J Pharm Sci* 1984, 73, 1080-1087.
- 44 Mathew, A. E.; Mejillano, M. R.; Nath, J. P.; Himes, R. H.; Stella, V. J. Synthesis and evaluation of some water-soluble prodrugs and derivatives of taxol with antitumor activity. *J Med Chem* 1992, 35, 145-151.
- 45 Chan, O. H.; Schmid, H. L.; Stilgenbauer, L. A.; Howson, W.; Horwell, D. C.; Stewart, B. H. Evaluation of a targeted prodrug strategy of enhance oral absorption of poorly water-soluble compounds. *Pharm Res* 1998, 15, 1012-1018.
- 46 Mamidi, R. N.; Mullangi, R.; Kota, J.; Bhamidipati, R.; Khan, A. A.; Katneni, K.; Datla, S.; Singh, S. K.; Rao, K. Y.; Rao, C. S.; Srinivas, N. R.; Rajagopalan, R. Pharmacological and pharmacokinetic evaluation of celecoxib prodrugs in rats. *Biopharm Drug Dispos* 2002, 23, 273-282.
- 47 Kondo, N.; Iwao, T.; Masuda, H.; Yamanouchi, K.; Ishihara, Y.; Yamada, N.; Haga, T.; Ogawa, Y.; Yokoyama, K. Improved oral absorption of a poorly water-soluble drug, HO-221, by wet-bead milling producing particles in submicron region. *Chem Pharm Bull (Tokyo)* 1993, 41, 737-740.
- 48 Kubo, H.; Osawa, T.; Takashima, K.; Mizobe, M. Enhancement of oral bioavailability and pharmacological effect of 1-(3,4-dimethoxyphenyl)-2,3-bis(methoxycarbonyl)-4-hydroxy-6,7,8-

- trimethoxynaphthalene (TA-7552), a new hypocholesterolemic agent, by micronization in co-ground mixture with D-mannitol. *Biol Pharm Bull* 1996, 19, 741-747.
- 49 Mosharraf, M. The effects of solid state properties on solubility and in vitro dissolution behaviour of suspended sparingly soluble drugs. Thesis 206 from The Faculty of Pharmacy, Uppsala University: Uppsala, 1999.
- 50 Pitha, J. Amorphous water-soluble derivatives of cyclodextrins: nontoxic dissolution enhancing excipients. *J Pharm Sci* 1985, 74, 987-990.
- 51 Szejtli, J. Medicinal applications of cyclodextrins. *Med Res Rev* 1994, 14, 353-386.
- 52 Khan, K. A.; Rooke, D. J. Effect of disintegrant type upon the relationship between compressional pressure and dissolution efficiency. *J Pharm Pharmacol* 1976, 28, 633-636.
- 53 Sixsmith, D. The properties of tablets containing microcrystalline cellulose. *J Pharm Pharmacol* 1977, 29, 82-85.
- 54 Hogben, C. A. M.; Tocco, D. J.; Brodie, B. B.; Schanker, L. S. On the mechanism of intestinal absorption of drugs. *J Pharmacol Exp Ther* 1959, 125, 275-282.
- 55 Palm, K.; Luthman, K.; Ros, J.; Gråsjö, J.; Artursson, P. Effect of molecular charge on intestinal epithelial drug transport: pH-dependent transport of cationic drugs. *J Pharmacol Exp Ther* 1999, 291, 435-443.
- 56 Diamond, J. M. The epithelial junction: bridge, gate, and fence. *Physiologist* 1977, 20, 10-18.
- 57 Pappenheimer, J. R.; Reiss, K. Z. Contribution of solvent drag through intercellular junctions to absorption of nutrients by the small intestine of the rat. *J Membr Biol* 1987, 100, 123-136.
- 58 Kubinyi, H. Nonlinear dependence of biological activity on hydrophobic character: the bilinear model. *Farmaco* 1979, 34, 248-276.
- 59 Tsuji, A.; Tamai, I. Carrier-mediated intestinal transport of drugs. *Pharm Res* 1996, 13, 963-977.
- 60 Yang, C. Y.; Dantzig, A. H.; Pidgeon, C. Intestinal peptide transport systems and oral drug availability. *Pharm Res* 1999, 16, 1331-1343.
- 61 Brodin, B.; Nielsen, C. U.; Steffansen, B.; Frokjaer, S. Transport of peptidomimetic drugs by the intestinal Di/tri-peptide transporter, PepT1. *Pharmacol Toxicol* 2002, 90, 285-296.
- 62 Tamai, I.; Tsuji, A. Transporter-mediated permeation of drugs across the blood-brain barrier. *J Pharm Sci* 2000, 89, 1371-1388.
- 63 Wachter, V. J.; Wu, C. Y.; Benet, L. Z. Overlapping substrate specificities and tissue distribution of cytochrome P450 3A and P-glycoprotein: implications for drug delivery and activity in cancer chemotherapy. *Mol Carcinog* 1995, 13, 129-134.
- 64 Gan, L. S.; Moseley, M. A.; Khosla, B.; Augustijns, P. F.; Bradshaw, T. P.; Hendren, R. W.; Thakker, D. R. CYP3A-like cytochrome P450-mediated metabolism and polarized efflux of cyclosporin A in Caco-2 cells. *Drug Metab Dispos* 1996, 24, 344-349.
- 65 Schuetz, E. G.; Beck, W. T.; Schuetz, J. D. Modulators and substrates of P-glycoprotein and cytochrome P4503A coordinately up-regulate these proteins in human colon carcinoma cells. *Mol Pharmacol* 1996, 49, 311-318.
- 66 Benet, L. Z.; Izumi, T.; Zhang, Y.; Silverman, J. A.; Wachter, V. J. Intestinal MDR transport proteins and P-450 enzymes as barriers to oral drug delivery. *J Control Release* 1999, 62, 25-31.
- 67 Benet, L. Z.; Cummins, C. L. The drug efflux-metabolism alliance: biochemical aspects. *Adv Drug Deliv Rev* 2001, 50 Suppl 1, S3-11.
- 68 Lin, J. H.; Chiba, M.; Baillie, T. A. Is the role of the small intestine in first-pass metabolism overemphasized? *Pharmacol Rev* 1999, 51, 135-158.
- 69 Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 2000, 44, 235-249.
- 70 Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeny, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997, 23, 3-25.
- 71 Pan, L.; Ho, Q.; Tsutsui, K.; Takahashi, L. Comparison of chromatographic and spectroscopic methods used to rank compounds for aqueous solubility. *J Pharm Sci* 2001, 90, 521-529.
- 72 Taub, M. E.; Kristensen, L.; Frokjaer, S. Optimized conditions for MDCK permeability and turbidimetric solubility studies using compounds representative of BCS classes I-IV. *Eur J Pharm Sci* 2002, 15, 331-340.
- 73 Avdeef, A. pH-metric solubility. 1. Solubility-pH profiles from Bjerrum plots. Gibbs buffer and pK_a in the solid state. *Pharm Pharmacol Commun* 1998, 4, 165-178.
- 74 Avdeef, A.; Berger, C. M.; Brownell, C. pH-Metric Solubility. 2: Correlation between the acid-base titration and the saturation shake-flask solubility-pH methods. *Pharm Res* 2000, 17, 85-89.

- 75 Avdeef, A.; Berger, C. M. pH-metric solubility. 3. Dissolution titration template method for solubility determination. *Eur J Pharm Sci* 2001, 14, 281-291.
- 76 Avdeef, A.; Testa, B. Physicochemical profiling in drug research: a brief survey of the state-of-the-art of experimental techniques. *Cell Mol Life Sci* 2002, 59, 1681-1689.
- 77 Yalkowsky, S. H.; Banerjee, S. Eds. *Aqueous solubility. Methods of estimation for organic compounds*. Marcel Dekker Inc.: New York, 1992.
- 78 Nyström, C.; Mazur, J.; Barnett, M. I.; Glazer, M. Dissolution rate measurements of sparingly soluble compounds with the Coulter Counter model TAIL. *J Pharm Pharmacol* 1985, 37, 217-221.
- 79 Nyström, C.; Bistrat, M. Coulter counter measurements of solubility and dissolution rate of sparingly soluble compounds using micellar solutions. *J Pharm Pharmacol* 1986, 38, 420-425.
- 80 Martin, Y. C. A practitioner's perspective of the role of quantitative structure-activity analysis in medicinal chemistry. *J Med Chem* 1981, 24, 229-237.
- 81 Anderson, B. D.; Raykar, P. V. Solute structure-permeability relationships in human stratum corneum. *J Invest Dermatol* 1989, 93, 280-286.
- 82 Artursson, P.; Karlsson, J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem Biophys Res Commun* 1991, 175, 880-885.
- 83 Seiler, P. Interconversion of lipophilicities from hydrocarbon/water systems into octanol/water system. *Eur J Med Chem* 1974, 9, 473-479.
- 84 Conradi, R. A.; Hilgers, A. R.; Ho, N. F.; Burton, P. S. The influence of peptide structure on transport across Caco-2 cells. *Pharm Res* 1991, 8, 1453-1460.
- 85 Pidgeon, C.; Ong, S.; Liu, H.; Qiu, X.; Pidgeon, M.; Dantzig, A. H.; Munroe, J.; Hornback, W. J.; Kashner, J. S.; Glunz, L.; Szczerba, T. IAM chromatography: an in vitro screen for predicting drug membrane permeability. *J Med Chem* 1995, 38, 590-594.
- 86 Ong, S.; Liu, H.; Pidgeon, C. Immobilized-artificial-membrane chromatography: measurements of membrane partition coefficient and predicting drug membrane permeability. *J Chromatogr A* 1996, 728, 113-128.
- 87 Kansy, M.; Senner, F.; Gubernator, K. Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *J Med Chem* 1998, 41, 1007-1010.
- 88 Hidalgo, I. J.; Raub, T. J.; Borchardt, R. T. Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability. *Gastroenterology* 1989, 96, 736-749.
- 89 Artursson, P. Epithelial transport of drugs in cell culture. I: A model for studying the passive diffusion of drugs over intestinal absorptive (Caco-2) cells. *J Pharm Sci* 1990, 79, 476-482.
- 90 Engman, H. A.; Lennernäs, H.; Taipalensuu, J.; Otter, C.; Leidvik, B.; Artursson, P. CYP3A4, CYP3A5, and MDR1 in human small and large intestinal cell lines suitable for drug transport studies. *J Pharm Sci* 2001, 90, 1736-1751.
- 91 Polli, J. W.; Wring, S. A.; Humphreys, J. E.; Huang, L.; Morgan, J. B.; Webster, L. O.; Serabjit-Singh, C. S. Rational use of in vitro P-glycoprotein assays in drug discovery. *J Pharmacol Exp Ther* 2001, 299, 620-628.
- 92 Tavelin, S.; Milovic, V.; Ocklind, G.; Olsson, S.; Artursson, P. A conditionally immortalized epithelial cell line for studies of intestinal drug transport. *J Pharm Exp Ther* 1999, 290, 1212-1221.
- 93 Tavelin, S.; Taipalensuu, J.; Söderberg, L.; Morrison, R.; Chong, S.; Artursson, P. Prediction of the oral absorption of low-permeability drugs using small intestine-like 2/4/A1 cell monolayers. *Pharm Res* 2003, 20, 397-405.
- 94 Tavelin, S.; Taipalensuu, J.; Hallböök, F.; Vellonen, K. S.; Moore, V.; Artursson, P. An improved cell culture model based on 2/4/A1 cell monolayers for studies of intestinal drug transport: characterization of transport routes. *Pharm Res* 2003, 20, 373-381.
- 95 Ajay, A.; Walters, W. P.; Murcko, M. A. Can we learn to distinguish between "drug-like" and "nondrug-like" molecules? *J Med Chem* 1998, 41, 3314-3324.
- 96 Oprea, T. I. Property distribution of drug-related chemical databases. *J Comput Aided Mol Des* 2000, 14, 251-264.
- 97 Frimurer, T. M.; Bywater, R.; Naerum, L.; Lauritsen, L. N.; Brunak, S. Improving the odds in discriminating "drug-like" from "non drug-like" compounds. *J Chem Inf Comput Sci* 2000, 40, 1315-1324.

- 98 Oprea, T. I.; Gottfries, J. Chemography: the art of navigating in chemical space. *J Comb Chem* 2001, 3, 157-166.
- 99 Walters, W. P.; Murcko, M. A. Prediction of 'drug-likeness'. *Adv Drug Deliv Rev* 2002, 54, 255-271.
- 100 Oprea, T. I.; Zamora, I.; Ungell, A.-L. Pharmacokinetically based mapping device for chemical space navigation. *J Comb Chem* 2002, 4, 258-266.
- 101 Pickett, S. D.; McLay, I. M.; Clark, D. E. Enhancing the hit-to-lead properties of lead optimization libraries. *J Chem Inf Comput Sci* 2000, 40, 263-272.
- 102 Corina, for information see; <http://www2.chemie.uni-erlangen.de/software/corina/index.html>.
- 103 Concord, for information see; <http://www.tripos.com/sciTech/inSilicoDisc/chemInfo/concord.html>.
- 104 SMILES, for information see; <http://www.daylight.com/dayhtml/smiles>.
- 105 Mannhold, R.; Rekker, R. F.; Sonntag, C.; Ter Laak, A. M.; Dross, K.; Polymeropoulos, E. E. Comparative evaluation of the predictive power of calculation procedures for molecular lipophilicity. *J Pharm Sci* 1995, 84, 1410-1419.
- 106 Kier, L. B.; Hall, L. H. Eds. *Connectivity in structure-activity analysis*; Research Studies Press, John Wiley and Sons: Letchworth, 1986.
- 107 Kier, L. B.; Hall, L. H. An electrotopological-state index for atoms in molecules. *Pharm Res* 1990, 7, 801-807.
- 108 Abraham, M. H. Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. *Chem Soc Rev* 1993, 22, 73-83.
- 109 Mannhold, R.; van de Waterbeemd, H. Substructure and whole molecule approaches for calculating log P. *J Comput Aided Mol Des* 2001, 15, 337-354.
- 110 Raevsky, O. A.; Skvortsov, V. S. 3D hydrogen bond thermodynamics (HYBOT) potentials in molecular modelling. *J Comput Aided Mol Des* 2002, 16, 1-10.
- 111 Palm, K.; Stenberg, P.; Luthman, K.; Artursson, P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm Res* 1997, 14, 568-571.
- 112 Stenberg, P.; Luthman, K.; Artursson, P. Prediction of membrane permeability to peptides from calculated dynamic molecular surface properties. *Pharm Res* 1999, 16, 205-212.
- 113 Stenberg, P.; Norinder, U.; Luthman, K.; Artursson, P. Experimental and computational screening models for the prediction of intestinal drug absorption. *J Med Chem* 2001, 44, 1927-1937.
- 114 van de Waterbeemd, H.; Kansy, M. Hydrogen-bonding capacity and brain penetration. *Chimia* 1992, 46, 299-303.
- 115 Winiwarter, S.; Bonham, N. M.; Ax, F.; Hallberg, A.; Lennernäs, H.; Karlén, A. Correlation of human jejunal permeability (in vivo) of drugs with experimentally and theoretically derived parameters. A multivariate data analysis approach. *J Med Chem* 1998, 41, 4939-4949.
- 116 Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J Med Chem* 2000, 43, 3714-3717.
- 117 Cruciani, G.; Pastor, M.; Guba, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur J Pharm Sci* 2000, 11 Suppl 2, S29-39.
- 118 Norinder, U.; Österberg, T.; Artursson, P. Theoretical calculation and prediction of Caco-2 cell permeability using MolSurf parametrization and PLS statistics. *Pharm Res* 1997, 14, 1786-1791.
- 119 Norinder, U.; Österberg, T.; Artursson, P. Theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parametrization and PLS statistics. *Eur J Pharm Sci* 1999, 8, 49-56.
- 120 Parrott, N.; Lave, T. Prediction of intestinal absorption: comparative assessment of GASTROPLUS and IDEA. *Eur J Pharm Sci* 2002, 17, 51-61.
- 121 Nelson, T. M.; Jurs, P. C. Prediction of aqueous solubility of organic compounds. *J Chem Inf Comput Sci* 1993, 34, 601-609.
- 122 Meylan, W. M.; Howard, P. H.; Boethling, R. S. Improved method for estimating water solubility from octanol/water partition coefficient. *Environ Toxicol Chem* 1996, 15, 100-106.
- 123 Huuskonen, J. Estimation of water solubility from atom-type electrotopological state indices. *Environ Toxicol Chem* 2001, 20, 491-497.
- 124 Liu, R.; Sun, H.; So, S. S. Development of quantitative structure-property relationship models for early ADME evaluation in drug discovery. 2. Blood-brain barrier penetration. *J Chem Inf Comput Sci* 2001, 41, 1623-1632.

- 125 Yaffé, D.; Cohen, Y.; Espinosa, G.; Arenas, A.; Giral, F. A fuzzy ARTMAP based on quantitative structure-property relationships (QSPRs) for predicting aqueous solubility of organic compounds. *J Chem Inf Comput Sci* 2001, *41*, 1177-1207.
- 126 Cheng, A.; Merz, K. M. Prediction of aqueous solubility of a diverse set of compounds using quantitative structure-property relationships. *J Med Chem* 2003, *46*, 3572-3580.
- 127 Huuskonen, J.; Salo, M.; Taskinen, J. Neural network modeling for estimation of the aqueous solubility of structurally related drugs. *J Pharm Sci* 1997, *86*, 450-454.
- 128 Hjorth Krarup, L.; Christensen, I. T.; Hovgaard, L.; Frokjaer, S. Predicting drug absorption from molecular surface properties based on molecular dynamics simulations. *Pharm Res* 1998, *15*, 972-978.
- 129 McFarland, J. W.; Avdeef, A.; Berger, C. M.; Raevsky, O. A. Estimating the water solubilities of crystalline compounds from their chemical structures alone. *J Chem Inf Comput Sci* 2001, *41*, 1355-1359.
- 130 Myrdal, P. B.; Manka, A. M.; Yalkowsky, S. H. Aquafac 3: Aqueous functional group activity coefficients; application to the estimation of aqueous solubility. *Chemosphere* 1995, *30*, 1619-1637.
- 131 AquaSol database compiled by S. H. Yalkowsky. For further information: www.pharm.arizona.edu/outreach/aquasol.
- 132 Egan, W. J.; Merz, K. M., Jr.; Baldwin, J. J. Prediction of drug absorption using multivariate statistics. *J Med Chem* 2000, *43*, 3867-3877.
- 133 Hansch, C.; Quinlan, J. E.; Lawrence, G. L. The linear free-energy relationship between partition coefficients and the aqueous solubility of organic liquids. *J Org Chem* 1968, *33*, 347-350.
- 134 Palm, K.; Luthman, K.; Ungell, A.-L.; Strandlund, G.; Beigi, F.; Lundahl, P.; Artursson, P. Evaluation of dynamic polar molecular surface area as predictor of drug absorption: comparison with other computational and experimental predictors. *J Med Chem* 1998, *41*, 5382-5392.
- 135 Motulsky, H. Ed.. *Intuitive Biostatistics*; Oxford University Press: New York, 1995.
- 136 Jackson, E. J. Ed. *A users guide to principal components*; Wiley: New York, 1991.
- 137 Höskuldsson, A. PLS regression methods. *J Chemometrics* 1988, *2*, 211-228.
- 138 Oprea, T. I.; Gottfries, J. Toward minimalistic modeling of oral drug absorption. *J Mol Graph Model* 1999, *17*, 261-274, 329.
- 139 Ichikawa, H. Hierarchy neural networks as applied to pharmaceutical problems. *Adv Drug Deliv Rev* 2003, *55*, 1119-1147.
- 140 Taskinen, J.; Yliruusi, J. Prediction of physicochemical properties based on neural network modelling. *Adv Drug Deliv Rev* 2003, *55*, 1163-1183.
- 141 Wessel, M. D.; Jurs, P. C.; Tolan, J. W.; Muskal, S. M. Prediction of human intestinal absorption of drug compounds from molecular structure. *J Chem Inf Comput Sci* 1998, *38*, 726-735.
- 142 Ghuloum, A. M.; Sage, C. R.; Jain, A. N. Molecular hashkeys: a novel method for molecular characterization and its application for predicting important pharmaceutical properties of molecules. *J Med Chem* 1999, *42*, 1739-1748.
- 143 Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold S. Eds. *Multi- and megavariable data analysis. Principles and applications*. Umetrics Academy; Umeå, 2001.
- 144 Yalkowsky, S. H.; Valvani, S. C. Solubility and partitioning I: Solubility of non-electrolytes in water. *J Pharm Sci* 1980, *69*, 912-922.
- 145 Hildebrand, J. C.; Scott, R. L. Eds. *The solubility of nonelectrolytes*; Reinhold Publishing Corporation: New York, 1950.
- 146 Yalkowsky, S. H. Solubility and partitioning V: Dependence of solubility on melting point. *J Pharm Sci* 1981, *70*, 971-973.
- 147 Ruelle, P.; Buchmann, M.; Ho, N. T.; Kesselring, U. W. The mobile order theory versus UNIFAC and regular solution theory-derived models for predicting the solubility of solid substances. *Pharm Res* 1992, *9*, 788-791.
- 148 Kamlet, M. J.; Abboud, J.-L. M.; Abraham, M. H.; Taft, R. W. Linear solvation energy relationships. 23. A comprehensive collection of the solvatochromic parameters, π^* , α and β , and some methods for simplifying the generalized solvatochromic equation. *J Org Chem* 1983, *48*, 2877-2887.
- 149 Kamlet, M. J.; Doherty, R. M.; Fiserova-Bergerova, V.; Carr, P. W.; Abraham, M. H.; Taft, R. W. Solubility properties in biological media 9. Prediction of solubility and partition of organic nonelectrolytes in blood and tissue from solvatochromic parameters. *J Pharm Sci* 1987, *76*, 14-17.

- 150 Bustamante, P.; Martin, A.; Gonzalez-Guisandez, M. A. Partial solubility parameters and solvatochromic parameters for predicting the solubility of single and multiple drugs in individual solvents. *J Pharm Sci* 1993, 82, 635-640.
- 151 Bustamante, P.; Escalera, B.; Martin, A.; Selles, E. A modification of the extended Hildebrand approach to predict the solubility of structurally related drugs in solvent mixtures. *J Pharm Pharmacol* 1993, 43, 253-257.
- 152 Abraham, M. H.; Le, J. The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship. *J Pharm Sci* 1999, 88, 868-880.
- 153 Absolv, for further information; <http://www.pion-inc.com/Absolv.htm>.
- 154 Yalkowsky, S. H.; Pinal, R.; Banerjee, S. Water solubility: A critique of the solvatochromic approach. *J Pharm Sci* 1988, 77, 74-77.
- 155 Yalkowsky, S. H.; Pinal, R. Estimation of the aqueous solubility of complex organic compounds. *Chemosphere* 1993, 26, 1239-1261.
- 156 Li, A.; Yalkowsky, S. H. Solubility of organic solutes in ethanol/water mixtures. *J Pharm Sci* 1994, 83, 1735-1740.
- 157 Raevsky, O. A.; Trepalin, S. V.; Trepalina, H. P.; Gerasimenko, V. A.; Raevskaja, O. E. SLIPPER-2001 -- software for predicting molecular properties on the basis of physicochemical descriptors and structural similarity. *J Chem Inf Comput Sci* 2002, 42, 540-549.
- 158 SLIPPER, for information see; <http://software.timtec.net/slipper.htm>.
- 159 Amidon, G. L.; Yalkowsky, S. H.; Leung, S. Solubility of nonelectrolytes in polar solvents II. Solubility of aliphatic alcohols in water. *J Pharm Sci* 1974, 63, 1858-1866.
- 160 Jorgensen, W. L.; Duffy, E. M. Prediction of drug solubility from Monte Carlo simulations. *Bioorg Med Chem Lett* 2000, 10, 1155-1158.
- 161 QikProp. For further information: <http://www.schrodinger.com/Products/qikprop.html>.
- 162 Mitchell, B. E.; Jurs, P. C. Prediction of aqueous solubility of organic compounds from molecular structure. *J Chem Inf Comput Sci* 1998, 38, 489-496.
- 163 Wanchana, S.; Yamashita, F.; Hashida, M. Quantitative structure/property relationship analysis on aqueous solubility using genetic algorithm-combined partial least squares method. *Pharmazie* 2002, 57, 127-129.
- 164 Bodor, N.; Huang, M.-J. A new method for the estimation of the aqueous solubility of organic compounds. *J Pharm Sci* 1992, 81, 954-960.
- 165 Sutter, J. M.; Jurs, P. C. Prediction of aqueous solubility for a diverse set of heteroatom-containing organic compounds using a quantitative structure-property relationship. *J Chem Inf Comput Sci* 1996, 36, 100-107.
- 166 Huuskonen, J.; Salo, M.; Taskinen, J. Aqueous solubility prediction of drugs based on molecular topology and neural network modeling. *J Chem Inf Comput Sci* 1998, 38, 450-456.
- 167 Huuskonen, J.; Rantanen, J.; Livingstone, D. Prediction of aqueous solubility for a diverse set of organic compounds based on atom-type electrotopological state indices. *Eur J Med Chem* 2000, 35, 1081-1088.
- 168 Huuskonen, J. Estimation of aqueous solubility in drug design. *Comb Chem High Throughput Screen* 2001, 4, 311-316.
- 169 Conradi, R. A.; Hilgers, A. R.; Ho, N. F.; Burton, P. S. The influence of peptide structure on transport across Caco-2 cells. II. Peptide bond modification which results in improved permeability. *Pharm Res* 1992, 9, 435-439.
- 170 Ren, S.; Das, A.; Lien, E. J. QSAR analysis of membrane permeability to organic compounds. *J Drug Target* 1996, 4, 103-107.
- 171 Camenisch, G.; Alsenz, J.; van de Waterbeemd, H.; Folkers, G. Estimation of permeability by passive diffusion through Caco-2 cell monolayers using the drugs' lipophilicity and molecular weight. *Eur J Pharm Sci* 1998, 6, 317-324.
- 172 Clark, D. E. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J Pharm Sci* 1999, 88, 807-814.
- 173 van de Waterbeemd, H.; el Tayar, N.; Carrupt, P. A.; Testa, B. Pattern recognition study of QSAR substituent descriptors. *J Comput Aided Mol Des* 1989, 3, 111-132.
- 174 van de Waterbeemd, H.; Camenisch, G.; Folkers, G.; Chretien, J. R.; Raevsky, O. A. Estimation of blood-brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. *J Drug Target* 1998, 6, 151-165.

- 175 Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 2002, 45, 2615-2623.
- 176 Österberg, T.; Norinder, U. Prediction of drug transport processes using simple parameters and PLS statistics. The use of ACD/logP and ACD/ChemSketch descriptors. *Eur J Pharm Sci* 2001, 12, 327-337.
- 177 Yamashita, F.; Wanchana, S.; Hashida, M. Quantitative structure/property relationship analysis of Caco-2 permeability using a genetic algorithm-based partial least squares method. *J Pharm Sci* 2002, 91, 2230-2239.
- 178 Stenberg, P.; Luthman, K.; Ellens, H.; Lee, C. P.; Smith, P. L.; Lago, A.; Elliott, J. D.; Artursson, P. Prediction of the intestinal absorption of endothelin receptor antagonists using three theoretical methods of increasing complexity. *Pharm Res* 1999, 16, 1520-1526.
- 179 Zhao, Y. H.; Le, J.; Abraham, M. H.; Hersey, A.; Eddershaw, P. J.; Luscombe, C. N.; Boutina, D.; Beck, G.; Sherborne, B.; Cooper, I.; Platts, J. A. Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors. *J Pharm Sci* 2001, 90, 749-784.
- 180 Mizutani, M. Die Dissoziation der schwachen Elektrolyte in wässrig-alkoholischen Lösungen. IV. Die Dissoziation der schwachen Elektrolyte in Methylalkohol. *Z Physik Chemie* 1925, 119, 318-326.
- 181 FDA Guidance for Industry. Waiver of in vivo bioavailability and bioequivalence studies for immediate-release solid oral dosage forms based on a biopharmaceutics classification system. For further information: <http://www.fda.gov/cder/guidance/index.htm>.
- 182 *Physicians' Desk Reference*; 55th ed.; Medical Economics Company: Montvale, 2000.
- 183 Hedstrand A.-G. Ed. *FASS Läkemedel i Sverige 2001*; Elanders: Kungsbacka, 2001.
- 184 Molconn-Z, 15S, Hall Associates Consulting, Quincy, MA.
- 185 Olsson, T.; Sherbukhin, V. Synthesis and Structure Administration (SaSA), AstraZeneca R&D Mölndal.
- 186 Pearlman, R. S.; Smith, K. M. Eds. *New software tools for chemical diversity*; Kluwer Academic: Dordrecht, 1997.
- 187 MAREA version 2.4. The program MAREA is available upon request from the authors. The program is provided free of charge for academic users. Contact Johan Gräsjo (e-mail johan.grasjo@galenik.uu.se).
- 188 Chen, X. Q.; Cho, S. J.; Li, Y.; Venkatesh, S. Prediction of aqueous solubility of organic compounds using a quantitative structure-property relationship. *J Pharm Sci* 2002, 91, 1838-1852.
- 189 Wegner, J. K.; Zell, A. Prediction of aqueous solubility and partition coefficient optimized by a genetic algorithm based descriptor selection method. *J Chem Inf Comput Sci* 2003, 43, 1077-1084.
- 190 Stahura, F. L.; Godden, J. W.; Bajorath, J. Differential Shannon entropy analysis identifies molecular property descriptors that predict aqueous solubility of synthetic compounds with high accuracy in binary QSAR calculations. *J Chem Inf Comput Sci* 2002, 42, 550-558.
- 191 Neuhoff, S.; Ungell, A.-L.; Zamora, I.; Artursson, P. pH-dependent bidirectional transport of weakly basic drugs across Caco-2 monolayers: implications for drug-drug interactions. *Pharm Res* 2003, 20, 1141-1148.
- 192 Parshad, H. Design of poorly soluble drug salts. Pharmaceutical chemical characterization of organic salts. Thesis from The Danish University of Pharmaceutical Sciences: Copenhagen, 2003.
- 193 Streng, W. H.; Hsi, S. K.; Helms, P. E.; Tan, H. G. General treatment of pH-solubility profiles of weak acids and bases and the effects of different acids on the solubility of a weak base. *J Pharm Sci* 1984, 73, 1679-1684.
- 194 Anderson, B. D.; Conradi, R. A. Predictive relationships in the water solubility of salts of a nonsteroidal anti-inflammatory drug. *J Pharm Sci* 1985, 74, 815-820.
- 195 Ravin, L. J.; Shami, E. G.; Rattie, E. S. Micelle formation and its relationship to solubility behavior of 2-butyl-3-benzofuranyl-4-(2-(diethylamino)ethoxy)-3,5-diiodophenyl ketone hydrochloride. *J Pharm Sci* 1975, 64, 1830-1833.
- 196 Smith, S. W.; Anderson, B. D. Salt and mesophase formation in aqueous suspensions of lauric acid. *Pharm Res* 1993, 10, 1533-1543.

- 197 Maurin, M. B.; Vickery, R. D.; Gerard, C. A.; Hussain, M. Solubility of ionization behavior of the antifungal α -(2,4-difluorophenyl)- α -[(1-(2-(2-pyridyl)phenylethynyl)]-1H-1,2,4-triazole-1-ethanol bismesylate (XD405). *Int J Pharm* 1993, 94, 11-14.
- 198 Dittert, L. D.; Higuchi, T.; Reese, D. R. Phase solubility technique in studying the formation of complex salts of triamterene. *J Pharm Sci* 1964, 53, 1325-1329.
- 199 Hendriksen, B. A.; Felix, M. V.; Bolger, M. B. The composite solubility versus pH profile and its role in intestinal absorption prediction. *AAPS PharmSci* 2003, 5, 35-49.
- 200 Schuetz, E. G.; Yasuda, K.; Arimori, K.; Schuetz, J. D. Human MDR1 and mouse mdr1a P-glycoprotein alter the cellular retention and disposition of erythromycin, but not of retinoic acid or benzo(a)pyrene. *Arch Biochem Biophys* 1998, 350, 340-347.
- 201 Takano, M.; Hasegawa, R.; Fukuda, T.; Yumoto, R.; Nagai, J.; Murakami, T. Interaction with P-glycoprotein and transport of erythromycin, midazolam and ketoconazole in Caco-2 cells. *Eur J Pharmacol* 1998, 358, 289-294.
- 202 Sandström, R.; Karlsson, A.; Knutson, L.; Lennernäs, H. Jejunal absorption and metabolism of R/S-verapamil in humans. *Pharm Res* 1998, 15, 856-862.
- 203 Tamai, I.; Tsuji, A. Carrier-mediated approaches for oral drug delivery. *Adv Drug Deliv Rev* 1996, 20, 5-32.
- 204 Seelig, A.; Landwojtowicz, E. Structure-activity relationship of P-glycoprotein substrates and modifiers. *Eur J Pharm Sci* 2000, 12, 31-40.
- 205 Seelig, A.; Blatter, X. L.; Wohnsland, F. Substrate recognition by P-glycoprotein and the multidrug resistance-associated protein MRP1: a comparison. *Int J Clin Pharmacol Ther* 2000, 38, 111-121.
- 206 QikProp version 1.2. See also reference 161.
- 207 Clark, D. E. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. *J Pharm Sci* 1999, 88, 815-821.
- 208 Kelder, J.; Grootenhuis, P. D.; Bayada, D. M.; Delbressine, L. P.; Ploemen, J. P. Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm Res* 1999, 16, 1514-1519.
- 209 Li, Z. J.; Ojala, W. H.; Grant, D. J. W. Molecular modeling study of chiral drug crystals: lattice energy calculations. *J Pharm Sci* 2001, 90, 1523-1539.
- 210 Adsmoond, D. A.; Grant, D. J. Hydrogen bonding in sulfonamides. *J Pharm Sci* 2001, 90, 2058-2077.
- 211 Jain, N.; Yalkowsky, S. H. Estimation of the aqueous solubility I: application to organic nonelectrolytes. *J Pharm Sci* 2001, 90, 234-252.
- 212 Ran, Y.; Yalkowsky, S. H. Prediction of drug solubility by the general solubility equation (GSE). *J Chem Inf Comput Sci* 2001, 41, 354-357.
- 213 Goodwin, J. T.; Conradi, R. A.; Ho, N. F.; Burton, P. S. Physicochemical determinants of passive membrane permeability: role of solute hydrogen-bonding potential and volume. *J Med Chem* 2001, 44, 3721-3729.
- 214 Marrink, S. J.; Berendsen, H. J. C. Simulation of water transport through a lipid membrane. *J Phys Chem* 1994, 98, 4155-4168.
- 215 Walter, E.; Janich, S.; Roessler, B. J.; Hilfinger, J. M.; Amidon, G. L. HT29-MTX/Caco-2 cocultures as an in vitro model for the intestinal epithelium: in vitro-in vivo correlation with permeability data from rats and humans. *J Pharm Sci* 1996, 85, 1070-1076.
- 216 van de Waterbeemd, H. The fundamental variables of the biopharmaceutics classification system (BCS): a commentary. *Eur J Pharm Sci* 1998, 7, 1-3.

Acta Universitatis Upsaliensis

*Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Pharmacy*

Editor: The Dean of the Faculty of Pharmacy

A doctoral dissertation from the Faculty of Pharmacy, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy*. (Prior to July 1985, the series was published under the title “Abstracts of Uppsala Dissertations from the Faculty of Pharmacy”.)

Distributor:
Uppsala University Library,
Box 510, SE-751 20 Uppsala, Sweden
www.uu.se, acta@ub.uu.se

ISSN 0282-7484
ISBN 91-554-5747-9