# Poorly Soluble Marketed Drugs Display Solvation Limited Solubility

Christel A. S. Bergström,\* Carola M. Wassvik, Kajsa Johansson, and Ina Hubatsch

Pharmaceutical Screening and Informatics, Department of Pharmacy, Uppsala University, Uppsala Biomedical Centre, Post Office Box 580, SE-751 23 Uppsala, Sweden

Received June 5, 2007

We determined the intrinsic aqueous solubility of 15 poorly soluble drugs with solubilities ranging from 2.9 nM to 1.1  $\mu$ M. We then analyzed the data from a physicochemical perspective, using experimentally determined solid-state properties and easily interpretable two-dimensional molecular descriptors, to better understand the factors underlying poor solubility. The analysis shows that poorly soluble drugs that have reached the market are solubility limited by solvation rather than by their solid state.

### Introduction

The methods applied in the drug discovery process in use nowadays often result in molecules of poor solubility. Low solubility can cause low bioavailability or give rise to large fluctuations in the fraction absorbed in humans that can often not be compensated by a high permeability. Furthermore, low solubility may be associated with stability problems and difficulties in developing an acceptable formulation.<sup>1</sup>

Early in the drug discovery process, compounds are often available only as virtual products or synthesized in limited quantities. As it is not always possible to measure the solubility directly, good predictive computational models are desirable.<sup>2</sup> Several computational approaches for predicting solubility have been developed, but many of these used training sets consisting primarily of nondrug-like molecules such as alcohols, pesticides, and herbicides,<sup>3-5</sup> as a result of which they may not be very accurate in predicting the solubility of drugs.<sup>1.6</sup> Additionally, there is a distinctive lack of quantitative experimental data for poorly soluble drugs and drug candidates with a solubility of less than 1  $\mu$ M. Because the drug space of poorly soluble compounds is essentially uncovered, difficulties in correctly predicting poorly soluble drug-like compounds prevail.

In this study, we have experimentally determined the solubility of a series of poorly soluble drugs, and solubility data ranging from 2.9 nM to 1.1  $\mu$ M are presented. This unique dataset has been used to analyze the following in detail: (i) which experimental setup is preferable for measuring poorly soluble compounds and (ii) which physicochemical properties are fundamental for the solubility of such compounds. Finally, using the data obtained, we discuss which poorly soluble compounds are most likely to successfully complete the drug development process.

# **Experimental Section**

Selection of Dataset. A series of 15 poorly soluble compounds was selected for this study (Figure 1). The choice was made from a literature search using PubMed and the search string "poor\* AND solubility AND drug\* AND formulation" to extract information on poorly soluble drug-like compounds. Compounds that were judged to (i) be more soluble than 1  $\mu$ M from the publications, (ii) be unstable, (iii) only exist as salt forms, (iv) not be commercially available, or (v) be very expensive were excluded. Principal component analysis (PCA) was conducted using Simca-P version 10 (Umetrics, Umeå, Sweden) to verify the structural diversity of

\* To whom correspondence should be addressed. Phone: +46184714645. Fax: +46184714223. E-mail: christel.bergstrom@farmaci.uu.se. the dataset. The solubility and solid-state characteristics of astemizole, glyburide, and indomethacin had been determined previously in our laboratory,<sup>7</sup> and these data were included in the present study.

Solid-State Characterization by Differential Scanning Calorimetry (DSC). Thermograms were recorded with a Seiko DSC220C analysis module with an automatic cooling controller (Seiko Instruments, Inc., Japan). Triplicate samples of 1–3 mg were weighed in sealed and pierced aluminum pans (TA Instruments, Delaware). Only one sample was run for felodipine and troglitazone, owing to the small quantities available. Samples of each compound were heated from room temperature to approximately 50 °C above their melting point at a rate of 10 °C/min and purged with nitrogen gas at a flow rate of 80 mL/min. The melting temperature, ( $T_m$ ), entropy of melting ( $\Delta S_m$ ), and enthalpy of melting ( $\Delta H_m$ ), were determined for each compound. Danazol decomposed directly after melting when measured at 10 °C/min and was, therefore, determined at 20 °C/min to allow the melting to finish completely before decomposition.

Solubility Determinations by Shake-Flask Method. Each drug was added in excess to 1.5 mL Eppendorf tubes containing 1 mL of Milli-Q water. The pH of the suspensions were adjusted to at least 2 pH units below  $pK_a$  (acids) or 2 pH units above  $pK_a$  (bases) with 0.5 M HCl or 0.5 M NaOH. This kept the drugs in their uncharged states in accordance with the Henderson-Hasselbalch equation and allowed the intrinsic solubility to be determined. Neutrals (bases with a  $pK_a$  below 2 and acids with a  $pK_a$  above 12) and zwitterions were not pH adjusted. The tubes were placed onto a plate shaker at 300 rpm at room temperature. The experiments (using 3 to 5 replicate samples) were terminated after 24, 72, and 192 h or later if equilibrium solubility had not been obtained after 192 h. The excess solid was separated from the solution by centrifugation in an Eppendorf Centrifuge 5403 (Eppendorf AG, Hamburg, Germany) at 23 000 g for 15 min. After the centrifugation, approximately 0.25 mL of the supernatant was sampled with Pasteur glass pipettes and dispensed into glass insert vials. These vials were placed in Eppendorf tubes and centrifuged for 15 min a second time to obtain complete separation of the solution and the remaining solid. After the centrifugation, the supernatant was withdrawn with Pasteur pipettes and dispensed into glass HPLC vials. Glass was used throughout this procedure to minimize the risk of underestimating the solubility owing to adsorption of the drugs to plastics.

**HPLC-MS/MS Analysis.** Directly after the termination of the solubility experiment, the concentration of the samples was determined with a ThermoFinnigan TSQ Quantum Discovery triplequadrupole mass spectrometer using electrospray ionization (ESI), coupled to a ThermoFinnigan Surveyor autosampler and Surveyor HPLC-MS pump (Thermo Electron Corp., Waltham, U.S.A.). For separation, an XTerra MSC18 column (3.5  $\mu$ m, 2.1 × 20 mm; Waters, Milford, U.S.A.) and a flow rate of 200  $\mu$ L/min were used. The samples (5  $\mu$ L) were injected and run with a gradient using



Figure 1. Chemical structures of the compounds studied. The compounds had the following physicochemical profile: MW 261.1–705.7 g/mol, ClogP 3.5–6.8, and PSA 2.4–137.2 Å<sup>2</sup>.

water and acetonitrile with or without formic acid in either the negative or positive polarity mode (see Supporting Information). For each compound, a standard curve consisting of seven concentrations was established, with a separate quality control being made of three concentrations.

**Statistics.** The solubility values are presented as means  $\pm$  SD. ANOVA was used to test whether the differences between means were statistically significant (p < 0.05). The coefficient of determination ( $R^2$ ) was used to assess the goodness-of-fit of standard curves of measured concentrations. Partial least-squares projection to latent structures (PLS; Simca-P v.10) was used to further analyze the importance of different physicochemical properties, as described previously.<sup>8</sup> Calculations of two-dimensional molecular descriptors reflecting among other the size, polarizability, hydrophobicity/ hydrophilicity and electron distribution were made with the program Selma.<sup>7.8</sup> Only nonskewed descriptors (n = 75) were included in the PLS analysis, and a variable selection was performed to exclude noise and increase the transparency of the model.

# **Results and Discussion**

**Experimental Results.** The solubility and solid-state characteristics are presented in Table 1. The solubility of the 15 drugs ranged from 2.9 nM to  $1.1 \,\mu$ M. The limit of detection of troglitazone was 15.9 nM, but even though the experiment was extended to 1104 h (46 days), all shake-flask samples were below this limit. It is unlikely that this was because of degradation of the sample, because an LC-MS scan of the m/z 200–500 of the samples did not reveal any other significant peaks. However, as an m/z scan of cinnarizine samples shaken for longer than 72 h indicated degradation, the solubility value from 72 h is reported.

Extended time studies were performed, but a long agitation time was not needed for all compounds to reach their solubility equilibrium (see the Supporting Information). The shaking time ranged from 24 to 1104 h, and for eight compounds, there was no statistically significant difference between the means of the first and last time point at the 95% CI. However, it was not possible to know beforehand which compounds would need a longer time. Danazol and tamoxifen had the largest difference between the first and last concentrations. It was not surprising that danazol required a long time to reach equilibrium, as previous results from our laboratory have shown that steroids may require a long time to attain equilibrium. This we have found for hydrocortisone<sup>6</sup> and corticosterone (unpublished results). The solubility of danazol increased 3.2-fold with time, and tamoxifen decreased 5-fold with time. Such variations in the solubility indicate that, to roughly estimate the equilibrium solubility of poorly soluble compounds, a 24 h equilibrium time period is long enough. However, if the solubility is supposed to be used as an input in computational solubility models, it is preferable to extend the time scale for the experiment to ensure that the equilibrium solubility is obtained. By adopting this approach, solubility data of higher quality are used in the model development, thereby increasing the chances of producing an accurate in silico model for the solubility and minimizing the risk of modeling noise.

**Physicochemical Properties and their Relation to Solubility.** The relation between physicochemical properties and solubility was investigated (Table 2 and Supporting Information Table S3). First, the relationship to the widely used molecular

Table 1. Experimental Solubility Values and Solid-State Characteristics

-	-				
	time <sup>a</sup> (h)	$S^b$ (nM)	$T_{\rm m}^{c}$ (°C)	$\Delta H_{ m m}^{d}$ (kJ/mol)	$\Delta S_{\rm m}^{e}$ (J/mol/K)
albendazole	216	$983.0 \pm 112.2$	$178.1 \pm 4.3$	$98.6 \pm 16.1$	$218.7 \pm 37.8$
astemizole <sup>f</sup>	24	$66.8 \pm 21.1$	$174.4 \pm 0.1$	$51.1 \pm 0.8$	$114.1 \pm 1.7$
carvedilol	744	$713.0 \pm 56.0$	$114.1 \pm 0.5$	$57.6 \pm 1.0$	$148.7 \pm 2.6$
cinnarizine	72	$18.6 \pm 1.7$	$120.2 \pm 1.0$	$45.7 \pm 2.3$	$116.1 \pm 5.6$
danazol	1032	$36.0 \pm 28.3$	$228.6 \pm 0.5$	$35.5 \pm 1.3$	$70.8 \pm 2.6$
felodipine	504	$276.0 \pm 17.7$	$139.1^{h}$	$34.8^{h}$	84.3 <sup>h</sup>
glimepiride	213	$12.7 \pm 2.9$	$212.5 \pm 0.6$	$53.3 \pm 4.5$	$109.9 \pm 9.2$
glyburide <sup>f</sup>	24	$89.1 \pm 40.6$	$173.6 \pm 0.1$	$46.3 \pm 0.1$	$103.7 \pm 0.2$
indomethacin <sup>f</sup>	24	$1133.3 \pm 251.7$	$159.8 \pm 0.0$	$37.9 \pm 0.2$	$87.6 \pm 0.4$
itraconazol	336	$3.3 \pm 4.6$	$165.4 \pm 0.9$	$69.9 \pm 3.5$	$159.5 \pm 7.8$
rimonabant	380	$96.5 \pm 12.5$	$154.7 \pm 0.4$	$36.1 \pm 1.7$	$84.4 \pm 4.2$
tamoxifen	888	$2.9 \pm 1.0$	$97.8 \pm 0.3$	$34.0 \pm 0.4$	$91.6 \pm 1.1$
terfenadine	864	$11.6 \pm 7.2$	$149.6 \pm 0.3$	$58.1 \pm 2.0$	$137.5 \pm 4.8$
tolfenamic acid	192	$13.6 \pm 1.5$	$212.1 \pm 0.3$	$41.2 \pm 1.3$	$84.9 \pm 2.7$
troglitazone	1104	<15.9g	$139.2^{h}$	$48.8^{h}$	$118.3^{h}$
mean		246.9	161.3	49.9	115.3
min		2.9	97.8	34.0	70.8
max		1133.3	228.6	98.6	218.7

<sup>*a*</sup> Time (h) shows the equilibrium time used in the solubility study. <sup>*b*</sup> S (nM) is the solubility value presented in nanomolar concentration. <sup>*c*</sup>  $T_m$  denotes the melting point in Celsius. None of the compounds displayed polymorphism. <sup>*d*</sup>  $\Delta H_m$  denotes the enthalpy of melting. <sup>*e*</sup>  $\Delta S_m$  denotes the entropy of melting. <sup>*f*</sup> Data taken from Wassvik et al., 2006, ref 7 in this paper. <sup>*g*</sup> Troglitazone could not be detected within the sample even after an equilibrium time of 1104 h. The presented value is the limit of detection for troglitazone using the applied LC-MS/MS settings. The sample showed no indication of degradation of troglitazone, in spite of the long time used for the study. <sup>*h*</sup> Only one sample was measured for solid-state characteristics as only a small amount of the substance was available.

Table 2. Physicochemical Characteristics of the Compounds Studied

MW	$PSA^{a}$	NPSA <sup>a</sup>			acid/
(g/mol)	(Å <sup>2</sup> )	(Å <sup>2</sup> )	$ClogP^{b}$	pKa <sup>b</sup>	base <sup>b</sup>
265.3	65.6	231.9	3.5	11.4; 5.6	a; b
458.6	35.4	488.5	6.1	9.0	b
406.3	78.4	381.4	4.0	8	b
368.5	2.4	452.4	6.1	7.5	b
337.5	45.9	356.1	3.9		neutral
384.3	68.7	346.9	5.6	2.7	b
490.6	137.2	439.3	4.2	5.0	а
494.0	126.9	415.0	4.2	5.0	а
357.6	68.8	306.8	4.2	4.0	а
705.7	84.7	654.2	6.5	4.0	b
463.8	50.7	418.2	6.6	3.6	b
371.5	10.3	464.2	6.8	8.5	b
471.7	46.3	561.1	6.1	9.5	b
261.1	54.8	227.1	5.7	3.7	а
441.6	92.2	400.6	5.6	6.1	а
418.5	64.6	409.6	5.3		
261.1	2.4	227.1	3.5		
705.7	137.2	654.2	6.8		
	MW (g/mol) 265.3 458.6 406.3 368.5 337.5 384.3 490.6 494.0 357.6 705.7 463.8 371.5 471.7 261.1 441.6 418.5 261.1 705.7	$\begin{array}{c c} MW & PSA^a \\ (g/mol) & (Å^2) \\ \hline 265.3 & 65.6 \\ 458.6 & 35.4 \\ 406.3 & 78.4 \\ 368.5 & 2.4 \\ 337.5 & 45.9 \\ 384.3 & 68.7 \\ 490.6 & 137.2 \\ 494.0 & 126.9 \\ 357.6 & 68.8 \\ 705.7 & 84.7 \\ 463.8 & 50.7 \\ 371.5 & 10.3 \\ 471.7 & 46.3 \\ 261.1 & 54.8 \\ \hline 441.6 & 92.2 \\ 418.5 & 64.6 \\ 261.1 & 2.4 \\ 705.7 & 137.2 \\ \hline \end{array}$	$\begin{array}{c cccc} MW & PSA^a & NPSA^a \\ (g/mol) & (Å^2) & (Å^2) \\ \hline \\ 265.3 & 65.6 & 231.9 \\ 458.6 & 35.4 & 488.5 \\ 406.3 & 78.4 & 381.4 \\ 368.5 & 2.4 & 452.4 \\ 337.5 & 45.9 & 356.1 \\ 384.3 & 68.7 & 346.9 \\ 490.6 & 137.2 & 439.3 \\ 494.0 & 126.9 & 415.0 \\ 357.6 & 68.8 & 306.8 \\ 705.7 & 84.7 & 654.2 \\ 463.8 & 50.7 & 418.2 \\ 371.5 & 10.3 & 464.2 \\ 471.7 & 46.3 & 561.1 \\ 261.1 & 54.8 & 227.1 \\ \hline \\ 441.6 & 92.2 & 400.6 \\ 418.5 & 64.6 & 409.6 \\ 261.1 & 2.4 & 227.1 \\ 705.7 & 137.2 & 654.2 \\ \hline \end{array}$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$

<sup>*a*</sup> Selma descriptors obtained from Dr. Ulf Norinder, AstraZeneca R&D, Södertälje. <sup>*b*</sup> Calculated from SciFinder Scholar.

descriptor ClogP, that is, the calculated partition coefficient between octanol and water was studied. We have previously shown that this relationship for 270 compounds spanning more than 9 log units in solubility, has an R value of  $0.74.^7$  For the 15 poorly soluble compounds studied in this dataset, the R was 0.71 (Figure 2a). However, the relationship obtained with the 270 compounds suggests that ClogP needs to be 6.6 or larger to enable a solubility of less than 1  $\mu$ M to be predicted, as a result of which most of the compounds investigated in the present work would not be predicted as poorly soluble. Instead, it was noted that all compounds with a ClogP value larger than 6 showed an intrinsic solubility less than 100 nM, that is, a solubility value more than 10-fold lower than that predicted by the general equation. One reason that the role played by lipophilicity in restricting solubility is underestimated could be that only a limited number of poorly soluble compounds were included in our previous study because of the lack of solubility



**Figure 2.** Graphs showing the correlation between the physicochemical properties and the solubility. (a) Correlation between the solubility and the calculated lipophilicity (R of -0.71). (b) Correlation between the solubility and the melting point (R of 0.13). (c) The correlation between the solubility and the nonpolar surface area. R went from -0.48 to -0.85 after exclusion of the outlier tolfenamic acid.

data for such compounds.<sup>7</sup> Therefore, poorly soluble compounds did not have the power to influence the general equation to any great extent.

Several calculated and experimentally determined properties were investigated to obtain a deeper knowledge of what other physicochemical properties influence the solubility of poorly soluble compounds. First, we performed linear regression analysis in which we correlated those descriptors, one at the



**Figure 3.** The poorly soluble compounds superimposed on the two first principal components (PC) describing 52% of the structural diversity of the oral drug space. The ellipse shows the 95% CI of the structural diversity of orally administered drugs registered in Sweden (n = 527, Selma descriptors used as input). PC1 mainly reflects the size (the higher the value, the larger the molecule) and PC2 reflects the lipophilicity (the lower the value, the higher the lipophilicity). The poorly soluble compounds are all clustered into one-quarter of the plot. The compounds are colored according to their solubility from the lowest solubility (red) to the highest (blue). A trend emerges whereby the closer a compound is to the ellipse in the lower right quadrant, the poorer the solubility. The arrow is drawn to indicate in which direction the solubility increases. This trend was confirmed by superimposing solubility data previously published by our laboratory,<sup>8</sup> resulting in that no soluble compounds were found in the lower right quadrant. The compounds are abbreviated to the four first letters of their names; for the full names, see Tables 1 or 2.

time, that have previously been identified as being important for solubility with the solubility values obtained (see Table S3 in the Supporting Information). No correlation was observed between the solubility and the solid-state properties investigated ( $T_m$ ,  $\Delta S_m$ , and  $\Delta H_m$ ), all of which resulted in *R* less than 0.13 (Figure 2b and Table S3, Supporting Information). Hence, the poor solubility of the compounds studied was not dependent on the stability of the crystal, although the compounds displayed a wide range of  $T_m$  (97.8–228.6 °C, Table 1). Furthermore, none of the hydrogen bond descriptors investigated (the number of acceptors and donors and the sum of these) proved to be important for this dataset. This is in agreement with the findings for the solid state, because increased polarity and the number of hydrogen bonds are important for making stable crystals.<sup>9</sup>

The molecular size proved to be important for decreasing the solubility for all compounds except for tolfenamic acid. Molecular descriptors such as molecular weight, polarizability, and nonpolar surface area, all of which are highly correlated to the size of the molecule, displayed *R* values of 0.61, 0.79, and 0.85 when correlated to the solubility after the exclusion of tolfenamic acid (Figure 2c and Table S3 in Supporting Information). The reason for tolfenamic acid being an outlier is probably the combination of its high lipophilicity, high melting point, and small size in comparison to the other compounds of the dataset. Hence, for this specific compound, both the lipophilicity and the stability of the crystal can be reasons for its poor solubility. However, it is unlikely that size itself is the limiting descriptor for tolfenamic acid.

It is well-known that molecular size and lipophilicity are highly correlated and that larger molecules generally display a higher ClogP value than smaller ones. However, the correlation (R) between ClogP and MW for the 15 compounds in this dataset was only 0.36, indicating that the size per se is also restricting the solubility of these molecules (see Table S3, Supporting Information).

In a second step, we analyzed the influence of the different descriptors on solubility using multivariate data analysis. The model obtained was not regarded as a model for the prediction of poor solubility because the dataset was too small. Instead, it was used as a tool to further investigate the properties of importance for poor solubility. The PLS analysis showed that descriptors related to lipophilicity, size, and polarizability (reflected in the molecular refractivity) were important for restricting solubility (Figure S2 in Supporting Information), confirming the results from the correlation matrix. Additionally, this analysis also showed that the larger the difference between the highest occupied and lowest unoccupied molecular orbital (HOMO and LUMO, respectively), the lower the solubility. The energy levels of these molecular orbitals have also previously been reported to influence the solubility.<sup>10</sup>

How can we interpret the results concerning the properties that underlie poor solubility? The results suggest that increased molecular size, increased polarizability, increased lipophilicity, and an increased energy gap between the HOMO and LUMO will decrease the solubility. First, the larger the size of the molecule, the larger the cavity in the water needs to be, which implies a greater amount of energy is required to break the tight structure of water. For this dataset, the molecular size was highly correlated with the polarizability, resulting in an R of 0.93. Hence, we believe that the negative influence of polarizability on solubility found in this dataset also is partly a reflection of the energy penalty involved in the cavity formation process. As the cavity is formed, the molecule is incorporated and needs to make bonds with the water molecules to remain in solution. The higher the lipophilicity of a compound, the less favorable the hydration, and, subsequently, the lower the solubility. Finally, the multivariate data analysis identified a descriptor related to the energy gap between the HOMO and LUMO as being important for poorly soluble compounds. Indeed, it was found that the larger the energy difference between these two orbitals, the poorer the solubility. It is known that the energy gap is related to the degree of conjugation and that a higher degree of conjugation results in smaller energy gaps. Hence, the analysis revealed that for this dataset the solubility will improve for compounds that are more conjugated than others. One explanation for this could be that conjugated systems are more rigid and often more compact than nonconjugated ones, and therefore, the cavity formation in water will not be as extensive. Hence, this finding further indicated that a majority of the compounds investigated in this study are solvation limited in their solubility rather than limited by their solid state. If the compounds were to be solubility limited by their solid state, it is more likely that a decrease in conjugation would result in an increase in solubility because flat and rigid compounds tend to form more stable crystals than flexible ones.<sup>11</sup>

All of the compounds examined have been developed as oral dosage forms (Table S4 and Table S5 in Supporting Information). We therefore superimposed the poorly soluble compounds onto the chemical space covered by oral drugs, as identified by a PCA based on all orally registered compounds in Sweden (n = 527), and found the poorly soluble compounds clustered in a dedicated chemical volume (Figure 3). Hence, by using physicochemical descriptors and the structural diversity of the oral drug space, an estimation of the likeliness of poor solubility could be obtained. Such a tool is convenient and easily applied in the early drug discovery process and can be used to guide the medicinal chemist away from poor solubility issues.

Poorly soluble compounds are often referred to as "brick dust" and "grease balls". Brick dust represents a stable crystal in which the strong intermolecular bonds within the crystal restrict the solubility of the compound in water, whereas grease balls represent highly lipophilic compounds that are unable to form bonds with the water molecules. The poorly soluble compounds investigated in this study have all passed through the development process and, by adopting strategic formulation procedures, they have reached the market in spite of their limited solubility (Table S4, Supporting Information). The thorough analysis of the physicochemical properties important for lowering solubility presented above did not support solid-state limited solubility for the dataset as a whole. Rather, the successful development of the investigated compounds is probably related to them acting as grease balls. For such compounds, the solubility can be improved largely by incorporating excipients, such as disintegrants, solubility enhancers/wetting agents, cyclodextrins, and lipids, in the formulation.<sup>12</sup> Furthermore, bile salts in the intestinal fluid will markedly improve the intestinal solubility of highly lipophilic compounds.<sup>13</sup> For highly lipophilic compounds with a ClogP value of three or more, we therefore recommend the solubility assay to be performed in the presence of an emulsifying agent or in a simulated intestinal fluid<sup>14</sup> to avoid underestimating the intestinal solubility. Moreover, the pH gradient in the intestine will affect the solubility of several of the compounds investigated, and their apparent solubility will be higher at a physiological pH than the intrinsic solubility. However, for most of the compounds investigated in this work, the solubility will still be in the lower micromolar range at physiological pH, confirming that solubilizing agents are of the utmost importance for further increasing the solubility and obtaining a therapeutic effect with these compounds.

## Conclusion

We have presented a unique dataset including experimental solubility values and solid-state characteristics for poorly soluble drugs. The thorough analysis of the factors underlying the poor solubility of drugs showed that, for this dataset, the solubility is restricted by solvation rather than the solid state. The most important properties causing poor solubility were a high lipophilicity, a large molecular size, and a poorly conjugated electron system. The study showed that the poor solubility caused by these properties can be treated by optimizing the formulations, but components of the intestinal fluid are also likely to play an important role in the successful development of these compounds. Finally, the statistical analysis showed that poorly soluble compounds are located in a specific volume of the oral drug space. This indicates that this simple tool can give direct feedback on solvation limited solubility even before drug synthesis.

Acknowledgment. We would like to acknowledge Dr. Ulf Norinder at AstraZeneca R&D Södertälje for providing us with the molecular descriptors. Financial support from the Knut and Alice Wallenberg Foundation, the Swedish Foundation for Strategic Research, and The Swedish Fund for Research without Animal Experiments is greatly appreciated.

**Supporting Information Available:** LC-MS/MS settings for analysis of the compounds studied; time versus solubility profiles of the compounds studied; a correlation matrix of physicochemical properties, molecular descriptors, and solubility; formulation approaches to increase solubility for the compounds studied; dose regimen of the compounds; and PLS model statistics and loading plots for the final model based on poorly soluble compounds. This information is available free of charge via the Internet at http:// pubs.acs.org

#### References

- (1) 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 Delivery Rev.* **1997**, *23*, 3–25.
- (2) Delaney, J. S. Predicting aqueous solubility from structure. Drug Discovery Today 2005, 10 (4), 289–295.
- (3) 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.
- (4) Sun, H. A universal molecular descriptor system for prediction of logP, logS, logBB, and absorption. J. Chem. Inf. Comput. Sci. 2004, 44 (2), 748–757.
- (5) Huuskonen, J. Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology. J. Chem. Inf. Comput. Sci. 2000, 40 (3), 773–777.
- (6) Bergström, C. A. S.; Norinder, U.; Luthman, K.; Artursson, P. Experimental and computational screening models for prediction of aqueous drug solubility. *Pharm. Res.* 2002, 19 (2), 182–188.
- (7) Wassvik, C. M.; Holmen, A. G.; Bergström, C. A. S.; Zamora, I.; Artursson, P. Contribution of solid-state properties to the aqueous solubility of drugs. *Eur. J. Pharm. Sci.* **2006**, *29* (3–4), 294–305.
- (8) Bergström, C. A. S.; Wassvik, C. M.; Norinder, U.; Luthman, K.; Artursson, P. Global and local computational models for prediction of aqueous solubility of drug-like molecules. *J. Chem. Inf. Comput. Sci.* 2004, 44 (4), 1477–1488.
- (9) Bergström, C. A. S.; Norinder, U.; Luthman, K.; Artursson, P. Molecular descriptors influencing melting point and their role in classification of solid drugs. *J. Chem. Inf. Comput. Sci.* 2003, 43 (4), 1177–1185.
- (10) Wang, Z.-W.; Feng, J.-L.; Wang, Z.-N.; Li, G.-Z.; Lou, A.-J. Quantitative structure-property relationship on prediction of the interaction parameters δ<sup>2t</sup> of organic compounds. *J. Dispersion Sci. Technol.* **2006**, 27, 11–14.
- (11) Wassvik, C. M. Computational analysis of aqueous drug solubilityinfluence of the solid state. Thesis from the Faculty of Pharmacy, Uppsala University, Uppsala, 2006.
- (12) Avdeef, A.; Bendels, S.; Tsinman, O.; Tsinman, K.; Kansy, M. Solubility-excipient classification gradient maps. *Pharm. Res.* 2007, 24 (3), 530–545
- (13) Persson, E. M.; Gustafsson, A. S.; Carlsson, A. S.; Nilsson, R. G.; Knutson, L.; Forsell, P.; Hanisch, G.; Lennernas, H.; Abrahamsson, B. The effects of food on the dissolution of poorly soluble drugs in human and in model small intestinal fluids. *Pharm. Res.* 2005, 22 (12), 2141–2151.
- (14) Galia, E.; Nicolaides, E.; Horter, D.; Lobenberg, R.; Reppas, C.; Dressman, J. B. Evaluation of various dissolution media for predicting in vivo performance of class I and II drugs. *Pharm. Res.* **1998**, *15* (5), 698–705.

JM0706416