molecular pharmaceutics

Article

pubs.acs.org/molecularpharmaceutics

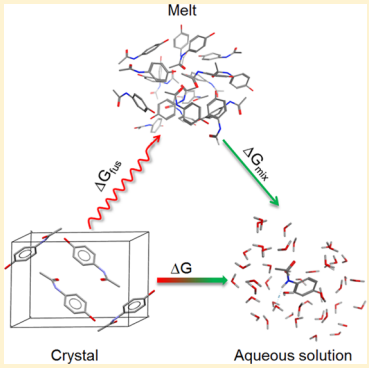
Major Source of Error in QSPR Prediction of Intrinsic Thermodynamic Solubility of Drugs: Solid vs Nonsolid State Contributions?

Yuriy A. Abramov*

Pfizer Global Research and Development, Groton, Connecticut 06340, United States

Supporting Information

ABSTRACT: The main purpose of this study is to define the major limiting factor in the accuracy of the quantitative structure−property relationship (QSPR) models of the thermodynamic intrinsic aqueous solubility of the drug-like compounds. For doing this, the thermodynamic intrinsic aqueous solubility property was suggested to be indirectly "measured" from the contributions of solid state, ΔG_{fus}, and nonsolid state, ΔG_{mix}, properties, which are estimated by the corresponding QSPR models. The QSPR models of ΔG_{fus} and ΔG_{mix} properties were built based on a set of drug-like compounds with available accurate measurements of fusion and thermodynamic solubility properties. For consistency ΔG_{fus} and ΔG_{mix} models were developed using similar algorithms and descriptor sets, and validated against the similar test compounds. Analysis of the relative performances of these two QSPR models clearly demonstrates that it is the solid state contribution which is the limiting factor in the accuracy and predictive power of the QSPR models of the thermodynamic intrinsic solubility. The performed analysis outlines a necessity of development of new descriptor sets for an accurate description of the long-range order (periodicity) phenomenon in the crystalline state. The proposed approach to the analysis of limitations and suggestions for improvement of QSPR-type models may be generalized to other applications in the pharmaceutical industry.



KEYWORDS: *thermodynamic intrinsic aqueous solubility, QSPR/QSAR, crystal packing contribution, free energy of fusion, free energy of mixing, error propagation*

1. INTRODUCTION

A thermodynamic aqueous solubility is one of the major factors in determining bioavailability of orally administrated drugs.[1] According to a recent report, over 75% of oral drug development candidates have a low solubility based on the Biopharmaceutics Classification System (BCS).[2] Therefore, the solubility behavior of drugs remains one of the most challenging aspects in modern drug design. The focus on drug solubility improvement emerges early in drug discovery at the lead optimization stage. The work continues further on at formulation design and solid form selection stages in drug development. From practical and regulatory considerations, an accurate experimental measurement is the preferred means of thermodynamic aqueous solubility determination. However, it is impossible to perform solubility measurements for tens or hundreds of virtual compounds typically considered by a drug discovery team during the lead optimization step. In addition, *in silico* approaches may provide a tool for rational solubility improvement of poorly soluble drugs.[3] These considerations motivated a large number of computational model developments for aqueous solubility prediction employing a vast variety of methods.[4−11] For the most part, these models were using large sets of experimental measurements coupled with statistical approaches to build QSPR models. The current state-of-the-art solubility prediction accuracy of such models is considered to be approximately 0.7−1.0 log solubility (referred to mol/L) for drug-like molecules.[4] This solubility error is noticeably higher than a systematic error which may be introduced by variation of polymorphic forms of the compound in different measurements. It was demonstrated[12] that there is a 95% probability that a solubility ratio between a pair of polymorphs is less than 2-fold, which translates into a log S error of below 0.3.

Two sources of the relatively high uncertainty of the QSPR solubility prediction models were proposed. One is thought to be related to a poor selection of experimental data set for QSPR model training. However, in a recent study[13] it was demonstrated that the models derived from the most accurate solubility measurements are not more accurate than those derived from the "noisy" literature data. The latter data set was extracted from several different sources from the published literature, for which the experimental uncertainty is estimated to be 0.6−0.7 log S units (referred to mol/L). The authors concluded that "it is the deficiency of QSPR methods (algorithms and/or descriptor sets)...which is the limiting factor in accurately predicting aqueous solubility for pharmaceutical molecules". Another proposed reason for the limited accuracy of the QSPR solubility models is related to the fact that these models are not typically based on any fundamental consideration of physics. To solve this problem, a thermody-

Received: February 7, 2015
Revised: April 11, 2015
Accepted: April 16, 2015

DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

namic cycle solubility approach was proposed for prediction of solubility of crystalline drug-like molecules from the first principal calculations of sublimation and hydration free energies.[14−17] With this approach, a crystal packing contribution to the drug solubility required either experimentally determined crystal structure (retrospective analysis)[14−16] or crystal structure prediction calculations (allows prospective prediction).[17] Nevertheless, it was demonstrated that such an approach does not allow solubility predictions superior to the QSPR models.[14−16] The major source of error in the thermodynamic cycle predictions in these studies was not specifically defined.

The main purpose of the current study is to define the major limiting factor in the accuracy of the QSPR modeling of the thermodynamic intrinsic aqueous solubility of crystalline drug-like compounds. Based on the outcome of the study, the means of improvement of the accuracy of the thermodynamic solubility prediction were proposed.

2. COMPUTATIONAL APPROACH

In order to determine what appears to be the limiting factor in the accuracy of the QSPR modeling of the intrinsic aqueous thermodynamic solubility, one needs to get back to the basics of solubility phenomenon. While only aqueous solubility is considered below, the same considerations are applicable to solubility in any solvent. There are only two contributions to the intrinsic thermodynamic solubility of crystalline compounds—a crystal packing contribution (solid state property) and a liquid or molecular (nonsolid state) property contribution. The exact description of each of these contributions depends on the thermodynamic cycle, which is used to describe the drug solubility phenomenon. Indeed, thermodynamic solubility is defined as a difference between the values of two thermodynamic state functions and, therefore, should be independent of the path taken between these two states. Typically, two thermodynamic cycles are considered to describe crystalline compound solubility—a fusion cycle[18] and a sublimation cycle.[14] In the former case, the solid state contribution is described by a free energy of fusion, ΔG_{fus}, of the crystalline drug projected to ambient temperature. Another contribution to the fusion cycle is presented by a free energy of mixing, ΔG_{mix}, of the created supercooled liquid (melt) with water (Figure 1a). In the sublimation cycle the solid state contribution is presented by a free energy of sublimation, ΔG_{sub}, at ambient temperature. In that case the second contribution to the sublimation cycle is presented by a free energy of molecular hydration, ΔG_{hydr} (Figure 1b).

While both thermodynamic cycles are exact, the sublimation cycle is further away from the actual solubilization mechanism than the fusion cycle. Moreover, the experimental estimation of free energy of sublimation involves complicated experiments,[19] while measurements of fusion properties are quite routine.[20]

The log S determination based on both thermodynamic cycles can be presented in the following general form (referred to molar fractions):

$$\log S = -\frac{(\Delta G_{solid} + \Delta G_{nonsolid})}{\ln(10)\, RT} \tag{1}$$

where ΔG_{solid} denotes the ΔG_{fus} or ΔG_{sub} property, while $\Delta G_{nonsolid}$ denotes the ΔG_{mix} or ΔG_{hydr} property, respectively.

A standard deviation of log S defined by eq 1 may be derived from propagation of errors as[21]

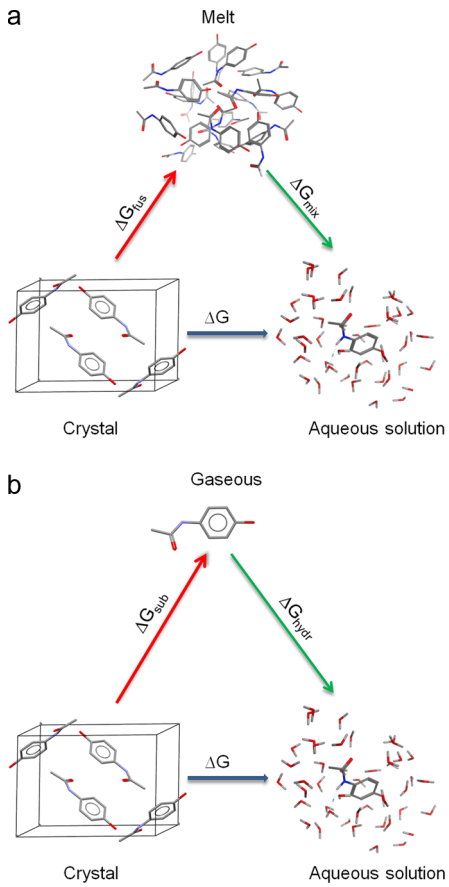


Figure 1. Fusion (a) and sublimation (b) thermodynamic cycles describing aqueous solubility of crystalline compounds.

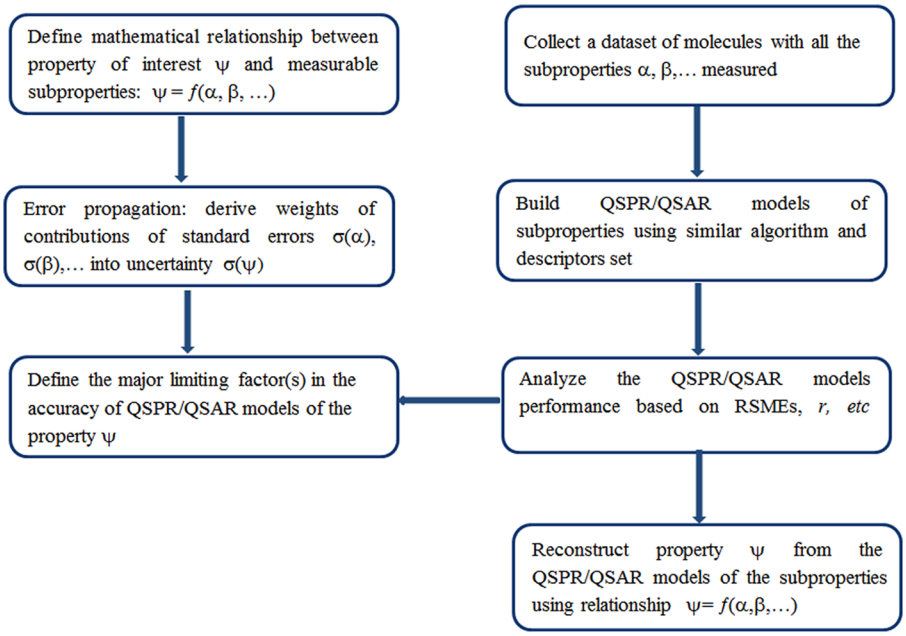
$$\sigma_{\log S} = \left[\left(\frac{\partial \log S}{\partial \Delta G_{solid}}\right)^2 \sigma^2_{\Delta G solid} + \left(\frac{\partial \log S}{\partial \Delta G_{nonsolid}}\right)^2 \sigma^2_{\Delta G nonsolid} + 2\left(\frac{\partial \log S}{\partial \Delta G_{solid}}\right)\left(\frac{\partial \log S}{\partial \Delta G_{nonsolid}}\right)\sigma_{solid\ nonsolid}\right]^{1/2}$$
$$= \frac{1}{\ln(10)\, RT}\left[\sigma^2_{\Delta G solid} + \sigma^2_{\Delta G nonsolid} + 2\sigma_{solid\ nonsolid}\right]^{1/2} \tag{2}$$

Here $\sigma_{\Delta G solid}$ and $\sigma_{\Delta G nonsolid}$ are standard errors (uncertainties) of solid and nonsolid contributions, respectively, while $\sigma_{solid\ nonsolid}$ is an estimated covariance between these two properties. Uncertainties of the solid and nonsolid properties being random and independent from each other, the third term of eq 2 (covariance) vanishes. Equation 2 demonstrates that contributions of the standard errors of the solid and nonsolid properties to the log S uncertainty are equally weighted and can be directly compared.

In general, it is possible to "measure" the thermodynamic intrinsic aqueous solubility and its standard error according to eqs 1 and 2, using the contributions of solid (ΔG_{fus} or ΔG_{sub}) and nonsolid (ΔG_{mix} or ΔG_{hydr}, respectively) properties, which are estimated by the corresponding QSPR models. Evaluation of the relative performance of these QSPR models would allow us to answer the question of which contribution, solid or nonsolid, is the limiting factor in the accuracy of the QSPR prediction of the intrinsic aqueous solubility. However, a data set of compounds of reasonable size, for which a pair of ΔG_{fus}

DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

Scheme 1. A General Workflow of the Proposed Approach of Defining the Major Limiting Factor(s) Regarding Accuracy of Quantitative Structure−Property or Structure−Activity Relationship (QSPR/QSAR) Models



and ΔG_{mix} or ΔG_{sub} and ΔG_{hydr} measurements are available for modeling, does not exist. In fact, the largest amount of measured data of drug-like compounds contains thermodynamic solubility and fusion (melting point, T_m, and heat of fusion, ΔH_{fus}) properties. Therefore, the following approach is proposed in this study.

In order to derive the ΔG_{fus} property at ambient temperature (T) from T_m and ΔH_{fus} measurements, the following approximations may be used:

$$\Delta G_{fus} \approx \Delta H_{fus}\left(1 - \frac{T}{T_m}\right) \tag{3}$$

$$\Delta G_{fus} \approx \Delta H_{fus}(T_m - T)\frac{T}{T_m^2} \tag{4}$$

$$\Delta G_{fus} \approx \Delta H_{fus}\frac{T}{T_m}\ln\frac{T_m}{T} \tag{5}$$

These approximations are based on various assumptions related to a difference in heat capacity, ΔC_p, between the compound liquid and solid phases.[12,22] It was recently demonstrated[12] that eqs 4 and 5 provide the best performance for drug-like molecules, and therefore, only these two approximations were adopted in the current study.

The ΔG_{mix} property was estimated backward from the relationship in eq 1 using accurate experimental measurements of the thermodynamic intrinsic aqueous solubility and ΔG_{fus} properties of the drug-like compounds. This made it possible to build QSPR models of the ΔG_{mix} and ΔG_{fus} properties and compare their accuracies based on test set predictions. The thermodynamic intrinsic aqueous solubility property was estimated from the ΔG_{mix} and ΔG_{fus} models according to eq 1 and compared with the experimental values.

A general workflow of the proposed approach is presented in Scheme 1.

3. DATA SET SELECTION

The data set of the drug-like compounds was created by combining publically available data on accurate intrinsic aqueous thermodynamic solubility and fusion properties (T_m and ΔH_{fus}) measurements. Perhaps the most reliable intrinsic aqueous thermodynamic solubility measurements were reported for 132 organic crystals.[23−25] Solubility data for an additional 18 drug compounds were taken from different sources.[26,27] The total solubility data set was further combined with available T_m and ΔH_{fus} properties, providing the final data set of solubility and fusion properties for 62 drug-like compounds (Table 1). The final set contained 69% of highly reliable CheqSol solubility measurements.[23−25] This data set was split into training (55) and test (7) sets using a maximum dissimilarity algorithm, which allowed selection of the representative subsets of the original data set. According to atom pair similarity analysis[28] with a threshold of 0.7, the selected test set compounds belonged to the chemical space defined by the training set. Therefore, no significant extrapolations were expected for validation of the QSPR models on the test set compounds.[29]

4. COMPUTATIONAL METHODS

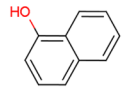
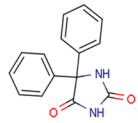
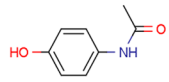
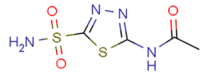
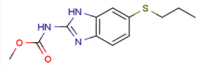
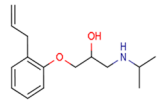
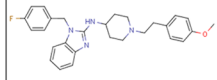
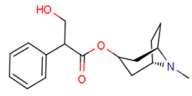
4.1. Modeling approaches. For consistency all the QSPR models were built based on the same training set compounds, using similar algorithms and descriptor sets, and validated against the similar test compounds.

Two different advanced machine learning regression methods—Random Forest (RF)[51] and Cubist[52] (https://www.rulequest.com)—were used in this study. Both methods were demonstrated to be suitable to model the data covering a very broad chemistry space with possible nonlinear relationships.[53−56] In addition, both methods utilize built-in tools for selection of the important descriptors and therefore are quite robust to the overfitting problem.

Cubist is a tool for generating rule-based QSPR models. Each rule is a conjunction of conditions associated with a linear

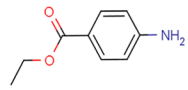
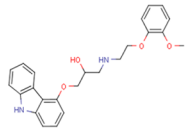
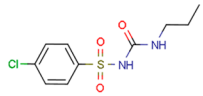
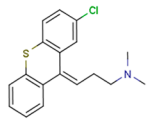
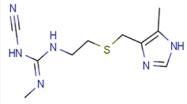
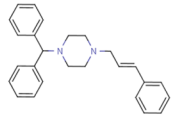
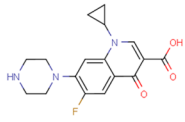
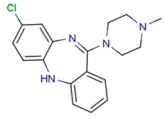
DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

Table 1. Experimental Solubility, log S (M), and Fusion Data, T_m and ΔH_{fus}, Used in This Study

Compound	Structure	$logS$(M)	T_m, K	ΔH_{fus}, kJ/mol	Test set	References
1-naphthol		-1.98	368.7	23.3	No	23,30
5,5-diphenylhydantoin		-3.86	568.8	40.1	No	23,27
acetaminophen		-1.06	442.0	28.1	No	23,31
acetazolamide		-2.43	532.2	28.6	No	23,32
albendazole		-6.01	451.3	98.6	No	26
alprenolol		-2.63	331.2	35.6	No	23,33
astemizole		-7.18	447.5	51.1	No	26
atropine		-2.00	388.5	35.5	No	23,34

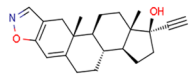
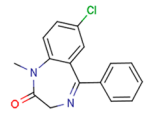
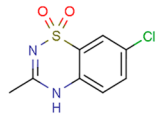
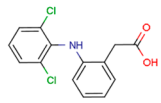
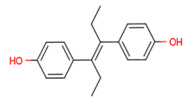
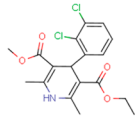
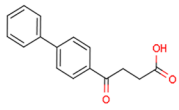
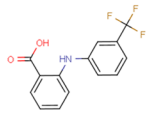
DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

Table 1. continued

Compound	Structure	*logS*(M)	T_m, K	ΔH_{fus}, kJ/mol	Test set	References
benzocaine		-2.33	362.6	24.6	No	24,35
carvedilol		-6.15	387.3	57.6	No	23,26
chlorpropamide		-3.25	401.0	25.7	No	23, 27
chlorprothixene, form II[a]		-5.87	370.3	27.8	No	23,36
cimetidine		-1.69	417.5	43.1	No	23,37
cinnarizine		-7.73	393.4	45.7	No	26
ciprofloxacin		-3.60	541.5	64.5	Yes	23,38
clozapine		-3.24	457.1	35.9	No	24, 27

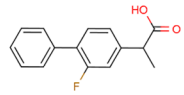
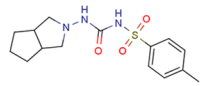
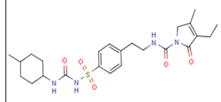
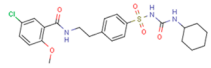
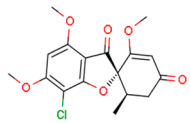
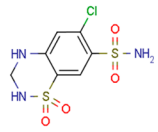
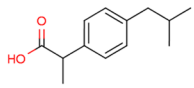
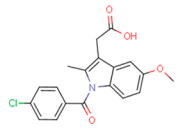
DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

Table 1. continued

Compound	Structure	*logS*(M)	T_m, K	ΔH_{fus}, kJ/mol	Test set	References
danazol		-7.44	501.8	35.5	No	26
diazepam		-3.85	404.8	24.7	No	27
diazoxide		-3.36	600.4	34.1	No	23,27
diclofenac		-5.46	453.0	40.9	No	23, 39
diethylstilbestrol		-4.42	451.0	33.4	No	24, 27
felodipine		-6.56	412.3	34.8	No	26
fenbufen		-5.19	459.3	46.2	No	27
flufenamic acid		-5.36	405.0	26.7	Yes	23,40

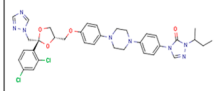
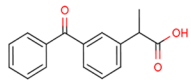
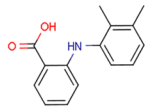
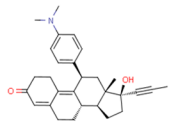
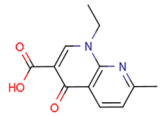
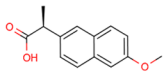
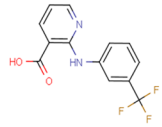
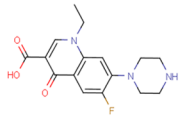
DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

Table 1. continued

Compound	Structure	*logS*(M)	T_m, K	ΔH_{fus}, kJ/mol	Test set	References
flurbiprofen		-4.15	386.7	27.9	No	23,41
gliclazide		-4.07	444.5	44.2	No	27
glimepiride		-7.90	485.6	53.3	No	26
glyburide		-7.05	446.8	46.3	No	26
griseofulvin		-4.83	491.1	44.7	No	27
hydrochlorothiazide		-2.68	540.8	33.6	No	24,27
ibuprofen		-3.59	346.4	26.6	No	23,27
indomethacin		-4.61	433.0	37.9	No	25,27

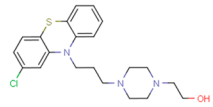
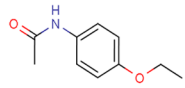
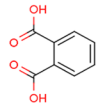
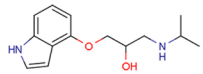
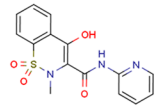
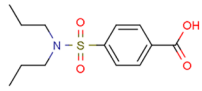
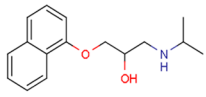
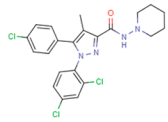
DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

Table 1. continued

Compound	Structure	*logS*(M)	T_m, K	ΔH_{fus}, kJ/mol	Test set	References
itraconazole		-8.48	438.5	69.9	No	26
ketoprofen		-3.21	368.0	37.3	No	24,27
mefenamic acid		-6.74	503.5	38.7	No	23,42
mifepristone		-5.75	467.0	31.7	No	27
nalidixic acid		-3.61	501.9	35.9	No	23,43
naproxen		-4.50	428.8	34.2	No	23,27
niflumic acid		-4.58	478.0	36.5	No	23,40
norfloxacin		-2.76	492.6	32.4	No	23,44

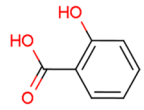
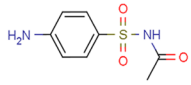
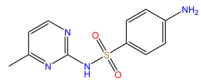
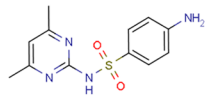
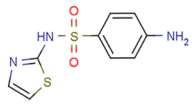
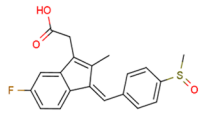
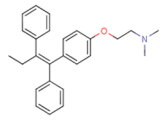
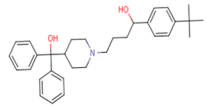
DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

Table 1. continued

Compound	Structure	*logS*(M)	T_m, K	ΔH_{fus}, kJ/mol	Test set	References
perphenazine		-4.62	370.0	41.8	No	27
phenacetin		-2.48	407.4	34.1	No	27
phthalic acid, form I		-1.61	463.5	36.5	Yes	23,45
pindolol		-3.79	423.6	60.6	No	23,46
piroxicam		-4.80	473.4	36.3	No	23,27
probenecid		-4.86	472.0	40.9	No	24,27
propranolol		-3.49	365.5	43.5	No	23,33
rimonabant		-7.01	427.9	36.1	No	26

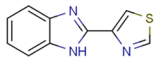
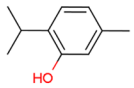
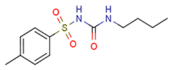
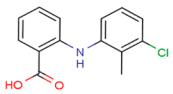
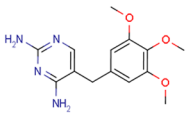
DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

Table 1. continued

Compound	Structure	*logS*(M)	T_m, K	ΔH_{fus}, kJ/mol	Test set	References
salicylic acid		-1.93	432.5	23.0	No	24,47
sulfacetamide		-1.52	455.2	29.8	No	23,48
sulfamerazine		-3.12	508.5	41.3	No	24,48
sulfamethazine		-2.73	469.0	39.2	Yes	23,48
sulfathiazole		-2.69	447.0	29.5	Yes	23,31
sulindac, form I[a]		-3.68	460.2	33.4	No	23,27
tamoxifen		-8.54	371.0	34.0	No	26
terfenadine		-7.94	422.8	58.1	No	24,26
testosterone		-4.20	426.5	28.2	No	27

DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

Table 1. continued

Compound	Structure	*logS*(M)	T_m, K	ΔH_{fus}, kJ/mol	Test set	References
thiabendazole		-3.48	573.2	35.2	No	24,49
thymol		-2.19	324.2	22.0	No	23,50
tolbutamide		-3.46	400.2	24.5	Yes	24,31
tolfenamic acid		-7.87	485.3	41.2	Yes	26
trimethoprim		-2.95	472.9	49.8	No	23, 27

[a]The reported solubility ratio between polymorphic forms I and II appeared to be untypically high, which may be an indication that the lower soluble form is a hydrate. Therefore, the highest solubility value was selected for the data set.

expression. Cubist can also use a boosting-like scheme called committees; each is made up of several rule-based models. Predictions made by each member of a committee for a target value are averaged to give the final prediction. The importance of the individual descriptors can be estimated from the frequency of their use in the final model. In this study, 20-member committee models with a maximum of 20 rules were built for each property of interest.

Random Forest (RF) is an ensemble of n_{tree} unpruned decision trees created by using bootstrap samples of the training data and random subset of m_{try} variables to define the best split at each node.[51] The bootstrap sample used during tree growth is a random selection with replacement from the molecules in the training set. Model performance for each tree is internally assessed with the prediction error of the data left-out in the bootstrap procedure (out-of-bag data). The average of these results for all trees provides an in situ cross-validation (out-of-bag validation). The RF prediction of new data is made by averaging the individual predictions of all the trees in the forest. The number of trees n_{tree} in the RF in this study was set to 1000. The m_{try} parameter was set to a default value of one-third of the whole descriptor set.[56]

4.2. Descriptors. Both Cubist and Random Forest methods utilize intrinsic (built-in) selection of important descriptors and are generally not sensitive to the presence of irrelevant features. Therefore, a relatively large set of descriptors was used in this study, including Dragon descriptors (www.talete.mi.it/products/dragon_description.htm), VolSurf+ descriptors (http://www.moldiscovery.com/software/vsplus/), and a set of in-house SMARTS keys.[57,58] The total number of descriptors was decreased by the exclusion of zero-variance and highly correlated descriptors—in the cases where the Pearson pairwise correlation coefficient exceeded the value of 0.85, one descriptor of the pair was removed.

4.3. Model selection and comparison. The model performance was evaluated using the predictions made for the test set. Five statistical measures were evaluated to compare the models: root-mean-square error, RMSE; a relative standard deviation, RSD; Pearson correlation coefficient, r; p value and Fisher's z-test (https://sites.google.com/site/fundamentalstatistics).

RSME determines an absolute standard error (σ) of predictions:

DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

$$RSME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred})^2} \quad (6)$$

Here n is the data set size, and y_i^{obs} and y_i^{pred} are the observed and predicted values for molecule i.

The relative standard deviation, RSD, is defined in percentage units as

$$RSD = 100\frac{RSME}{|\overline{y_i^{obs}}|} \quad (7)$$

Here $|\overline{y_i^{obs}}|$ is the mean absolute value of the observed property y. Therefore, the RSD value demonstrates how well property y is predicted by the corresponding QSPR equation when compared to the mean value of the property.

The Pearson correlation coefficient measures the strength and direction of a linear relationship between observed and predicted properties and may vary between values of +1 and −1:

$$r = \frac{\sum_{i=1}^{n}(y_i^{obs} - \overline{y_i^{obs}})(y_i^{pred} - \overline{y_i^{pred}})}{\sqrt{\sum_{i=1}^{n}(y_i^{obs} - \overline{y_i^{obs}})^2}\sqrt{\sum_{i=1}^{n}(y_i^{pred} - \overline{y_i^{pred}})^2}} \quad (8)$$

The performance of each model in application to the test set was also validated by a p value, which measures the probability of the null hypothesis that the observed and predicted properties are not related. Consequently the smaller the p value, the greater probability that the null hypothesis may be rejected and the model may be used to predict the corresponding property.

In addition, a statistical significance of difference between the Pearson coefficients of two models of interest was evaluated using Fisher's z-test. For this the correlation coefficients r were transformed into r' according to the relationship: $r' = 0.5\ \ln((1 + r)/(1 - r))$. A z-score for the difference between the r' value for models 1 and 2 was evaluated as

$$z = \sqrt{(n-3)}\,(r_1' - r_2') \quad (9)$$

The difference between the Pearson coefficients of the two models is considered to be statistically significant if the absolute value of the z-score exceeded a critical value of 1.96, corresponding to a 95% confidence level.

5. RESULTS AND DISCUSSIONS

An initial analysis of the relationship between experimental log S, ΔG_{fus}, and ΔG_{mix} properties demonstrates that the ΔG_{mix} is the determinant factor of the intrinsic thermodynamic solubility for the model compounds used in this study (Figure 2a). The Pearson correlation coefficient between these properties is as high as 0.96, with an outlier brick dust compound albendazole for which the contribution of the ΔG_{fus} is the limiting factor of its poor aqueous solubility. A correlation between experimental log S and ΔG_{fus} properties is quite poor, displaying an r value of only 0.45 (Figure 2b). These observations demonstrate that the solubility trend in the compound set selected for the current study is controlled predominantly by nonsolid state properties.

The results of statistical performance of the QSPR models of the ΔG_{fus}, ΔG_{mix}, and log S properties in application to the training and test sets are presented in Tables 2 and 3, respectively. The Cubist algorithm gave a better overall performance than the RF approach. Therefore, the analysis of the results presented below is mostly based on the Cubist QSPR models, though similar conclusions could be derived from the RF models. The five most important descriptors for each Cubist model are listed in Table 4. As could be expected,

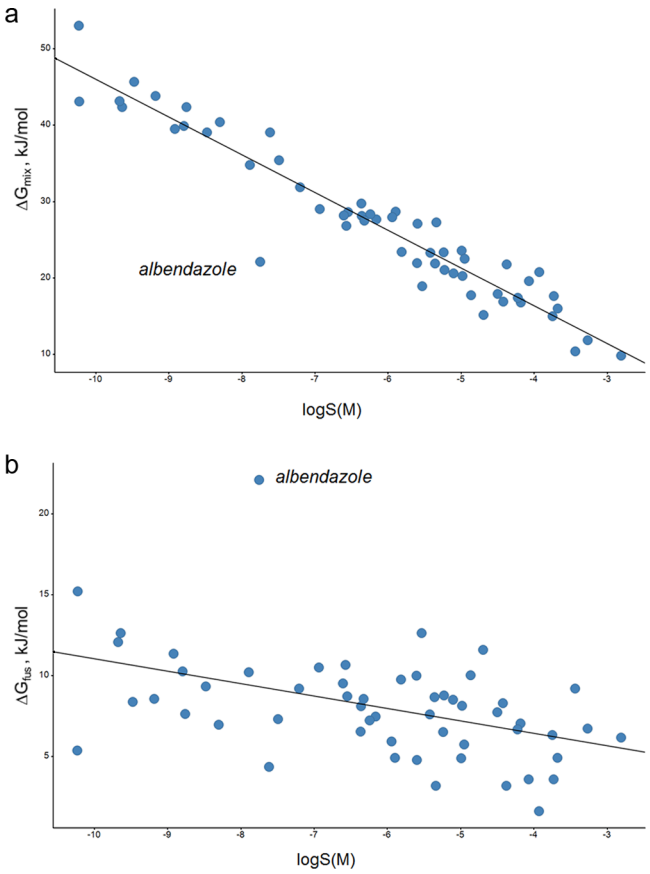


Figure 2. Correlation between experimental log S *vs* (a) ΔG_{mix} and (b) ΔG_{fus} properties for the total compound set. ΔG_{fus} properties were calculated based on eq 4.

Table 2. Statistical Performance of the QSPR Models of the ΔG_{fus}, ΔG_{mix}, and log S Properties in Application to the Training Set[a]

Based on Eq	Property	r	RSME	|RSD| (%)	QSPR method
4	ΔG_{fus}	0.961	1.1	8.1	Cubist
		0.980	1.2	7.8	RF
	ΔG_{mix}	0.999	0.5	2.0	Cubist
		0.983	2.5	9.2	RF
	log S_{TC}	0.994	0.2	3.5	Cubist
		0.982	0.5	7.8	RF
5	ΔG_{fus}	0.946	1.5	15.2	Cubist
		0.983	1.5	15.2	RF
	ΔG_{mix}	0.997	0.8	3.2	Cubist
		0.981	2.7	10.8	RF
	log S_{TC}	0.985	0.3	5.5	Cubist
		0.979	0.5	8.5	RF
NA	log S_{QSPR}	0.996	0.2	2.7	Cubist
		0.983	0.5	7.6	RF

[a]The log S_{QSPR} property was fitted to the solubility training set, while the log S_{TC} property was calculated *via* a thermodynamic cycle approach (1), using corresponding ΔG_{fus} and ΔG_{mix} QSPR models. RSME values of free energies are presented in kJ/mol. The log S values are referred to molar fractions.

DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

Table 3. Statistical Performance of the QSPR Models of the ΔG_{fus}, ΔG_{mix}, and log S Properties in Application to the Test Set[a]

Based on Eq	Property	r	p	RSME	|RSD| (%)	QSPR method
4	ΔG_{fus}	0.20	0.665	3.8	44.6	Cubist
		0.22	0.635	3.5	41.1	RF
	ΔG_{mix}	0.96	0.007	6.6	27.7	Cubist
		0.92	0.004	8.3	34.9	RF
	log S_{TC}	0.97	0.000	0.8	13.7	Cubist
		0.94	0.001	1.3	22.6	RF
5	ΔG_{fus}	0.22	0.636	5.1	48.0	Cubist
		0.20	0.670	4.9	45.8	RF
	ΔG_{mix}	0.95	0.001	5.1	23.6	Cubist
		0.90	0.006	8.5	39.7	RF
	log S_{TC}	0.96	0.001	0.7	12.6	Cubist
		0.94	0.001	1.3	22.2	RF
NA	log S_{QSPR}	0.91	0.005	1.0	17.1	Cubist
		0.93	0.003	1.3	22.4	RF

[a]The log S_{QSPR} property was fitted to the solubility training set, while log S_{TC} property was calculated *via* thermodynamic cycle approach (1), using corresponding ΔG_{fus} and ΔG_{mix} QSPR models. RSME values of free energies are presented in kJ/mol. The log S values are referred to molar fractions.

the most important descriptor of the ΔG_{mix} and log S_{QSPR} models is related to the octanol−water partition coefficient, log P.

It was found that the standard errors (RSME) of the best QSPR models of the ΔG_{fus} and ΔG_{mix} properties in application to both training (Table 2) and test (Table 3) sets are quite comparable. However, these absolute error parameters may provide a misleading message. According to the r, p, and RSD values for the test set validation (Table 3), only the ΔG_{mix} model is statistically significant and is the truly predictive one ($r > 0.9$, $p < 0.01$, RSD < 30%). In contrast, all the ΔG_{fus} models display extremely low r coefficients of about 0.2, high p values of greater than 0.6, and high RSD values of 45−48%. Therefore, the ΔG_{fus} models are not statistically significant and there is no structure−property relationship which could be derived from these QSPR models. Comparison of the statistical performance of the ΔG_{fus} models in application to the training (Table 2) and test (Table 3) sets demonstrates that these models are overfitted. These observations are indicative that all the QSPR approaches with the current selection of machine learning algorithms and the descriptor set are treating the ΔG_{fus} property as a noise.

The above conclusions were further supported by the Fisher's z-test of the statistical significance of the difference between the Pearson coefficients of the ΔG_{fus} and ΔG_{mix} QSPR models in the application to the test set. The absolute values of the z-scores (eq 9) for the models based on eqs 4 and 5 are equal to 3.5 and 3.2, respectively, significantly exceeding the critical value of 1.96.

Consequently, in this study the ΔG_{mix} property was demonstrated to be a trainable and predictive one, while the ΔG_{fus} property is shown to be nonpredictive. Therefore, it is the ΔG_{fus} (solid state) contribution which is the limiting factor in the accuracy and predictive power of QSPR models of intrinsic thermodynamic solubility. Based on this finding, it is possible to assume that the up-to-date successes of the log S QSPR modeling are related predominantly to an improved description of the nonsolid term. That assumption may be indirectly supported by a noticeably better performance of the log S_{TC} solubility model, built via the thermodynamic cycle (eq 1), relative to the log S_{QSPR} solubility model built directly from the solubility observations (Table 3). Indeed, one possible explanation of this observation is that the computational algorithms and descriptor set used in this study were able to do a better job in fitting the ΔG_{mix} property by itself, rather than fitting this property in combination with the "unfittable" ΔG_{fus} contribution to the thermodynamic solubility observations.

The failure of QSPR modeling of the ΔG_{fus} property is related to the lack of physically meaningful descriptor sets relevant to the fusion phenomenon.[59] Upon fusion of a crystalline compound, it is only periodicity (or long-range order) that is being destroyed, while close-range intermolecular interactions still remain in the melted form. However, the majority of the currently available descriptors is rather applicable to description of close-range, or even close contact (surface properties descriptors), interactions, which are relevant

Table 4. Five Most Important Descriptors in Each Cubist QSPR Model

Property	Descriptor symbol	Importance[a] (%)	Description
ΔG_{mix}	Dragon ALOGP2	49	Squared Ghose−Crippen octanol−water partition coeff (log P^2)
	Dragon RDF050m	30	Radial Distribution Function-050/weighted by mass
	Dragon Mor28e	19	Signal 28/weighted by Sanderson electronegativity
	Dragon MATS7p	18	Moran autocorrelation of lag 7 weighted by polarizability
	Dragon VEA1	16	Eigen vector coefficient sum from adjacency matrix
ΔG_{fus}	[N,n]~*~[N,n,O,o]	38	SMARTS key
	Dragon Mor07m	23	Signal 07/weighted by mass
	Dragon RDF145u	20	Radial Distribution Function-145/unweighted
	VolSurf+ DRACDO	18	Dry-Acceptor−Donor triplet pharmacophoric descriptor
	Dragon ASP	16	Asphericity
log S_{QSPR}	Dragon ALOGP2	47	Squared Ghose−Crippen octanol−water partition coeff (log P^2)
	Dragon BEHm4	28	Highest eigenvalue no. 4 of Burden matrix/weighted by atomic masses
	Volsurf+ CD3	23	The ratio of the hydrophobic volume over the total molecular surface at the 3rd energy level.
	Volsurf+ LgD6	23	The log P (octanol/water) computed via the sum of the log P and the fraction of every species at pH 6.
	Volsurf+ LgS7	23	The logarithm of solubilities computed at pH 7.

[a]For the Cubist model the descriptor importance is presented via percentage of cases in the training data for which the descriptor appears in a model of an applicable rule.

DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

to many physiologically related phenomena in the liquid state. The long-range order in drug-like crystals is imposed by a combination of the following major long-, medium-, and short-range interactions: extended H-bonding network throughout the crystal; electrostatic interaction of preferably ordered molecular dipole moments, π-stacking interaction, and a shape complementarity (close packing; molecular symmetry).

In the case of the sublimation thermodynamic cycle of the aqueous solubility (Figure 1b), the long-range-order is typically a smaller contribution to the sublimation energy. The sublimation phenomenon (which was not closely considered in this study due to the lack of data) is described by destruction of all of the intermolecular interactions in the crystalline state, among which the short-range interactions are the strongest. This consideration accounts for the multiple successes in the QSPR modeling of the sublimation enthalpy (or lattice energy) of organic crystals.[3,60−62] However, the long-range order still remains an inherent contribution to the sublimation energy and therefore behaves as the limiting factor in accuracy of the QSPR models of the sublimation energy. Indeed, it was found in a recent study that inclusion of the lattice energy (sublimation enthalpy) descriptor, which captures only short-range interactions, did not significantly improve the quality of the aqueous solubility QSPR models.[11] In addition to this, there are important entropic contributions, which are noticeably more pronounced in absolute value in the sublimation rather than in the fusion thermodynamic cycle and should be appropriately considered for an accurate QSPR modeling of sublimation free energy, ΔG_{sub}.[15,63] As a result, all these considerations support the conclusion that the solid state contribution is the limiting factor in accuracy and predictive power of the QSPR models of the intrinsic aqueous solubility of crystalline drug-like compounds independently of the thermodynamic cycle considered.

From general considerations, an impact of the limitations imposed by the poor performance of the ΔG_{solid} QSPR models on the accuracy of the thermodynamic solubility predictions would depend on the interplay between the solid and nonsolid contributions to the log S in each specific case. For example, for the majority of the drug-like compounds considered in the current study, the ΔG_{mix} property displays the dominant contribution to the log S (Figure 2). The above consideration allowed mitigation of the impact of the failure of the ΔG_{fus} model on the absolute accuracy of the log S QSPR predictions. However, in the case of the brick dust compounds for which the intrinsic thermodynamic solubility is driven by the noticeably higher solid state contributions, the limitations of predicting this property are expected to have a more dramatic effect on the performance of the log S models. This effect can be understood in terms of projected higher RSME values of the ΔG_{fus} QSPR predictions, which may reach 45−48% of the mean absolute values (Table 3).

The above considerations outline the limited nature of the chemical and physical space of the solid drugs considered in this work. Future study on a diverse set of drug-like compounds will address these limitations.

6. CONCLUSION

The QSPR/QSAR modeling is one of the most popular computational approaches for a fast and reliable estimation of various properties and end-points related to the pharmaceutical industry, such as for example ADMET properties including aqueous solubility. One of the typical justifications of choosing such a model relative to any other is related to the complexity of the underlying phenomenon, for which a direct first principle type approach is not applicable. However, in order for the QSPR/QSAR model to be more accurate and predictive, it is important to incorporate into selection of descriptor sets and computational schemes as much physical or mechanistic information as possible.

In order to define the major limiting factors of the accuracy of the solubility QSPR models it was proposed in this study to get back to the basics of the solubility phenomenon and to deconvolute the thermodynamic solubility into the solid state, ΔG_{fus}, and nonsolid state, ΔG_{mix}, contributions. This approach allowed indirect "measurement" of the intrinsic thermodynamic aqueous solubility based on the developed QSPR models of the ΔG_{fus} and ΔG_{mix} properties. Propagation of error consideration showed that contributions of standard errors of the ΔG_{fus} and ΔG_{mix} properties to the log S uncertainty are equally weighted and can be directly compared. Analysis of the relative performances of the ΔG_{fus} and ΔG_{mix} QSPR models demonstrated that it is the solid state contribution which is the limiting factor in the accuracy and predictive power of the QSPR models of the thermodynamic intrinsic solubility. The performed analysis outlines a necessity of development of new descriptor sets for an accurate description of the long-range order (periodicity) phenomenon in the crystalline state.

The proposed approach (Scheme 1) of the prediction of an end-point or property of interest based on QSPR/QSAR models of contributing properties may be generalized to other applications in the pharmaceutical industry, such as, for example, ADME properties. That will allow us to define the limiting factors in the accuracy of predictions. In addition, the final model may be superior to the one built directly from the end-point observations.

■ ASSOCIATED CONTENT

Supporting Information

SMILES strings and all the experimental properties listed in Table 1. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: yuriy.a.abramov@pfizer.com.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The author is grateful to Mr. Brian Samas for the valuable discussions. The author would like to thank Dr. Simone Sciabola for assistance in interpretation of the VolSurf+ descriptors.

■ REFERENCES

(1) Yu, L. X.; Amidon, G. L.; Polli, J. E.; Zhao, H.; Mehta, M. U.; Conner, D. P.; Shah, V. P.; Lesko, L. J.; Chen, M.-L.; Lee, V. H. Biopharmaceutics classification system: the scientific basis for biowaiver extensions. *Pharm. Res.* **2002**, *19* (7), 921−925.

(2) Di, L.; Kerns, E. H.; Carter, G. T. Drug-like property concepts in pharmaceutical design. *Curr. Pharmaceut. Des.* **2009**, *15* (19), 2184−2194.

(3) Docherty, R. P.; Pencheva, K.; Abramov, Y. A. Low solubility in drug development: de-convoluting the relative importance of solvation and crystal packing. *J. Pharm. Pharm.* **2015**, DOI: 10.1111/jphp.12393.

DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

(4) Jorgensen, W. L.; Duffy, E. M. Prediction of drug solubility from structure. *Adv. Drug Delivery Rev.* **2002**, *54* (3), 355−366.
(5) Gudmundsson, O.; Venkatesh, S. Strategies for in silico and experimental screening of physicochemical properties. *Biotechnol.: Pharm. Aspects* **2004**, *1*, 393−412.
(6) Delaney, J. S. Predicting aqueous solubility from structure. *Drug Discovery Today* **2005**, *10* (4), 289−295.
(7) Johnson, S. R.; Zheng, W. Recent progress in the computational prediction of aqueous solubility and absorption. *AAPS J.* **2006**, *8* (1), E27−E40.
(8) Johnson, S. R.; Chen, X.-Q.; Murphy, D.; Gudmundsson, O. A computational model for the prediction of aqueous solubility that includes crystal packing, intrinsic solubility, and ionization effects. *Mol. Pharmaceutics* **2007**, *4* (4), 513−523.
(9) Balakin, K. V.; Savchuk, N. P.; Tetko, I. V. In silico approaches to prediction of aqueous and DMSO solubility of drug-like compounds: trends, problems and solutions. *Curr. Med. Chem.* **2006**, *13* (2), 223−241.
(10) Wang, J.; Hou, T. Recent advances on aqueous solubility prediction. *Comb. Chem. High Throughput Screening* **2011**, *14* (5), 328−338.
(11) Salahinejad, M.; Le, T. C.; Winkler, D. A. Aqueous solubility prediction: Do crystal lattice interactions help? *Mol. Pharmaceutics* **2013**, *10* (7), 2757−2766.
(12) Abramov, Y. A.; Pencheva, K. Thermodynamics and relative solubility prediction of polymorphic systems. In *Chemical Engineering in the Pharmaceutical Industry: R&D to Manufacturing*; am Ende, D. J., Ed.; John Wiley and Sons: 2011; pp 477−490.
(13) Palmer, D. S.; Mitchell, J. B. Is experimental data quality the limiting factor in predicting the aqueous solubility of druglike molecules? *Mol. Pharmaceutics* **2014**, *11* (8), 2962−2972.
(14) Palmer, D. S.; Llinàs, A.; Morao, I.; Day, G. M.; Goodman, J. M.; Glen, R. C.; Mitchell, J. B. Predicting intrinsic aqueous solubility by a thermodynamic cycle. *Mol. Pharmaceutics* **2008**, *5* (2), 266−279.
(15) Palmer, D. S.; McDonagh, J. L.; Mitchell, J. B.; van Mourik, T.; Fedorov, M. V. First-Principles Calculation of the Intrinsic Aqueous Solubility of Crystalline Druglike Molecules. *J. Chem. Theory Comput.* **2012**, *8* (9), 3322−3337.
(16) McDonagh, J. L.; Nath, N.; De Ferrari, L.; Van Mourik, T.; Mitchell, J. B. Uniting Cheminformatics and Chemical Theory To Predict the Intrinsic Aqueous Solubility of Crystalline Druglike Molecules. *J. Chem. Inf. Model.* **2014**, *54* (3), 844−856.
(17) Abramov, Y. A. Computational modeling of drug solubility. *IQPC Improving Solubility Forum*, Princeton, NJ, USA, 2008
(18) Grant, D. J.; Higuchi, T. *Solubility behavior of organic compounds*; Wiley: New York, 1990.
(19) Zielenkiewicz, X.; Perlovich, G.; Wszelaka-Rylik, M. The vapour pressure and the enthalpy of sublimation: determination by inert gas flow method. *J. Therm. Anal. Calorim.* **1999**, *57* (1), 225−234.
(20) Clas, S.-D.; Dalton, C. R.; Hancock, B. C. Differential scanning calorimetry: applications in drug development. *Pharm. Sci. Technol. Today* **1999**, *2* (8), 311−320.
(21) Taylor, J. *Introduction to error analysis, the study of uncertainties in physical measurements*; University Science Books: 1997.
(22) Hoffman, J. D. Thermodynamic driving force in nucleation and growth processes. *J. Chem. Phys.* **1958**, *29* (5), 1192−1193.
(23) Llinàs, A.; Glen, R. C.; Goodman, J. M. Solubility challenge: can you predict solubilities of 32 molecules using a database of 100 reliable measurements? *J. Chem. Inf. Model.* **2008**, *48* (7), 1289−1303.
(24) Hopfinger, A. J.; Esposito, E. X.; Llinàs, A.; Glen, R. C.; Goodman, J. M. Findings of the challenge to predict aqueous solubility. *J. Chem. Inf. Model.* **2008**, *49* (1), 1−5.
(25) Comer, J.; Judge, S.; Matthews, D.; Towes, L.; Falcone, B.; Goodman, J.; Dearden, J. The intrinsic aqueous solubility of indomethacin. *ADMET and DMPK* **2014**, *2* (1), 18−32.
(26) Bergström, C. A.; Wassvik, C. M.; Johansson, K.; Hubatsch, I. Poorly soluble marketed drugs display solvation limited solubility. *J. Med. Chem.* **2007**, *50* (23), 5858−5862.
(27) Wassvik, C. M.; Holmén, A. G.; Bergström, C. A.; Zamora, I.; Artursson, P. Contribution of solid-state properties to the aqueous solubility of drugs. *Eur. J. Pharm. Sci.* **2006**, *29* (3), 294−305.
(28) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure-activity studies: definition and applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25* (2), 64−73.
(29) Sheridan, R. P.; Feuston, B. P.; Maiorov, V. N.; Kearsley, S. K. Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (6), 1912−1928.
(30) Rai, U.; Pandey, P.; Rai, R. Physical chemistry of binary organic eutectic and monotectic alloys; 1, 2, 4, 5-tetrachlorobenzene−α-naphthol and TCB−resorcinol systems. *Mater. Lett.* **2002**, *53* (1), 83−90.
(31) Yu, L. Inferring thermodynamic stability relationship of polymorphs from melting data. *J. Pharm. Sci.* **1995**, *84* (8), 966−974.
(32) Baraldi, C.; Gamberini, M.; Tinti, A.; Palazzoli, F.; Ferioli, V. Vibrational study of acetazolamide polymorphism. *J. Mol. Struct.* **2009**, *918* (1), 88−96.
(33) Li, Z. J.; Zell, M. T.; Munson, E. J.; Grant, D. J. Characterization of racemic species of chiral drugs using thermal analysis, thermodynamic calculation, and structural studies. *J. Pharm. Sci.* **1999**, *88* (3), 337−346.
(34) Domanska, U.; Pobudkowska, A.; Pelczarska, A.; Gierycz, P. pKa and Solubility of Drugs in Water, Ethanol, and 1-Octanol. *J. Phys. Chem. B* **2009**, *113* (26), 8941−8947.
(35) Wassvik, C. M.; Holmén, A. G.; Draheim, R.; Artursson, P.; Bergström, C. A. Molecular characteristics for solid-state limited solubility. *J. Med. Chem.* **2008**, *51* (10), 3035−3039.
(36) Domalski, E. S.; Hearing, E. D. Heat capacities and entropies of organic compounds in the condensed phase. Volume III. *J. Phys. Chem. Ref. Data* **1996**, *25* (1), 1−525.
(37) Crafts, P. The role of solubility modeling and crystallization in the design of active pharmaceutical ingredients. *Comput.-Aided Chem. Eng.* **2007**, *23*, 23−85.
(38) Yu, X.; Zipp, G. L.; Davidson, G. R., III The effect of temperature and pH on the solubility of quinolone compounds: estimation of heat of fusion. *Pharm. Res.* **1994**, *11* (4), 522−527.
(39) Surov, A. O.; Voronin, A. P.; Manin, A. N.; Manin, N. G.; Kuzmina, L. G.; Churakov, A. V.; Perlovich, G. L. Pharmaceutical Cocrystals of Diflunisal and Diclofenac with Theophylline. *Mol. Pharmaceutics* **2014**, *11* (10), 3707−3715.
(40) Perlovich, G. L.; Surov, A. O.; Bauer-Brandl, A. Thermodynamic properties of flufenamic and niflumic acids—Specific and non-specific interactions in solution and in crystal lattices, mechanism of solvation, partitioning and distribution. *J. Pharm. Biomed. Anal.* **2007**, *45* (4), 679−687.
(41) Henck, J. O.; Kuhnert-Brandstatter, M. Demonstration of the terms enantiotropy and monotropy in polymorphism research exemplified by flurbiprofen. *J. Pharm. Sci.* **1999**, *88* (1), 103−108.
(42) Surov, A. O.; Terekhova, I. V.; Bauer-Brandl, A.; Perlovich, G. L. Thermodynamic and structural aspects of some fenamate molecular crystals. *Cryst. Growth Des.* **2009**, *9* (7), 3265−3272.
(43) Romero, S.; Bustamante, P.; Escalera, B.; Mura, P.; Cirri, M. Influence of solvent composition on the solid phase at equilibrium with saturated solutions of quinolones in different solvent mixtures. *J. Pharm. Biomed. Anal.* **2004**, *35* (4), 715−726.
(44) Oliveira, P.; Bernardi, L.; Murakami, F.; Mendes, C.; Silva, M. A. Thermal characterization and compatibility studies of norfloxacin for development of extended release tablets. *J. Therm. Anal. Calorim.* **2009**, *97* (2), 741−745.
(45) Sabbah, R.; Perez, L. Étude thermodynamique des acides phtalique, isophtalique et téréphtalique. *Can. J. Chem.* **1999**, *77* (9), 1508−1513.
(46) Perlovich, G. L.; Volkova, T. V.; Bauer-Brandl, A. Thermodynamic study of sublimation, solubility, solvation, and distribution processes of atenolol and pindolol. *Mol. Pharmaceutics* **2007**, *4* (6), 929−935.

DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX

(47) Pena, M.; Escalera, B.; Reíllo, A.; Sánchez, A.; Bustamante, P. Thermodynamics of cosolvent action: phenacetin, salicylic acid and probenecid. *J. Pharm. Sci.* **2009**, *98* (3), 1129−1135.
(48) Martínez, F.; Gómez, A. Estimation of the solubility of sulfonamides in aqueous media from partition coefficients and entropies of fusion. *Phys. Chem. Liq.* **2002**, *40* (4), 411−420.
(49) Muela, S.; Escalera, B.; Peña, M. Á.; Bustamante, P. Influence of temperature on the solubilization of thiabendazole by combined action of solid dispersions and co-solvents. *Int. J. Pharm.* **2010**, *384* (1), 93−99.
(50) Chickos, J. S.; Braton, C. M.; Hesse, D. G.; Liebman, J. F. Estimating entropies and enthalpies of fusion of organic compounds. *J. Org. Chem.* **1991**, *56* (3), 927−938.
(51) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.
(52) Quinlan, J. R. In *Combining Instance-Based and Model-Based Learning*; ICML: 1993; pp 236−243.
(53) Gao, H.; Yao, L.; Mathieu, H. W.; Zhang, Y.; Maurer, T. S.; Troutman, M. D.; Scott, D. O.; Ruggeri, R. B.; Lin, J. In silico modeling of nonspecific binding to human liver microsomes. *Drug Metab. Dispos.* **2008**, *36* (10), 2130−2135.
(54) Gupta, R. R.; Gifford, E. M.; Liston, T.; Waller, C. L.; Hohman, M.; Bunin, B. A.; Ekins, S. Using open source computational tools for predicting human metabolic stability and additional absorption, distribution, metabolism, excretion, and toxicity properties. *Drug Metab. Dispos.* **2010**, *38* (11), 2083−2090.
(55) Palmer, D. S.; O'Boyle, N. M.; Glen, R. C.; Mitchell, J. B. Random forest models to predict aqueous solubility. *J. Chem. Inf. Model.* **2007**, *47* (1), 150−158.
(56) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1947−1958.
(57) Lee, P. H.; Cucurull-Sanchez, L.; Lu, J.; Du, Y. J. Development of in silico models for human liver microsomal stability. *J. Comput.-Aided Mol. Des.* **2007**, *21* (12), 665−673.
(58) Tu, M. L. D. *An in Silico Model to Predict P-gp Substrate*; The World Pharmaceutical Congress. Cambridge Healthtech Institute: Needham, MA, Philadelphia, PA, 2004.
(59) Hughes, L. D.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. Why are some properties more difficult to predict than others? A study of QSPR models of solubility, melting point, and Log P. *J. Chem. Inf. Model.* **2008**, *48* (1), 220−232.
(60) Charlton, M. H.; Docherty, R.; Hutchings, M. G. Quantitative structure−sublimation enthalpy relationship studied by neural networks, theoretical crystal packing calculations and multilinear regression analysis. *J. Chem. Soc., Perkin Trans. 2* **1995**, No. 11, 2023−2030.
(61) Ouvrard, C.; Mitchell, J. B. Can we predict lattice energy from molecular structure? *Acta Crystallogr., Sect. B: Struct. Sci.* **2003**, *59* (5), 676−685.
(62) Salahinejad, M.; Le, T. C.; Winkler, D. A. Capturing the crystal: prediction of enthalpy of sublimation, crystal lattice energy, and melting points of organic compounds. *J. Chem. Inf. Model.* **2013**, *53* (1), 223−229.
(63) Abramov, Y. A. *QSPR modeling of chemical and physical stability of pharmaceuticals, 20th EuroQSAR symposium*, St. Petersburg, Russia. 2014.

DOI: 10.1021/acs.molpharmaceut.5b00119
Mol. Pharmaceutics XXXX, XXX, XXX−XXX