

A free energy perturbation approach to estimate the intrinsic solubilities of drug-like small molecules

Sayan Mondal¹, Gary Tresadern², Jeremy Greenwood¹, Byungchan Kim¹, Joe Kaus¹, Matthew Wirtala¹, Thomas Steinbrecher³, Lingle Wang¹, Craig Masse⁴, Ramy Farid¹, and Robert Abel^{1*}

1. Schrödinger, 120 W 45th Street, New York, NY 10036, USA
2. Computational Chemistry, Janssen Research & Development, Janssen Pharmaceutica N. V., Turnhoutseweg 30, B-2340 Beerse, Belgium.
3. Schrödinger GmbH, Q7 23, 68161 Mannheim, Germany
4. Nimbus Therapeutics, 130 Prospect St, Cambridge, MA 02139, USA

*To whom correspondence should be addressed. E-mail: robert.abel@schrodinger.com

Abstract

Optimizing the solubility of small molecules is important in a wide variety of contexts, including in drug discovery where the optimization of aqueous solubility is often crucial to achieve oral bioavailability. In such a context, solubility optimization cannot be successfully pursued by indiscriminate increases in polarity, which would likely reduce permeability and potency. Moreover, increasing polarity may not even improve solubility itself in many cases, if it stabilizes the solid-state form. Here we present a novel physics-based approach to predict the solubility of small molecules, that takes into account three-dimensional solid-state characteristics in addition to polarity. The calculated solubilities are in good agreement with experimental solubilities taken both from the literature as well as from several active pharmaceutical discovery projects. This computational approach enables strategies to optimize solubility by disrupting the three-dimensional solid-state packing of novel chemical matter, illustrated here for an active medicinal chemistry campaign.

I. Introduction

Solubility is an important characteristic of a compound for various applications, including pharmaceutical, environmental, and industrial uses. For example, adequate aqueous solubility is required for pharmaceutical compounds to be orally bioavailable – it has been estimated that almost 40% of new chemical entities are practically insoluble (1) and many drug discovery programs fail to reach the market due to low solubility. Low solubility has also been associated with stability issues and often leads to difficulties in the development of acceptable formulations for toxicology studies. (2) Therefore, improving solubility can become a focal point of lead optimization in drug discovery with the goal of designing compounds that would move predominantly insoluble series into more soluble regimes.

The most common approach to modeling solubility has been empirical, via fitting to datasets of experimentally measured solubility using different regression methods including machine learning techniques. (3–6) Some of these efforts have shown correlations with descriptors such as logP and polar surface area. (5) However, solubility cannot be universally improved simply by adding polar groups, particularly when the polar moiety happens to stabilize the solid state,

increasing the melting point (m.p.), and thereby counteracting increased favorable interactions with water. This has been elegantly demonstrated for benzodiazepine molecules where different substitutions, sometimes increasing logP, disturbed solid-state interactions leading to lower m.p. and increased solubility (7). Moreover, in a drug discovery context, solubility needs to be balanced with other ADME properties, for example permeability, as well as potency of the chemical entity, which can deteriorate with indiscriminate increases in polarity. In addition, approaches utilizing machine learning or other empirical methods trained to an experimental dataset are inherently limited by the chemical space in the training sets. While these methods interpolate very well, they tend to extrapolate poorly to novel chemotypes – a situation typical for chemical matter to be the subject of a lead optimization campaign. Therefore, empirical methods trained on public datasets, where the observed changes in solubility are often driven primarily by changes in hydrophobicity and aromaticity, often have limitations in predicting solubility of novel chemical matter while aligning other drug-like properties. To address such cases in a more satisfactory fashion, we expect consideration of the behavior chemical matter in both aqueous and solid states will be required.

Physics-based incorporate the three-dimensional characteristics of molecules and the environment with which the molecule interacts. Such methods may be able to model effects involving disrupting of solid state packing, as well as provide a more accurate estimation of the solvation energy. It is expected such methods should provide a more complete picture of aqueous solubility, and facilitate the balancing of solubility with potency and other ADME properties for novel chemical matter. Further, an atomistic and physically rigorous solubility prediction method might also provide a structural explanation for the observed solubility of a compound, which may facilitate the generation of hypotheses to improve solubility. A physics-based method could also be extended to work for additional solvents other than water. This is particularly relevant in late-stage pharmaceutical development when hundreds of different solvents may be considered in order to maximize synthetic ease and yield. Lastly, we expect atomistic solubility prediction methods are likely to better generalize to novel co-solvents and unusual small molecules, such as macrocycles, than the more commonly used empirical ligand-based techniques.

In this work we introduce an atomistic free energy method to estimate the aqueous solubility of neutral drug-like small molecules in an amorphous form, which does not require the small molecule crystal structure to be prespecified or predicted prior to calculating the solubility

estimate. This is an important consideration for informing design and synthesis efforts for a typical drug discovery program, because the crystal structure data (or related physical properties like melting point used in the general solubility equation(8)) are unavailable prior to synthesis. We have here employed the commonly used thermodynamic cycle(9) for aqueous solubility depicted in Figure 1, whereby the solubility is decomposed into a sublimation process and a solvation process--i.e. first, transferring a ligand from the solid aggregate into vacuum (sublimation), and then transferring the ligand it from vacuum into the solvent. The key insight leading to the present work was the hypothesis that the average free energy of sublimation of particles decoupled from the surface of amorphous material in a neat water box would be highly correlated with the implied sublimation free energies of solid materials commonly subjected to solubility measurements in medicinal chemistry campaigns. Moreover, when alchemically computing the sublimation free energy of the particles, a fast diffusing liquid is used to alleviate the artifacts arising from the long time scales of crystal healing, and to make convergent simulations feasible within the design and modeling cycles of a typical drug discovery program. We have developed and tested this novel approach within the state of the art FEP+ framework providing an easy to run streamlined protocol using sampling enhancements (10), absolute free energy perturbation (11, 12), and the latest generation OPLS3e forcefield (13). We have found solubilities calculated utilizing this approach to be in good agreement with experimental solubilities taken both from the literature as well as from several active pharmaceutical discovery projects.

II. Methods

The solubility free energy $\Delta G_{\text{solubility}}$ can be computed as the sum of the solvation free energy ΔG_{solv} and the sublimation free energy $\Delta G_{\text{sublimation}}$ (see Figure 1)

$$\Delta G_{\text{solubility}} = \Delta G_{\text{solv}} + \Delta G_{\text{sublimation}}$$

The solvation free energy of a single molecule is computed using commonly employed alchemical techniques(11, 12) and the latest version of the OPLS forcefield.(11–13) $\Delta G_{\text{sublimation}}$ is computed in several steps, as follows. First, a dense amorphous solid is constructed with 64 molecules packed with Monte Carlo simulations using periodic boundary conditions. The amorphous solid is inserted into a neat water box and further equilibrated with Molecular

Dynamics simulations, using periodic boundary conditions. If the amorphous solid is observed to spontaneously dissolve as determined by ligands flaking off the aggregate, the solubility is estimated to be very high ($\gg 200 \mu\text{M}$), and the workflow is terminated (see Supporting Information for the list of very highly soluble compounds tested). At this level of solubility, a quantitative measure of solubility is not required for guiding drug discovery project decision making. Such high solubility values would be beyond the upper threshold of typical kinetic solubility assays used for screening in many drug discovery programs. If the amorphous solid is stable and does not dissolve under the simulation conditions, 5 molecules of the amorphous solid are selected for sublimation at the aggregate--water interface for free energy calculation methods. The 5 molecules to be sublimated are chosen such that at least 20% of the heavy atoms of the chosen molecules are solvent exposed in terms of Solvent Accessible Surface Area (SASA). These five molecules at the water/solid interface are alchemically decoupled in five independent free energy calculations, and a median sublimation free energy is estimated from these five computational sublimations. The reason multiple independent ligands are alchemically decoupled is due to the heterogeneity of the amorphous material, so that a median sublimation free energy can be estimated for the aggregate as a whole. The solubility of 'intermediate' and 'low' solubility compounds is then quantitatively estimated from the difference in the computed solvation free energy and the average sublimation free energy for the evaluated compounds. Although water is used in all calculations reported here, we emphasize that other "fast diffusing" solvents could be used so as to rapidly fill the spaces created by the molecule annihilated in the sublimation step. This can be achieved by choosing a solvent with a relaxation time much faster than the time scale of the simulation, which allows one to avoid the long-time scale rearrangement of the aggregate that would be required to re-equilibrate after the alchemical annihilation of the molecule. Please see Supporting Information for further methodological details, including simulation lengths and convergence.

III. Results and Discussion

To facilitate testing of this approach with public data (Figure 2), it was tested against the experimental data reported for a series of 19 neutral benzodiazepines reported by Svensson and

co-workers (14) distributed over 3 log orders of solubility, including examples of matched pairs of molecules where decreasing logP does not improve solubility. The solubility calculation method reported here was found to have a mean unsigned error of 0.4 log units (0.6 kcal/mol) and an R^2 value of 0.7 compared to the experimental data. The calculations were further compared for the 14 neutral compounds measured in a publication from Taylor and co-workers,(15) leaving out the charged iopanoic acid compound. These comprise a group of complex and diverse molecules that are of pharmaceutical interest. The solubility calculations here were found to have a mean unsigned error of 0.8 log units (1.1 kcal/mol) and an R^2 value of 0.5 compared to the experimental data. Overall, for the 33 public compounds taken together (Figure 2), the calculated solubility was in good agreement with the reported experimental solubility, with a mean unsigned error of 0.6 log units and an R^2 value of 0.54.

Beyond quantifying the prediction accuracy, we sought to investigate if the calculations could elucidate molecular features underlying lower or higher observed solubility in molecular matched pairs. Determination of such features may lead to hypotheses for improving solubility and ideation of new molecules, even though the synthesis decisions should ideally be guided by actual calculations taking into account the complete energetics, rather than such qualitative analysis. The trajectories for the benzodiazepines explain why the more polar uncapped amides do not actually improve solubility, with the uncapped NH stabilizing the solid-state aggregate via H-bonding with the carbonyl of adjacent molecules (Figure 3A) thus nullifying any solubility enhancement from the increased polarity. To take another example, $-\log S$ for nifedipine is 4.9 experimentally and 4.8 by FEP+ solubility, which is only sparingly soluble and only modestly improved over the insoluble felodipine which has a $-\log S$ of 6.0 experimentally and 5.4 by FEP+ solubility despite the introduction of the nitro group. As detailed in Table 1, the calculations reported herein indeed do recapitulate the introduction of the highly polar nitro-group leading to a substantially more favorable solvation free energy by several kcal/mol; however, the calculations also show the nitro-group leading to tight interactions within the solid form leading to a highly unfavorable sublimation free energy (ΔG_{sub}). The balance of these two competing terms leads to the introduction of the nitro-group only modestly increasing the solubility of the core. Evidence of the interactions formed by the nitro-groups in the amorphous solid used for the sublimation free energy calculations is readily apparent in the trajectories (Figure 3A). The interplay between ligand

polarity and stabilization of the solid form, is further illustrated with the molecular matched pairs of 2-hydroxybiphenyl, 4-hydroxybiphenyl, and 4,4'-dihydroxybiphenyl where there is a ligand crystal structure available. Counterintuitively, the addition of an extra hydroxyl in 4,4'-dihydroxybiphenyl results in a loss of solubility, whereas a gain of solubility would have been expected from a simple consideration of the increased polarity. The amorphous solubility computed by FEP+ correlates with the solubility observed experimentally for these matched pairs (Table 2), with the lower solubility of 4,4'-dihydroxybiphenyl arising from a substantially higher stabilization of the aggregate form (ΔG_{sub}) compared to that for 2-hydroxybiphenyl or 4-hydroxybiphenyl. Thus, the increase in polarity is more than offset via the additional H-bonding stabilization of the aggregate from the two hydroxyl groups, adding to the π - π stacking stabilization from the biphenyl moiety (Figure 3C). The positioning of the hydroxyl also impacts solubility, e.g., the experimental solubility of 4-hydroxybiphenyl is lower than that of 2-hydroxybiphenyl, with a $-\log S$ of 3.5 vs 2.4 despite the two molecules having similar polarity. It appears from an analysis of the trajectories that the 4-hydroxyl molecule is able to form stabilizing interactions in the solid via ordered intermolecular H-bonds and π - π stacking interactions that are apparently not possible in the 2-hydroxyl molecule. These three molecules serve as good examples of the complex interplay between molecule/water interactions in solvent and molecule/molecule interactions in the solid, which would be very difficult for a purely ligand-based method to resolve.

We next report expanded testing of the methodology. A larger set of 103 compounds were submitted to the standard solubility FEP+ protocol as described in the methods. The calculations were performed internally and independently at Janssen R&D. The compounds originated from recent late lead optimization programs and cover diverse chemical scaffolds and range of physicochemical properties: MW 254 to 654, ClogP -1.1 to 6.0, number of rotatable bonds from 0 to 12. A consistent experimental approach was used to measure the thermodynamic solubility of each. In brief, 2.5 mg of compound in 500 μ l of solvent was submitted to 24h shaking, filtered, and diluted. Filtrate samples were quantified by LC/UV. X-ray diffraction data was measured up front to categorize samples as either crystalline, amorphous or mixed. The set of compounds was limited to those predicted to be neutral at pH 7.4 based on calculated pKa and results for the whole set are shown in Figure 4A. The FEP+ solubility showed good correlation with the experimental data (R^2 0.57). For the compound samples that were known to be in either an amorphous or mixed

crystalline/amorphous state, analogous to the amorphous system prepared in silico, the correlation was stronger with an observed R^2 value of 0.72 (Figure 4B).

This level of accuracy makes solubility FEP+ modeling suitable for medicinal chemistry programs faced with insoluble chemical matter where solubility improvements must be carefully balanced to achieve alignment of ADMET and potency properties. The application of FEP+ solubility within a typical medicinal chemistry effort is illustrated for a program which was carried out as part of a closely integrated partnership between Schrödinger and Nimbus Therapeutics (Figure 5A). For the Nimbus program, solubility FEP+ modeling was applied in conjunction with prospective FEP+ modeling for potency and selectivity to accelerate early stage discovery efforts. Within the priority series for this program, most of the analogs proved to be highly insoluble ($<2 \mu\text{M}$), which could be recapitulated by the FEP+ solubility model in retrospective testing. As a result, the program team designed several ideas centered around the established concept of disrupting planarity of the scaffolds with the goal of decreasing crystal packing energies (16). The majority of these ideas were predicted to be insoluble, but the FEP+ solubility model did identify an analog that with a predicted solubility of $\sim 10 \mu\text{M}$. The increased solubility of this analog was confirmed experimentally and represented a substantial improvement from the typical “brick dust” compounds identified earlier in the Nimbus program. The solubility FEP+ model also identified a number of analogs with more significantly improved aqueous solubility, but these compounds were also predicted to lose potency in ligand FEP+ modeling and, therefore, were not pursued as targets. Gratifyingly, the solubility FEP+ model identified a side chain modification that was predicted to improve solubility while maintaining the potency and selectivity of the original analogs. This compound was synthesized and experimentally confirmed to possess a $>90\text{x}$ improvement in aqueous solubility relative to the initial analogs while preserving the potency and kinase selectivity of the series. By using this iterative design process enabled by the solubility FEP+ modeling, it was possible to quickly identify analogs with improved solubility which lead to compounds with improved pharmacokinetic profiles.

A specific use case for this approach is illustrated for a congeneric series (Figure 5B) within a Nimbus program that was optimized by FEP+ solubility modeling from inception. For this series, aqueous kinetic solubility was routinely obtained for all compounds and a few key analogs were also selected for thermodynamic solubility measurement in SIF and SGF media. The prospective

performance displayed a MUE of 0.6 in logS (0.8 kcal/mol) for FEP+ solubility vs kinetic aqueous solubility. Overall, 42 of the 52 compounds synthesized based on promising FEP+ solubility predictions, displayed High (>50 μM) or Medium (between 10 and 50 μM) aqueous solubility. A small number of compounds (10) were made with Low (<10 μM) predicted solubility as key negative controls or to explore modification of other ADME properties. The overall classification accuracy for the congeneric series is presented in Table 3. The first compound prepared in the series (**1**, Figure 5B) possessed modest kinetic solubility (24 μM). As might be expected, FEP+ solubility predicted a boost in aqueous solubility for analog **2** in Fig. 5B driven by reduced logP and removal of a symmetry element in the pyrazole side chain. Strikingly, however, FEP+ solubility predicted a substantial boost in solubility for analog **3** in Fig. 5B that incorporates a 1,4-oxazepanyl side chain. This compound was synthesized and confirmed to have a high solubility experimentally. The thermodynamic solubility of analog **3** in SIF was also measured and confirmed to be high as well (~300 μM), making it one of the most soluble analogs prepared within the series. Counterintuitively, FEP+ solubility had also predicted low solubility for a matched pair analog (**4** in Fig. 5B), which incorporated an azepanyl side chain containing a polar diol moiety. Compound **4** was synthesized as a negative control and confirmed to have low solubility experimentally.

Critical to improving the probability of success within the multi-parameter optimization context of lead optimization, are: 1) access to non-trivial modes of gaining solubility via the solid state without degrading cellular permeability which can often occur with increasing polarity, and 2) combining solubility predictions at scale with modeling of other end-points such as permeability and potency. With this domain of applicability and 80% classification accuracy (MUE ~1 kcal/mol) into High, Medium, and Low solubility categories, a discovery team should be able to rapidly improve solubility in 7-9 out of every 10 compounds synthesized.

The domain of applicability of FEP+ solubility for NCEs will benefit from the extensive chemotype coverage and transferability of latest forcefields like OPLS3e.(13) Additionally, the accuracy of this method and the domain of applicability will continually improve along with refinements in forcefields. FEP+ solubility can be used early in a drug discovery campaign and for diverse compounds, because it utilizes *absolute* free energy calculations and does not require

reference experimental data in any related chemical series. This is in contrast to the *relative* free energy calculations utilized for potency predictions. (17)

A current limitation in the domain of applicability of the current method is that the predictions are for the neutral form of the compound, i.e., its intrinsic solubility. As the insoluble chemical matter in most drug discovery programs tends to be predominantly neutral compounds, predicting the intrinsic solubility should suffice for many lead optimization campaigns where solubility is an issue. The solubility of compounds charged at the pH of physiological interest includes a solubilizing contribution based on the pKa. In principle, this contribution can be estimated using the Henderson-Hasselbach equation,

$$S_{\text{total}} = S_{\text{intrinsic}}(1 + 10^{(\text{pKa} - \text{pH})})$$

Thus, there is a solubilizing contribution for pKa > pH, such that -logS decreases linearly with decreasing pH for, e.g., basic amines. This is supported by some empirical data, c.f. (18)¹⁶. Combining FEP+ solubility with pKa predictions for charged compounds, would be a natural extension of the domain of applicability to charged compounds, especially as accurate pKa prediction methods continue to be developed and validated for wider sets of chemotypes (19).

IV. Conclusion

We have here presented a method for calculating intrinsic solubility, where the sublimation and hydration contributions to the solubility are evaluated with absolute free energy calculations. Our approach makes use of a simulated amorphous solid, constructed by successive Monte Carlo and Molecular Dynamics calculations, from which we decouple five copies of the molecule from the surface and generate an average sublimation energy. The protocol was implemented within the FEP+ framework and benefits from the state of the art OPLS3e force field. Accurate correlations with the experimental data were observed for challenging test cases, such as the benzodiazepine series designed to probe aspects of solid-state limited solubility, which provides confidence the method is capturing the essential physics. The approach is also expected to benefit from the OPLS3e force field which has a substantial emphasis on drug-like small molecules, with extensive parameterization and ease of updating missing torsional terms. This provides a greater chance of success and enables incorporating both conformational factors and polarity in achieving potency/ ADME balance in lead optimization. The utility of this solubility

FEP+ methodology has also been demonstrated in several medicinal chemistry lead optimization efforts where it has been leveraged, in concert with other modeling tools, to rapidly align the ADME properties for priority series.

Figures and Tables

Figure 1. Thermodynamic cycle for modeling solubility and the calculation strategy in FEP+ solubility. The solubility is computed via a sublimation and solvation leg. The implied sublimation free energy is computed via FEP+ as described in the Methods section, starting from a disordered system built using Monte Carlo and stabilized with MD simulations.

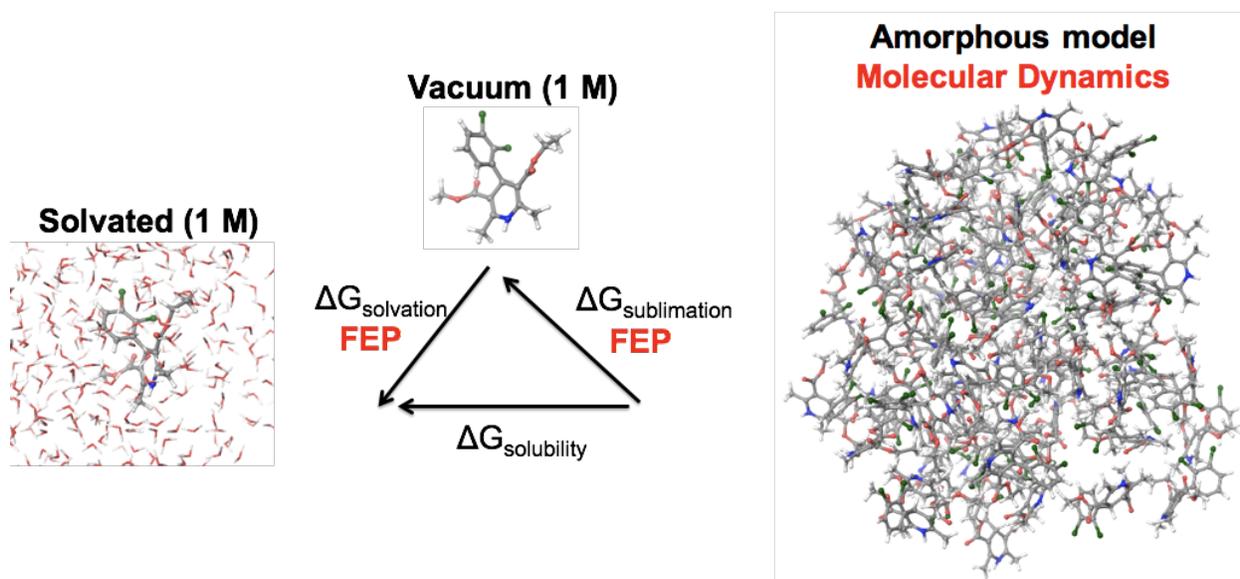


Figure 2 Computed solubility is in good agreement with the experimental data obtained from public datasets, compared here for two published experimental datasets (14, 15). The overall R² between computed and measured logS is 0.5, and mean unsigned error is 0.6 in logS.

FEP, -LogS vs. Experimental, -Log S

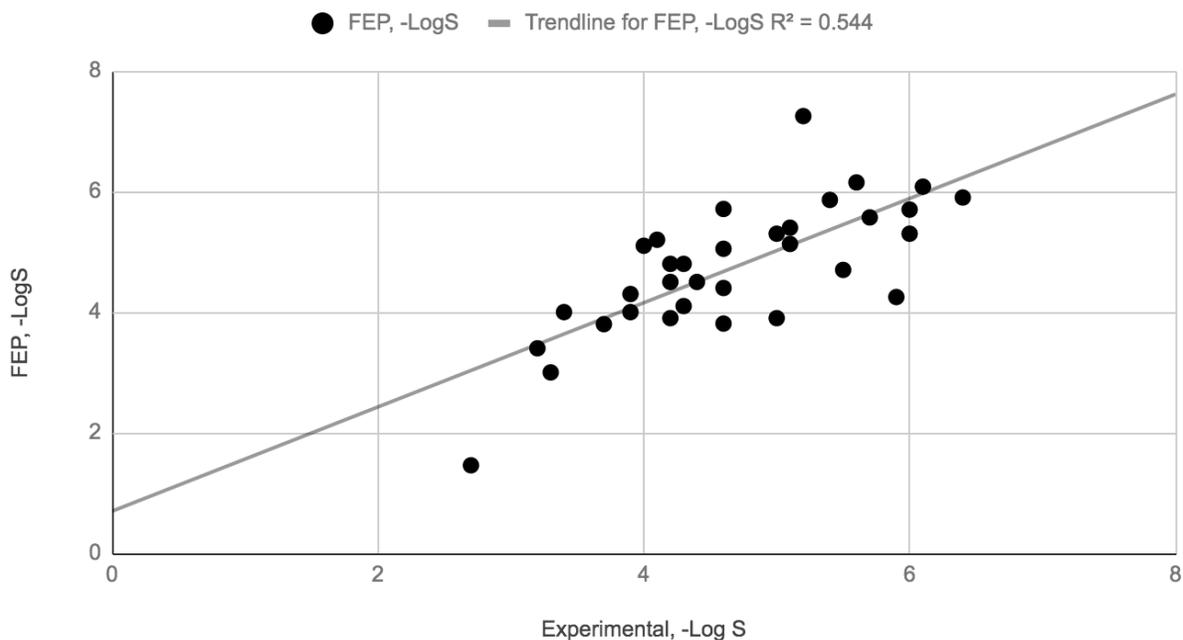


Figure 3 Stabilization of the aggregate via H-bonding involving a polar substituent that nullifies any solubility enhancement from the polar piece, illustrated by 3D structures of the aggregates. (A) A simulation snapshot of a benzodiazepine, showing H-bonding between the uncapped NH of the cyclic amide with the carbonyl from adjacent molecules adding to the stabilization from the intermolecular pi-stacking. (B) A simulation snapshot of nifedipine aggregate, showing stabilization via H-bonding involving the nitro group. (C) X-ray structure of 4,4'-dihydroxybiphenyl, showing the stabilization from the H-bonding involving the additional hydroxyl compared to 2-hydroxybiphenyl. The solid aggregate of 4,4'-dihydroxybiphenyl is stabilized via the hydroxyls in addition to the pi-stacking of the biphenyl moiety in the two compounds. Thus, the addition of the polar hydroxyl, in fact, results in a loss of solubility rather than a gain of solubility as might be expected looking just at the increase in polarity.

Figure 3A

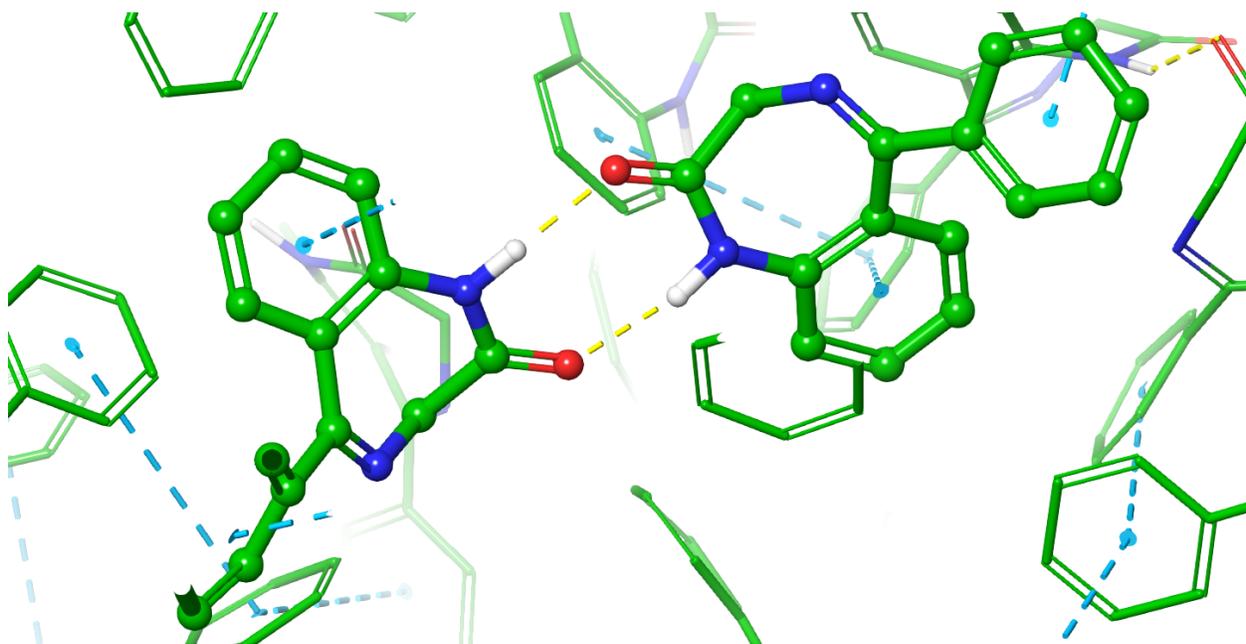


Figure 3B

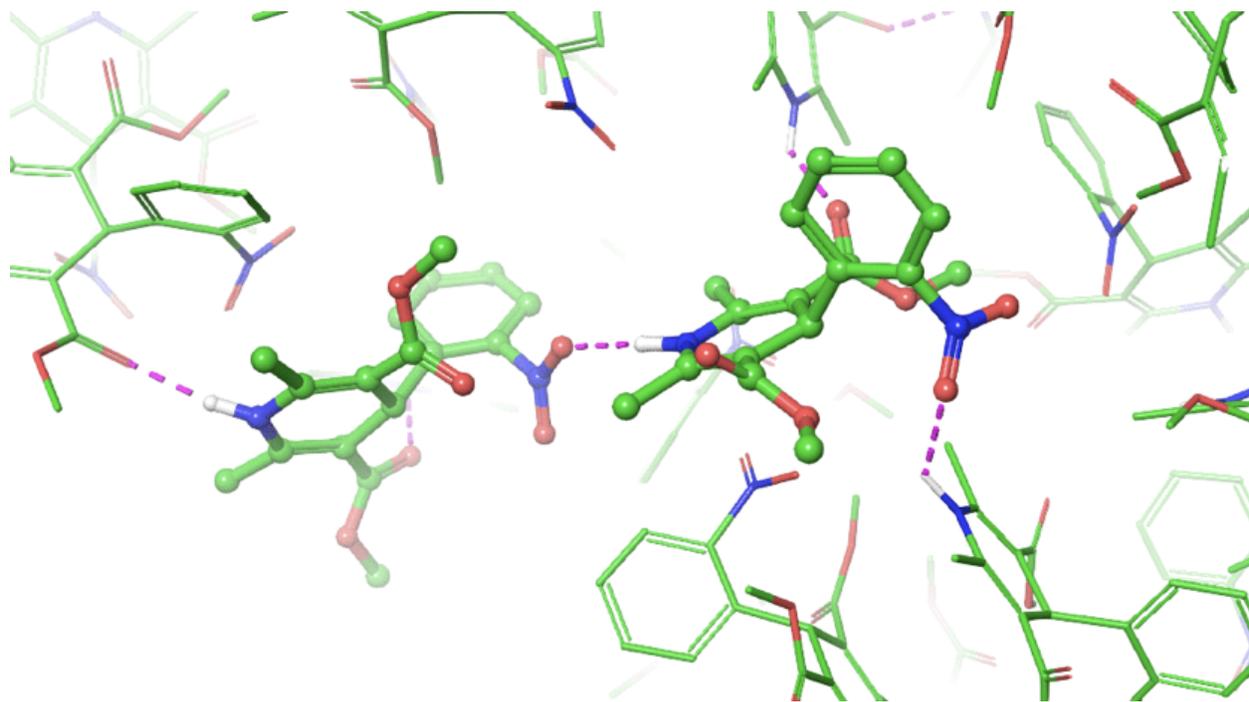


Figure 3C

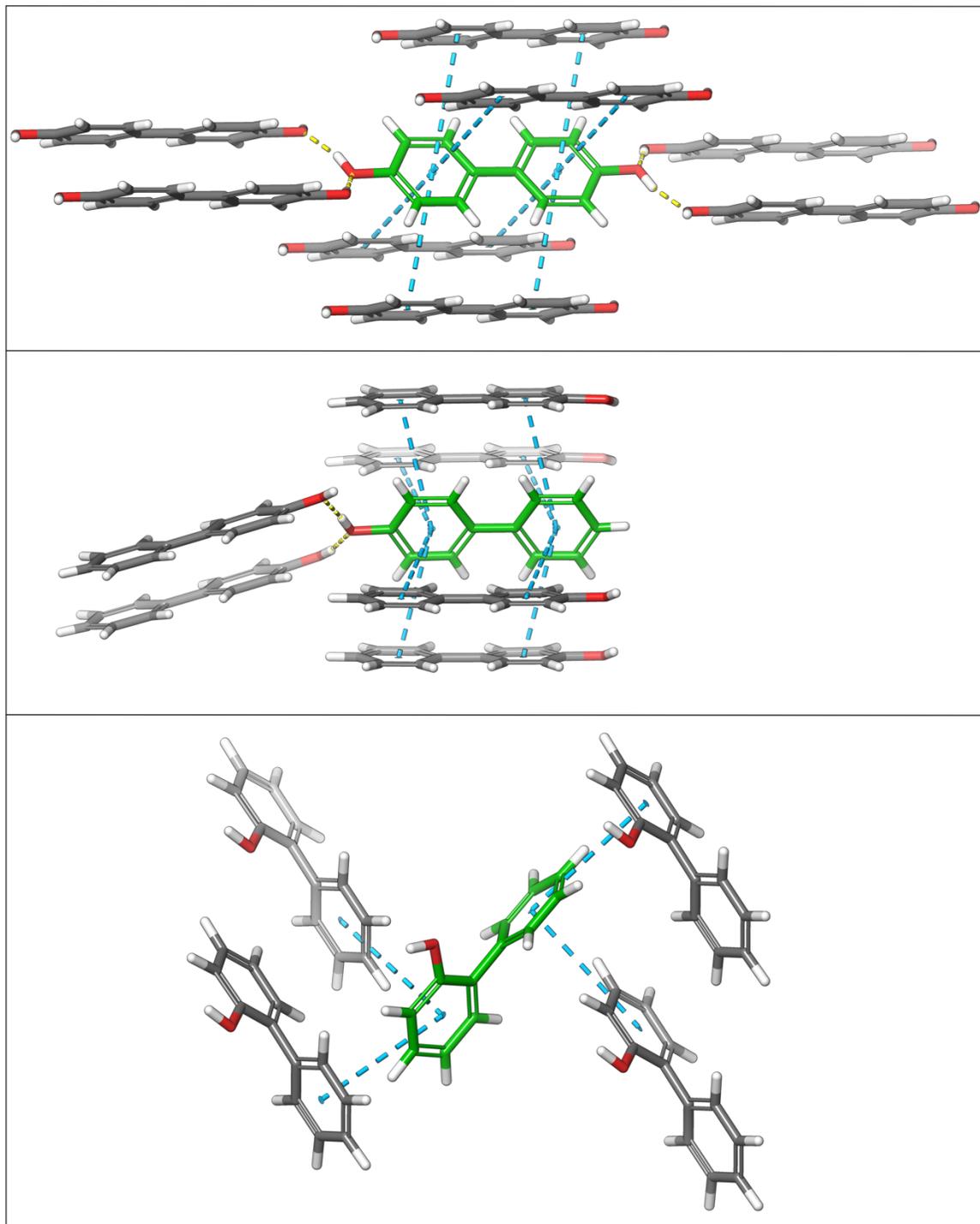


Table 1. Detailed FEP+ solubility data for felodipine and nifedipine

Felodipine		Nifedipine		Difference	
ΔG_{hyd} kcal/mol	ΔG_{sub} kcal/mol	ΔG_{hyd} kcal/mol	ΔG_{sub} kcal/mol	$\Delta\Delta G_{\text{hyd}}$ kcal/mol	$\Delta\Delta G_{\text{sub}}$ kcal/mol
-11.0	18.4	-15.8	22.3	-4.8	3.9

Table 2. Detailed FEP+ solubility data for 2-hydroxybiphenyl, 4-hydroxybiphenyl, and 4,4'-dihydroxybiphenyl

2-hydroxybiphenyl		4-hydroxybiphenyl		4,4'-dihydroxybiphenyl	
ΔG_{hyd} kcal/mol	ΔG_{sub} kcal/mol	ΔG_{hyd} kcal/mol	ΔG_{sub} kcal/mol	ΔG_{hyd} kcal/mol	ΔG_{sub} kcal/mol
-6.3	8.7	-8.1	11.0	-12.6	15.9

Figure 4 Performance of FEP+ solubility vs measured thermodynamic solubility in an expanded set comprising 103 compounds from recent late lead optimisation programs at Janssen R&D, spanning diverse chemotypes and physico-chemical properties, for (A) all 103 compounds, irrespective of crystalline or amorphous solid state, and (B) 34 compounds in amorphous or mixed amorphous/crystalline forms.

Figure 4A

FEP, -LogS vs. Experimental, -LogS

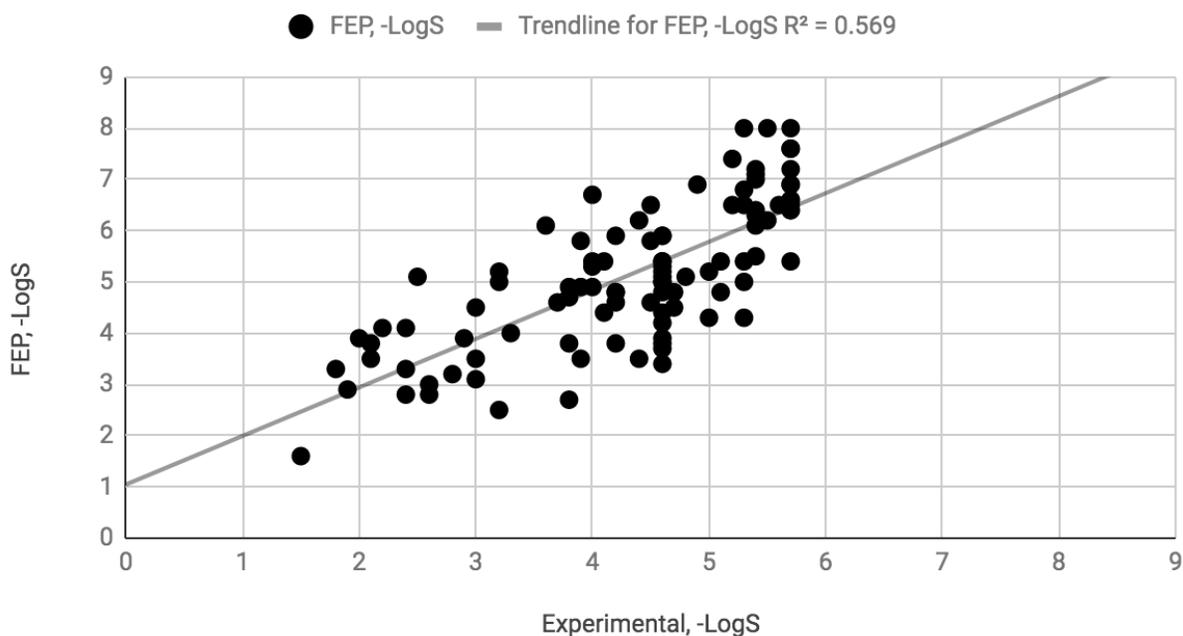


Figure 4B

FEP, -LogS vs. Expt, -LogS

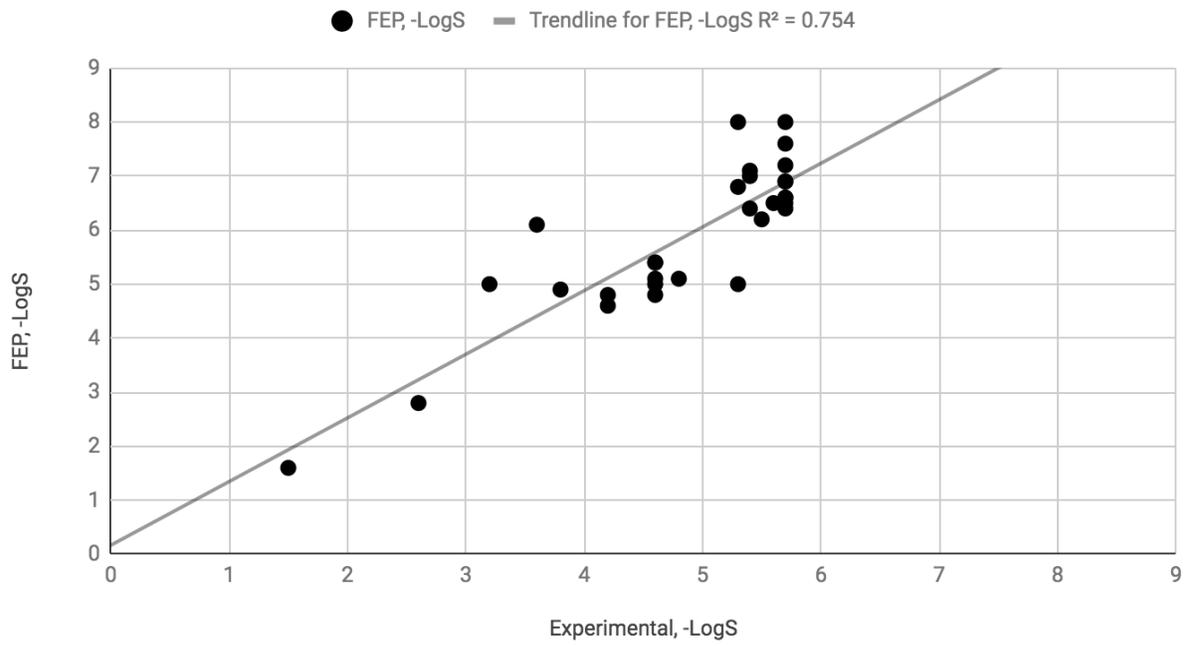


Figure 5 Illustrative use of FEP+ solubility to help attain an ADME/ potency balance in a program pursued collaboratively by Nimbus and Schrödinger. (A) Illustration of the discovery of soluble matter against the background of insoluble predicted matter in a lead chemical series, (B) Molecular matched pairs in a congeneric chemical series, illustrating the use of FEP+ solubility to improve solubility without the need to add significant polarity.

Figure 5A

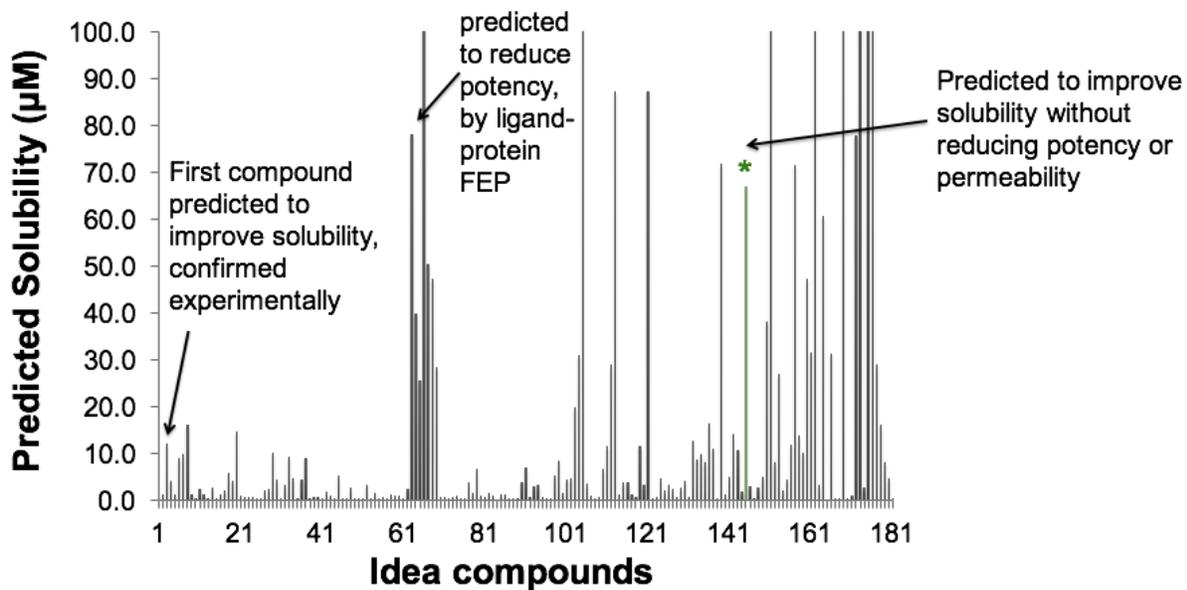
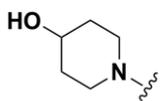


Table 3. Prospective performance of FEP+ solubility vs measured kinetic aqueous solubility in a congeneric chemical series, in a live drug discovery project pursued collaboratively by Nimbus and Schrödinger.

MUE in logS = 0.6		Experimental Measurements (Kinetic Solubility)		
		Insoluble (< 10 μ M)	Medium (10 - 50 μ M)	High (>50 μ M)
N = 52				
Prospective FEP+ Predictions	Insoluble (< 10 μ M)	5	3	2
	Medium (10 - 50 μ M)	0	8	9
	High (>50 μ M)	2	7	16

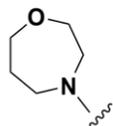
Figure 5B

First R-group in Series

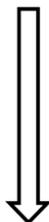


FEP solubility = 4 μ M
Kinetic solubility = 24 μ M

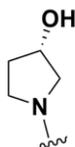
Solubilizing R-group



FEP solubility = 361 μ M
Kinetic solubility = 82 μ M
Thermodynamic solubility (SIF) = 303 μ M

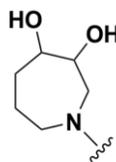


Solubilizing R-group



FEP solubility = 68 μ M
Kinetic solubility = 64 μ M

Negative Control



FEP solubility = 6 μ M
Kinetic solubility = 8 μ M

References

1. K. T. Savjani, A. K. Gajjar, J. K. Savjani, Drug Solubility: Importance and Enhancement Techniques. *ISRN Pharmaceutics* **2012**, 1–10 (2012).
2. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **23**, 3–25 (1997).
3. T. Cheng, Q. Li, Y. Wang, S. H. Bryant, Binary classification of aqueous solubility using support vector machines with reduction and recombination feature selection. *J. Chem. Inf. Model.* **51**, 229–236 (2011).
4. D. S. Palmer, N. M. O’Boyle, R. C. Glen, J. B. O. Mitchell, Random forest models to predict aqueous solubility. *J. Chem. Inf. Model.* **47**, 150–158 (2007).
5. C. A. S. Bergström, C. M. Wassvik, U. Norinder, K. Luthman, P. Artursson, Global and local computational models for aqueous solubility prediction of drug-like molecules. *J. Chem. Inf. Comput. Sci.* **44**, 1477–1488 (2004).
6. C. A. S. Bergstrom, In silico Predictions of Drug Solubility and Permeability: Two Rate-limiting Barriers to Oral Drug Absorption. *Basic Clinical Pharmacology Toxicology* **96**, 156–161 (2005).
7. L.-E. Briggner, R. Hendrickx, L. Kloo, J. Rosdahl, P. H. Svensson, Solid-State Perturbation for Solubility Improvement: A Proof of Concept. *ChemMedChem* **6**, 60–62 (2011).
8. Y. Ran, N. Jain, S. H. Yalkowsky, Prediction of aqueous solubility of organic compounds by the general solubility equation (GSE). *J. Chem. Inf. Comput. Sci.* **41**, 1208–1217 (2001).
9. M. J. Schnieders, *et al.*, The Structure, Thermodynamics and Solubility of Organic Crystals from Simulation with a Polarizable Force Field. *J. Chem. Theory Comput.* **8**, 1721–1736 (2012).
10. L. Wang, B. J. Berne, R. A. Friesner, On achieving high accuracy and reliability in the calculation of relative protein–ligand binding affinities. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1937–1942 (2012).
11. D. Shivakumar, E. Harder, W. Damm, R. A. Friesner, W. Sherman, Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *J. Chem. Theory Comput.* **8**, 2553–2558 (2012).
12. D. Shivakumar, *et al.*, Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **6**, 1509–1519 (2010).
13. K. Roos, *et al.*, OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **15**, 1863–1874 (2019).
14. L.-E. Briggner, L. Kloo, J. Rosdahl, P. H. Svensson, In silico solid state perturbation for solubility improvement. *ChemMedChem* **9**, 724–726 (2014).
15. G. A. Ilevbare, L. S. Taylor, Liquid–Liquid Phase Separation in Highly Supersaturated Aqueous Solutions of Poorly Water-Soluble Drugs: Implications for Solubility Enhancing Formulations. *Crystal Growth & Design* **13**, 1497–1509 (2013).

16. M. Ishikawa, Y. Hashimoto, Improvement in aqueous solubility in small molecule drug discovery programs by disruption of molecular planarity and symmetry. *J. Med. Chem.* **54**, 1539–1554 (2011).
17. L. Wang, *et al.*, Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **137**, 2695–2703 (2015).
18. Y. Ran, A. Jain, S. H. Yalkowsky, Solubilization and preformulation studies on PG-300995 (an anti-HIV drug). *J. Pharm. Sci.* **94**, 297–303 (2005).
19. D. M. Philipp, M. A. Watson, H. S. Yu, T. B. Steinbrecher, A. D. Bochevarov, Quantum chemical pKa prediction for complex organic molecules. *International Journal of Quantum Chemistry* **118**, e25561 (2018).