Solubility Prediction, Solvate and Cocrystal Screening as Tools for Rational Crystal Engineering

C. Loschen[a*], A. Klamt[a,b]

[a]COSMOlogic GmbH & Co. KG, Imbacher Weg 46, 51379 Leverkusen, Germany;

[b]Institute of Physical and Theoretical Chemistry, University of Regensburg, 93053 Regensburg, Germany

Andreas Klamt, +49-2171-731681, email: klamt@cosmologic.de

Correspondence: Christoph Loschen, Tel: +49-2171-3636671, email: loschen@cosmologic.de

Abstract

Objectives The fact that novel drug candidates are becoming increasingly insoluble is a major problem of current drug development. Computational tools may address this issue by screening for suitable solvents or by identifying potential novel cocrystal formers that increase bioavailability. In contrast to other more specialized methods the fluid phase thermodynamics approach COSMO-RS allows for a comprehensive treatment of drug solubility, solvate and cocrystal formation and many other thermodynamics properties in liquids. The basic idea of COSMO-RS consists of using the screening charge density as computed from first principles calculations in combination with fast statistical thermodynamics in order to compute the chemical potential of a compound in solution.

This article gives an overview of recent COSMO-RS developments that are of interest for drug development and contains several new application examples for solubility prediction and solvate/cocrystal screening.

Key finding The fast and accurate assessment of drug solubility and the identification of suitable solvents, solvate or cocrystal formers is nowadays possible and may be used to complement modern drug development.

Summary COSMO-RS theory can be applied to a range of physico-chemical properties which are of interest in rational crystal engineering. Most notably, in combination with experimental reference data accurate

quantitative solubility predictions in any solvent or solvent mixture are possible. Additionally, COSMO-RS can be extended to the prediction of cocrystal formation which results in considerable predictive accuracy concerning coformer screening. In its recent variant COSMO*quick* costly quantum chemical calculations are avoided resulting in a significant speed-up and ease-of-use.

Keywords: Solubility Prediction, COSMO-RS, Liquid Phase Thermodynamics, Cocrystals. Solvates

Introduction

The fact that new drug candidates as generated by high-throughput screening methods are of increasingly poor aqueous solubility poses a major problem for modern drug research.[1] Thus, an important task during drug development is to find ways to solubilize such compounds. For computational methods to be supportive in this regard they need to be of reliable predictivity and at the same time be efficient enough to save real lab work. Although many *in-silico* models and prediction methods have been developed targeting different individual aspects in drug development, only a few approaches allow for a comprehensive treatment of a broader range of thermodynamic properties. Since its development in the early 1990ies the liquid phase thermodynamics theory COSMO-RS (<u>C</u>onductor like <u>S</u>creening <u>Mo</u>del for <u>R</u>eal <u>S</u>olvents) has been applied in variety of fields such as solubility prediction,[2] solvent screening,[3] excipient ranking,[4] computation of micellar systems[5] and ionic liquids,[6] prediction of pKa[7] and redox potentials[8] and calculation of partitioning coefficients.[9]

Recently, crystal engineers re-discovered the technique of co-crystallizing an excipient compound with an active pharmaceutical ingredient (API) as an alternative route to address the problem of poor drug solubility.[10] In this context, the COSMO-RS method has been demonstrated to be able to guide the choice of promising co-crystal formers (coformers) by the prediction of the relative probability that such a coformer forms a cocrystal with a given active.[11] The same approach could also be applied to estimate the tendency of solvents to form crystalline solvates.

In addition and also with direct relation to crystal engineering, the effect of milling on the surface energetics of molecular crystals was analyzed with COSMO-RS.[12]

A few other quite specialized methods have been developed to screen for potential coformers.[13,14] However, apart from the fact that the COSMO-RS approach performs at least comparably[15] they have the disadvantage of not being embedded in a general theory, and thus are not transferable to other thermodynamic properties.

Overall, it has been shown in many cases that COSMO-RS theory offers a rich choice of methods to aid drug developers finding a suitable solid form for a new drug. This article presents for the first time a general overview of recent COSMO-RS developments which may be of particular interest for crystal engineering and solid form selection based on many novel application examples. It is sectioned as follows. First, a short overall introduction into COSMO-RS theory is given. Then, recent examples for solubility prediction of some typical drug and coformers in different solvents and solvent mixtures are evaluated. This is followed by a case study concerning solvate screening and finally results of several cocrystal screenings are presented.

COSMO-RS Theory

COSMO-RS theory was developed by Klamt on top of the quantum chemical COSMO solvation model (Conductor like Screening Model) in order to overcome some of the limitations inherent to all dielectric continuum solvation models. One shortcoming of the continuum models is for example to account for differences between molecules with identical dielectric constants but quite different solvent properties like benzene and cyclohexane.[16] This finding led to the novel concept of COSMO-RS, a combination of COSMO with the statistical thermodynamics treatment of interacting surface segments.[17]

COSMO-RS starts from the polarization charges on the COSMO surfaces, as obtained by individual COSMO calculations for solute and solvent molecules. Those surfaces are defined by the molecular cavity separating the solute volume from the embedding dielectric continuum. Intermolecular interactions in COSMO-RS are calculated as local contacts of COSMO surfaces and quantified by the polarization charge densities σ and σ' of the two surface segments. For this purpose histograms of polarization charge densities, the so-called σ-profile (Figure 1), are considered.

Instead of deriving macroscopic properties from an ensemble of interacting molecules like in a Molecular Dynamics simulation for example, COSMO-RS uses an ensemble of interacting *surface segments*. This approximation leads to a significant improvement of the overall efficiency: Once the σ-profile is determined

from a quantum-chemical calculation (usually density functional theory (DFT) is employed) the intermolecular thermodynamics is evaluated within fractions of a second. Nevertheless, due to the rigorous statistical thermodynamics applied, one obtains a good and thermodynamically consistent approximation for the free energy and the chemical potential of the compound of interest and hence has access to all related physico-chemical properties in the liquid phase.
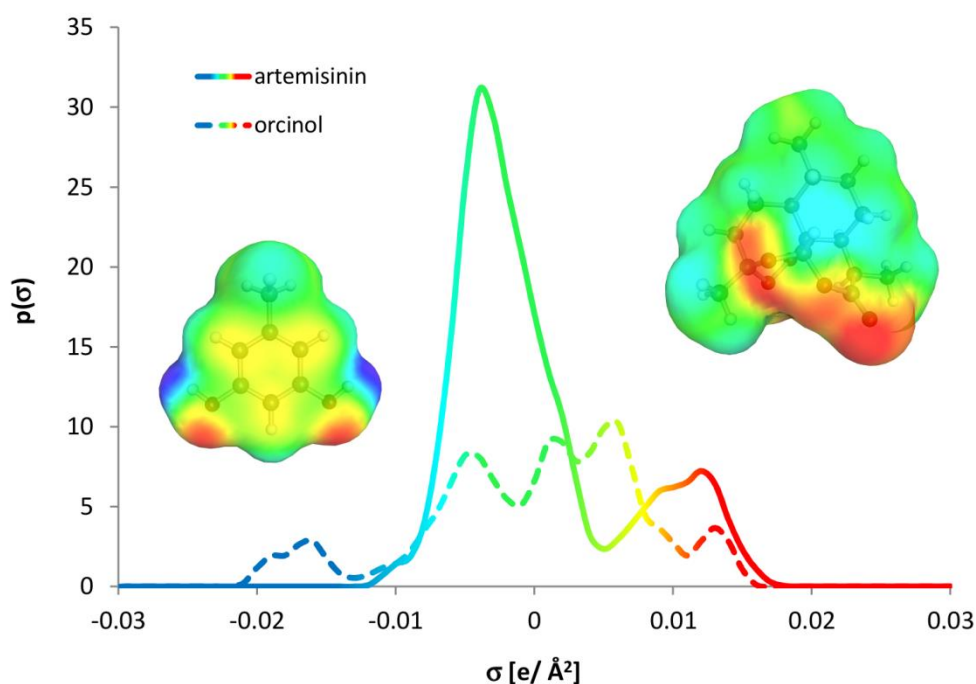

Figure 1 σ-profiles for some drug (artemisinin) and its coformer (5-methylresorcinol). Hydrogen bond donor regions are coloredblue, whereas hydrogen bond acceptor regions are colored red. The overlay of the two profiles shows the complementarity of the hydrogen bond donating coformer and the hydrogen bond accepting drug.

COSMO-RS theory takes into account the most important modes of molecular interactions, electrostatics, hydrogen bonding and slightly more empirically also Van der Waals interactions. The so-called hydrophobic interactions are well represented as a result of electrostatics and hydrogen bonding. For more details of the approach we refer to a recent review.[16]

Although nowadays the necessary quantum-chemical results are readily available for small to medium sized molecules, the generation of the σ-profiles is still the rate-determining step of the overall COSMO-RS workflow. However, this issue can be mitigated by deploying precomputed quantum-chemical COSMO calculation in a large database. Currently, a database with more than 100000 molecules is available, which allows the instantaneous generation of σ-profiles via the composition out of molecular fragments (COSMO*quick* approach).[18] This comes at the cost of a somewhat reduced accuracy, but further improves the efficiency and the ease of use of the approach. It allows for thermodynamic property prediction using as input just the molecular topology as generated from a usual 2-dimensional molecule sketching program.

Computational details

All COSMO-RS calculations presented in this article have been carried out with COSMO*quick*. COSMO*quick* is a JAVA-based software tool and graphical user interface which internally calls FORTRAN routines for the generation of σ-profiles and the COSMO*therm* code[21] for the subsequent thermodynamic computations.[22] Approximate σ-profiles are generated by accessing a database containing about 100,000 diverse chemical compounds, which avoids costly quantum-chemical calculations. Structures for those database constituents have been obtained by AM1/COSMO geometry optimization followed by a single point DFT/COSMO calculation (BP-SVP level of theory) with TURBOMOLE[23] in order to obtain the σ-surface.

The final σ-profile is then generated by matching one or several database fragments to the topology of the molecule under scrutiny. The chemical potentials in the liquid phase are then calculated according to standard COSMO-RS theory.

Solubility prediction of drugs and coformers

The solubility of liquid compounds can be readily obtained from the chemical potential in the liquid phase. For solids, i.e. for almost all drug compounds, additional information about the transfer from the liquid to the

solid state is necessary. This information is contained in the free energy of fusion ΔG_{fus} which is related to the melting point T_m and the melting enthalpy (enthalpy of fusion) ΔH_{fus}, under the assumption that the change in heat capacity may be neglected:

$$\Delta G_{fus} = \Delta H_{fus} - T\frac{\Delta H_{fus}}{T_m}$$

The mole fraction solubility is then calculated from the free energy difference::

$$\ln(x_S) = \left[\mu_X - \mu_S - \Delta G_{fus}\right]/RT$$

Here, μ_X and μ_S are the chemical potentials of the pure drug and the drug in solution, i.e. in some solvent, respectively; R is the gas constant and T the temperature. The chemical potential of the drug in solution is of course concentration dependent and thus the last equation has to be solved iteratively. However, working with a chemical potential μ_S at infinite dilution and solving for x_s is a sensible approximation as long as the drug solubility is sufficiently low as it is the case for most current drugs anyway.

The liquid phase chemical potentials μ_X and μ_S can be computed with a high accuracy by COSMO-RS theory. ΔG_{fus} may be attained either via experimental melting point and fusion enthalpies, via the combination of a reference solubility measurement and a COSMO-RS computation (i.e. solving the last equation for ΔG_{fus}), or via a quantitative structure property relationship (QSPR) model. The use of experimental data is certainly to be preferred and delivers more accurate results. Using a QSPR model for the fusion data should be applied only with care as it is accompanied usually by a large error bar.

Recently, a new solubility prediction algorithm based on COSMO-RS theory was introduced, using experimental reference solubilities in several different solvents in order to get the free energy of fusion and additionally a correction term for the chemical potential of the drug in solution for each specific solvent.[18]

Using the information from 3 or 4 reference solubility measurements this approach allows for a very accurate prediction of solubility in additional solvents or solvent mixtures.

There exists plenty of literature data on solubility measurements in different solvents. An excellent collection of solubility datasets has been compiled by the open notebook sciene (ONS) project[24] and may be used to get access to the original literature sources. Table 1 shows some results of drug and coformer solubility predictions using 3 to 4 reference solvents as compared with experimental data. Reference solvent selection was guided by the objective to include a balanced set consisting of hydrogen bond acceptors, hydrogen bond donors and non-polar solvents. Apart from that the selection was carried out arbitrarily. As an error measure the root mean squared error (RMSE) in relation to the experiment is specified. The average RMSE is of similar magnitude for drugs and for coformers and amounts to about ~0.5 log units. This is compares quite favorable to other state-of-the-art methods which rarely exceed this threshhold.[18] Details of the solubility predictions are listed in the supplement (Table S1).

The large RMSE of cinchonidine (0.87) is mostly due to its solubility in triethylamine, showing a strong deviation between experiment ($\log10(x)$=-2.84) and prediction ($\log10(x)$=0.0, i.e. complete solubility). This strong deviation is in part due to shortcomings of the prediction method, which at the employed level of theory (BP-SVP) is known to be less accurate for tertiary amines. Increasing the theory level by using COSMO*therm* at the TZVP-FINE level[21] reduce the deviation significantly for triethylamine ($\log10(x)$=-3.0), but the overall deviation over all solvents still amounts to RMSE=0.78. Finally, omitting this single value from the error analysis yields a significantly reduced RMSE for this dataset of 0.67.

An extension to solvent mixtures is straight-forward and in the following a few representative cases are presented. For solvent mixtures, the compositions with mole fraction x=0.0 and x=1.0 are fed as reference into the model in order to estimate the free energy of fusion and the corresponding correction to the chemical

potential. Other more empirical methods like NRTL-SAC[19] need more than just two reference points for a reliable parameter fit and thus additional measurements in other solvents are necessary before a specific solvent mixture can be addressed.

Figure 2 and Figure 3 present solubility data for sulfadiazine, salicylic acid, prednisolone and paracetamol in different binary solvent mixtures. As it is evident from Figure 2 and 3 the shape of the solubility curves in binary mixtures can be qualitatively very different. The system sulfadiazine in water/DMF[39] is most easily handled as its curve is monotonically decreasing (Figure 2 (a)). The systems salicylic acid[40,41] and prednisolone[42] in water/ethanol show a shallow maximum in the experiments, which cannot be reproduced as such by the prediction method, however the overall shape of the curve is met quite well (Figure 2 (c) and (d)). The systems paracetamol in water/dioxane[43,44,45] and sulfadiazine in water/dioxane[46] exhibit pronounced solubility maxima, which are predicted qualitatively correct (Figure 2 (b) and Figure 3), though the location of the maxima is not met exactly. Water rich phases are predicted not as well, which is mostly due to the intricacies of the strong hydrogen bonding network of water, which is not described perfectly at the employed level of theory. The case of paracetamol as shown in Figure 3 is in particular interesting, as three sets of experiments from different groups have been reported and the difference among those is already quite large. This shows, that concerning the evaluation of predictions one has to consider the fact that experiments may be subject to a somebackground noise, for example due to not well specified experimental conditions.
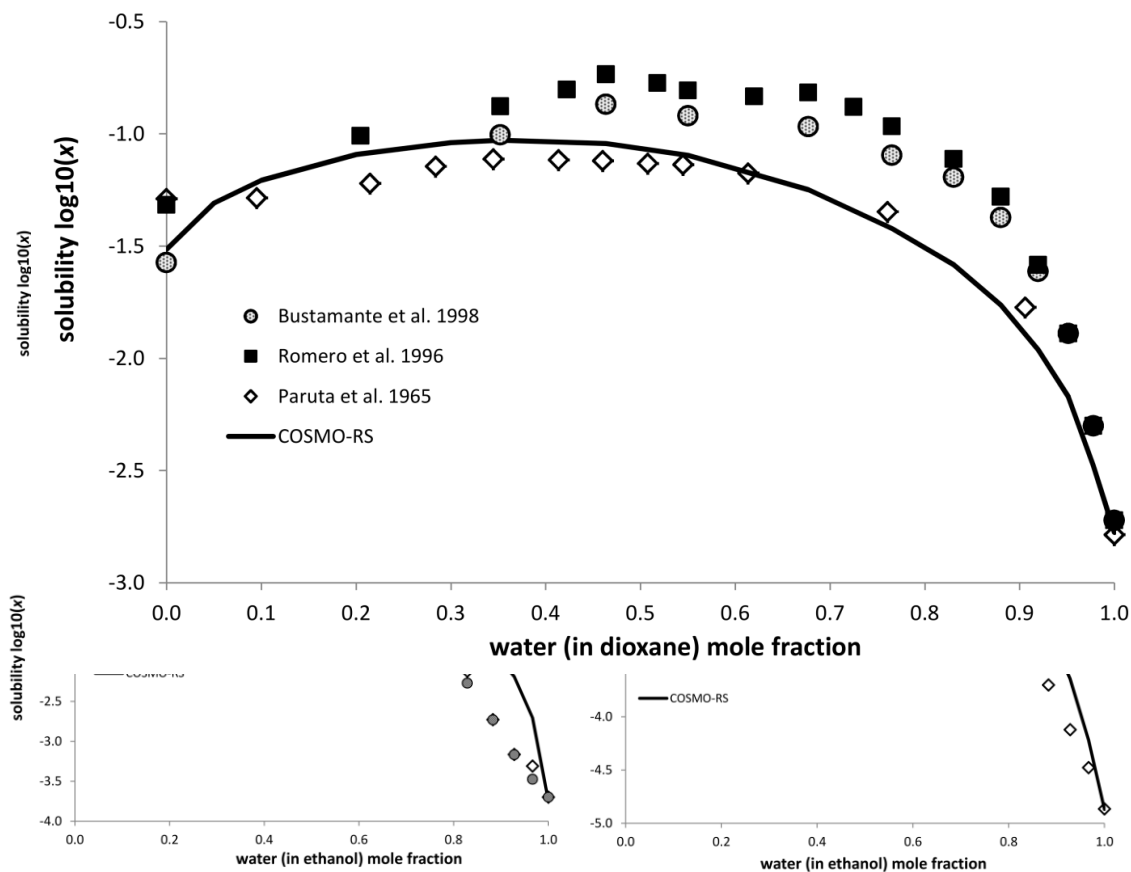
Figure 2 Predicted versus experimental solubility of sulfadiazine in a water/DMF mixture (a), sulfadiazine in a water/dioxane mixture (b), salicylic acid in a water/ethanol mixture (c), and prednisolone in water/ethanol (d).

An interesting application of this pure and mixed solvent prediction approach is the computational screening over a large set of solvents in order to identify the best possible solvent mixture for a given drug. This is combinatorial problem and even for binary solvents can become extremely costly. However, as solubility predictions can be done within fractions of a second, such vast screenings become feasible at a reasonable amount of time and resources.

Figure 3 Predicted versus experimental solubility of paracetamol in a water/dioxane mixture.

Solvate and cocrystal screening

Solvate and cocrystal formation are closely related phenomena. In both cases crystals are formed from two or more organic moieties with a defined stoichiometry that mostly have modified solubility and dissolution properties. For solvates one of the constituents is however liquid at room temperature ($\Delta G_{\text{fus}}(298K)=0$), whereas for cocrystal both components are solid ($\Delta G_{\text{fus}}(298K)>0$). The formation of solvates and cocrystals can be described via liquid phase thermodynamics according to the assumption that the interactions in the crystal are similar to a virtual supercooled liquid. Thus, the strength of the interactions in the cocrystal as compared with the pure reactants can be estimated via the mixing enthalpy (or equivalently the excess enthalpy) ΔH_{mix} of the API-coformer or API-solvent mixture with given stoichiometry. In other words, the mixing enthalpy is a rough approximation to the free energy of cocrystal formation ΔG_{cc}:

$$\Delta G_{cc} = \Delta H_{mix} - T\Delta S_{mix} - \Delta\Delta G_{fus} \approx \Delta H_{mix}$$

ΔS_{mix} is the entropy of mixing and $\Delta\Delta G_{\text{fus}}$ is the difference between the free energy of fusion of the cocrystal and the reactants. The overall entropy change between reactant and cocrystal and also $\Delta\Delta G_{\text{fus}}$ are assumed to be close to zero. For the overall entropy change between related crystalline states (of reactants and cocrystal) this seems to be reasonable assumption at least concerning the accuracy of the approach. For the remaining fusion enthalpy part this is a far more severe approximation because this corresponds to neglecting any enthalpic change due to the solid state order. The mixing enthalpy is given according to:

$$\Delta H_{mix} = H_{AB} - x_A H_A - x_B H_B$$

With H_{AB} being the enthalpy of the supercooled stoichiometric mixture of API A and coformer B and H_A (H_B) the enthalpy of the pure (supercooled) liquid A (B) and x being the respective mole fractions.

Though this approach neglects the solid state order it performs surprisingly well and usually an excellent enrichment is obtained as compared to a random trial of manually selected coformers. [11]
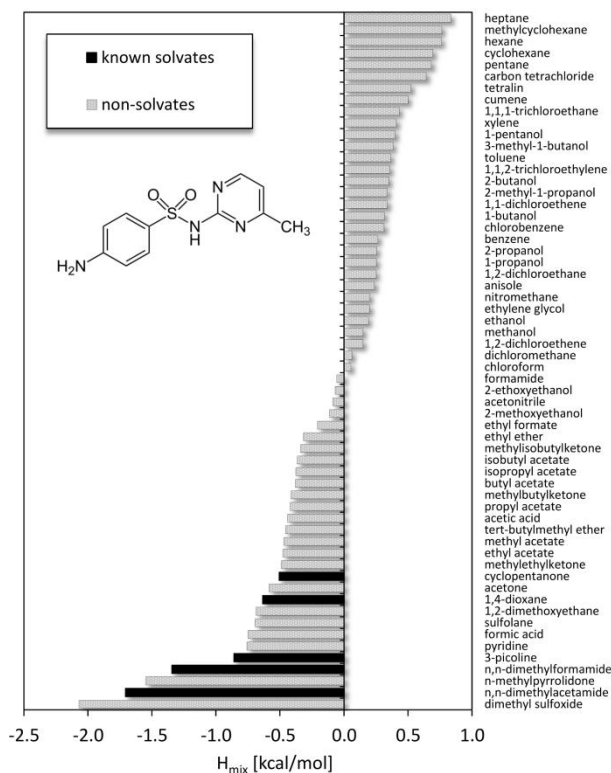

Figure 4 Solvate screening case study for sulfamerazine using experimentally known solvates[47] and FDA class 1,2 and 3 solvents.

Figure 4 shows computed excess enthalpies of the experimentally identified solvates of the sulfonamide antibacterial sulfamerazine (marked red) together with FDA class 1, 2 and 3 solvents.[48] The reported solvates cyclopentanone, 1,4-dioxane, dimethylacetamide, dimethylformamide and 3-picoline all show a strongly negative H_{mix} due to their strong enthalpic interaction with sulfamerazine. Thus, in this hypothetical screening, a significant enrichment is obtained, and for example just by trying out the five top solvents having the most negative excess enthalpy one would have found already three solvates. The fact that dimethyl sulfoxide and N-methylpyrrolidone are not known to form solvates in spite of their large negative excess enthalpies is probably due to an unfavorable crystal packing which is not taken into account by the excess enthalpy screening.

12

As a quantitative measure to assess the performance of a screening a receiver operating characteristics (ROC) diagram is quite useful. An ROC curve plots the true positive rate (number of true positive predictions/ total number of positive observations) versus false positive rate (number of false-positive predictions/ total number of negative observations) for a binary classifier system (cocrystal screening results) as its discrimination threshold (ΔH_{mix} cutoff) is varied from small to higher values. The area under the curve (AUC) measures the overall performance of the model. Predictions with higher AUCs are generally better with an AUC=1.0 being a perfect prediction and should always be higher than 0.5, indicating the model is better than random selection. An example on how to evaluate the statistics for the case of piracetam is given in the supplement (Table S1).

In addition to the surprisingly well results of just using the excess enthalpy / mixing enthalpy for screening we empirically found out that the number of rotatable bonds of API and coformers n is a further significant descriptor concerning the formation of cocrystals. We currently incorporate n by fitting a single parameter a to a set of experimental cocrystal screening results to obtain an effective screening function:

$$f\,(\Delta H_{mix}, n) = \Delta H_{mix} + a\max(1, n)$$

For the sake of simplicity the parameter a also takes into account stoichiometry of the system which for an unknown cocrystal is assumed to be always 1:1. The training set consists mostly out of experimental data we have used in one of our previous studies.[11] The coformer ranking according to f usually gives somewhat improved results in particular if the coformers vary with respect to the number of rotatable bonds.
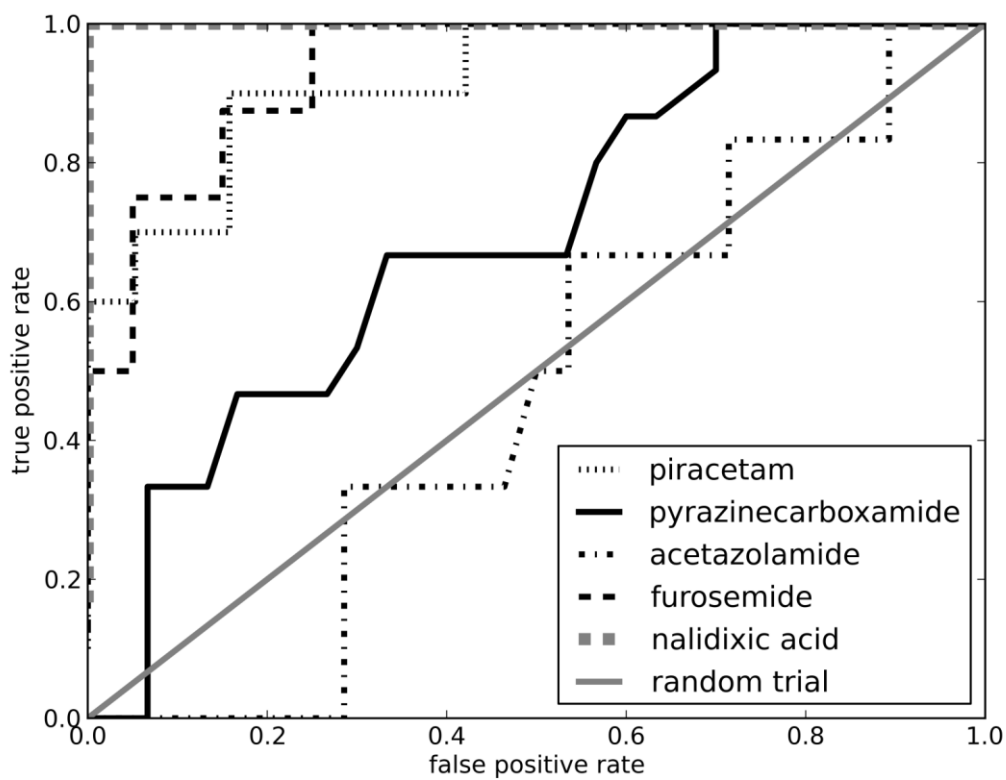
Figure 5 ROC plot of COSMO-RS cocrystal screenings using the mixing enthalpy and the number of rotatable bonds for some drugs. Experimental data as used in reference 15.

In Figure 5 shows some recent examples of COSMO-RS based cocrystal screenings using data from the validation study of Hunter and co-workers[15]. Table 2 shows some additional details about the calculations, in particular a comparison between screenings using quantum-chemical derived σ-profiles (AUC,FULL-CRS) and instantaneously, i.e. database generated σ-profiles (COSMO*quick* approach, AUC,CQ).

Please note that the number of tested coformers for acetazolamide is only 34 in Table 2, because from the original set comprising 36 coformers as given in reference 15 two duplicates have been removed (nicotinic acid=3-pyridinecarboxylic acid and isonicotinic acid=4-pyridinecarboxylic acid).

A significant enrichment is obtained in all cases (the AUC value is larger than 0.5), except for acetazolamide using COSMO*quick*, where a virtual screening would not have been better than a random trial. This is partially due to the fact that the thiadiazole group of the acetazolamide is not properly represented in the COSMO*quick* database used to generate the σ-profile. In most cases the software allows to identify such an underrepresented molecule/functional group by a score related to the quality of the as generated σ-profile prior to property calculation. The missing group/molecule can then eventually be added to the database. For the case of acetazolamide this gives a slight improvement (AUC=0.56). The remaining deviation from the experiment is due to the underlying approximations made, i.e. neglect of the fusion enthalpy differences between drug and coformers.

Conclusion

COSMO-RS as a thermodynamic theory provides a number of useful tools in order to support solid form selection and crystal engineering. With the ability to compute the chemical potential of a drug in its pure liquid state, in a solvent, or a solvent mixture, a broad variety of thermodynamic properties are accessible.

15

The prediction of solubility needs additional information about the solid state of a drug. Usually some experimental reference data as solubilities in one or several solvents or as the fusion enthalpy/melting point of the drug is sufficient to obtain quite accurate quantitative predictions.

Applications to solvate and cocrystal screening using the excess/mixing enthalpy of a supercooled API-coformer mixture have been proven to be surprisingly predictive, yielding good enrichments in most cases as compared to manual coformer selection.

The capability to screen for suitable solvents, solvates or cocrystals, accurately and efficiently, within a single theoretical framework and software, is particularly well suited to second rational drug development.

Funding

Notes

The authors declare the following competing financial interest(s): Andreas Klamt is CEO and Christoph Loschen is an employee of COSMO*logic*. COSMO*logic* commercially distributes the COSMO*quick* and COSMO*therm* software packages used in this paper.

References

1. Lipinski CA et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997;23: 3 - 25.

2. Klamt A et al. Prediction of aqueous solubility of drugs and pesticides with COSMO-RS. *J Comput Chem* 2002;23: 275-281.

3. Eckert F and Klamt A. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J* 2002;48: 369-385.

4. Pozarska A et al. Application of COSMO-RS- as an excipient ranking tool in early formulation development. *Eur J Pharm Sci* 2013;49: 505 - 511.

5. Klamt A et al. COSMOmic: a mechanistic approach to the calculation of membrane−water partition coefficients and internal distributions within membranes and micelles. *J Phys Chem B* 2008;112: 12148-12157.

6. Diedenhofen M et al. Prediction of infinite dilution activity coefficients of organic compounds in ionic liquids using COSMO-RS†. *J. Chem. Eng. Data* 2003;48: 475-479.

7. Eckert F and Klamt A. Accurate prediction of basicity in aqueous solution with COSMO-RS. *J. Comp. Chem.* 2006;27: 11-19.

8. Guerard JJ and Arey JS. Critical evaluation of implicit solvent models for predicting aqueous oxidation potentials of neutral organic compounds. *J Chem Theory Comput* 2013;9: 5046-5058.

9. Buggert M et al. COSMO-RS calculations of partition coefficients: different tools for conformation search. *Chem Eng Technol* 2009;32: 977-986.

10. Blagden N et al. Crystal engineering of active pharmaceutical ingredients to improve solubility and dissolution rates. *Adv Drug Deliv Rev* 2007;59: 617-630.

11. Abramov YA et al. Rational coformer or solvent selection for pharmaceutical cocrystallization or desolvation. *J. Pharm. Sci.* 2012;101: 3687.

12. Luner PE et al. Evaluation of milling method on the surface energetics of molecular crystals using inverse gas chromatography. *Cryst Growth Des* 2012;12: 5271-5282.

13. Fabian L. Cambridge structural database analysis of molecular complementarity in cocrystals. *Cryst Growth Des* 2009;9: 1436-1443.

14. Musumeci D et al. Virtual cocrystal screening. *Chem Sci* 2011;2: 883-890.

15. Grecu T et al. Validation of a computational cocrystal prediction tool: comparison of virtual and experimental cocrystal screening results. *Crystal Growth & Design* 2013;14: 165-171.

16. Klamt A. The COSMO and COSMO-RS solvation models. *WIREs: Comput. Mol. Sci.* 2011;1: 699-709.

17. Klamt A. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* 1995;99: 2224-2235.

18. Loschen C and Klamt A. COSMO*quick*: a novel interface for fast σ-profile composition and its application to COSMO-RS solvent screening using multiple reference solvents. *Ind Eng Chem Res* 2012;51: 14303-14308.

19. Chen C-C and Crafts PA. Correlation and prediction of drug molecule solubility in mixed solvent systems with the nonrandom two-liquid segment activity coefficient (NRTL-SAC) model. *Ind Eng Chem Res* 2006;45: 4816-4824.

20. Loschen C and Klamt A. New Developments in Prediction of Solid-State Solubility and Cocrystallization using COSMO-RS Theory. In: Abramov YA, ed. *Computational Approaches in Pharmaceutical Solid State Chemistry*. 2014-2015. In preparation.

21. Eckert, F and Klamt, A. COSMO*therm*, Version C3.0-Revision 13.01; COSMO*logic* GmbH & Co.KG, Leverkusen, Germany, 2013.

22. Loschen C and Klamt A, COSMO*quick*, Version 1.2; COSMO*logic* GmbH & Co. KG, Leverkusen, Germany, 2013.

23. TURBOMOLE V6.5 2013. A development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; Available at: http://www.turbomole.com.

24. Bradley, JC et al. Open notebook science challenge. http://www.onschallenge.wikispaces.com/ (accessed 19/05/2014)

25. Bustamante P et al. Partial solubility parameters and solvatochromic parameters for predicting the solubility of single and multiple drugs in individual solvents. *J Pharm Sci* 1993;82: 635.

26. Granberg RA and Rasmuson ÅC. Solubility of paracetamol in pure solvents. *J Chem Eng Data.* 1999;44: 1391-1395.

27. MonÁrrez CI et al. Solubility of 9-fluorenone in organic nonelectrolyte solvents: comparison of observed versus predicted values based upon mobile order theory. *Phys Chem Liq* 2003;41: 73-80.

28. Rager T and Hilfiker R. Cocrystal formation from solvent mixtures. *Cryst Growth Des* 2010;10: 3237-3241.

29. Ma Z and Zaera F. Role of the solvent in the adsorption−desorption equilibrium of cinchona alkaloids between solution and a platinum surface: correlations among solvent polarity, cinchona solubility, and catalytic performance. *J Phys Chem B* 2005;109: 406-414.

30. Daniels CR et al. Mathematical correlation of 4-aminobenzoic acid solubilities in organic solvents with the Abraham solvation parameter model. *Phys Chem Liq* 2004;42: 633-641.

31. Barra J et al. Proposition of group molar constants for sodium to calculate the partial solubility parameters of sodium salts using the van krevelen group contribution method . *Eur J Pharm Sci* 2000;10: 153 - 161.

32. Starr JN and King CJ. Water-enhanced solubility of carboxylic acids in organic solvents and its application to extraction processes. *Ind Eng Chem Res* 1992;31: 2572-2579.

33. Wei D and Cao W. solubility of adipic acid in acetone, chloroform, and toluene. *J Chem Eng Data* 2009;54: 152-153.

34. Lihua F et al.. measurement and correlation for solubility of adipic acid in several solvents . *Chin J Chem Eng* 2007;15: 110 - 114.

35. Gaivoronskii A and Granzhan V. solubility of adipic acid in organic solvents and water. *Russ J Appl Chem* 2005;78: 404-408.

36. Beerbower A et al. Expanded solubility parameter approach I: naphthalene and benzoic acid in individual solvents. *J Pharm Sci* 1984;73: 179-188.

37. De Fina KM et al. Solubility of 2-hydroxybenzoic acid in select organic solvents at 298.15K. *J Chem Eng Data* 1999;44: 1262-1264.

38. Martin A et al. Expanded solubility parameter approach II: p-hydroxybenzoic acid and methyl p-hydroxybenzoate in individual solvents. *J Pharm Sci* 1984;73: 188-194.

39. Elworthy PH and Worthington HEC. The solubility of sulphadiazine in water-dimethylformamide mixtures. *J Pharm Pharmacol* 1968;20: 830-835.

40. Peña M et al. Solubility parameter of drugs for predicting the solubility profile type within a wide polarity range in solvent mixtures. *Int J Pharm* 2006;321: 155 - 161.

41. Jouyban A et al. Solubility prediction of salicylic acid in water-ethanol-propylene glycol mixtures using the Jouyban-Acree model. *Pharmazie*; 61: 318-321.

42. Ali HSM et al. Solubility of budesonide, hydrocortisone, and prednisolone in ethanol + water mixtures at 298.2 K. *J Chem Eng Data* 2010;55: 578-582.

43. Paruta AN and Irani SA. Dielectric solubility profiles in dioxane–water mixtures for several antipyretic drugs. Effect of substituents. *J Pharm Sci* 1965;54: 1334-1338.

44. Romero S et al. Rhe behavior of paracetamol in mixtures of amphirotic and amphiprotic-aprotic solvents. Relationship of solubility curves to specific and nonspecific interactions. *Chem Pharm Bull* 1996;44: 1061-1064.

45. Bustamante PP et al. Enthalpy–entropy compensation for the solubility of drugs in solvent mixtures: paracetamol, acetanilide, and nalidixic acid in dioxane–water. *J Pharm Sci* 1998;87: 1590-1596.

46. Bustamante PP et al. A modification of the extended hildebrand approach to predict the solubility of structurally related drugs in solvent mixtures. *J Pharm. Pharmacol* 1993;45: 253-257.

47. Aitipamula S et al. The solvates of sulfamerazine: structural, thermochemical, and desolvation studies. *CrystEngComm* 2012;14: 691-699.

48. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073395.pdf (accessed 19/05/2014)

Table 1 Results of solubility predictions for different drugs and typical cocrystal formers. Deviation from experiment is given as root mean squared error (RMSE). The RMSE is given relative to solubilities measured in logarithmic mole fractions ($\log10(x)$). Number of reference solvents (#no refs) and total solvents (#no solvents) are also specified.

drug/coformer	#no refs	#no solvents	RMSE	source	exptl. data
sulfadiazine	3	19	0.42	18	25
paracetamol	4	23	0.48	18	26
fluorenone	3	21	0.34	18	27
saccharin	3	9	0.42	18	28
cinchonidine	3	23	0.87	18	29
4-aminobenzoic acid	4	23	0.58	this work	30,31
adipic acid	4	16	0.51	this work	32,33,34,35
benzoic acid	4	29	0.27	this work	36
salicylic acid	4	26	0.51	this work	31,37
4-hydroxy benzoic acid	4	20	0.34	this work	38

Table 2 Results of different computational cocrystal screenings. Experimental data as used in reference 15. Screening scores have been measured using the area under the ROC curve (AUC) value. Results for the method as used in reference 15 (SSIP), the full COSMO-RS level using quantum-chemical computed σ-profiles (CRS) and the COSMO*quick* level using database generated σ-profiles are given (CQ). CRS and CQ take into account the number of rotatable bonds of API and coformers. The final two columns give the number of coformers in the test set and the number of experimentally confirmed cocrystals, respectively.

drug	AUC,SSIP	AUC, CRS	AUC,CQ	n,test set	n,cocrystals
piracetam	0.93	0.88	0.92	29	10
pyrazinecarboxamide	0.79	0.78	0.69	45	15
acetazolamide	0.78	0.62	0.47	34	6
furosemide	0.65	0.95	0.94	28	8
nalidixic acid	1.0	1.00	1.00	22	6
AUC,mean	0.83	0.85	0.80		