This is a postprint of the Industrial Engineering Chemistry Research, Volume 51, 12.October 2012, Pages 14303-14308.
The original article can be found under http://pubs.acs.org/doi/abs/10.1021/ie3023675

COSMO*quick*: a Novel Interface for Fast σ-Profile Composition and its Application to COSMO-RS Solvent Screening using Multiple Reference Solvents

*Christoph Loschen**[1], Andreas Klamt*[1,2]

[1]COSMOlogic GmbH & Co. KG, Leverkusen, Germany

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

KEYWORDS: COSMO-RS, Solubility Prediction, Fluid Phase Thermodynamics, Reference Solvents, NRTL-SAC, COSMOquick, COSMOfrag

ABSTRACT: We present a novel, simpler to use modification of the standard COSMO-RS solubility prediction scheme which in addition can achieve higher accuracy by the usage of multiple experimental reference solubilities. When using only one reference solvent the approach reduces to the original COSMO-RS based solubility prediction. A considerable speedup and simplification compared to the original COSMO-RS arises from the usage of approximate σ-profiles generated from a database of COSMO-files from 65000 diverse molecules. This method

enables fast and accurate solvent screening. Solubility predictions using the novel approach on pure solvents perform favorable when compared to NRTL-SAC calculations. The new method is accessible via a graphical user-interface (COSMO*quick*) and combines the reliability and broad applicability of COSMO-RS theory with some practical advantages of more empirical solubility models.

1. Introduction

Solubility prediction is an important task in pharmaceutical process and drug development and in chemical engineering. Computational methods can be useful for example when the available amounts of a new substance are small and costly. A crucial problem is the ever decreasing water solubility of novel drugs, and predictive tools may be used to screen for solvents or solvent mixtures with optimal solubility. Applications of solubility calculations are manifold and quite a few methods exist addressing this issue. Widely used solubility estimation schemes are the traditional Hansen parameter approach[(1)] and the NRTL-SAC model[(2),(3)]. Amongst others, the UNIFAC[(4)] and PC-SAFT[(5)] models are also used for solubility prediction. Those methods have a common disadvantage: they perform well in their core region of parameterization, but often have severe problems with extrapolation, if less common or novel functional groups or group combinations appear in the solutes or solvents. The COSMO-RS approach as introduced by Klamt[(6),(7)] is known to be much more predictive towards such situations, since it is based on molecular polarity information derived from first-principles calculations. The respective screening charge density histogram (σ-profile) of a molecule is then used to compute liquid-phase properties via statistical thermodynamics. In general, this results in a wide applicability and extrapolative ability of the method; COSMO-RS theory has been applied successfully for many different physico-chemical properties in addition to solubility

prediction and solvent screening[8],[9], like ionic-liquids[10], pK_a-prediction[11], cocrystal formation[12] and many more. Recently even a σ-profile based algorithm for quantifying ligand-receptor interactions in proteins has been introduced.[13] In spite of the generality of the COSMO-RS theory, there may be practical cases, where strongly parameterized methods may be seen as advantageous. Given the case that there is sufficient experimental data available, those methods are not only more easily applied by a scientist without background in computational chemistry, they may also give superior results within the limited scope of the given parameterization. To address this issue we are introducing in this paper a modified COSMO-RS based solubility prediction scheme. It uses a correction of the chemical potentials of the solute in solution which is capable of interpolating between multiple reference solvents. Moreover a considerable reduction of the computational demands and usage complexity is achieved by allowing for the replacement of the quantum chemical calculations by approximated σ-profiles based on a large database of pre-calculated compounds. Hence, the novel approach is especially suited for efficient solvent screening where a few experimental data points are available, a typical scenario for an early stage drug development process.

2. Theoretical Basis

The COSMO-RS (Conductor like Screening MOdel for Real Solvents) theory is basically a combination of the dielectric continuum solvation model COSMO[14] with the statistical thermodynamics treatment of interacting surfaces. It is implemented in the program code COSMO*therm* which is able to calculate the chemical potential and of almost arbitrary solutes in almost any pure or mixed solvent. Within several other applications in many areas of chemistry, chemical engineering and pharmaceutical chemistry, COSMO*therm* has been proven to be a

valuable tool for solvent screening, i.e. for screening for a suitable solvent for a given solute X. A detailed description of COSMO-RS and its applications can be found in a recent review.[7]

2.1. Multiple Reference Solvents

From the σ-profiles, or more generally from the COSMO surface charge information contained in the COSMO files, COSMO-RS readily computes the chemical potentials of a solute X in any pure or mixed solvent i. With the chemical potential of the pure and solvated state the mole fraction solubility x_i may be computed according to:

$$(1)\quad \log_{10}(x_i) = \left[\mu_X^X - \mu_i^X - \Delta G_{fus}^X\right] / RT \ln(10)$$

μ_X^X and μ_i^X are the (pseudo) chemical potentials[15] of the pure liquid or sub-cooled liquid X and the solute X in solvent i, respectively. The chemical potential μ_i^X depends on the concentration of the solute, and thus equation (1) usually is solved iteratively by the original COSMO*therm* program. For low solubilities the infinite dilution estimate of the chemical potential can be used, avoiding the iterative procedure. This approximation is used in COSMO*quick* and throughout this paper.

The free energy of fusion ΔG_{fus} usually is not known and cannot directly be calculated, since an accurate prediction of ΔG_{fus} from the structure in particular would require the prediction of the crystal structure. Therefore, for the purpose of solvent screening ΔG_{fus} can be fitted to experimental solubility data in one or more reference solvents. While ΔG_{fus} by definition does not depend on the solvent, in practice some variation of the fitted ΔG_{fus} value will occur depending on the solvent used for the adjustment, since the solvent-dependent prediction error will be subsumed into the fitted ΔG_{fus}. Therefore, in this paper we introduce the possibility to

allow for multiple reference solvents by using the differences in the free energy of fusion to correct the chemical potentials μ_i^X of the solute. The correction term is interpolated based on the similarity between the reference solvents and the solvent under scrutiny. In detail, the average free energy of fusion $< \Delta G_{fus} >$ is calculated from the references and a correction term for each reference solvent i is obtained:

$$(2)\quad \Delta\mu_{cor,i}^{X} = \Delta G_{fus,i}^{X} - \left\langle \Delta G_{fus}^{X} \right\rangle \quad i = 1...n$$

Then, the correction terms for each new solvent j is weighted with a weight factor c_{ij} by its similarity with each reference i and an individual solvent specific correction is calculated:

$$(3)\quad \Delta\mu_{cor,j}^{X} = \sum_{i}^{references} c_{ij}\Delta\mu_{cor,i}^{X}$$

$$(4)\quad c_{ij} = w_{ij}^{A} \Big/ \sum_{i}^{references} w_{ij}^{A}$$

The weighting factors c_{ij} are determined by the so-called σ-potential similarity of solvent j and reference i, w_{ij}:

$$(5)\quad w_{ij} = \exp\left(- \sum_{m=-0.02}^{m=+0.02} \left| \mu_j^X(\sigma_m) - \mu_i^X(\sigma_m) \right| \right)$$

The reference solvent with the most similar σ–potential and thus the closest physicochemical resemblance gets the highest weight in equation (3). To avoid the dominance of just one reference, and to mitigate the risk of overemphasizing a single, potentially questionable measurement, the weighting factor w_{ij} is enhanced with an exponent A and subsequently

normalized. Currently good results have been obtained with an *ad-hoc* choice of A=0.5 for mixtures, which is also the internal default. For an *a posteriori* justification for this choice of the parameter and for variation of computed solubility with varying A please refer to Figure S1 in the supplement.

Finally, we obtain the solubility of the solute in solvent j by the following modification of equation (1):

$$(6)\quad x_j = \exp\left[\frac{(\mu_X^X - \mu_j^X - \Delta\mu_{cor,j}^X - <\Delta G_{fus}^X>)}{RT}\right]$$

The approach works best if a diverse set of reference solvents is used. For example one may use a non-polar solvent like hexane, a donor-acceptor solvent like water and an acceptor solvent like acetone. Usual three balanced solvents are sufficient; however, even a smaller set of solvents will give satisfying results. Without any reference solvent it is still possible to get relative solubilities with the method.

2.2. Approximate σ-Profiles

In order to shortcut the usual quantum-chemical calculations we have used a fragmentation approach to generate σ-profiles as implemented in our COSMO*frag* code.(16) The underlying concept here is the approximate composition of the σ-profile of a new molecule from existing σ-profiles of molecules that have already been pre-calculated quantum-chemically. Currently the COSMOfrag database holds more than 65,000 molecules with diverse functional groups ranging from solvents to complex drugs within a database. In this way, there is no need for quantum chemical calculations prior to COSMO-RS calculations of a new molecule at the

expense of a small loss of accuracy. Technically, only a 2D structure information (e.g. SMILES or SD file) is sufficient as input. Thus, within a fraction of a second σ-profiles of nearly quantum-chemical accuracy are available, allowing for the high throughput screening of large number of molecules.

3. Computational Details

COSMO-RS based multi-reference solubility calculations have been carried out with COSMO*quick*. COSMO*quick* is a JAVA-based software tool which internally calls COSMO*frag*(16) for the generation of σ-profiles and COSMO*therm* for thermodynamic computations. Approximated σ-profiles are generated by accessing the COSMO*frag* database with about 65.000 diverse chemical compounds and their respective COSMO files, which avoids costly quantum-chemical calculations. Structures for those COSMO files have been obtained by AM1/COSMO geometry optimization followed by a single point DFT/COSMO calculation (BP-SVP) with TURBOMOLE.[17] From the σ-profiles chemical potentials in the liquid phase are calculated according to standard COSMO-RS theory.[6],[7]

NRTL-SAC computations have been obtained with our own implementation following the equations given by Chen and co-workers.[2] After the original publication they have published modified solvent parameters which have been used in this work instead of their 2004 parameter set.[3] Solute parameters, i.e. the conceptual segments X,Y-,Y+ and Z have been computed by minimizing the root mean squared error (RMSE) error between experimental and computed logarithmic solubilities using a downhill SIMPLEX algorithm. Several starting vectors [X,Y-

,Y+,Z] were tried in order to avoid getting stuck in a local minimum of the optimization space. The code is written in Python and makes use of the numerical Python extension NumPy.(18)

4. Results and Discussion

4.1. COSMO-RS Solubility Prediction Using Multiple References and Comparison with NRTL-SAC

To better assess the performance of the COSMO-RS multiple reference approach we have compared our solubility results with the NRTL-SAC method. First, we deal with a recently published study[3] concerning the solubility prediction of the three drugs: paracetamol[19], sulfadiazine[20] and cimetidine.[21] In this work Chen *et al.* have used their previously published NRTL-SAC method with a modified set of solvent parameters using the following reference solvents for the three solutes: water, dioxane, toluene, DMA, DMSO, acetone and benzene for sulfadiazine; toluene, chloroform, water, acetone, ethanol, DMSO, acetonitrile and THF for paracetamol; n-octane, acetonitrile, water, ethanol, MEK and ethylacetate for cimetidine. For the COSMO*quick* calculations the following reduced set of references proved to be sufficient: water, dioxane and toluene for sulfadiazine; toluene, chloroform, water and acetone for paracetamol; n-octane, acetonitrile, water and ethanol for cimetidine. Table 1 shows their results as taken from the original reference (3), the results of our own NRTL-SAC implementation using the original parameters and the results of the multiple reference calculations using the COSMO*quick*

software. For all solutes a fragmentation was enforced, even in cases where the solute exists as a whole in the database. The fragmentation effect on the solubility is separately discussed in the section below. For better comparability solubility is given in logarithmic units, i.e. ln(S), with S in mg/g solvent, as specified in the original reference.

Table 1. Comparison of NRTL-SAC and COSMO*quick* solubility prediction results on three different drugs.

solute	method	no. reference solvents	tot no. of solvents	RMSE[a]
sulfadiazine	NRTL-SAC	7	19	2.872 (2.950)[b]
sulfadiazine	COSMO*quick*	3	19	1.331
paracetamol	NRTL-SAC	8	23	0.993 (1.075)[b]
paracetamol	NRTL-SAC	8	23[c]	1.272
paracetamol	COSMO*quick*	4	23[c]	1.085
cimetidine	NRTL-SAC	6	11	1.910 (0.799)[b]
cimetidine	COSMO*quick*	4	11	1.137

[a] root mean squared error (RMSE) of ln(S), S in mg/g solvent, [b] in parenthesis original results from reference (3). [c] using the newer and probably more accurate CCl_4 experimental value from reference (22).

For the drug sulfadiazine, a sulfonamide antibiotic, the performance of COSMO*quick* (RMSE=1.331) on a set of 19 reference solvents[20] is clearly superior to NRTL-SAC (RMSE=2.872), in spite of using only 3 reference solvents instead of 7. Indeed, even the authors state in the 2006 paper[3] on the sulfadiazine results: "The difficulty with some of the sulfadiazine solubility data creates the possibility that the current NRTL-SAC model formulation and parameters may be sub-optimal for certain classes of solvents or solutes." The main problem

for the NRTL-SAC method in this case seems to be the incorrect descriptions of the solubility in alcoholic solvents.

The case of paracetamol is particular interesting because the COSMO*quick* method did allow to identify an experimental flaw concerning the measurement of CCl_4. Worried by the strong deviations of more than two magnitudes between the experimental value of Granberg and Rasmuson[19] and our predicted results we contacted one of the authors who stated that more recent measurements of the paracetamol solubility in carbon tetrachloride of Mota *et al.*[22] are probably more reliable.[23] And indeed, there is an excellent agreement of our predicted solubility (log10(w)=-5.3) with the recent value from Mota *et al*.(log10(w)=-5.5). The NRTL-SAC value (log10(w)=-3.7) is much closer to the older and thus meanwhile most likely to be considered wrong value for the solubility of paracetamol in CCl_4.

Table 2. Comparison of measured and predicted solubilities of paracetamol in carbon tetrachloride.

Solubility log10(w), w in g/g	comment	reference
-3.0	experimental, T=333K	Granberg et al., 1999 [(19)]
-5.5	experimental, T=298K	Mota et al., 2009 [(22)]
-3.7	predicted	NRTL-SAC, this work
-5.3(-4.9)	predicted, T=298K (T=333K)	COSMO*quick*, this work

Table 2 gives an overview over different literature and predicted values of the paracetamol/CCl_4 system. Even though there is obviously a strong temperature dependence of the paracetamol solubility this fact cannot explain the strong deviations between our and Mota's results as compared to Granberg's and the NRTL-SAC results. Although the paracetamol/CCl_4 system may be of subordinate practical relevance, this examples nicely demonstrates the extrapolative capability and reliability of the new model also for cases which are difficult to measure either due to their low solubility and/or due to their toxicity.

The last compound of Table 1 under investigation is cimetidine, a histamine receptor antagonist. According to Chen and coworkers the RMSE on 11 solvents including a parameter fit on 6 solvents yields an RMSE=0.80 (solubilities measured as ln(S) with S in mg/g) on the experimental data.(21) However, we cannot reproduce their published results, and we get a much larger RMSE=1.91 on the overall dataset. The strong deviation is mostly due to the solubility in octane, our computed NRTL-SAC value being log10(x) = -9.3 [ln(S)=-21.3] against experimental log10(x) = -6.8 [ln(S)=-15.6], i.e. being more than 2 magnitudes off at the log10(x) scale. (with x in mole fraction and S in mg/g). Leaving the octane value out we get a closer agreement with the original data (RMSE=0.86, ln(S), S in mg/g). From this finding we infer that the computed value for octane must have been omitted in the values published in reference (3). The COSMO*quick* value for octane using only 4 reference solvents agrees well with the experiment (log10(x)=-6.8 in mole fraction or ln(S)=-15.6, S in mg/g), therefore the overall deviation from the experimental dataset is significantly smaller with an RMSE=1.137.

A few solvents from the experimental sources have not been computed with NRTL-SAC due to the fact that no solvent parameters are tabulated for those. For example from the 26 solvents of the Granberg and Rasmuson dataset(19) only 23 could be computed in reference (3) with the

NRTL-SAC method because for 1-hexanol, 1-heptanol, and diethylamine no parameters were available. For the COSMO*quick* there are no such limitations as the approach knows no solvent parameterization, the σ-profiles for any solvent can be taken instantaneously from the database fragments. Thus, this approach is readily applicable even for novel or rare solvents without exceptions.

Using our own NRTL-SAC code we have made further comparisons between the two solubility prediction methods on a series of different drugs where the experimental data has been taken from literature. Table 3 gives a summary of those calculations.

Table 3. Comparison of the performance of solubilities as computed by COSMO*quick* and by NRTL-SAC. Root mean squared error (RMSE) of the computations has been calculated from the decadic logarithm of the experimental mole fraction solubility log10(x). Number of references (#no ref.) and total number of solvents including the references (#no solvents) are also given.

solute	RMSE (COSMO*quick*)	#no ref.	RMSE (NRTL-SAC)	#no ref.	#no solvents	exp. source
fluorenone	0.34	3	0.39 (0.37)[a]	5	21	[(24)]
xanthene	0.27	3	0.53(0.66)[a]	5	19	[(25)]
monuron	0.25	3	0.25(0.33)[a]	5	24	[(26)]
cinchonidine	0.87	3	0.94(0.87)[a]	5	23	[(27)]
saccharin	0.42	3	0.95(0.59)[a]	5	9	[(28)]

[a]results for optimized parameter A in parenthesis.

Table 4. NRTL-SAC parameters and reference solvents used to generate the results of Table 3.

molecule	X	Y-	Y+	Z	A	reference solvents
fluorenone	0.73	0	1.5605	0.0456	6.110[a]	n-hexane, acetonitrile, ethanol, dichloromethane, 1-octanol
fluorenone	1.5531	0	1.6442	0.3178	5.456[b]	
xanthene	1.739	0	1.389	0	6.181[a]	n-hexane, acetonitrile, ethanol, 1,2-dichloroethane, 1-octanol
xanthene	3.906	0.669	3.146	0	4.395[b]	
monuron	0.205	0.548	0.904	0.324	7.920[a]	n-hexane, ethylacetate, ethanol, 1,2-dichloroethane, 1-octanol
monuron	1.296	1.267	1.265	0	5.519[b]	
cinchonidine	2.220	1.757	0.937	0.173	7.621[a]	n-hexane, ethyl acetate, acetonitrile, chloroform, 1-butanol
cinchonidine	2.137	0.747	0.225	0.994	10.124[b]	
saccharin	0.461	1.562	0.372	0.000	7.677[a]	water, acetone, glycol, dioxane, acetic acid
saccharin	0.432	0.000	0.000	0.000	10.197[b]	

[a]Parameter A determined from $A=\Delta S_{fus}/T$ [b]Parameter A has been fully optimized

In their 2006 paper Chen and coworkers used an additional parameter A in order to improve the NRTL-SAC results, where ΔH_{fus} was still taken from the literature:

$$(7)\quad \ln K_{sp} = -\frac{\Delta G_{fus}}{RT} = \frac{\Delta S_{fus}}{R} - \frac{\Delta H_{fus}}{RT} = A + \frac{\Delta H_{fus}}{RT}$$

In other words, the entropy of fusion was used as a (partially) adjustable parameter. Because NRTL-SAC in its original form had at least four parameters to be fitted, four for each of the conceptual segments, plus the optional value of A, in total five references were chosen for the NRTL-SAC parameterization here. Table 4 shows the optimized solute parameters for both cases, with A being fixed, i.e. computed from the experimental entropy of fusion or alternatively being treated as free parameter. References have been selected to get a balanced set of solvents, i.e. non-polar, polar, H-bond accepting and H-bond donating ones, if available. Note that again from the available experimental data only those observations could be presented in Table 3 for which NRTL-SAC parameters do exist. While COSMO*quick* works in principle with only one reference, a recommended setup may make use of three reference solvents which are sufficient to obtain good results, thus the first three solvents of Table 4 have been used: n-hexane, acetonitrile, ethanol (flurorenone, xanthene); n-hexane, ethylacetate, ethanol (monuron); n-hexane, ethyl acetate, acetonitrile (cinchonidine) and water, acetone and glycol (saccharin).

Results for fluorenone, monuron and cinchonidine are of comparable accuracy for both methods, whereas for xanthene COSMO*quick* results are more accurate (RMSE=0.27 versus RMSE=0.53). Concerning saccharin, NRTL-SAC results improve significantly when the additional parameter A is relaxed, with the RMSE dropping from 0.95 to 0.59, however COSMO*quick* being still more accurate (RMSE=0.41). Since the introduction of the additional

flexibility by optimizing A does in some cases even yield worse results, as for example for the cases of xanthene and monuron, the improvement in the case of saccharin may be considered as accidental. Optimizing A basically means adapting the entropy of fusion in the given NRTL-SAC framework. Table 4 reveals that upon full relaxation of A the NRTL-SAC parameters may change dramatically, which questions the interpretability of the conceptual segments in this case.

4.2. The Effect of σ-Profile Approximation

Figure 1 compares the σ-profile of some drug molecules used for the solubility predictions presented herein, calculated by a DFT/COSMO computation versus its composition from database fragments. In general, the composition of σ-profiles works quite reliable for a broad class of molecules due to the diversity of molecules represented in the database. If a molecule exists in the database, as it may be very well the case for typical drugs or solvents, they are used directly and one obtains the original COSMO file at the SVP-AM1 level. Difficulties in σ-profile generation can arise if rare functional groups are not properly represented in the database, in such cases a warning message is generated. Moreover, an inaccurate composition may be caused by internal hydrogen bonds that are split among fragments and thus are not properly reproduced. Conformations of molecules are not taken into account yet, as well as ionic compounds which are not representable currently. Nevertheless, it should be noted, that molecules with bad compositions still can be added as full COSMO files. For further details of σ-profile composition and its performance we refer to reference (16).

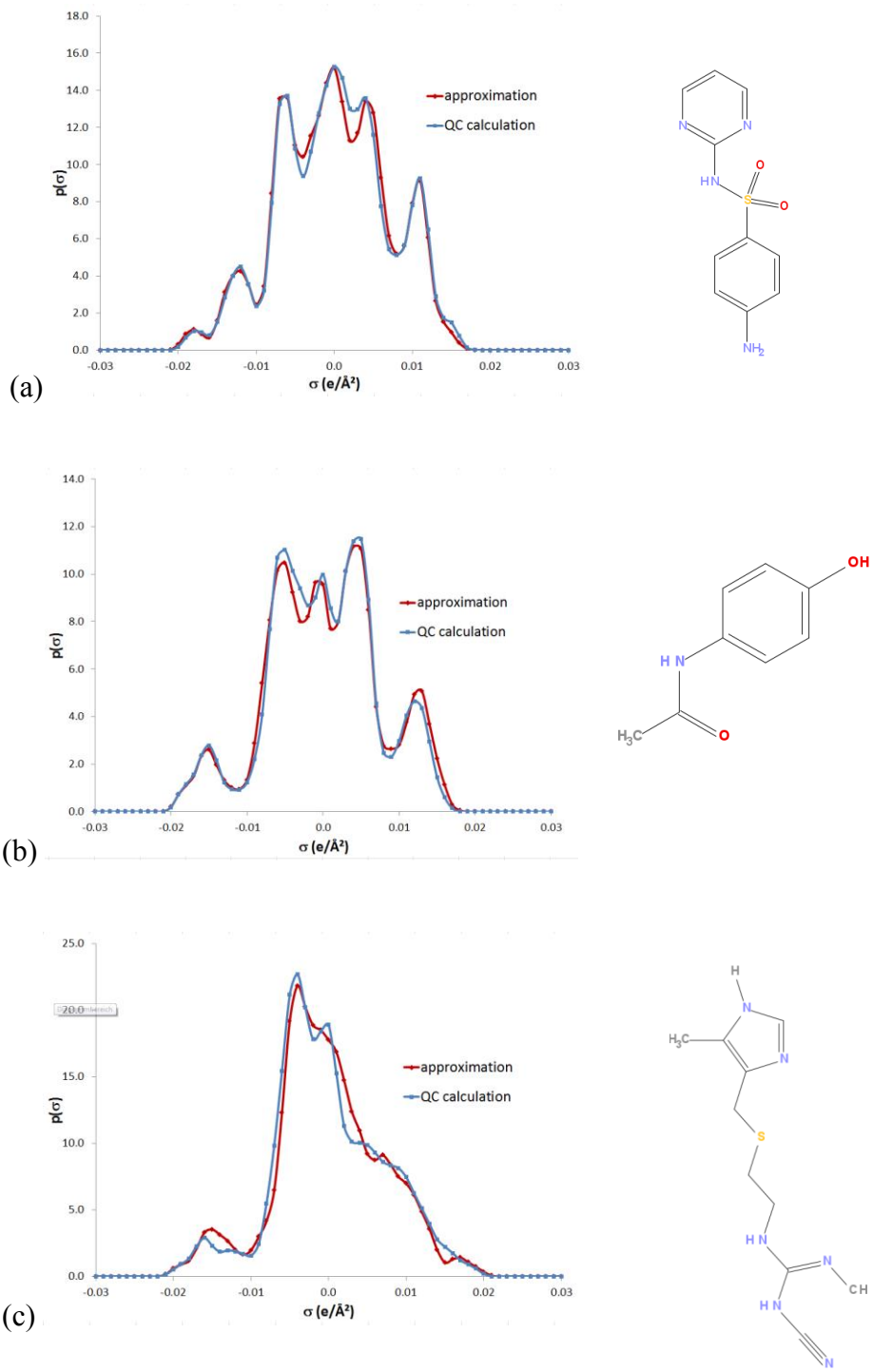

Figure 1. Comparison of quantum-chemically computed (blue line) versus approximated σ-profiles (red line), both generated at the BP-SVP-AM1 level for the drugs sulfadiazine (a), paracetamol (b) and cimetidine (c) as used in the solubility calculations above.

Table 5. Comparison of root mean squared errors (RMSE) from solubility predictions based on original σ-profiles at the SVP/AM1 level versus approximate σ-profiles composed from the COSMO*frag* database.

	RMSE on log10(x)		
solute	**original σ-profile**	**approximate σ-profile**	**no. of fragments**
sulfadiazine	0.41	0.42	2
cimetidine	0.49	0.49	4
paracetamol	0.48	0.48	2
fluorenone	0.34	0.34	3
xanthene	0.25	0.27	3
monuron	0.30	0.25	2
cinchonidine	0.80	0.87	3
saccharin	0.42	0.42	1

In Table 5 we have carved out the effect of the σ-profile generation on the computed solubilities. For each of the solutes once the original σ-profile obtained by a DFT calculation at the SVP/AM1 level was compared to the respective approximate, fragment composed σ-profile as taken from the database. From the results it is obvious that the approximation has a negligible influence on the final results presented here: the approximated σ-profiles of the 8 solutes used in this study yield the same accuracy as the ones from the original COSMO files. We consider the somewhat smaller RMSE using the approximated σ-profiles of monuron for fortuitous. The

fragmentation approach is restricted currently to the SVP-AM1 level. However there are no principle obstacles hindering its extension to a higher level of theory. Limitations of the COSMOfrag approach currently are the neglect of conformational flexibility and the restriction to neutral solutes and solvents.

5. Conclusion

We have introduced a novel method for efficient estimation of solubilities of organic compounds in different solvents. This multi-reference solubility approach makes use of one or more experimental data points, allowing for an empirical correction to chemical potentials and solubilities computed by COSMO-RS theory. Time-consuming quantum calculations can be avoided to a large degree by the composition of the σ-profiles required for the COSMO-RS calculations from pre-computed database fragments (COSMO*frag*). The multi-reference solubility prediction based on approximated σ-profiles was tested on several organic drugs and drug-like solutes. The general performance for pure solvents is slightly favorable compared to NRTL-SAC, and it needs less experimental data. Among the advantages of the new method is the general applicability, as no solvent parameters have to be determined beforehand. Moreover, very few references are required. Typically two or three diverse solvents are enough and even without any reference relative solubilities can be predicted. The whole workflow is integrated into COSMO*quick*, a JAVA based graphical user interface, allowing for an easy setup of solubility calculation and input of experimental data. The COSMO*quick* tool offers additional features as cocrystal screening[12] and the quick calculation of several ADME properties, partitioning coefficients and also a whole set of useful QSPR descriptors.

Supporting Information Available: Experimental and computed solubility data obtained with NRTL-SAC and COSMO*quick*. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Christoph Loschen. COSMOlogic GmbH & Co. KG, Leverkusen, Germany. Email to: loschen@cosmologic.de

Literature Cited

(1) Hansen, C. M. *Hansen Solubility Parameters: A User's Handbook*; CRC Press: Boca Raton, FL, 2000.
(2) Chen, C.-C.; Song, Y. Solubility Modeling with a Nonrandom Two-Liquid Segment Activity Coefficient Model. *Ind. Eng. Chem. Res.* **2004**, *43*, 8354.
(3) Chen, C.-C.; Crafts, P. A. 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.
(4) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures. *AIChE Journal* **1975**, *21*, 1086.
(5) Gross, J.; Sadowski, G. Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. *Ind. Eng. Chem. Res.* **2001**, *40*, 1244.
(6) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074.
(7) Klamt, A. The COSMO and COSMO-RS Solvation Models. *WIREs: Comput. Mol. Sci.***2011**, *1*, 699.
(8) Klamt, A.; Eckert, F.; Hornig, M.; Beck, M. E.; Bürger, T. Prediction of Aqueous Solubility of Drugs and Pesticides with COSMO-RS. *J. Comput. Chem.* **2002**, *23*, 275.
(9) Eckert, F.; Klamt, A. Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach. *AIChE J* **2002**, *48*, 369-385.
(10) Kahlen, J.; Masuch, K.; Leonhard, K. Modelling Cellulose Solubilities in Ionic Liquids Using COSMO-RS. *Green Chem.* **2010**, *12*, 2172.
(11) Eckert, F.; Klamt, A. Accurate Prediction of Basicity in Aqueous Solution with COSMO-RS. *J. Comp. Chem.* **2006**, *27*, 11.
(12) Abramov, Y. A.; Loschen, C.; Klamt, A. Rational Coformer or Solvent Selection for Pharmaceutical Cocrystallization or Desolvation. *J. Pharm. Sci.* **2012**, *101*,3687.
(13) Klamt, A.; Thormann, M.; Wichmann, K.; Tosco, P. COSMOsar3D: Molecular Field Analysis Based on Local COSMO σ-Profiles. *J. Chem. Inf. Model.* **2012**, *52*, 2157.
(14) Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799-805.

(15) Naim, A. B. *Solvation Thermodynamic;* Plenum Press: New York, NY, 1987.
(16) Hornig, M.; Klamt, A. COSMOfrag: A Novel Tool for High-Throughput ADME Property Prediction and Similarity Screening Based on Quantum Chemistry. *J. Chem. Inf. Model* **2005**, *45*, 1169.
(17) TURBOMOLE, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.

(18) Oliphant T. E. *Guide to NumPy*; Trelgol Publishing: USA 2006. http://numpy.scipy.org.
(19) Granberg, R. A.; Rasmuson, Å. C. Solubility of Paracetamol in Pure Solvents *J. Chem. Eng. Data.* **1999**, *44*, 1391.
(20) Bustamante, P.; Martin, A.; Gonzalez-Guisandez, M. A. Partial Solubility Parameters and Solvatochromic Parameters for Predicting the Solubility of Single and Multiple Drugs in Individual Solvents. *J. Pharm. Sci.* **1993**, *82*, 635.
(21) Westwood, M. C. *A Study on the Polymorphic Forms of Cimetidine*, M.Philos. Thesis, De Montfort University: U.K. 1998.
(22) Mota, F. L.; Carneiro, A. P.; Queimada, A. J.; Pinho, S. P.; Macedo, E. A. Temperature and Solvent Effects in the Solubility of Some Pharmaceutical Compounds: Measurements and Modeling. *Eur. J. Pharm. Sci.* **2009**, *37*, 499.
(23) Personal Communication from Prof. Åke C. Rasmuson, 2011.
(24) MonArrez, C.; Stovall, D.; Woo, J.; Taylor, P.; Acree W. E. 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.
(25) MonÁrrez, C. I.; Stovall, D. M.; Woo, J. H.; Taylor, P.; Acree, W. E. Solubility of Xanthene in Organic Nonelectrolyte Solvents: Comparison of Observed Versus Predicted Values Based Upon Mobile Order Theory. *Phys. Chem. Liq.* **2002**, *40*, 703.
(26) De Fina, K. M.; Sharp, T. L.; Chuca, I.; Spurgin, M. A.; Acree, W. E.; Green, C. E.; Abraham, M. H. Solubility of the Pesticide Monuron in Organic Nonelectrolyte Solvents. Comparison of Observed Versus Predicted Values Based upon Mobile Order Theory. *Phys. Chem. Liq.* **2002**, *40*, 255.
(27) Ma, Z.; 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.
(28) Rager, T.; Hilfiker, R. Cocrystal Formation from Solvent Mixtures. *Cryst. Growth Des.* **2010**, *10*, 3237.