In Silico Prediction of Medium Effects on Esterification Equilibrium Using the COSMO-RS Method

Maurizio Fermeglia,[†] Paolo Braiuca,[‡] Lucia Gardossi,[‡] Sabrina Pricl,[†] and Peter J. Halling*[,§]

Molecular Simulation Engineering (MOSE) Laboratory, DICAMP, and Dipartimento di Scienze Farmaceutiche, Università degli Studi di Trieste, P.le Europa 1 34127 Trieste, Italy, and WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, U.K.

This paper presents a new approach for predicting solvent effects on esterification reactions of industrial importance in the field of biocatalysis. The COSMO-RS method has been used to calculate the activity coefficients of the chemical species involved in various reactions, carried out in different solvents. For comparison we also used the traditional UNIFAC method. Three lipase-catalyzed esterifications were considered: (1) 1-dodecanoic acid with menthol in *n*-hexane, *n*-heptane, cyclohexane, 2,2,4-trimethylpentane, toluene, acetonitrile, and 2-methyl-2-butanol; (2) 1-dodecanoic acid and 1-dodecanol in *n*-hexane, *n*-heptane, cyclohexane, 2,2,4-trimethylpentane, and toluene; and (3) glycerol and *n*-octanoic acid in acetonitrile, benzene, and toluene and in the neat reaction mixture (without any solvent). Predicted activities were used to calculate the thermodynamic equilibrium ratio. This should be independent of medium, and the variation in COSMO-RS values is at most 9-fold (corresponding to a $\Delta G°$ of about 5.5 kJ/mol, which would still be a very useful prediction) and often only 2-fold (corresponding to less than 2 kJ/mol or 0.5 kcal/mol, therefore comparable with experimental error). UNIFAC is weaker, especially when important roles are played by conformational freedom, intramolecular interactions, strong polar effects, and charge distribution of molecules in the mixture. The relative percent deviations from the mean of equilibrium constants in different solvents range between 17 and 49 for COSMO-RS versus 32 to 65 for UNIFAC. The COSMO-RS method opens up new perspectives for the development of theoretical models for solvent selection with general applicability.

Introduction

Biocatalysis is becoming an important tool for the production of fine chemicals, because of several drivers. From the environmental point of view, the chemical industry is exploring bioprocesses to improve long-term sustainability. Attractive features of biocatalysts include versatility; substrate selectivity; regio-, chemo-, and enantioselectivity; limited use of hazardous reagents; reduction of byproducts; and catalysis at moderate temperature and pressure.

Since the pioneering work of Klibanov (*1*), it has been clear that the technological utility of biocatalysts can be widened further by carrying out some reactions in organic media. This allows the possibility of dissolving hydrophobic molecules, avoiding bacterial contamination, and improving product yield in processes such as ester synthesis, where the position of equilibrium is reversed with respect to aqueous solutions. The several advantages of biotransformations are discussed in refs *1*−*4*.

Unlike most other synthetic methods, many enzymatic processes operate under conditions where the position of chemical equilibrium may control the final yield achieved. In part this reflects the avoidance of forcing conditions and aggressive reagents; it is these that make many purely chemical transformations essentially irreversible thermodynamically. For example, the chemical synthesis of an ester would normally use strongly acidic conditions, water removal, or an activated acyl donor like an acid chloride. In such reactions yield is limited by kinetic factors including side reactions. In contrast, biocatalytic ester synthesis often uses direct condensation of the acid and alcohol under neutral conditions with significant levels of water present (giving a water activity not much less than 1). Nevertheless, very high conversions can be achieved in such thermodynamically controlled syntheses, for example, in nonpolar organic solvents (*5*−*7*).

When the conversion attainable is close to the thermodynamic equilibrium, understanding and predicting effects on this become important. Hence biocatalysis has stimulated a number of studies of how the equilibria of organic reactions are affected by the conditions, particularly the choice of solvent (*6*, *8*−*12*). Careful measurements of equilibrium positions are quite laborious but are crucial for understanding reaction energetics and for process optimization. In this context, the most accurate and extensive studies are those of Tewari and colleagues (*6*, *10*−*12*) on a variety of ester synthesis equilibria.

It is clear that a reliable theoretical technique for predicting the reaction thermodynamics would be a valuable tool for biocatalysis in nonconventional media. Correct predictions of the activity coefficients of each reactant in each solvent lead to prediction of the solvent dependence of the equilibrium constant, thus allowing the prediction of the feasibility of a biocatalytic synthesis.

* To whom correspondence should be addressed. Ph: +44 (0)141 548 2683. Fax: +44 (0)141 548 4822. Email: p.j.halling@strath.ac.uk.

[†] DICAMP, Università degli Studi di Trieste.

[‡] Dipartimento di Scienze Farmaceutiche, Università degli Studi di Trieste.

[§] University of Strathclyde.

10.1021/bp060081o CCC: $33.50
Published on Web 00/00/0000

It is possible to make predictions of solvent effects on dilute solution equilibria using experimental partition coefficients. Application of this approach to esterification reactions that might be catalyzed enzymatically (*8*) also revealed a useful correlation with solvent properties. However, these methods are not applicable to more concentrated systems in which ideal dilute behavior no longer applies and of course require data on partition coefficients that is usually not available.

Group contribution methods, especially UNIFAC (*13*), have found successful industrial applications, especially to mixtures of hydrocarbons and other nonpolar compounds. Some biocatalysis-related applications can also be found in the literature. UNIFAC has been examined for the prediction of esterification equilibria (*14*) and was found to perform quite well for some simple esters. However, predictions were poor for glycerol esters, probably because of the limitations in the treatment of intramolecular interactions and other proximity effects when polar groups are close together. UNIFAC predictions are independent of the location of the functional groups in the molecules and cannot account for the effect that one group has on another group if they are close in the molecule.

Predictions of activity coefficients in enzyme reaction mixtures can be useful for understanding kinetics as well. There are often large solvent effects on the reaction rate and the enzyme's affinity for the reactants. However, it is clear that these effects can often be largely explained by solvent effects on solvation of the reactants themselves, without needing to invoke effects on the enzyme (*15*−*17*). Even if there are some effects on the enzyme, these reactant solvation effects will always occur as well. Hence understanding the effects of solvent on substrate solvation offers useful predictions about kinetics as well. UNIFAC has also found a useful application in the prediction of solvent dependence of enzymatic prochiral selectivity, by calculating the activity coefficients of the solvated portions of substrates after their binding to the enzyme (*18*).

The basis of the UNIFAC method is to consider molecules as an aggregation of functional groups. This approach offers the great advantage of being able to describe a virtually infinite number of different compounds with a small number of experimentally estimated parameters, but this approximation often limits dramatically the accuracy of results. Recent expansion of the parameter tables has improved the calculation of intermolecular interactions. However, the inaccuracy of the basic assumption of the additivity of group contribution cannot be overcome, as discussed further herein.

Searching for an alternative theoretical method we now describe the use, in the field of biocatalysis, of the COSMO-RS method. This is based upon the approximated continuum description of solvent, in particular on the "conductor-like screening model", which has proved to be a powerful technique for the prediction of thermo physical properties of mixtures. Therefore, the aim of the present paper is to evaluate the applicability of the COSMO-RS model to the prediction of activities for systems of interest to biocatalysis, especially in those cases where the UNIFAC method showed poor accuracy.

Theory

Excess Gibbs energies, and consequently activity coefficients, for condensed phases can be calculated by means of quantum chemistry calculations. First principles quantum mechanical methods are capable of correctly describing the electronic and geometrical structures of solute molecules, including polarization effects due to the presence of the solvent. However, in a quantum mechanical approach, only a few solvent molecules can be considered explicitly because of due the high computational cost. The use of Density Functional Theory (DFT), as implemented in many commercially available codes (Turbomole, Gaussian, DMol[3], and others), allows us to treat larger molecular systems than in ab initio methods such as Hartree−Fock or Møeller−Plesset perturbation theory (MP2). However, even within the framework of DFT, a simplified model has to be used, and the most successful theories so far are all based on the polarizable continuum model (PCM) proposed originally by Miertus et al. (*19*). In this model, the solute molecule is embedded in a dielectric continuum of permittivity $\in$. The dielectric continuum is polarized by the charge distribution of the solute, which results in a charge distribution on the cavity surface (also called solvent accessible surface, SAS). For a given charge distribution within the cavity, the surface screening charges are then calculated by solving the Poisson equation and hence used as a perturbation included in the Hamiltonian for electronic structure calculations of the solute. The new, perturbed density of the solute is used again for Poisson calculations, and the procedure is iterated until self-consistency is achieved. At costs comparable to gas-phase calculations, these methods are capable of giving a surprisingly good description of the properties and energetics of molecules in various solvents, including water. Despite the considerable success of PCM-based models, they are hard to justify theoretically, mainly because of the high electric fields on the molecular surfaces of even fairly polar solutes. These fields are so strong that the major part of the solvent polarizability, i.e., the reorientaton of static dipoles, no longer behaves linearly, as it does in the macroscopic limit, but is almost at saturation.

Starting from these considerations and with the aim of solving the inherent inconsistencies of the original method, Klamt and co-workers (*20*−*22*) have proposed a theory, called the conductor-like screening model for real solvent (COSMO-RS). This is based on screening in conductors, is noniterative, and allows for the calculation of accurate gradients without cavity shape constraints. The theory has been extended and enhanced for treating complex systems (*23*−*26*). The COSMO-RS theory takes the ideally screened molecules as starting points for the description of molecules in solution. The deviations from ideal screening, which occur in any real solvent, are described as pair wise misfit interactions of the ideal screening charges on parts of the molecules in contact in the fluid. Since COSMO-RS does not depend on experimental data or any parametrization for the solvent, it efficiently enables the calculation of the chemical potential of almost any solute in almost any solvent. Thus, it is capable of treating almost the entire equilibrium thermodynamics of fluid systems and can constitute a powerful alternative to group/fragment-based methods such as UNIFAC.

Basically, the COSMO-RS method assumes that the molecule is made up of small surface elements rather than functional groups. The charge density of each of these elements is evaluated by means of quantum mechanics calculations. An ensemble of interacting molecules is replaced by the corresponding ensemble of interacting surface pieces characterized by a screening charge density, namely, the contribution of all of the differences in polarity obtained when the molecules are not isolated. The characteristic distribution of charges obtained by considering all of these surface elements is the "sigma profile" of the molecule, which describes the amount of surface in the ensemble having a screening charge density between σ and $\sigma + d\sigma$. Starting from the sigma profile and applying a straightforward statistical mechanics treatment, one obtains directly the excess Gibbs energy as a function of composition.

Many are the advantages of the method. First, because each molecule is characterized by a sigma profile, the method can treat proximity effects, can distinguish among isomers, and can also take into account the result of a conformational analysis.

Furthermore, unlike UNIFAC, the method is not based on an analytical form of the excess Gibbs energy, and the model contains only a very few adjustable parameters at the atomistic level, such as the atomic dimensions. If the parameters are known for each element present, predictions can be made for any compounds or mixtures made up from them. There is no requirement for large databases of experimental data for characterizing the binary interactions between functional groups, as needed for UNIFAC. The method is presently applicable only to liquids, since the basic theory has been derived at liquid-like densities.

With COSMO-RS we use DFT calculations, which yield ground-state properties as reliably as Hartree−Fock methods with higher order correlation corrections but at much lower cost. DFT methods, even at the default level, have sufficiently good tails to reproduce reliably quantities such as dipole moments and polarizabilities, which are of crucial importance for any solvation calculation.

Compared with COSMO-RS, UNIFAC has a number of limitations, to set against its great calculation speed. The assumption that a given group makes the same contribution in different molecules implies that it is not influenced by the nature of the atoms that surrounds it. Similarly, positional isomers are necessarily considered identical (e.g., the different mono- and dioctanoyl derivatives of glycerol), and the effects of different conformations are also ignored. Moreover UNIFAC shows a significant difficulty in the treatment of mixtures of molecules with strong size and shape differences. COSMO-RS is an efficient method for calculating equilibrium properties for liquid mixtures that shows several advantages with respect to UNIFAC. Nonetheless, being based on approximations, it has some limitations such as the impossibility of describing structural changes induced by hydrogen bond formation. Fortunately these limitations are not crucial for the applications considered.

Calculations

In this work we consider three lipase-catalyzed esterification reactions, carried out in different solvents: (1) esterification of 1-dodecanoic acid with menthol in *n*-hexane, *n*-heptane, cyclohexane, 2,2,4-trimethylpentane, toluene, acetonitrile, and 2-methyl-2-butanol (*10*); (2) esterification of 1-dodecanoic acid and 1-dodecanol in *n*-hexane, *n*-heptane, cyclohexane, 2,2,4-trimethylpentane, and toluene (*6*); and (3) esterification of glycerol and *n*-octanoic acid in acetonitrile, benzene, and toluene and in the neat reaction mixture (without any solvent) (*12*). In the case of this last reaction, there are five possible products, which lead to 12 different reactions. Experimental concentrations at equilibrium were taken directly from the work of Tewari et al. (*6*, *10*, *12*).

Both UNIFAC and COSMO-RS offer predictions of activity coefficients (γ, defined such that $a_i = \gamma_i \cdot x_i$) in a mixture of any defined composition. For any esterification reaction, there will be a thermodynamic equilibrium constant (K_{th}) given by a ratio of activities:

$$K_{th} = \frac{a_{ester} \cdot a_{water}}{a_{alcohol} \cdot a_{acid}} \quad (1)$$

whose value is independent of the composition of the medium (including the nature of any solvent). It is necessary that the same standard state is used for these activities under all conditions; both UNIFAC and COSMO-RS satisfy this by using the pure (super cooled) liquid standard state.

K_{th} can be also defined as

$$K_{th} = K_x \cdot K_\gamma \quad (2)$$

where K_x is the mole fraction ratio as reactions approach equilibrium and K_γ is

$$K_\gamma = \frac{\gamma_{ester} \cdot \gamma_{water}}{\gamma_{alcohol} \cdot \gamma_{acid}} \quad (3)$$

Since all of the reactions considered here have the same number of molecules among the starting materials and products, the concentration-based equilibrium ratio K_c is dimensionless and equal to K_x.

The data of Tewari et al. offer a series of reported compositions as reactions approach equilibrium. The reported mole fractions can be used as an input to a UNIFAC or COSMO-RS calculation, leading to predicted activity coefficients for each reactant at the composition given.

Now, if the reactions measured were at equilibrium and the activity coefficient predictions are correct, all K_{th} values for a given reaction should be equal regardless of the solvent, composition, etc. (The K_c and K_γ. values are of course different.) Therefore, correct predictions of the activity coefficients of each reactant in each solvent lead simply to prediction of the solvent dependence of the equilibrium constant (K_c or K_x).

In the case of glycerol esterification, the analytical methods used by Tewari et al. (*12*) were unable to distinguish the positional isomers of monoacylglycerols. Those authors therefore used statistical arguments to assign isomer ratios. However, other experiments have shown that isomer ratios can be substantially different (e.g., Dudal and Lortie (*33*)). We have not attempted to adjust the reported concentrations, in part because the isomer ratio is probably medium-dependent, as suggested by the activity coefficients calculated by COSMO-RS. This should be taken into account in considering the reactions involving monoacylglycerols (3.1, 3.2, 3.6−3.10). It should also be noted that achievement of full equilibrium in the glycerol system probably involved nonenzymatic acyl migration reactions, because the lipase used was 1,3-specific.

For UNIFAC calculation the implementation available in the Aspen Properties version 11.1 program (*31*) with standard parametrization was used (original UNIFAC model).

The COSMO-RS statistical mechanics by Klamt (*27*) was implemented in the program COSMOtherm (*28*) from COSMOlogic (version C 2.1 rel 01/04) with the parametrization as implemented in the commercial Turbomole code (*29*) version 5.7 (*30*). The gas-phase reference energies for all the structures in the data sets were obtained from nonlocal DFT with VWN-BP (= B88-VWN-P86) gas-phase optimization applying the Ahlrichs-TZVP basis set (*30*). The resolution of identity (RI) method was used, although it proved to be irrelevant for the properties of interest in this investigation. All structures were then reoptimized in a continuum conductor (i.e., with Turbomole/COSMO and $f(\epsilon) = 1$, using a number of segments = 92). The values of the area and volume of the dielectric cavity in which the solute molecule is embedded can be extracted directly from the corresponding files. The COSMO-RS code reports in the output file the values of the activity coefficients under the experimental conditions.

A detailed conformational analysis has been carried out for those components for which different conformers show evident differences in the sigma profiles. Long chain aliphatic compounds (such as dodecanoic acid and dodecanol) do not show any appreciable effect on the sigma profile, since this is mainly influenced by the polar functional group. The same reasoning may easily be extended to all the solvents and to menthol, which has a limited conformational mobility. Therefore, the conformational analysis has been carried out for glycerol only, and the results have been used in the COSMO-RS predictions: four

Table 1. Equilibrium Constants for Menthol Reaction; Comparison between the COSMO-RS and UNIFAC Models for Prediction of Activities[a]

		COSMO-RS		UNIFAC	
	K_x	K_γ	K_{th}	K_γ	K_{th}
n-hexane	6.4	9.2	58.6	31.5	200.7
n-heptane	18.2	11.8	215.0	30.6	556.5
cyclohexane	25.9	12.3	319.3	33.1	856.7
2,2,4-trimethylpentane	16.9	11.5	194.2	30.4	515.2
toluene	10.6	14.9	157.2	24.3	257.1
acetonitrile	3.4	34.4	115.9	22.4	75.4
2-methyl-2-butanol	4.3	21.6	93.1	25.8	110.8
SSD			**87.7**		**285.2**
RAD % (eq 4)			**41**		**64**

[a] K_x is the composition equilibrium constant, K_γ is the activity coefficient ratio, and K_{th} is the thermodynamic equilibrium constant.

Table 2. Equilibrium Constants for Dodecanol Reaction; Comparison between the COSMO-RS and UNIFAC Models for Prediction of Activities[a]

		COSMO-RS		UNIFAC	
	K_x	K_γ	K_{th}	K_γ	K_{th}
n-hexane	42.6	3.0	127.7	34.5	1470.1
n-heptane	35.5	3.8	134.4	33.9	1202.9
cyclohexane	24.2	3.6	85.9	36.1	873.3
2,2,4-trimethylpentane	24.5	3.6	86.9	33.5	820.2
toluene	15.3	6.8	104.1	26.2	400.0
SSD			**22.6**		**406.1**
RAD % (eq 4)			**17**		**32**

[a] K_x is the composition equilibrium constant, K_γ is activity coefficient ratio, and K_{th} is thermodynamic equilibrium constant.

discrete conformers with equal probability have been used in the calculations.

All simulations were performed on Intel Xeon 3.03 MHz bi-processors with 1Gbyte RAM each connected to a storage area network of up to 2 Tbyte capacity with Linux Red Hat 7.2 or Windows XP operating systems. UNIFAC calculation time was negligible, and COSMO-RS calculation time was between 1 and 2 h for the Turbomole quantum mechanics and of the order of seconds for the statistical mechanics calculations on each mixture composition. It should be noted that quantum mechanics calculations are done just once for each molecule and the results are stored in a database for subsequent calculations on any mixture containing that compound.

Results and Discussion

Table 1 shows the results of the calculations for the esterification of menthol. The concentration equilibrium constant is reported in column 2 and clearly varies considerably between the different media. This must reflect differing solvation of the reactants, which the treatment in terms of thermodynamic activities should allow for. Columns 3 and 4 report the values of the activities and thermodynamic equilibrium constants obtained by the COSMO-RS model and columns 5 and 6 the same quantities calculated by UNIFAC. We can note that the average relative percent deviation and the standard deviations (SSD) of the K_{th} values are substantially lower for the COSMO-RS method than for UNIFAC. The relative average deviations are calculated with respect to the mean value, as follows:

$$\mathrm{RAD\%} = 100\,\frac{1}{n}\,\frac{\sum|x_i - \bar{x}|}{\bar{x}} \qquad (4)$$

The deviations give an indication of the departure from the mean, constant values; therefore, they are good measures of the quality of the model predictions, since K_{th} should be constant on changing solvents, as noted above.

Table 1 shows a 3-fold difference in K_{th} for the esterification of menthol in hexane and heptane, which is quite odd behavior considering the similarity of these two solvents. This happens using either COSMO-RS or UNIFAC: in both cases K_γ values are almost identical for the two solvents, while the significant difference is completely ascribable to the experimental K_x. Treating the hexane value as an outlier would significantly reduce the SSD and RAD values for COSMO-RS, but not for UNIFAC.

Table 2 shows similar results obtained for the esterification of dodecanol: the statistical indexes show clearly the significant difference in the performance of the models.

The higher accuracy of the COSMO-RS method is also evident for the esterification of glycerol (Table 3), in which all of the reactions are considered separately. The better performance of the COSMO-RS model is particularly evident for reactions 3.1, 3.2, 3.3, 3.4, 3.5, 3.9, and 3.10. Figure 1 shows a graphical comparison of K_{th} values calculated for each reaction. K_{th} values calculated by COSMO-RS (white bars) and UNIFAC (black bars) are reported, on a logarithmic scale, for all of the subreactions (see Table 3). The graphical representation makes it visually evident that for most of the 12 reactions K_{th} in the four different solvents is much more constant (similar height of the bars) when calculated by COSMO-RS.

To help the reader evaluate the differences between the two models, Figure 2 reports the details of four selected reactions (3.1, 3.3, 3.9, and 3.10). Comparison for all reactions is reported in Supporting Information. In the reactions in Figure 2, the COSMO-RS predicted values of K_{th} vary by only about 2-fold. It is generally considered hard to measure equilibrium constants more accurately than this, corresponding to a change in $\Delta G°$ value of less than 2 kJ or 0.5 kcal/mol. This can be seen in the differences between replicate measurements for the same reaction in the data of Tewari et al. Hence in these cases COSMO-RS predictions for medium effects would be as accurate as experiment, certainly in hands less skilled than the NIST group. In the worse cases the variation is about 9-fold, corresponding to about 5.5 kJ/mol. This would still be a very useful prediction that would point to the best solvents for experimental test.

The good predictions made using the COSMO-RS model are very promising. COSMO-RS uses the easier conductor solution to approximate the complicated solution at the limit for dielectrics. This approximation, which considers the solvent as an ideal conductor, makes COSMO-RS calculations exact for the infinite limit and keeps a very high accuracy in the case of strong dielectrics, such as water. Even in the case of the lowest dielectric for an organic solvent, the accuracy is still good.

The activity coefficients calculated by the two methods in some cases differ by more than 10-fold, up to more than 100-fold (see Supporting Information for tables of activity coefficients). Species that show these high differences are mostly reactants and products of the esterification of glycerol and octanoic acid. In particular, activity coefficients of glycerol in the hydrophobic solvents (toluene and benzene) differ by more than 10-fold, whereas an even higher variation is evident in the case of the correspondent esters in acetonitrile. Probably this is due to the important role that proximity and conformational effects play in the solvation process. Moreover COSMO-RS calculates different activity coefficients for the different dioctanoyl esters, which are indeed different chemical species, whereas UNIFAC cannot distinguish between them. However these differences cannot be discussed in detail without a comparison with experimental data for individual activity coefficients, which will be the subject of future work.

Table 3. Equilibrium Constants for Esterification of Glycerol and *n*-octanoic Acid. Comparison between the COSMO-RS and UNIFAC Models for the Prediction of Activities[a]

		neat reaction	aceto-nitrile	tol-uene	ben-zene	SSD	RAD% (eq 4)
3.1: glycerol + *n*-octanoic acid = 1-monooctanoyl glycerol + H_2O							
	K_x	1.6	5.8	0.8	0.6		
COSMO-RS	K_γ	1.5	0.3	4.2	2.7		
	K_{th}	2.4	1.8	3.3	1.6	0.8	25
UNIFAC	K_γ	0.4	0.6	0.7	0.6		
	K_{th}	0.6	3.3	0.5	0.4	1.4	85
3.2: glycerol + *n*-octanoic acid = 2-monooctanoyl glycerol + H_2O							
	K_x	1.6	5.8	0.8	0.6		
COSMO-RS	K_γ	1.5	0.3	4.2	2.7		
	K_{th}	2.4	1.8	3.3	1.6	0.8	25
UNIFAC	K_γ	0.4	0.6	0.7	0.6		
	K_{th}	0.6	3.3	0.5	0.4	1.4	85
3.3: glycerol + 2 *n*-octanoic acid = 1,2-dioctanoyl glycerol + 2 H_2O							
	K_x	0.9	13.9	1.2	1.1		
COSMO-RS	K_γ	5.3	0.5	1.6	0.7		
	K_{th}	4.9	7.1	1.8	0.8	2.9	64
UNIFAC	K_γ	1.3	2.7	4.0	3.4		
	K_{th}	1.2	37.1	4.7	3.8	17.0	109
3.4: glycerol + 2 *n*-octanoic acid = 1,3-dioctanoyl glycerol + 2 H_2O							
	K_x	1.3	20.4	1.7	1.6		
COSMO-RS	K_γ	2.0	0.4	1.5	0.7		
	K_{th}	2.7	9.0	2.6	1.1	3.5	67
UNIFAC	K_γ	1.3	2.7	4.0	3.4		
	K_{th}	1.8	54.5	7.0	5.6	25.0	109
3.5: glycerol + 3 *n*-octanoic acid = 1,2,3-trioctanoyl glycerol + 3 H_2O							
	K_x	0.5	16.0	2.4	1.8		
COSMO-RS	K_γ	89.9	2.7	35.1	9.7		
	K_{th}	47.6	42.5	82.6	17.8	26.7	37
UNIFAC	K_γ	14.5	42.3	76.8	59.9		
	K_{th}	7.7	676.3	180.8	109.9	297.1	89
3.6: 1-monooctanoyl glycerol + *n*-octanoic acid = 1,2-dioctanoyl glycerol + H_2O							
	K_x	0.6	2.4	1.5	1.8		
COSMO-RS	K_γ	3.6	1.6	0.4	0.3		
	K_{th}	2.0	3.9	0.6	0.5	1.6	70
UNIFAC	K_γ	3.3	4.7	5.8	5.3		
	K_{th}	1.9	11.4	8.7	9.8	4.2	38
COSMO-RS average						**6.8**	**49**
UNIFAC average						**55.2**	**65**

		neat reaction	aceto-nitrile	tol-uene	ben-zene	SSD	RAD% (eq 4)
3.7: 1-monooctanoyl glycerol + *n*-octanoic acid = 1,3-dioctanoyl glycerol + H_2O							
	K_x	0.8	3.5	2.2	2.7		
COSMO-RS	K_γ	1.4	1.4	0.4	0.2		
	K_{th}	1.1	5.0	0.8	0.7	2.1	82
UNIFAC	K_γ	3.3	4.7	5.8	5.3		
	K_{th}	2.7	16.7	12.8	14.4	6.2	38
3.8: 2-monooctanoyl glycerol + *n*-octanoic acid = 1,2-dioctanoyl glycerol + H_2O							
	K_x	0.6	2.4	1.5	1.8		
COSMO-RS	K_γ	3.6	1.6	0.4	0.3		
	K_{th}	2.0	3.9	0.6	0.5	1.6	70
UNIFAC	K_γ	3.3	4.7	5.8	5.3		
	K_{th}	1.9	11.4	8.7	9.8	4.2	38
3.9: 1-monooctanoyl glycerol + 2 *n*-octanoic acid = 1,2,3-trioctanoyl glycerol + 2 H_2O							
	K_x	0.3	2.8	3.0	3.0		
COSMO-RS	K_γ	61.0	8.5	8.4	3.6		
	K_{th}	19.8	23.6	25.0	10.9	6.3	22
UNIFAC	K_γ	36.5	74.8	111.3	94.3		
	K_{th}	11.9	207.6	332.6	285.5	141.4	48
3.10: 2-monooctanoyl glycerol + 2 *n*-octanoic acid = 1,2,3-trioctanoyl glycerol + 2 H_2O							
	K_x	0.3	2.8	3.0	3.0		
COSMO-RS	K_γ	61.0	8.5	8.4	3.6		
	K_{th}	19.8	23.6	25.0	10.9	6.3	22
UNIFAC	K_γ	36.5	74.8	111.3	94.3		
	K_{th}	11.9	207.6	332.6	285.5	141.4	48
3.11: 1,2-dioctanoyl glycerol + *n*-octanoic acid = 1,2,3-trioctanoyl glycerol + H_2O							
	K_x	0.6	1.2	2.0	1.6		
COSMO-RS	K_γ	16.9	5.2	22.6	13.9		
	K_{th}	9.8	6.0	44.8	22.7	17.5	62
UNIFAC	K_γ	11.1	15.8	19.3	17.8		
	K_{th}	6.4	18.2	38.3	29.1	13.8	46
3.12: 1,3-dioctanoyl glycerol + *n*-octanoic acid = 1,2,3-trioctanoyl glycerol + H_2O							
	K_x	0.4	0.8	1.3	1.1		
COSMO-RS	K_γ	44.8	6.0	23.6	14.8		
	K_{th}	17.6	4.7	31.8	16.5	11.1	40
UNIFAC	K_γ	11.1	15.8	19.3	17.8		
	K_{th}	4.4	12.4	26.0	19.8	9.3	46

[a] K_x is the composition equilibrium constant, K_γ is the activity coefficient ratio, and K_{th} is the thermodynamic equilibrium constant. Note that several of these reactions may be expressed as the sum of two or three others. Hence K_x, K_γ, and K_{th} values are equal to the product of those for the other reactions.

Conclusions

The calculations reported in this study show how the COSMO-RS model gives good results in predicting activities and hence equilibrium positions for the esterification reactions investigated. Recently (*32*), similar results indicating the superiority of the COSMO-RS model with respect to classical Excess Gibbs energy models have been reported in the literature, for a simpler system. Consequently, by means of the proposed method, it is possible to select the appropriate solvent for a given enzymatic reaction on the basis of ab initio calculations rather than experimental trials.

The predictions reported are based on DFT gas-phase calculations and the subsequent application of COSMO-RS statistical mechanics for the estimation of the Gibbs energy of the condensed mixtures. The COSMO-RS method proved to be very accurate and superior to UNIFAC. This is probably because COSMO-RS allows the treatment of long chain molecules, conformers, and isomers, while taking into account proximity effects of the functional groups in the molecules. The COSMO-RS method is computationally more intensive than UNIFAC because gas-phase DFT calculations are required for the isolated molecules.

We conclude that COSMO-RS should be a useful tool in biocatalysis for making in silico predictions for the selection of the optimal solvent for a given reaction, with two immediate applications. The first one is the prediction of equilibrium position: if experimental data are available for the equilibrium position in one solvent, one can use activity coefficients calculated from COSMO-RS to estimate K_{th}. This same K_{th} value should apply for the same reaction in any other medium of interest, and hence further iterative COSMO-RS calculations can be used to find the equilibrium position corresponding to any given starting mixture. An efficient algorithm for this is under development. The second application is the prediction or rationalization of medium effects on enzyme kinetics: as noted in the Introduction (*16*−*18*), these are often dominated by effects on reactant solvation. When this is the case and kinetic parameters (K_m or V_m/K_m) have been measured for one solvent, COSMO-RS calculations can give predictions for any other medium of interest.

The method can also aid the rational design of reaction systems in the case of water−organic solvent mixtures, where pK_a values can be affected. This appears particularly relevant since the protonation state of reactants and products can strongly

F

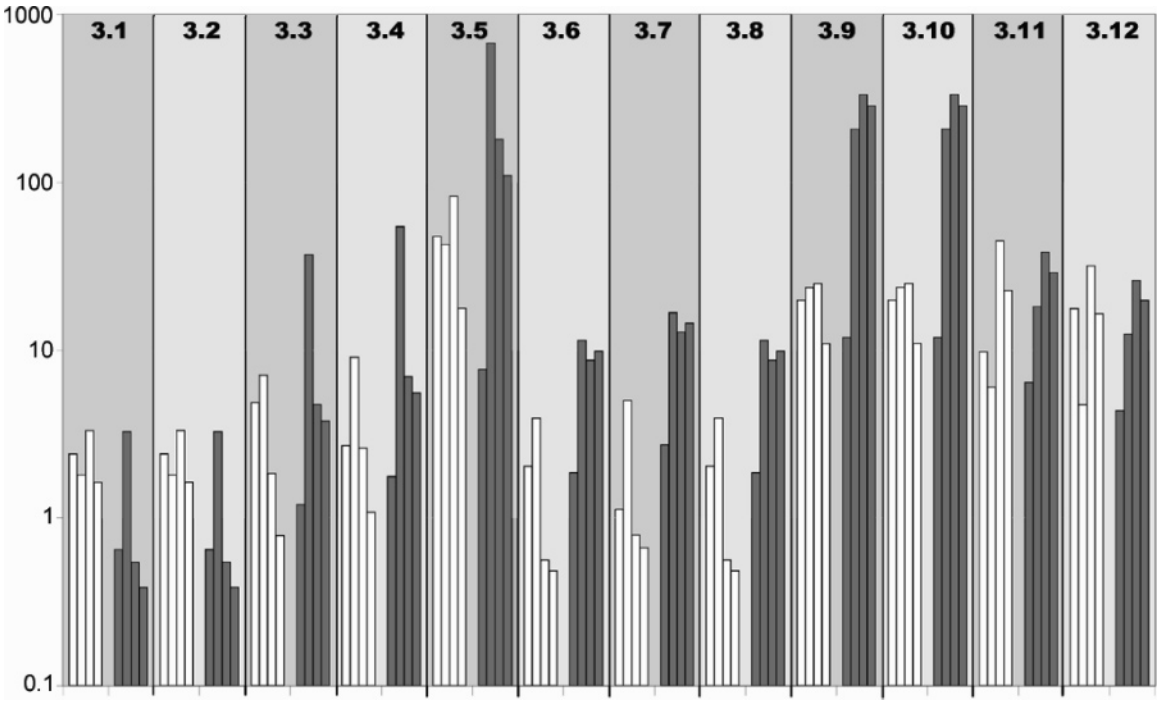

Figure 1. Comparison of K_{th} calculated by COSMO-RS and UNIFAC for the entire set of reactions in the esterification of glycerol and *n*-octanoic acid. Each group of four bars shows predictions, from left to right, for the neat reaction mixture, acetonitrile, toluene, and benzene. K_{th} is reported on a logarithmic scale, and the 12 reactions are indicated at the top, numbered as in Table 3. White bars represent the K_{th} calculated by COSMO-RS, and the UNIFAC results are reported in black. Accurate predictions are indicated by similar heights for all four bars in a group.

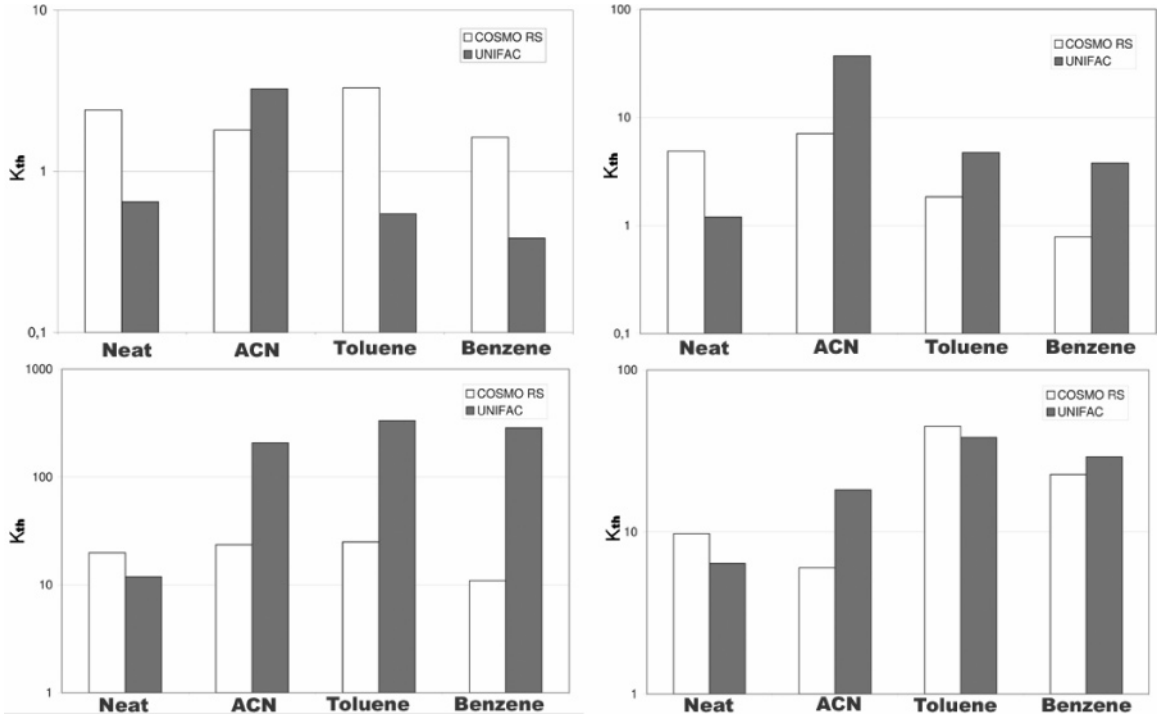

Figure 2. Detailed comparison of K_{th} calculated by COSMO-RS and UNIFAC for selected reactions in the esterification of glycerol and *n*-octanoic acid. Upper left: (3.1) glycerol + *n*-octanoic acid = 1-monooctanoyl glycerol + H_2O. Upper right: (3.3) glycerol + 2 *n*-octanoic acid = 1,2-dioctanoyl glycerol + 2 H_2O. Lower left: (3.9) 1-monooctanoyl glycerol + 2 *n*-octanoic acid = 1,2,3-trioctanoyl glycerol + 2 H_2O. Lower right: (3.11) 1,2-dioctanoyl glycerol + *n*-octanoic acid = 1,2,3-trioctanoyl glycerol + H_2O.

affect the reaction system. Moreover, enzyme catalytic activity is strongly influenced by pH, and often the pH range for enzyme activity is very narrow.

In the future, the method will be tested against diverse biocatalyzed reactions to assess its performance with very different chemical systems. The final goal is to make a COSMO-RS-based theoretical method for solvent selection with general applicability. In principle it is applicable also to biocatalysis in ionic liquids.

Acknowledgment

The authors are grateful to Andreas Klamt and Frank Eckert for help in the calculations and stimulating discussions. MIUR (Rome, Italy) and Università degli Studi di Trieste are gratefully

acknowledged for financial support to L.G., and P.B. is grateful to COST Action D25 (Working group on "Solving the Problems Enzymes Encounter in Organic Solvents") for having financed a STSM.

Supporting Information Available: Activity coefficients for all the species of the reactions. Figures summarizing the results of the complete subset of reactions of esterification of glycerol and octanoic acid. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Klibanov, A. M. Improving enzymes by using them in organic solvents. *Nature* **2001**, *409*, 241−246.

(2) Thomas, S. M.; DiCosimo, R.; Nagarajan, V. Biocatalysis: applications and potentials for the chemical industry. *Trends Biotechnol.* **2002**, *20*, 238−242.

(3) Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. Industrial biocatalysis today and tomorrow. *Nature* **2001**, *409*, 258−268.

(4) Zaks, A. Industrial biocatalysis. *Curr. Opin. Chem. Biol.* **2001**, *5*, 130−136.

(5) Claon, P. A.; Akoh, C. C. Enzymatic-synthesis of geraniol and citronellol esters by direct esterification in *n*-hexane. *Biotechnol. Lett.* **1993**, *15*, 1211−1216.

(6) Tewari, Y. B. Thermodynamics of the lipase-catalyzed esterification of 1-dodecanoic acid and 1-dodecanol in organic solvents. *J. Chem. Eng. Data* **1998**, *43*, 750−755.

(7) Molinari, F.; Villa, R.; Aragozzini, F. Production of geranyl acetate and other acetates by direct esterification catalyzed by mycelium of *Rhizopus delemar* in organic solvent. *Biotechnol. Lett.* **1998**, *20*, 41−44.

(8) Halling, P. J. Solvent selection for biocatalysis in mainly organic-systems−predictions of effects on equilibrium position. *Biotechnol. Bioeng.* **1990**, *35*, 691−701.

(9) Valivety, R. H.; Johnston, G. A.; Sucking, C. J.; Halling, P. J. Solvent effects on biocatalysis in organic-systems−equilibrium position and rates of lipase catalyzed esterification. *Biotechnol. Bioeng.* **1991**, *38*, 1137−1143.

(10) Tewari, Y. B. Thermodynamics of the lipase-catalyzed transesterification of (−)-menthol and dodecyl dodecanoate in organic solvents. *J. Mol. Catal. B* **2000**, *9*, 83−90.

(11) Tewari, Y. B.; Schantz, M. M.; Vanderah, D. J. Thermodynamics of the lipase-catalyzed esterification of 1-dodecanoic acid with (−)-menthol in organic solvents. *J. Chem. Eng. Data* **1999**, *44*, 641−647.

(12) Tewari, Y. B.; Bunk, D. M. Thermodynamics of the lipase-catalyzed esterification of glycerol and *n*-octanoic acid in organic solvents and in the neat reaction mixture. *J. Mol. Catal. B* **2001**, *15*, 135−145.

(13) Poling, B. E.; Prausnitz, J. M.; O'Connell, J. M. *The Properties of Gases and Liquids*, 5th ed.; McGraw-Hill Professional: New York, 2001.

(14) (a) Janssen, A. E. M.; Van Der Padt, A.; Van't Riet, K. Prediction of the reaction equilibrium of biocatalytic reactions in aqueous−organic 2-phase systems. *Biocatal. Biotransform.* **1995**, *16*, 223−240. (b) Janssen, A. E. M.; Van Der Padt, A.; Van't Riet, K. Solvent effects on lipase-catalyzed esterification of glycerol and fatty-acids. *Biotechnol. Bioeng.* **1993**, *42*, 953−962. (c) Janssen, A. E. M.; Van Der Padt, A.; Van Sonsbeek, H. M.; Van't Riet, K. The effect of organic-solvents on the equilibrium position of enzymatic acylglycerol synthesis. *Biotechnol. Bioeng.* **1993**, *41*, 95−103.

(15) Wescott, C. R.; Klibanov, A. M. Predicting the solvent dependence of enzymatic substrate-specificity using semiempirical thermodynamic calculations. *J. Am. Chem. Soc.* **1993**, *115*, 10362−10363.

(16) (a) Reimann, A.; Robb, D. A.; Halling, P. J. Solvation of cbz-amino acid nitrophenyl esters in organic media and the kinetics of their transesterification by subtilisin. *Biotechnol. Bioeng.* **1994**, *43*, 1081−1086. (b) Van Tol, J. B. A.; Stevens, R. M. M.; Veldhuizen, W. J.; Jongejan, J. A.; Duine, J. A. Do organic-solvents affect the catalytic properties of lipase−intrinsic kinetic-parameters of lipases in ester hydrolysis and formation in various organic-solvents. *Biotechnol. Bioeng.* **1995**, *47*, 71−81.

(17) Ebert, C.; Gardossi, L.; Linda, P.; Vesnaver, R.; Bosco, M. Influence of organic solvents on enzyme chemoselectivity and their role in enzyme−substrate interaction. *Tetrahedron* **1996**, *13*, 4867−4876.

(18) (a) Ke, T.; Wescott, C. R.; Klibanov, A. M. Prediction of the solvent dependence of enzymatic prochiral selectivity by means of structure-based thermodynamic calculations. *J. Am. Chem. Soc.* **1996**, *118*, 3366−3374. (b) Ke, T.; Klibanov, A. M. Insights into the solvent dependence of chymotryptic prochiral selectivity. *J. Am. Chem. Soc.* **1998**, *120*, 4259−4263.

(19) Miertus, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117−129.

(20) 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.

(21) 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.

(22) Klamt, A. COSMO and COSMO-RS. In *Encyclopedia of Computational Chemistry*; Wiley: New York, 1998.

(23) Eckert, F.; Klamt, A. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J.* **2002**, *48*, 369−385.

(24) Putnam, R.; Taylor, R.; Klamt, A.; Eckert, F.; Schiller, M. Prediction of infinite dilution activity coefficients using COSMO-RS. *Ind. Eng. Chem. Res.* **2003**, *42*, 3635−3641.

(25) Fermeglia, M.; Pricl, S. Prediction of phase equilibria for binary mixtures by molecular modeling. *AIChE J.* **2001**, *47*, 2371−2382.

(26) Lin, S. T.; Sandler, S. I. A priori phase equilibrium prediction from a segment contribution solvation model. *Ind. Eng. Chem. Res.* **2002**, *41*, 899−913.

(27) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. Refinement and parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074−5085.

(28) COSMOtherm program. COSMOlogic GmbH & Co. KG: Leverkusen, Germany, 1999−2004 (www.cosmologic.com).

(29) Ahlrichs, R.; Baer, M.; Haeser, M.; Horn, H.; Koelmel, C. Electronic structure calculations on workstation computers: The program system TURBOMOLE. *Chem. Phys. Lett.* **1989**, *162*, 165−169.

(30) Grimme, S.; Furche, F.; Ahlrichs, R. An improved method for density functional calculations of the frequency-dependent optical rotation. *Chem. Phys. Lett.* **2002**, *361*, 321−328 (www.turbomole-.com).

(31) Aspen Technology, Inc., Cambridge, MA, 1981−2001 (www.aspentech.com).

(32) Grob, S.; Hasse, H., Thermodynamics of phase and chemical equilibrium in a strongly nonideal esterification system, *J. Chem. Eng. Data* **2005**, *50*, 92−101.

(33) Dudal, Y.; Lortie, R. Influence of water activity on the synthesis of triolein catalyzed by immobilized *Mucor miehei* lipase. *Biotechnol. Bioeng.* **1995**, *45*, 129−134.

Accepted for publication May 31, 2006.

BP060081O