



## Computer-Aided Solvent Screening for Biocatalysis

**Abildskov, Jens; Leeuwen, M.B. van; Boeriu, C.G.; Broek, L.A.M. van den**

*Published in:*  
Journal of Molecular Catalysis B: Enzymatic

*Link to article, DOI:*  
[10.1016/j.molcatb.2012.09.012](https://doi.org/10.1016/j.molcatb.2012.09.012)

*Publication date:*  
2013

[Link back to DTU Orbit](#)

*Citation (APA):*  
Abildskov, J., Leeuwen, M. B. V., Boeriu, C. G., & Broek, L. A. M. V. D. (2013). Computer-Aided Solvent Screening for Biocatalysis. *Journal of Molecular Catalysis B: Enzymatic*, 85-86, 200-213. DOI: 10.1016/j.molcatb.2012.09.012

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Computer-Aided Solvent Screening for Biocatalysis

J. Abildskov<sup>a</sup>, M.B. van Leeuwen<sup>b</sup>, C.G. Boeriu<sup>b</sup>, L.A.M. van den Broek<sup>b</sup>

<sup>a</sup>CAPEC, Department of Chemical and Biochemical Engineering, Building 229, Soeltofts Plads, DTU, 2800 Kgs. Lyngby, Denmark

<sup>b</sup>Wageningen UR Food & Biobased Research, P.O. Box 17, 6700AA Wageningen, The Netherlands

## ABSTRACT

A computer-aided solvent screening methodology is described and tested for biocatalytic systems composed of enzyme, essential water and substrates/products dissolved in a solvent medium, without cells. The methodology is computationally simple, using group contribution methods for calculating constrained properties related to chemical reaction equilibrium, substrate and product solubility, water solubility, boiling points, toxicity and others. Two examples are provided, covering the screening of solvents for lipase-catalyzed transesterification of octanol and inulin with vinyl laurate. Esterification of acrylic acid with octanol is also addressed. Solvents are screened and candidates identified, confirming existing experimental results. Although the examples involve lipases, the method is quite general, so there seems to be no preclusion against application to other biocatalysts.

**Keywords:** Solvents, Screening, Lipases, Activity Coefficients, Inulin.

## 1. Introduction

Modern chemical synthesis involves frequently also biocatalytic steps. Since the pioneering work of Klivanov and co-workers on enzyme-catalyzed reactions in organic solvents [1], the application of enzymes in organic solvents has increased tremendously [2,3,4,5,6]. Enzyme function, and the equilibrium states of a reaction, can vary substantially when the solvent composition changes [7]. Thus, to improve enzyme function and reaction yield, solvent engineering may be a very useful tool. It is relatively easy to manipulate the solvent properties for enzymatic reactions, as hundreds of solvents and many more mixtures can be considered for biocatalysis. The freedom to select among several alternatives, however, also places demands on the selection of an ‘advantageous’ solvent. Outside research laboratories, introduction of a solvent into a process, will not only affect the chemical reaction. When processing is considered, solvent recycling will inevitably be required. Therefore expenses and ease of recycling must also be taken into account in solvent assessment. Since many future biobased processes will be solvent intensive, a systematic method to identify solvents will be desirable. Rather than mix the substrate, solvent plus enzyme and then test if it is suitable, the objective would *ideally* be to fix a performance target (defined through solvent properties and process constraints) and identify solvents that match the target. Then, less trials are required, and resources may be spent more strategically on designing an economically feasible process.

## 2. Background

In recent decades great advances have been made in modeling and computation of the properties of complex chemical systems. Solvent effects on biocatalytic reactions have been characterized frequently in literature, and there is considerable appreciation of how thermodynamic activities affect the reaction equilibrium [8,9,10,11,12,13]. In spite of this, computer-aided solvent selection methods seem not to have penetrated the biocatalysis community yet. In fact, biocatalytic reaction media are mostly selected based on past experience and ‘mix-and-test’ experiments [14,15,16]. Progress in modeling enzymatic solutions will undoubtedly continue, but achieving reliable prediction without measured data is unlikely. In particular, activities (or rates) of biocatalytic reactions are not at this stage predictable using purely computational approaches, although some fundamental results have been reported [17]. The influence of unknown side reactions and effects of impurities cannot be predicted from computations either. Therefore computational methods will for long need to be coupled with at least a limited set of measurements for validation. The question is how to use most conveniently the benefits of computational approaches, such that it can realistically be coupled with practice?

Screening wide ranges of solvent candidates is possible if simple methods exist for checking solvent properties against performance criteria. A method for identifying molecules (or mixtures) with desirable properties is Computer-Aided Molecular Design (CAMD). Previously CAMD has been used to design a range of chemical products, e.g., polymers [18,19,20], refrigerants [21,22], microparticles with desired loadings [23], extraction solvents [24] and reaction solvents [25] for synthetic chemical reactions.

Many biocatalytic systems - from growing cells to purified enzymes - exist [26,27,28]. Design criteria, rationalizing the solvent effects on biocatalytic reactions, depend upon the system [9], i.e., if the biocatalyst is an enzyme with an interphase, dissolved in a nearly anhydrous solvent medium, part of a cell, solvent free or other.

Laane et al. [9] gave criteria for some cases, with emphasis on a (bacterial) cell-containing aqueous phase, plus an immiscible and biocompatible organic phase that partitions (possibly) toxic substrates to the cells based on metabolic demands and the thermodynamic equilibrium of the system. Laane’s work was based on an early study of biocatalysis inhibition caused by solvents [29]. That work concluded that high biocatalytic rates result when the solvent polarity is low and its molecular weight exceeds 150 g/mol. Laane and co-workers [9] claimed a stronger relationship between bioactivity and the logarithm of the octanol/water partition coefficient ( $\log P$ ) of the solvent. In short, biocatalysis is low when  $\log P < 2$ , moderate/unpredictable when  $2 < \log P < 4$  and high when  $\log P > 4$ . The underlying observations showed sigmoidal plots of bioactivity versus  $\log P$ , with a transition region containing an inflection point (critical  $\log P$  or  $\log P_{\text{crit}}$ ) above which organism growth in the aqueous phase was not adversely affected by the solvent. This correlation was rationalized in terms of the solvents ability to *distort* the essential water layer that stabilizes the biocatalyst.

A comprehensive solvent screening approach for living cells-containing systems was later developed and applied by Daugulis and co-workers in a series of articles [30]. Their requirements to a suitable solvent are a favorable *distribution coefficient* for the product (ratio of product concentration in the solvent to product concentration in the

aqueous culture medium) at equilibrium. Also high *selectivity* is required, so that the solvent preferentially removes the product over water. Low aqueous solubility is required to minimize the solvent *loss*. Chemical and thermal *stability* is required, since over the course of long or continuous fermentations the solvent may be recycled many times. In order to facilitate product recovery constraints may also be placed upon density, viscosity and boiling point. Since the solvent will be in direct contact with the biocatalyst, *non-biodegradability* is also considered, so that the organism does not use the solvent as a substrate. The solvent must also be completely *biocompatible* so that bioactivity is not compromised. In addition, the choice may be guided by demanding that the solvent must be non-hazardous, non-toxic, in-expensive, not form emulsions and be available in bulk quantity. Treatment of toxicity on a quantitative basis seems to be the most troublesome step. Their analysis seems to suggest that toxicity increases when  $\log P$  is high and when water solubility is high. The approach is centered about an extractant screening program (ESP) database. To screen a large number of extractants, physicochemical properties, availability, price, and any existing toxicity data (such as  $LD_{50}$ ) are looked up in the ESP database. Quantities, such as the distribution coefficient, selectivity and aqueous solubility, are calculated from activity coefficient models. Biocompatibility needs to be determined experimentally, since  $\log P_{crit}$  varies among organisms, and its prediction for mixtures is more complex than for pure solvents.

Wang and Achenie [31] also considered extractive fermentations: While their criteria were closely related to those of Daugulis and co-workers, their method relied more on mathematical programming and optimization. They constrained biocompatibility (in terms of lethal concentration,  $LC_{50}$ , for the fresh water fish fathead minnow), separation factors, selectivity for product, solvent loss (in terms of activity coefficients) and ease of separation of solvent from product (in terms of boiling points). A CAMD approach can identify solvents from a far larger space of candidates than would be in a typical database. However, solvents generated with a CAMD method may not be commercially available, or may not be easily synthesized. What is needed, however, is not only the *best* candidate, but a *range* of promising candidates. Therefore, optimization methods must be extended to finding more than one solution [32]. This is not prohibitive, but must be taken into account when using optimization methods.

Extraction based (biphasic) systems seem most suited for reactions with hydrophobic products [33,34,35]. Laane et al. [9] generalized the validity of the correlations between free- and immobilized cell activities and  $\log P$  to enzymatic reactions in (nearly) anhydrous organic media, based on data of Zaks and Klibanov [36]. In what follows we will develop and describe a CAMD methodology for solvent screening for such a system (enzyme, water and substrates/products dissolved in a solvent medium, without cells). Some constraints resemble those of our predecessors, but we will also modify and develop other constraints relevant to the problem.

### 3. Materials and Methods

#### 3.1. CAMD methodology

A CAMD framework for enzymatic reactions is shown in Fig. 1. If we liken the search for a reaction medium to the search for a needle-in-a-haystack, first we must get rid of a

lot of hay. That is, eliminate non-promising solvents, and retain promising ones for further consideration. When limited candidates are left, more intensive and time-demanding steps (such as a set of strategically selected measurements, to validate the algorithm, and perhaps even molecular simulations) can follow. Ultimately scale-up studies on the most promising candidates can follow.

<Fig. 1>

### 3.1.1 CAMD Problem Formulation

Often the physical and chemical properties of a solvent can be resolved into terms related to its chemical groups. CAMD methodologies can be built to "exploit" such relationships. Here we will generate solvent structures by generating combinations of chemical groups. When a solvent is represented by its chemical groups, its properties can be predicted from group contribution methods resolving molecular properties into contributions of groups. Then, by articulating demands on the solvent as mathematical constraints on the property values, one can test if the solvent satisfies the constraints. The essential steps are:

- Generate structures (or combinations of groups),
- Predict properties (using group contribution methods), and
- Test if constraints on properties are satisfied

This can be termed a 'generate-and-test' method [37]. Group contribution methods do not quantitatively represent the properties of all solvents. However, the methods typically capture the gross features of property variation by its molecular structure. This is sufficient, so long as CAMD methods are not used for final decisions, but to limit a search space. Here we use constraints on solvent performance related to

- i. Reaction equilibrium conversion (in solvent)
- ii. Substrate solubility (in solvent)
- iii. Water/solvent mutual solubility
- iv. Solvent toxicity and possibly safety
- v. Ease of separation (after reaction)
- vi. Structural feasibility (of group sets)

considering a *disadvantageous* solvent to be one that,

- allows a low reaction conversion at equilibrium,
- dissolves insufficient amounts of substrates,
- distorts the essential water layer at the enzyme surface,
- is very toxic and/or flammable, or
- is difficult to separate/recycle subsequently

The theoretical concepts used in constraints formulation are outlined below.

### 3.1.1.1 Reaction Equilibrium

For a single chemical reaction (in a single phase) with stoichiometric coefficients,  $\mathbf{v}$ ,

$$|v_A|A + |v_B|B + \dots \rightleftharpoons |v_M|M + |v_N|N + \dots \quad (1)$$

a reaction coordinate at equilibrium,  $\epsilon$ , is defined [38] by

$$d\epsilon = \frac{dN_A}{v_A} = \frac{dN_B}{v_B} = \dots = \frac{dN_M}{v_M} = \frac{dN_N}{v_N} = \dots \quad (2)$$

Here,  $\mathbf{N}$  are the mole numbers. The reaction coordinate at equilibrium will depend upon the reaction medium. It is given implicitly by a non-linear equation,

$$f = \ln K - \sum_j v_j \ln(x_j \gamma_j) = 0 \quad (3)$$

In equation (3)  $\mathbf{x}$  are the mole fractions (expressible in terms of  $\epsilon$ ) and  $\boldsymbol{\gamma}$  are Lewis/Randall normalized activity coefficients (depending upon  $\mathbf{x}$  and temperature). The equilibrium constant,  $K$ , can be expressed,

$$-RT \ln K = \Delta G_f^\circ = \sum_{j=1} v_j G_{fj}^\circ \quad (4)$$

The sum on the right hand side involves Gibbs free energies of formation for each species in its standard state.  $K$  may also be resolved into

$$K = \prod_j (x_j \gamma_j)^{v_j} = \prod_j x_j^{v_j} \prod_j \gamma_j^{v_j} \Rightarrow K_{eq} = \prod_j x_j^{v_j} = K \prod_j \gamma_j^{-v_j} \quad (5)$$

In other treatments  $K_{eq}$  is referred to as the equilibrium constant. The application of this classical thermodynamic approach to biocatalytic reactions has been discussed by several investigators [33,34,39]. Calculations have been described by Shen et al. [40] and later by Stamatis et al. [39]. The most rigorous application is to solve  $f = 0$  (eq. (3)) with respect to  $\epsilon$ . Here we will transform the problem, in order to strengthen its applicability within CAMD. Firstly, in eq. (3) all solvent effects are on the summation, whereas  $K$  is solvent independent. Thus, the error made when employing an imprecise value of  $K$  is similar in all solvents considered. Therefore *ranking* of a set of solvents, in terms of their ability to promote a given reaction, is possible even in the absence of  $K$ . Secondly, solving eq. (3) with respect to  $\epsilon$  involves iterations. In CAMD we need to consider (perhaps) tens of thousands of alternative reaction media. Solving non-linear equations for so many alternatives is undesirable. Fortunately simpler, non-iterative methods can be developed. Linearizing eq. (3) with respect to  $\epsilon$ , one can estimate a reaction coordinate from:

$$0 \approx f_{\epsilon=\epsilon_0} + \left( f'_{\epsilon} \right)_{\epsilon=\epsilon_0} (\epsilon_{lin} - \epsilon_0) \quad (6)$$

Since this is a linear(ized) analysis of a non-linear problem,  $\epsilon_{lin}$  obtained from eq. (6) is only an approximation to the value obtained from solving eq. (3). If  $\epsilon_0$  can be selected near the  $\epsilon$  satisfying eq. (3), the two results will agree quite well though. Unlike solving eq. (3) by iteration, calculating  $\epsilon$  from eq. (6) is non-iterative. That simplifies the work, particularly when a great set of solvents are involved.

In practice biocatalytic reactants will in some cases preferably be quite concentrated. In fact, one reaction system that is often attractive when using liquid reactants is a solvent-free liquid mixture of reactants. A fully general solvent screening method would allow reactant concentration to be a variable, and selecting optimal

concentrations could be as much a part of process design as the selection of solvent. For that purpose the use of eq. (6) may be viable.

However, in many other cases, such as transesterification of inulin with lauryl groups [41], substantial amounts of solvent are required for dissolution of one or more reactants. Here, the reactants are present at very low concentrations. At this stage much convenience is gained from focusing on the situation where concentrations are low. If reactants are present at concentrations less than 1 mole percent, i.e.  $x_i < 0.01$ , one can without loss of accuracy replace activity coefficients by their values at infinite dilution. Then the criterion of equilibrium reduces to

$$0 = \ln K - \sum_j \nu_j \ln x_j - \sum_j \nu_j \ln \gamma_j^\infty \quad (7)$$

In eq. (7), the first two terms on the right-hand side, i.e.  $\ln K - \sum \nu_j \ln x_j$ , form a decreasing function of  $\epsilon$ , so when the rightmost sum decreases, the reaction coordinate increases. Thus for ranking of solvents, evaluation of  $S$  in eq. (8) is sufficient,

$$S = \sum_j \nu_j \ln \gamma_j^\infty \quad (8)$$

Eq. (8) is computationally simpler than eq. (6), but produces no reaction coordinate value. Thus the equilibrium reaction mixture composition is not determined. However, eq. (8) requires no  $K$ -value. This is significant, since frequently  $K$  is unavailable, or uncertain at least. For solvent-based processes, where the reactants concentration is low, eq. (8), seems to provide the most convenient measure of conversion. In fact, we have found for all cases studied that when concentrations are low, the two quantities  $S$  and  $\epsilon_{\text{lin}}$  rank solvents identically. That is, the solvent giving the lowest  $S$  generally also gives the greatest  $\epsilon_{\text{lin}}$ , and vice versa. We expect this to be so up to substantial concentration levels, so in what follows we focus on eq. (8). Daugulis and co-workers [30] and Wang and Achenie [31] seem not to employ constraints explicitly related to solvent effects on chemical reaction conversion at equilibrium.

### 3.1.1.2 Substrate Solubility

Infinite dilution activity coefficients are useful for systematic methods, because these quantities not only represent convenient common points, but in general represent conditions of maximum deviations from ideality and therefore possess maximum sensitivity to structural factor changes. Miscibility (or solubility) is promoted by low non-ideality or low (Lewis/Randall normalized) activity coefficients. Thus, to ensure substrate solubility in a solvent, it makes sense to place upper limits on the infinite dilution activity coefficients of the substrates in a solvent for consideration. For less diluted systems, i.e. substrate mole fractions above 0.01, activity coefficients will of course differ from their corresponding values at infinite dilution. Nevertheless, the ranking is normally similar. An important observation can be made from eq. (8): For sufficient equilibrium conversion,  $S$  in eq. (8), must be small, implying that activity coefficients will be small for products and large for substrates. However, sufficient substrate solubility implies small substrate activity coefficients. Thus, what is desired is a solvent where the substrate activity coefficients are low enough for sufficient substrate solubility, and where product activity coefficients are even lower. This can be difficult to find.

### 3.1.1.3 Essential Water

Water activity ( $a_w$ ) is an important parameter in biocatalysis [8,9,42,43], since the hydration of the enzyme surface depends on the water activity in the solvent (solvent plus water) medium. Constraints are imposed in order to preserve this hydration level.

When the tendency of a solvent to partition between phases or sites of different polarity is a relevant phenomenon,  $\log P$  has been widely used. When a distinct aqueous phase (even microscopic) is present ( $a_w \approx 1$ ), solubility in it probably lies behind successful use of  $\log P$  to correlate solvent effects on biocatalysts [9]. In low-water reaction mixtures ( $a_w < 1$ , and no distinct aqueous phase exists) solvent effects are as much related to water as to the solvent itself.

More hydrophilic solvents will have more capacity to dissolve water. Hence if solvents are compared at constant water content, hydrophilic solvents will leave less water on the biocatalyst surface. In this case  $\log P$  seems less obvious for correlation of solvent effects, since the appropriate parameter is water solubility in the solvent. If solvents are compared at constant water activity - as is frequently seen in scientific work e.g. [34,44] - the water stripping effect is eliminated. Yet the adverse effect of solvents seems still to partly reflect their tendency to partition into the relatively polar environment around the biocatalyst. Therefore, even when  $\log P$  is questionable,  $\log P$  may still give reasonable correlations - perhaps reflecting the general cross-correlation of all solvent polarity scales [8]. Thus, it makes sense to exclude solvents with low  $\log P$  and solvents in which water is readily dissolved.

As above we will base constraints on infinite dilution activity coefficients in pure solvents, although in reality, surface water is exposed to a mixture of water and solvent. High water activity coefficients at infinite dilution in solvent ensure that the water does not easily dissolve in the solvent. There is a very strong correlation between  $\log P$  and the simple solubility of the organic species in water [8]. Thus, constraining infinite dilution activity coefficients of solvents in water to high values, will imply high values of solvent  $\log P$ , as suggested by previous treatments [9,30,31]. This is due to the fact that  $\log P$  is related to the difference between the log of infinite dilution activity coefficients of solvent in water and in 1-octanol, respectively. The water term is typically dominant and shows the greatest variation. This is consistent with Halling [8], and it makes this treatment consistent with the recommendations of Laane [9] in this respect. Lower solvent activity coefficients (in water) will also imply less hydrophilic solvent, and less distortion of the essential layer [9]. Altogether, constraining infinite dilution activity coefficients of solvents in water, will lead to exclusion of solvents with low  $\log P$ . This is equivalent to the  $\log P$  constraints employed by Daugulis [27] and Wang and Achenie [28]. In addition we constrain the water in solvent solubility by constraining the water activity coefficients at infinite dilution.

### 3.1.1.4 Toxicity/Separability/Flammability

LC<sub>50</sub> is the lethal concentration in the sense of Martin and Young [45]. It is a measure developed from animal studies. Translation to human responses is not uncontroversial; yet we assume that it ranks solvents appropriately in terms of their toxicity. Thus, demanding a minimum on LC<sub>50</sub> of a solvent puts an upper limit on the



allowed toxicity of the solvent candidate. This treatment of toxicity resembles that of Wang and Achenie [31] but differs from that of Daugulis and co-workers [30].

Placing lower limits on a solvents boiling point has a number of consequences. In particular a high boiling solvent may facilitate removal of a volatile product by stripping. Addressing the downstream processing in more details may affect this constraint, though. Boiling temperatures also are closely related to a solvents flash point. A high boiling point corresponds to a high flash point [46], so in some (though not all) situations this may also represent a constraint on the flammability of a candidate solvent.

### 3.1.1.5 Structural Feasibility

Not all sets of molecular groups can be joined to form a feasible molecule. For example, two methyl groups can be joined to form ethane, but three cannot be joined to form “(CH<sub>3</sub>)<sub>3</sub>”. To facilitate the handling of constraints related to structural feasibility of a set of molecular groups, it is customary to restrict the problem to either *cyclic* or *acyclic* structures [32,47]. Here, we will consider only acyclic structures.

## 3.2. Experimental Methods

Experimental studies of the transesterification of vinyl laurate and 1-octanol were carried out as follows: A commercial lipase B (0.2 mg/mL) from *Candida antarctica* (Novozyme 453; Novozymes) was incubated at 40 °C for 22 h in organic solvent (100 mL) containing 0.1 mmol/mL vinyl laurate and 0.2 mmol/mL 1-octanol. Hexadecane (0.02 mmol/mL) was added as internal standard. At regular time intervals (typically ¼, ½, ¾, 1, 2, 4, 6 and 22 hours) 50 µL samples were diluted in 950 µL organic solvent in which the incubation was performed. A Focus GC (Interscience) equipped with a Restec Rxi®-5ms 30m × 0.25 mm × 0.25 µm column and a FID detector was used for determination of the formed ester. One µL of the diluted sample was injected into the GC and the applied linear temperature program started at 80 °C and was increased to 300 °C with 7.5 °C/min. The carrier gas was helium with a flow pressure of 150kPa. The incubations were performed in duplicate and the standard deviation of the results was less than 5%. The synthesis of the transesterification product between vinyl laurate and 1-octanol was calculated from the conversion of 1-octanol. All organic solvents used were of analytical grade.

## 4. Case Studies

Increasingly complicated biocatalytic reactions evolve during the years, where solvent selection is not straightforward. Decoration of inulin with lauryl esters [41] using a commercial lipase B preparation from *C. antarctica*, is an example of this. As a substrate, inulin needs to be dissolved. This is not an easy step, since only few solvents seem able to do this satisfactorily. One is DMSO [10], but it is not a desirable solvent. It smells, it is difficult to remove from the reaction mixture and some enzymes lose their catalytic activity at high DMSO concentrations. We have considered this system as an example to be screened using the CAMD procedure outlined above. As a prelude to this we use a simpler ‘test’ reaction system (called Case 1) where vinyl laurate is transesterified with 1-

octanol, using the same lipase. The transesterification of inulin with vinyl laurate is described in Case 3.

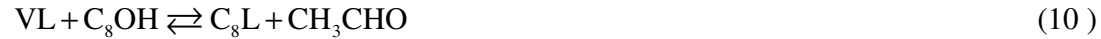
Equilibrium conversions often do not vary strongly with solvent for transesterification reactions [39], so an esterification reaction which often shows greater effects may serve as another test of the methodology. Therefore, we address an esterification reaction (Case 2) and have taken the data presented by Nordblad and Adlercreutz [44]. Solvent selection is not the goal in this case but the data are very useful for illustrating the need for experimental validation using the CAMD methodology.

#### 4.1. Case 1; Transesterification of Octanol/Vinyl Laurate

The transesterification reaction equilibrium of octanol with vinyl laurate using lipase B from *C. antarctica* can be written as ('V' is vinyl and 'L' is laurate),



The last reaction is an interconversion of ethenol to acetaldehyde by tautomerization. In solutions where tautomerization is possible, a chemical equilibrium of the tautomers will be reached. The exact ratio of the tautomers depends on several medium dependent factors, including temperature, solvent composition, and pH. Since the tautomerization is shifted strongly to the right, we may – as an approximation - write the *net* reaction equilibrium on the form (eq. 10 ),



Equilibrium conversion of transesterification reactions are often not very solvent sensitive [39]. In the absence of the vinyl group – with no tautomerization step – this is consistent with the fact that S in eq. (8 ) calculated, using group contribution methods such as UNIFAC [48], has the same value in all solvents. However, for vinyl esters, due to the tautomerization step, appreciable solvent dependence is possible.

With either of the two esters (VL and C<sub>8</sub>L), one common side reaction is hydrolysis, involving the water adsorbed on the protein surface or dissolved in the solvent medium. For studies directed towards final selection, the effects of a potential hydrolytic side reaction need to be taken into account, both in terms of reduced yield and consumption of water affecting the biocatalyst environment. For initial screenings, as here, we have not considered this aspect, although the extension is not too difficult.

##### 4.1.1 CAMD Problem Formulation; Case 1

We reject solvents that violate the constraints,

$$\begin{aligned} \text{i) } & S < \delta_c \\ \text{ii) } & \ln \gamma_{\text{VL},S}^\infty < \delta_{\text{VL}} \quad ; \quad \ln \gamma_{\text{C}_8\text{OH},S}^\infty < \delta_{\text{C}_8\text{OH}} \\ \text{iii) } & \ln \gamma_{\text{w},S}^\infty > \delta_{\text{w},S} \quad ; \quad \ln \gamma_{\text{S},\text{w}}^\infty > \delta_{\text{S},\text{w}} \\ \text{iv) } & -\log(\text{LC}_{50,S}) < \delta_{\text{LC}} \quad ; \quad \text{v) } T_{\text{b},S} > \delta_{\text{b},S} \end{aligned} \quad (11)$$

$$\text{a) } \sum_j^{\dim(\mathbf{n})} (2 - v_j) n_j = 2m, \quad \text{b) } \forall i : \sum_j^{\dim(\mathbf{n})} n_j \geq n_i (v_i - 2), \quad \text{c) } n_{\max} \geq \sum_j^{\dim(\mathbf{n})} n_j \geq n_{\min}$$

In (11)  $n_i$  is the number of times group  $i$  appears in the candidate solvent,  $\mathbf{v}$  is an array of group valency numbers and  $m = 1$  for acyclic molecules. Subscript  $S$  denotes the solvent sought and  $W$  denotes water. Constraint values,  $\delta$ , appear on the right hand side. The doubly subscripted activity coefficients,  $\gamma_{i,j}$ , are of solute  $i$  in solvent  $j$ .

Constraint i) excludes solvents giving low equilibrium conversions. Constraints ii) exclude solvents that do not dissolve sufficient substrate (VL/C<sub>8</sub>OH). Constraints iii) and iv) exclude the most hydrophilic and the most toxic solvents, respectively. v) constrains the boiling point, in order to facilitate stripping of acetaldehyde ( $T_b \approx 294$  K).

We limit the allowed set of pre-specified structural groups to five different basis sets (Table 1) for each of which the molecular design problem i)-v) plus a)-c) is solved. The basis sets in Table 1 consists of 9 groups, so  $\dim(\mathbf{n}) = 9$ .

<Table 1>

The basis sets, raise the issue of solvent reactivity (possibly enzyme-catalyzed). Using lipase reactions, a primary or secondary alcohol may become a reactant, as may any ester of such an alcohol. Amines will also often act as lipase substrates, producing amides. Furthermore, non-enzymatic reactions of amines with vinyl esters may be significant under mild conditions, and there are even examples where this is true for alcohols. Such knowledge must be used as a sensible input to the process. One could therefore argue that the alcohol group should be excluded from the basis sets, in order to have no primary and secondary alcohols included. However, in group contribution based CAMD, that would also exclude tertiary alcohols - a frequent biocatalytic medium [14,41,44]. Therefore, the alcohol group is retained. Thus, there are situations where certain groups are included in the basis sets, even though some candidates formed from these groups can be undesirable from a chemical point of view. Therefore, solvent selection requires knowledge of chemistry related to solvent reactivity – even when assisted by computer-aided methods. By working with smaller basis sets, we lower the possibility of candidates with more than single functional groups. Property predictions from group contributions are not particularly reliable for such cases [49,50,51,52]. Table 2 summarizes the constraint values employed with the five different basis sets shown in Table 1.

<Table 2>

#### *4.1.2 Method of Solution and Computations*

Activity coefficients are calculated using the UNIFAC form of Hansen et al. [48]. The method has been extended and revised a number of times in recent years [53,54]. The existence of different versions has led to many studies comparing the versions. While the versions with extended temperature parameterizations have advantages in representation of heats of mixing, less difference is seen in the case of activity coefficients. For CAMD, we need the gross features of activity coefficient variations, for which temperature independent parameters are reasonable. We use consistent group tables for predicting boiling points [55] and LC<sub>50</sub> [45].

#### *4.1.3 Results*

Table 3 summarizes the numbers of candidates. Table A.1 lists the non-excluded solvents (in terms of groups) satisfying the specifications in Tables 1-2.

<Table 3>

From Basis Set 1, thirteen generated esters and ketones satisfy the criteria in (11 ). Substantial hydrocarbon segments are needed for a candidate to satisfy all constraints. From Basis Set 2, the ethers listed typically have substantial hydrocarbon portions, compared to typical ethers. From Basis Set 3 no generated structure satisfied the constraints. Basis set 4 gave only repetitions from Basis Sets 1 – 2. Basis Set 5 is a set of molecular groups (molecules small enough to be considered groups). None of these satisfy all constraints. It is useful to check what are the active constraints in eq. (11 ) that exclude particular candidates. This information is listed in Table 4.

<Table 4>

As mentioned, primary and secondary alcohols have a risk of becoming a reactant, as does any ester of such an alcohol and amines (including non-enzymatic reactions under certain conditions). Therefore, losing the primary alcohols, esters and amines as shown in Table 4 is not a problem. These would be eliminated anyway. However, here the solvents are excluded by other criteria as becoming a reactant.

#### 4.1.4 Experimental Testing

The property prediction methods used for the constraints i) – v) are quite reliable for ranking solvent properties. A key prediction of the model is the chemical equilibrium positions dependence on solvent, as explicit in constraint i). The ranking of solvents in terms of S assumes the presence of the enzyme has no impact on the equilibrium. This is not always valid. Therefore the ranking based on S is tested against a few experimental measurements. Fig. 2 shows the fraction of vinyl laurate converted as a function of time in *tert*-butanol. The curve fitted represents a fit of a second order kinetic model to the data. The correlation seems satisfactory, although some scatter is seen.

<Figure 2>

For all solvents tested, octyl ester production was at least linear for 1 h. The limiting conversion is near 69 % using *tert*-butanol as solvent. Results for the other solvents, are shown in Table 5. When ethyl acetate was used as solvent hardly any trans-esterification product between vinyl laurate and 1-octanol was found. This is due to reactions of 1-octanol with ethyl acetate which is the most abundant compound present.

<Table 5>

Fig. 3 shows the correlation among S and %-conversion of vinyl laurate. There is a clear trend with some scatter, but the gross features of the experimental and computational

results are consistent. Therefore, when  $S$  is small, experiments produce high conversions. This suggests that the CAMD procedure is valid.

<Figure 3>

#### 4.1.5 Conclusions/Discussions; Case 1

Solving a CAMD problem requires assumptions regarding the chemical reaction(s) taking place. The conclusions reached depend upon the validity of these assumptions, both regarding the main reaction and possible side reactions. It is vital to validate the reaction equilibria using (limited) experimental data. Case 2 will emphasize this point.

#### 4.2. Case 2; Esterification of Octanol/Acrylic Acid

Esterification reactions are often more sensitive to the solvent chosen. Moreover, esterification data are useful to demonstrate an important feature: For some reactions, the presence of the enzyme seems to affect the equilibrium conversion to products. Adsorption of water by the catalyst particles is one of the circumstance under which this can happen. Initial rate data of Nordblad and Adlercreutz [44] are considered here for esterification of acrylic acid (2-propenoic acid)/octanol to form octyl acrylate/water:



The reaction mixtures contained all 60 mM of acrylic acid and octanol, and the reactions were performed at 40°C. Kinetic initial rate are not simply related to a property at equilibrium, such as  $S$ , so perfect correlation among the two can hardly be expected. Nevertheless, speculation in correlations between kinetic properties (e.g. initial rates) and properties characterizing equilibrium states (such as  $\log P$ ) are common in the literature [9,34,44]. We therefore have computed  $S$  values (as above) for the solvents considered in these experiments and compared the results to the initial rates shown in Fig. 4.

<Figure 4>

In calculation of  $S$  (Fig. 4B) one ignores the presence of the enzyme. Perhaps not unexpectedly the  $S$  values suggest that esterification is not substantial in hydrocarbons - because formation of water in hydrocarbons is unfavorable. However, the experiment suggests that initial rates in hydrocarbons are great. Intuitively water - when formed as a reaction product - needs not be dissolved in the 'bulk' solvent phase, but can instead be adsorbed by the aqueous surface layer on the protein, consistent with previous experimental [56] and theoretical [57,58] results. That allows water to 'escape' from the medium, and it may promote water formation. Then a situation is created where the medium which expels water (or removes product) most effectively is the most efficient medium - even though water needs to be formed. We have therefore compared the values of  $S$  without the water term. This corresponds to a situation where the reaction mixture is in contact with a water phase where the water activity is constant ( $a_w$  near 1), though not necessarily unity. This leads to the ranking shown in Fig. 4A, where things fall into place much better, and low  $S$  values (indicating a favorable solvent) in fact are seen to produce high initial rates. The authors [44] noted an increased reaction rate when solvent  $\log P$

increased, consistent with previous data on esterification [34]. Thus, in the cases of transesterification as well as esterification,  $\log P$  values show the same trend as initial rate data. However, the underlying explanations based on a molecular level perception can be entirely different. Valivety et al. [34] rationalized esterification data at equilibrium by transforming the 'fundamental' equilibrium constant,  $K$  (similar to eq.(5)), into a so-called 'practically useful' equilibrium constant,  $K_0$ , defined by

$$K_0 = K \prod_{j \neq \text{water}} (\gamma_j)^{-v_j} = a_{\text{water}} \prod_{j \neq \text{water}} (x_j)^{v_j} \quad (13)$$

While the water activity was unity throughout, they found that high water solubility in the solvent gave a small  $K_0$  (poor esterification). They attempted correlation with other measures of hydrogen bonding capability. While their treatment appears different, it seems that recognizing that (nearly) constant water activity eliminates water from consideration is similar to our treatment. Since systems where water is a reactant require special attention, modified constraints can create conflicting situations. Valivety [34] noted that a hydrolysis reaction will be favored in solvents where water solubility is high. This appears at first glance to be in conflict with the nature of our constraint iii). However, if water is a substrate, we will also demand water solubility to be sufficiently high. This will imply upper limits in constraint ii) on substrate/water activity coefficients in solvents. This should allow for counterbalancing this effect, although admittedly some re-evaluation of constraints seems necessary in such cases. Thus, in both Cases considered we need the experimental results to check if computations underlying the CAMD algorithm rank the solvents appropriately. If the measurements do not confirm the calculations, the CAMD algorithm must be revised.

### 4.3. Case 3; Transesterification of Inulin/Vinyl Laurate

Inulin is a polysaccharide extractable from chicory or dahlia roots. It is sweet, non-fattening and can be used as a prebiotic. Decoration of inulin with laurate groups, in order to improve its functionality [41], is an interesting reaction. In the process the product accumulates both in solution and in the solid phase. Halling and co-workers [59] found, the highest overall yields (liquid plus solid phases) in solvents where the substrate solubilities are minimized. The best yields of solid product were found in solvents where both product and substrate solubilities are low. Our reaction scheme is somewhat different. Here, we will optimize the liquid product, based on  $S$  of eq.(8).

#### 4.3.1. CAMD Problem Formulation; Case 3

We have identified solvent structures by solving a CAMD problem, as before:

$$\begin{aligned} & \text{i) } S < \delta_C \\ & \text{ii) } \ln \gamma_{VL,S}^\infty < \delta_{VL} \quad ; \quad \ln \gamma_{I,S}^\infty < \delta_I \\ & \text{iii) } \ln \gamma_{W,S}^\infty > \delta_{W,S} \quad ; \quad \ln \gamma_{S,W}^\infty > \delta_{S,W} \\ & \text{iv) } -\log(LC_{50,S}) < \delta_{LC} \quad ; \quad \text{v) } T_{b,S} > \delta_{b,S} \\ & \text{a) } \sum_j^{\dim(n)} (2 - v_j) n_j = 2m, \quad \text{b) } \forall i : \sum_j^{\dim(n)} n_j \geq n_i (v_i - 2), \quad \text{c) } n_{\max} \geq \sum_j^{\dim(n)} n_j \geq n_{\min} \end{aligned} \quad (14)$$

The criteria applied to the octanol problem in eq. (11 ) are all applicable to the inulin reaction. As mentioned, when the solubility of the substrates increases, the conversion decreases, since the terms forced to be small by constraint ii) increase the sum in constraint i). Therefore, if the solubility of inulin is to be larger, conversion will be lower. Likewise, if the equilibrium reaction coordinate is to be large, sufficient inulin may not dissolve. As a compromise we have solved the CAMD problem with two different sets of inputs: One set with strong demands on inulin solubility, relaxing constraints on conversion, and another set with strong demands on conversion, relaxing demands on inulin solubility. Table 6 lists the constraint values used with Basis Sets 1-4 in Table 1.

<Table 6>

Experience [41] suggests that the reaction temperature for lipase catalyzed inulin transesterification may exceed 80°C, so the boiling point is constrained to above 360 K, to also facilitate acetaldehyde removal. Constraints on LC<sub>50</sub> and water solubility are not changed.

#### 4.3.2 Method of Solution and Computations

Thermodynamic treatment of inulin is not straightforward, since there are no standard models for predicting absolute values of its solubility in different solvents. With focus on the relative ranking of solvents, sensible approximations may be made to obtain useful results. To do so, the absolute solubility values need to be ignored. Because the procedure is non-standard, some explanation follows. When solubility of a solid solute in a solvent (mixture or pure) is small (less than 1 mole percent,  $x_i < 0.01$ ), the expression

$$\ln x_i \approx \ln x_i^{\text{id}} - \ln \gamma_i^\infty \approx \frac{\Delta H_{\text{mi}}}{RT_{\text{mi}}} \left(1 - \frac{T_{\text{mi}}}{T}\right) - \ln \gamma_i^\infty \quad (15)$$

is valid [60]. The ideal solubility is based on the properties of pure crystals, involving the temperature and melting heat of inulin. Using equation (15 ) for different solvents requires that inulin forms identical crystal lattices in all solvents considered. This is often the case, for small molecular solutes. For larger molecules, it is questionable. Although a fully reversible melting/fusion process is unrealistic (inulin partially decomposes (thermally) below the melting point), Table 7 lists experimental [61,62] melting enthalpy and temperature data reported on inulin. Table 7 also gives UNIFAC group assignments and molar masses of the different species relevant to the process. Also, inulin is a large molecule with a large number of groups. Group contribution methods are typically most reliable when used for systems with a moderate number of groups [51,52].

<Table 7>

Another issue which complicates the thermodynamic treatment of inulin is that it is polydisperse, so the chain length has no clear cut value, but follows some distribution. The inulin used here had an average length of 10 carbohydrate units. Thus its interaction with solvents will be dominated by the hydroxyl, hydrocarbon, and ether groups in the repeating unit. In what follows we compare the inulin repeating unit ('I-OH' in Table 7) solubility as if the repeating unit was a distinct molecule, assuming that the solvents that

dissolve the inulin repeating unit to highest extent will also be the best inulin solvents. We consider the decorated inulin to be the inulin repeating unit transesterified with vinyl laurate ('I-O-L' in Table 7). These assumptions and eq. (15) form the basis of the three rightmost columns of Table A.1 and columns 2-4 (from right) in Table A.2 and A.3. On this basis, it is clear that the mole fraction solubilities of inulin,  $x_I$ , can only be considered on as relative values. For example, considering 1 mole percent repeating unit as basis, the 'true' inulin mole fraction would be about 10 times less.

#### 4.3.3. Results

One important difference between Case 1 and Case 3 is the ease of dissolving octanol and inulin. We have examined the inulin solubility in the solvents considered for Case 1. The three rightmost columns in Table A.1 compare the inulin solubility in the respective solvents, as calculated using eq. (15). It would be desirable to base the solvent selection for Case 3 on experiences with Case 1, assuming that a good solvent in Case 1 will also be good in Case 3. Unfortunately, the three rightmost columns (giving  $\ln\gamma_{IS}^\infty$ ,  $\ln x_{IS}$ ,  $x_{IS}$ , where 'I' denotes inulin) in Table A.1 suggest that inulin is barely soluble in the solvents identified in Case 1. This suggests that Case 1 might be useful as a proof-of-concept reaction, but it provides limited information on a useful solvent in Case 3.

We have examined the twelve pure solvents shown in Table 8. With the exception of *tert*-butanol, these are considered molecular groups with UNIFAC, and do not resolve into smaller groups. Therefore these will not be generated by a CAMD method. Unfortunately most solvents in Table 8 ( $\text{CH}_3\text{OH}$ ,  $\text{CH}_3\text{NH}_2$ ,  $\text{CH}_3\text{CN}$ ,  $\text{CHCl}_3$ ,  $\text{CCl}_4$ ,  $\text{CS}_2$  and  $\text{CH}_2\text{Cl}_2$ ) have low boiling points. A number of these (except furfural) are much better inulin dissolvers than those in Table A.1. Unfortunately, the effective inulin dissolving solvents seem to be relatively hydrophilic, and vinyl laurate solubility is compromised.

<Table 8>

In Table A.2 - A.3 'W' is water, S is the solvent, I is inulin and E is the ester substrate (vinyl laurate). For each basis set the constrained quantities are reported in columns below the constraint value applied in the CAMD problem. Table A.2 displays results obtained from solving the CAMD problem in eq. (14), using constraints that (relatively) emphasize conversion over substrate solubility (Specification 1 in Table 6). Thus, these candidates give high conversions, but low amounts of inulin can be dissolved in these solvents, as shown in columns 2-4 from right. Certain group sets correspond to commonly known solvents, such as propionitrile, butyronitrile (Basis Set 3) and hexanone and 3-pentanone (Basis Set 4). All have low inulin solubilities except a few from Basis Set 3.

All entries in Table A.2 have a constraint i) sum,  $S < -3$ . The best inulin dissolver generated from Table A.2 is propionitrile. Inspection of all solvents in Table 8 having a constraint i) sum less than -3 shows that only methanol, acetonitrile and DMF produce gamma infinity values for inulin less than that of 1.45 found in propionitrile. Table A.3 lists results obtained when emphasizing high inulin solubility over conversion (Specification 2 in Table 6). The numbers in the rightmost column are greater in A.3 than in A.2. Thus, the solvent candidates in A.3 dissolve high amounts of inulin, but do not



allow as great conversion as those in A.2. Certain group arrays correspond to commonly known solvents, such as *tert*-butanol, 3-methyl-2-butanol, *tert*-pentanol and 3-methylbutanol (Basis Set 1), among which some are primary alcohols. All candidates of Basis Set 2 are amines. These can act as lipase substrates, producing amides, and non-enzymatic reactions with vinyl esters may be significant even under mild conditions. Propionitrile is regenerated from Basis Set 3. All entries in Basis Set 4 are repetitions. The candidates generated have low inulin solubilities.

Certainly Table 8 has entries with greater inulin solubilities than those reported in Table A.2 and A.3. DMF, acetonitrile, 1,2-ethanediol and DMSO dissolve inulin well, but not vinyl laurate. DMSO and 1,2-ethanediol do not provide great conversions either. All seem to be quite hydrophilic.

#### 4.3.4. Conclusions/Discussion; Case 3

It seems that the options among pure solvents are exhausted, since no single candidate seems to offer sufficient inulin solubility, sufficient hydrophobicity, high boiling temperature and sufficient conversion. This may only be feasible with mixed solvents, as experience shows.

The candidates of Table A.2 are propio- and butyronitrile, provided inulin solubility can be enhanced. This will not be achieved by combination with candidates from Table A.3, since the solubility enhancement seems insufficient for that. The candidates of Table A.3 are *tert*-butanol, *tert*-pentanol and propionitrile. Inulin solubility enhancement is needed in all cases. This can only be done by candidates outside Table A.2-A.3. Turning attention to Table 8, it seems to be at the expense of vinyl laurate solubility and hydrophobicity. Again, vinyl laurate solubility seems highest in propionitrile and hydrophobicity of propionitrile seems somewhat greater than that of the tertiary alcohols. From Table 8, DMSO and DMF look like realistic additives.

Previous works on transesterification of inulin [41] report the feasibility of carrying out the reaction in a mixture of *tert*-butanol and DMSO. The results of this section seem consistent with this experimental finding. It is clear that employing a mixed solvent of tertiary alcohol or propionitrile plus DMSO or DMF will require that the mixed solvent forms a stable single phase. According to UNIFAC, all such mixtures seem to form stable single phases at all compositions. Likewise, the hydrophobicity concerns can be evaluated. This makes us lean slightly towards the propionitrile systems. The less DMF or DMSO added, the closer the system will resemble the alcohol or nitrile systems in terms of hydrophobicity. The inulin solubility concerns will also point towards propionitrile, since addition of small amounts of DMSO or DMF will increase inulin solubility faster than in alcohols. On the other hand, the nitrile systems are more toxic.

## 5. Discussion

Automation of CAMD for biocatalysis is a challenge. Certain constraints seem to be similar for different reactions. However, the use of CAMD in biocatalysis is an iterative process with the steps: 1) Problem formulation; 2) Solve to obtain a sensible solution; 3) Reconcile the solution with previously known facts; 4) Revise the problem and solve again. Obtaining useful results requires a good understanding of the process, its

chemistry and its objectives. Different reactions may require different constraints, and constraints revisions may be necessary, experiments are necessary and constraints values need to be adjusted and so forth. For example, the two transesterification reactions, involving different substrates, led to quite different considerations. Here, we have investigated the applicability of CAMD as a methodology in biocatalysis.

The current methodology has some limitations. For example it does not directly incorporate the influence of multiple (or side) reactions. Inclusion of the effects of side reactions is possible. However, these must be known in advance. If they are, another constraint resembling i) on the main reaction can be included. Also, single phase solvents are considered only. We do not believe extensions to treat such cases as solvent free, biphasic systems will be too difficult, but revised constraints will be required to do so. Formal extension of this framework to solvent mixture screening would be an obvious next step. By inspection of the list of constraints in eq. (14) it seems that constraints i, ii, iii and a) - b) are straightforwardly extended to mixtures. The only exception is the solvent solubility in water in constraint iii), which is a single component property. The boiling points and  $LC_{50}$  estimates can be approximated by linear interpolations of pure components values – at least for screening purposes.

The constraints in eq. (11) contain thermodynamic statements, phrasing the solvent screening problem as an essentially thermodynamic problem. As is known [8,13], one can question whether thermodynamics is applicable to reaction systems for biocatalysis, since such systems are really not at equilibrium in all respects, as required for rigorous application of thermodynamics. Reaction rates are not considered explicitly. Only constraint iii) considers solvent/catalyst interactions. Thus, there is no guarantee that a solvent that satisfies all criteria in eq. (11) gives a high rate. Solvents that satisfy eq. (11) will enable high conversion. How fast the reaction approaches equilibrium is a different matter. The answer depends upon the efficiency of the enzyme in the medium. However, solvents not satisfying eq. (11) will not allow a high conversion, irrespective of catalyst efficiency. Therefore we may as well eliminate most of such solvents at an early stage. Therefore, while eq. (11) does not incorporate all important statements on enzyme performance, this is not a deficiency of the approach. Eq. (11) screens out poorly performing solvents, without final selections.

The choice of the UNIFAC forms may conceivably influence the decisions. Using the forms chosen here ensures that all properties predicted from group contributions are possible using a consistent group table. Accommodating multiple group tables and multiple definitions for describing the same molecule is thereby avoided. The most established UNIFAC versions mainly differ by the

- form of the temperature dependency of the interaction parameters,
- choice of which atom collections will be formed into groups,
- method of determining the volume and surface area parameters for groups, and
- parameter availability

The result is a suite of methods where the differences appear to be minimal, particularly for activity coefficients and free energies, in many organic and non-polymeric systems. Consistent with the fact that the composition dependence of  $H^E$  can be dramatically more complex than that of  $G^E$  [63], significant errors mostly appear only in excess enthalpies and entropies of systems with strongly polar, associating and/or non-spherical substances, and if the properties of isomers [50,51] are compared. Illustratively, Tassios and co-

workers [64] needed two different correlations to describe alcohol and non-alcohol systems. Obviously, the choice of which collections of atoms to define as groups, needs to be made with care so that these are independent of neighboring groups [51]. Differences among the versions mainly relate to whether or not cyclic and non-cyclic methylenes and hydroxyl groups of different orders are distinguished. The original UNIFAC forms [48] used volume and surface area parameters for the groups, proportional to values derived from the Bondi [65] tables, so the volume parameter is typically greater than the surface area parameter. Later versions [66] have routinely fitted these parameter values, while relaxing restrictions on their relative magnitudes. That gives suspicious results in some cases [67]. Finally, the modified form is more widely applicable, since it has parameters for more groups. Traditionally, extensions of parameter tables require experimental vapor-liquid equilibrium data. These are frequently unavailable. Much work has been done lately to try to derive the parameters from other sources than phase equilibrium data, including

- fitting the model to non-experimental ‘pseudo data’ generated from either molecular dynamics [68] simulations or calculations using conductor-like screening models [69]
- connectivity relationships [70]

It remains to be seen how this problem will be dealt with in future. Either way, at ambient conditions, there should not be significant differences in the results, since only activity coefficients have been employed.

## 6. Conclusions

A computational procedure is developing that can potentially assist the biochemist in the solvent selection step, since it

- helps rationalizing experimental results,
- helps strategic selection of experimental efforts,
- makes it more transparent why certain candidates are not selected,
- allows the consideration of a great search space of molecular candidates

and produces results consistent with experiments. Although the examples provided deal with lipases, there is no preclusion against application of the modeling framework to other enzymes, such as polyhydroxyalkanoates depolymerases in solvents. Therefore the results have a quite general value.

## Acknowledgements

This work was performed in the framework of the IP Bioproduction (NMP-2CF-2007-026515), within the 6<sup>th</sup> Framework Programme of the European Commission.

## Notation

a	:	activity (thermodynamic)
f	:	function defined in eq. (3 )
G	:	Gibbs free energy as in eq. (4 )
H	:	enthalpy as in eq. (15 )
K	:	reaction equilibrium constant

m	:	method specific number
$N_i$	:	Mole number of molecule i
$n_i$	:	number of times group i appears in a molecule
$n_{\max}, n_{\min}$	:	Upper and lower bounds on number of groups in a molecule
P	:	Partition coefficient (octanol/water)
R	:	Gas constant
S	:	Summation, defined for a given reaction as in eq. (8)
T	:	Temperature
$v_i$	:	group i valency
$x_i$	:	mole fraction of molecule i

#### Greek Symbols

$\varepsilon$	:	Reaction coordinate, as in eq. (2)
$\delta$	:	Constraint values (subscripted as in eq. (11)/(14))
$\gamma$	:	activity coefficient
$\nu$	:	stoichiometric coefficient
$\Delta$	:	change (in a function associated with a reaction)

#### Abbreviations

CAMD	:	<u>computer-aided molecular design</u>
VL	:	<u>vinyl laurate</u>
DMF	:	<u>dimethyl formamide</u>
DMSO	:	<u>dimethyl sulfoxide</u>
MTBE	:	<u>methyl-tert-butyl ether</u>
I-OH	:	<u>inulin</u>
I-O-L	:	esterified <u>inulin</u> (laurate)
C <sub>8</sub> OH	:	octanol
LC <sub>50</sub> (within 96 hrs.)	:	<u>lethal concentration</u> killing <u>50</u> % of a fathead minnow population

#### Subscripts

b,S	:	<u>boiling point of solvent S</u>
C	:	<u>conversion</u>
f	:	<u>formation</u>
i	:	component i
i,j	:	component i in j
LC	:	<u>lethal concentration</u>
S	:	<u>solvent</u>
W	:	<u>water</u>
0	:	reference condition

#### Superscripts

$\infty$	:	Infinite dilution
o	:	standard state of formation: ideal gas, 298.15 K and 1 bar
id	:	ideal solution as in eq. (15)
E	:	'excess' function

## Appendix A Solvent candidate structures generated

**Table A.1**  
Solvent candidates for vinyl laurate/octanol transesterification (T = 313.15 K)

								Constraint Values										
								3.5	350	2	2	2	2	-0.415				
Basis Set 1	CH <sub>2</sub>	CH	C	CH <sub>3</sub> CO	CH <sub>2</sub> CO	CH <sub>3</sub> COO	CH <sub>2</sub> COO	CH <sub>3</sub>	-log(LC <sub>50</sub> )	T <sub>b</sub> /K	lnγ <sub>WS</sub> <sup>∞</sup>	lnγ <sub>SW</sub> <sup>∞</sup>	lnγ <sub>ES</sub> <sup>∞</sup>	lnγ <sub>AS</sub> <sup>∞</sup>	S	lnγ <sub>IS</sub> <sup>∞</sup>	lnx <sub>IS</sub>	X <sub>IS</sub>
0	0	1	0	0	1	0	0	3	3.2	385.4	2.5	6.4	0.1	0.7	-0.46	4.9	-5.1	0.006
0	0	1	0	1	0	0	0	4	2.7	418.3	3.2	7.1	0.0	0.6	-0.42	5.3	-5.5	0.004
0	1	0	0	0	0	0	1	3	3.4	387.7	2.8	6.9	-0.2	0.7	-0.51	5.5	-5.7	0.003
0	1	0	0	0	0	1	0	2	2.9	366.8	2.4	5.4	0.1	0.7	-0.44	4.3	-4.5	0.01
0	2	0	0	1	0	0	0	4	3.2	441.2	3.3	8.2	0.0	0.6	-0.48	5.7	-5.9	0.003
1	0	0	0	0	0	0	1	2	3.1	369.4	2.7	5.8	-0.2	0.6	-0.46	5.0	-5.2	0.006
1	0	1	0	1	0	0	0	4	3.2	441.3	3.3	8.1	0.0	0.6	-0.47	5.7	-5.9	0.003
1	1	0	0	0	1	0	0	2	3.4	395.9	2.5	6.5	0.0	0.7	-0.48	4.9	-5.0	0.006
1	1	0	0	1	0	0	0	3	2.9	427.3	3.2	7.2	0.0	0.6	-0.43	5.3	-5.4	0.004
2	0	0	0	0	1	0	0	1	3.1	378.4	2.4	5.4	0.1	0.7	-0.44	4.3	-4.5	0.01
2	1	0	0	1	0	0	0	3	3.3	449.3	3.3	8.2	0.0	0.6	-0.48	5.7	-5.9	0.003
3	0	0	0	1	0	0	0	2	3.0	436.0	3.2	7.2	0.0	0.6	-0.43	5.3	-5.5	0.004
4	0	0	0	1	0	0	0	2	3.5	457.2	3.3	8.2	0.0	0.6	-0.48	5.7	-5.9	0.003
Basis Set 2	CH <sub>2</sub>	CH	C	CH <sub>2</sub> O	CH-O	CH <sub>2</sub> NH <sub>2</sub>	CHNH <sub>2</sub>	CH <sub>3</sub>	-log(LC <sub>50</sub> )	T <sub>b</sub> /K	lnγ <sub>WS</sub> <sup>∞</sup>	lnγ <sub>SW</sub> <sup>∞</sup>	lnγ <sub>ES</sub> <sup>∞</sup>	lnγ <sub>AS</sub> <sup>∞</sup>	S	lnγ <sub>IS</sub> <sup>∞</sup>	lnx <sub>IS</sub>	X <sub>IS</sub>
0	0	0	0	2	0	0	0	4	2.3	361.7	3.1	7.3	-0.3	0.7	-0.61	6.8	-7.0	0.0009
0	0	1	0	1	0	0	0	5	2.8	362.4	4.3	9.2	-0.1	1.2	-0.90	9.1	-9.3	9 10 <sup>-5</sup>
0	2	0	1	0	0	0	0	4	3.0	379.0	3.5	8.4	-0.2	0.9	-0.74	7.7	-7.8	0.0004
1	0	0	0	2	0	0	0	4	2.7	391.5	3.3	8.4	-0.2	0.8	-0.68	7.2	-7.4	0.0006
1	0	1	1	0	0	0	0	4	2.9	379.1	3.5	8.3	-0.2	0.9	-0.72	7.7	-7.9	0.0004
1	1	0	0	1	0	0	0	4	3.0	374.1	4.4	9.2	-0.1	1.2	-0.91	9.1	-9.3	9 10 <sup>-5</sup>
1	1	0	1	0	0	0	0	3	2.7	359.9	3.4	7.3	-0.4	0.8	-0.67	7.2	-7.4	0.0006
2	0	0	0	1	0	0	0	3	2.7	354.6	4.3	8.1	-0.3	1.1	-0.85	8.7	-8.9	0.0001
2	1	0	1	0	0	0	0	3	3.1	390.0	3.5	8.4	-0.2	0.9	-0.73	7.7	-7.8	0.0004
3	0	0	0	1	0	0	0	3	3.1	385.4	4.4	9.2	-0.1	1.2	-0.91	9.1	-9.3	9 10 <sup>-5</sup>
3	0	0	1	0	0	0	0	2	2.8	371.9	3.4	7.3	-0.4	0.8	-0.66	7.2	-7.4	0.0006
4	0	0	1	0	0	0	0	2	3.2	400.4	3.5	8.4	-0.2	0.9	-0.73	7.7	-7.8	0.0004
Basis Set 4	CH <sub>2</sub>	CH	C	CH <sub>2</sub> CO	CH <sub>3</sub> O	CH <sub>3</sub> N	CH <sub>2</sub> N	CH <sub>3</sub>	-log(LC <sub>50</sub> )	T <sub>b</sub> /K	lnγ <sub>WS</sub> <sup>∞</sup>	lnγ <sub>SW</sub> <sup>∞</sup>	lnγ <sub>ES</sub> <sup>∞</sup>	lnγ <sub>AS</sub> <sup>∞</sup>	S	lnγ <sub>IS</sub> <sup>∞</sup>	lnx <sub>IS</sub>	X <sub>IS</sub>
0	0	1	1	0	0	0	0	4	2.7	418.3	3.2	7.1	0.0	0.6	-0.42	5.3	-5.5	0.004
0	1	1	0	1	0	0	0	4	2.8	388.8	2.9	7.6	-0.2	0.7	-0.59	6.6	-6.8	0.001
0	2	0	0	1	0	0	0	3	2.5	370.5	2.8	6.5	-0.3	0.6	-0.53	6.1	-6.2	0.002
0	2	0	1	0	0	0	0	4	3.2	441.2	3.3	8.2	0.0	0.6	-0.48	5.7	-5.9	0.003
1	0	1	0	1	0	0	0	3	2.5	370.6	2.8	6.5	-0.3	0.6	-0.52	6.1	-6.3	0.002
1	0	1	1	0	0	0	0	4	3.2	441.3	3.3	8.1	0.0	0.6	-0.47	5.7	-5.9	0.003
1	1	0	0	1	0	0	0	2	2.2	350.5	2.6	5.4	-0.4	0.5	-0.44	5.5	-5.6	0.004
1	1	0	1	0	0	0	0	3	2.9	427.3	3.2	7.2	0.0	0.6	-0.43	5.3	-5.5	0.004
1	2	0	0	1	0	0	0	3	3.0	399.1	2.9	7.6	-0.3	0.7	-0.60	6.6	-6.7	0.001
2	0	1	0	1	0	0	0	3	2.9	399.3	2.9	7.6	-0.2	0.7	-0.59	6.6	-6.8	0.001
2	1	0	0	1	0	0	0	2	2.7	381.9	2.8	6.5	-0.3	0.6	-0.53	6.1	-6.2	0.002
2	1	0	1	0	0	0	0	3	3.3	449.3	3.3	8.2	0.0	0.6	-0.48	5.7	-5.9	0.003
3	0	0	0	1	0	0	0	1	2.3	363.1	2.6	5.4	-0.4	0.5	-0.44	5.5	-5.6	0.004
3	0	0	1	0	0	0	0	2	3.0	436.0	3.2	7.2	0.0	0.6	-0.43	5.3	-5.5	0.004
3	1	0	0	1	0	0	0	2	3.1	409.1	2.9	7.6	-0.2	0.7	-0.60	6.6	-6.7	0.001
4	0	0	0	1	0	0	0	1	2.8	392.7	2.8	6.5	-0.3	0.6	-0.52	6.1	-6.2	0.002
4	0	0	1	0	0	0	0	2	3.5	457.2	3.3	8.2	0.0	0.6	-0.48	5.7	-5.9	0.003
5	0	0	0	1	0	0	0	1	3.2	418.6	2.9	7.6	-0.2	0.7	-0.60	6.6	-6.7	0.001

'A': Alcohol (octanol); 'E': Ester (vinyl laurate); 'W': Water; 'S': Solvent, 'I': Inulin.  
lnx<sub>i</sub><sup>∞</sup> (313.15 K) = -0.17 (for i = I).

T<sub>b</sub> is the first-order result obtained with the method of Constantinou/Gani [55].

**Table A.2**  
Solvent candidates for vinyl laurate/inulin transesterification; specification 1 (as in Table 6) (T = 313.15 K)

										Constraint Values							
Basis Set 1										3.5	360	3	3	2	5		
CH <sub>2</sub>	CH	C	OH	CH <sub>3</sub> CO	CH <sub>2</sub> CO	CH <sub>3</sub> COO	CH <sub>2</sub> COO	CH <sub>3</sub>	-log(LC <sub>50</sub> )	T <sub>b</sub> /K	lnγ <sub>WS</sub> <sup>∞</sup>	lnγ <sub>SW</sub> <sup>∞</sup>	lnγ <sub>ES</sub> <sup>∞</sup>	lnγ <sub>IS</sub> <sup>∞</sup>	lnx <sub>l</sub>	x <sub>l</sub>	S
0	1	0	0	0	1	0	0	3	2.45	402.5	3.03	6.11	-0.02	4.84	-5.0	0.007	-4.31
2	0	0	0	0	1	0	0	2	2.57	412.3	3.03	6.10	-0.02	4.84	-5.0	0.007	-4.31
Basis Set 2										3.5	360	2	2	2	5		
CH <sub>2</sub>	CH	C	CH <sub>3</sub> O	CH <sub>2</sub> O	CH-O	CH <sub>2</sub> NH <sub>2</sub>	CHNH <sub>2</sub>	CH <sub>3</sub>	-log(LC <sub>50</sub> )	T <sub>b</sub> /K	lnγ <sub>WS</sub> <sup>∞</sup>	lnγ <sub>SW</sub> <sup>∞</sup>	lnγ <sub>ES</sub> <sup>∞</sup>	lnγ <sub>IS</sub> <sup>∞</sup>	lnx <sub>l</sub>	x <sub>l</sub>	S
0	1	0	0	2	0	0	0	3	2.42	383.2	2.10	5.80	-0.08	4.97	-5.1	0.006	-4.33
0	1	0	1	0	1	0	0	3	2.29	388.1	2.11	5.81	-0.08	4.98	-5.1	0.006	-4.33
2	0	0	0	2	0	0	0	2	2.54	393.9	2.10	5.80	-0.08	4.97	-5.1	0.006	-4.33
2	0	0	1	0	1	0	0	2	2.42	398.6	2.11	5.81	-0.08	4.98	-5.2	0.006	-4.33
Basis Set 3										3.5	360	2	2	2	5		
CH <sub>2</sub>	CH	C	CHO	CH <sub>2</sub> NH	CHNH	CH <sub>2</sub> CN	DMF-2	CH <sub>3</sub>	-log(LC <sub>50</sub> )	T <sub>b</sub> /K	lnγ <sub>WS</sub> <sup>∞</sup>	lnγ <sub>SW</sub> <sup>∞</sup>	lnγ <sub>ES</sub> <sup>∞</sup>	lnγ <sub>IS</sub> <sup>∞</sup>	lnx <sub>l</sub>	x <sub>l</sub>	S
0	0	0	0	0	0	1	0	1	1.71	364.2	2.48	3.55	0.68	1.45	-1.6	0.2	-4.06
0	0	1	0	0	0	1	0	3	2.76	425.2	2.84	6.47	0.35	3.46	-3.6	0.03	-4.25
0	0	1	0	0	0	2	0	2	3.23	510.7	2.06	6.30	1.91	2.09	-2.3	0.1	-4.19
0	1	0	0	0	0	1	0	2	2.48	410.0	2.77	5.55	0.42	2.89	-3.1	0.05	-4.20
0	2	0	0	0	0	1	0	3	3.25	447.4	2.97	7.59	0.31	3.89	-4.1	0.02	-4.32
1	0	0	0	0	0	1	0	1	2.15	393.7	2.64	4.55	0.52	2.25	-2.4	0.09	-4.13
1	0	1	0	0	0	1	0	3	3.20	447.4	2.93	7.49	0.30	3.92	-4.1	0.02	-4.31
1	1	0	0	0	0	1	0	2	2.92	433.9	2.88	6.57	0.36	3.43	-3.6	0.03	-4.27
1	1	0	0	0	0	2	0	1	3.40	516.5	2.09	6.42	1.96	2.09	-2.3	0.1	-4.20
2	0	0	0	0	0	1	0	1	2.60	419.4	2.77	5.55	0.42	2.89	-3.1	0.05	-4.20
2	1	0	0	0	0	1	0	2	3.37	455.3	2.96	7.59	0.31	3.90	-4.1	0.02	-4.32
3	0	0	0	0	0	1	0	1	3.05	442.3	2.88	6.57	0.36	3.43	-3.6	0.03	-4.27
4	0	0	0	0	0	1	0	1	3.49	462.9	2.96	7.59	0.31	3.90	-4.1	0.02	-4.32
Basis Set 4										3.5	360	2	2	2	5		
CH <sub>2</sub>	CH	C	OH	CH <sub>2</sub> CO	CH <sub>3</sub> O	CH <sub>3</sub> N	CH <sub>2</sub> N	CH <sub>3</sub>	-log(LC <sub>50</sub> )	T <sub>b</sub> /K	lnγ <sub>WS</sub> <sup>∞</sup>	lnγ <sub>SW</sub> <sup>∞</sup>	lnγ <sub>ES</sub> <sup>∞</sup>	lnγ <sub>IS</sub> <sup>∞</sup>	lnx <sub>l</sub>	x <sub>l</sub>	S
0	0	1	0	2	0	0	0	4	3.18	501.6	2.15	6.88	0.72	3.89	-4.1	0.02	-4.27
0	1	0	0	1	0	0	0	3	2.45	402.5	3.03	6.11	-0.02	4.84	-5.0	0.007	-4.31
0	1	0	0	2	0	0	0	3	2.90	491.2	2.07	6.01	0.88	3.57	-3.7	0.02	-4.25
1	0	0	0	1	0	0	0	2	2.13	385.6	2.85	5.04	-0.05	4.26	-4.4	0.01	-4.24
1	1	0	0	2	0	0	0	3	3.34	507.6	2.17	6.98	0.73	3.88	-4.1	0.02	-4.28
2	0	0	0	1	0	0	0	2	2.57	412.3	3.03	6.10	-0.02	4.84	-5.0	0.007	-4.31
2	0	0	0	2	0	0	0	2	3.02	497.6	2.07	6.01	0.88	3.57	-3.7	0.02	-4.25
3	0	0	0	2	0	0	0	2	3.47	513.5	2.17	6.97	0.73	3.88	-4.1	0.02	-4.28

'E': Ester (vinyl laurate); 'W': Water; 'S': Solvent, 'l': Inulin.

$\ln x_i^{\text{id}}(313.15 \text{ K}) = -0.17$  (for  $i = \text{l}$ ).

T<sub>b</sub> is the first-order result obtained with the method of Constantinou/Gani [55].

**Table A.3**

Solvent candidates for vinyl laurate/inulin transesterification; specification 2 (as in Table 6) (T = 313.15 K)

										Constraint Values							
Basis Set 1										4	360	1	1	2	2		0
CH <sub>2</sub>	CH	C	OH	CH <sub>3</sub> CO	CH <sub>2</sub> CO	CH <sub>3</sub> COO	CH <sub>2</sub> COO	CH <sub>3</sub>	-log(LC <sub>50</sub> )	T <sub>p</sub> /K	lnγ <sub>WS</sub> <sup>∞</sup>	lnγ <sub>SW</sub> <sup>∞</sup>	lnγ <sub>ES</sub> <sup>∞</sup>	lnγ <sub>IS</sub> <sup>∞</sup>	lnx <sub>i</sub>	x <sub>i</sub>	S
0	0	1	1	0	0	0	0	3	1.45	371.9	1.33	3.85	1.91	1.49	-1.7	0.2	-3.64
0	2	0	1	0	0	0	0	3	1.94	400.3	1.49	4.90	1.59	1.87	-2.0	0.1	-3.69
1	0	1	1	0	0	0	0	3	1.90	400.4	1.46	4.83	1.64	1.93	-2.1	0.1	-3.67
1	1	0	1	0	0	0	0	2	1.62	383.1	1.36	3.92	1.85	1.41	-1.6	0.2	-3.66
2	1	0	1	0	0	0	0	2	2.07	410.2	1.49	4.90	1.59	1.87	-2.0	0.1	-3.69
3	0	0	1	0	0	0	0	1	1.74	393.9	1.36	3.92	1.85	1.41	-1.6	0.2	-3.66
4	0	0	1	0	0	0	0	1	2.19	419.6	1.48	4.90	1.60	1.87	-2.0	0.1	-3.69
										Constraint Values							
Basis Set 2										4	360	1	1	2	2		0
CH <sub>2</sub>	CH	C	CH <sub>3</sub> O	CH <sub>2</sub> O	CH-O	CH <sub>2</sub> NH <sub>2</sub>	CHNH <sub>2</sub>	CH <sub>3</sub>	-log(LC <sub>50</sub> )	T <sub>p</sub> /K	lnγ <sub>WS</sub> <sup>∞</sup>	lnγ <sub>SW</sub> <sup>∞</sup>	lnγ <sub>ES</sub> <sup>∞</sup>	lnγ <sub>IS</sub> <sup>∞</sup>	lnx <sub>i</sub>	x <sub>i</sub>	S
0	0	1	0	0	0	0	1	4	2.95	380.7	1.11	5.72	-0.18	1.79	-2.0	0.1	-3.16
1	1	0	0	0	0	0	1	3	3.11	391.5	1.13	5.80	-0.20	1.74	-1.9	0.1	-3.17
3	0	0	0	0	0	0	1	2	3.24	401.8	1.13	5.80	-0.20	1.74	-1.9	0.1	-3.17
										Constraint Values							
Basis Set 3										4	360	1	1	2	2		0
CH <sub>2</sub>	CH	C	CHO	CH <sub>2</sub> NH	CHNH	CH <sub>2</sub> CN	DMF-2	CH <sub>3</sub>	-log(LC <sub>50</sub> )	T <sub>p</sub> /K	lnγ <sub>WS</sub> <sup>∞</sup>	lnγ <sub>SW</sub> <sup>∞</sup>	lnγ <sub>ES</sub> <sup>∞</sup>	lnγ <sub>IS</sub> <sup>∞</sup>	lnx <sub>i</sub>	x <sub>i</sub>	S
0	0	0	0	0	0	1	0	1	1.71	364.2	2.48	3.55	0.68	1.45	-1.6	0.2	-4.06

'E': Ester (vinyl laurate); 'W': Water; 'S': Solvent, 'I': Inulin.

$\ln x_i^{\text{id}}$  (313.15 K) = -0.17 (for i = 'I').

T<sub>b</sub> is the first-order result obtained with the method of Constantinou/Gani [55].

## Tables

Set	$n_{\max}$	$n_{\min}$	Groups	Type
1	7	2	-CH <sub>3</sub> , -CH <sub>2</sub> -, >CH-, >C<, -OH, -CH <sub>2</sub> COO-, CH <sub>3</sub> COO-, CH <sub>3</sub> CO-, -CH <sub>2</sub> CO-	Acyclics
2	7	2	-CH <sub>3</sub> , -CH <sub>2</sub> -, >CH-, >C<, >CH-O-, -CH <sub>2</sub> -O-, CH <sub>3</sub> -O-, -CH <sub>2</sub> NH <sub>2</sub> , >CHNH <sub>2</sub>	Acyclics
3	7	2	-CH <sub>3</sub> , -CH <sub>2</sub> -, >CH-, >C<, -CHO, -CH <sub>2</sub> NH-, >CHNH-, DMF-2, -CH <sub>2</sub> CN	Acyclics
4	7	2	-CH <sub>3</sub> , -CH <sub>2</sub> -, >CH-, >C<, -OH, -CH <sub>2</sub> CO-, CH <sub>3</sub> O-, CH <sub>3</sub> N<, -CH <sub>2</sub> N<	Acyclics
5	1	1	CH <sub>3</sub> OH, CH <sub>3</sub> NH <sub>2</sub> , CH <sub>3</sub> CN, CH <sub>2</sub> Cl <sub>2</sub> , CHCl <sub>3</sub> , CCl <sub>4</sub> , CS <sub>2</sub> , Furfural, DMSO, DMF	Molecular Groups

Basis set	1,3 ,4 , 5	2
$\delta_C$	-0.415	-0.415
$\delta_{VL,S}$	2	2
$\delta_{C8OH,S}$	2	2
$\delta_{W,S}$	2	3
$\delta_{S,W}$	2	3
$\delta_{LC}$	3.5	3.5
$\delta_{b,S}$	350 K	350 K

Basis Set	Candidates
1	13
2	12
3	0
4	18 (all repetitions from Set 1/Set 2)
5	0



**Table 4**

Active constraints in for group sets resembling real substances.

Resembling substance	Active constraint(s)
1-propanol, DMF, furfural	iii), ii)
<i>tert</i> -butanol, 1-pentanol, hexylamine	i), iii)
butyl acetate	iv)
acetonitrile, propionitrile, ethyl acetate	i)
methylamine, propionic aldehyde	v), iii), i)
methanol	v), iii)
ethylene glycol, DMSO	i), ii), iii)
alkanes, MTBE , dichloromethane, chloroform, CS <sub>2</sub>	v)

**Table 5**Converted fraction of vinyl laurate by lipase B from *Candida Antarctica* incubated with octanol in different solvents.

Solvent	Converted fraction of vinyl laurate after 22 hours	S Eq. (8 )
Hexane	0.92	-1.33
Isooctane	1.00	-1.37
Toluene	0.96	-1.27
2-Methyl-2-butanol	0.80	0.270
<i>tert</i> -butanol	0.69	0.302
2-Butanone	0.88	-0.191
Acetone	0.95	-0.128
Pyridine <sup>a</sup>	0.57	-0.159
Dimethylsulfoxide (DMSO)	n.d.	1.64

a: Substantial activity observed – even after 22 hours

n.d.: no product detectable

**Table 6**  
Constraints values for different basis sets for the transesterification of inulin/vinyl laurate.

Basis set	1	2-4	1-4
$\delta_C$	-3	-3	0
$\delta_{VL}$	2	2	2
$\delta_I$	5	5	2
$\delta_{w,s}$	3	2	1
$\delta_{s,w}$	3	2	1
$\delta_{LC}$	3.5	3.5	4
$\delta_{b,s}$	360 K	360 K	360 K
Specification	1		2

**Table 7**  
Melting Property Data and Group Assignments of Inulin (and decorated derivative).

Species	n	$M_w$ g/mol	$T_m$ K	$\Delta H_m$ J/g	$\Delta H_m$ J/mol	Group Assignment							
						CH <sub>3</sub>	CH <sub>2</sub>	CH	C	OH	CH <sub>2</sub> COO	CH <sub>2</sub> O	CH-O
Inulin	10	1963.7	451.2	9	17673.6	0	13	25	11	38	0	10	13
<sup>i</sup> I-OH	1	162.1	451.2	9	1459.3	0	1	2	1	3	0	1	1
<sup>i</sup> I-OL	1	344.5				1	10	2	1	2	1	1	1
<sup>ii</sup> t <sub>u</sub>	1	163.2					2	2	1	4			1
<sup>ii</sup> t <sub>l</sub>	1	179.2					1	3		4			2

<sup>i</sup>I-OH and I-OL denote species in the reaction: VL(1) + I-OH(2)  $\rightleftharpoons$  I-OL(3) + CH<sub>3</sub>CHO(4)

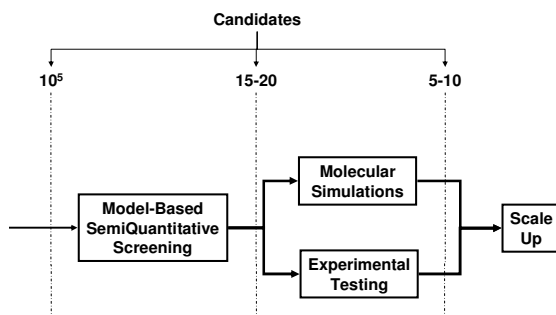
<sup>ii</sup>Upper and lower terminal units of inulin.

**Table 8**  
Quantities for solvents for lauryl esterification of inulin (T = 313.15 K).

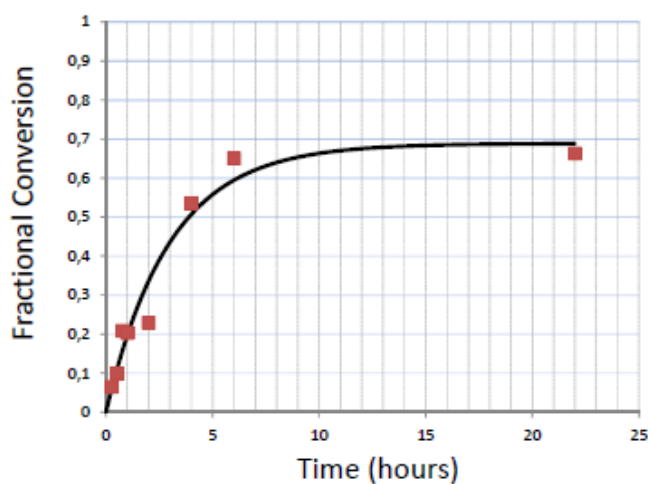
Solvent, S.	$T_{bS}/K$	$\ln \gamma_{ws}^\infty$	$\ln \gamma_{sw}^\infty$	$\ln \gamma_{is}^\infty$	$\ln \gamma_{2s}^\infty$	$\ln \gamma_{3s}^\infty$	$\ln \gamma_{4s}^\infty$	S
<i>tert</i> -butanol	371.9	1.33	3.85	1.91	1.49	-0.31	0.07	-3.64
CH <sub>3</sub> OH	337.8	0.49	0.82	2.71	-1.58	-2.49	-1.14	-4.76
CH <sub>3</sub> NH <sub>2</sub>	266.8	-0.78	-1.91	0.06	-4.83	-6.88	-0.47	-2.57
CH <sub>3</sub> CN	354.8	2.06	2.52	2.24	0.32	-1.26	-0.19	-4.01
CH <sub>2</sub> Cl <sub>2</sub>	313.0	4.71	5.37	-1.60	4.18	-2.32	0.02	-4.89
CHCl <sub>3</sub>	334.3	5.21	6.60	-1.72	4.49	-2.39	-3.59	-8.74
CCl <sub>4</sub>	349.8	6.86	9.00	-0.91	9.92	2.65	-1.05	-7.41
CS <sub>2</sub>	319.2	7.07	7.96	-0.12	8.53	2.07	-0.68	-7.01
Furfural	434.7	1.46	3.92	2.29	3.38	0.50	-0.43	-5.61
1,2-Ethandiol	470.5	-0.19	-0.21	5.79	-4.28	0.14	0.22	-1.16
DMSO	464.0	-1.40	-2.59	4.33	-4.67	-2.45	-0.19	-2.30
DMF	426.0	-0.04	-0.06	2.37	-2.65	-2.97	-1.73	-4.42

Indices on activity coefficients as in VL(1) + I-OH(2)  $\rightleftharpoons$  I-OL(3) + CH<sub>3</sub>CHO(4)

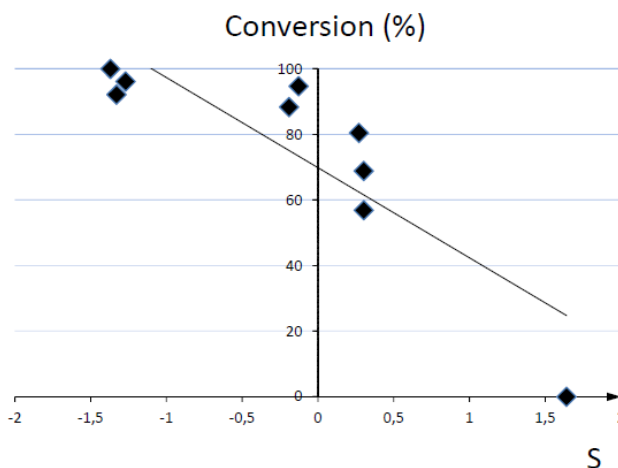
## Figures



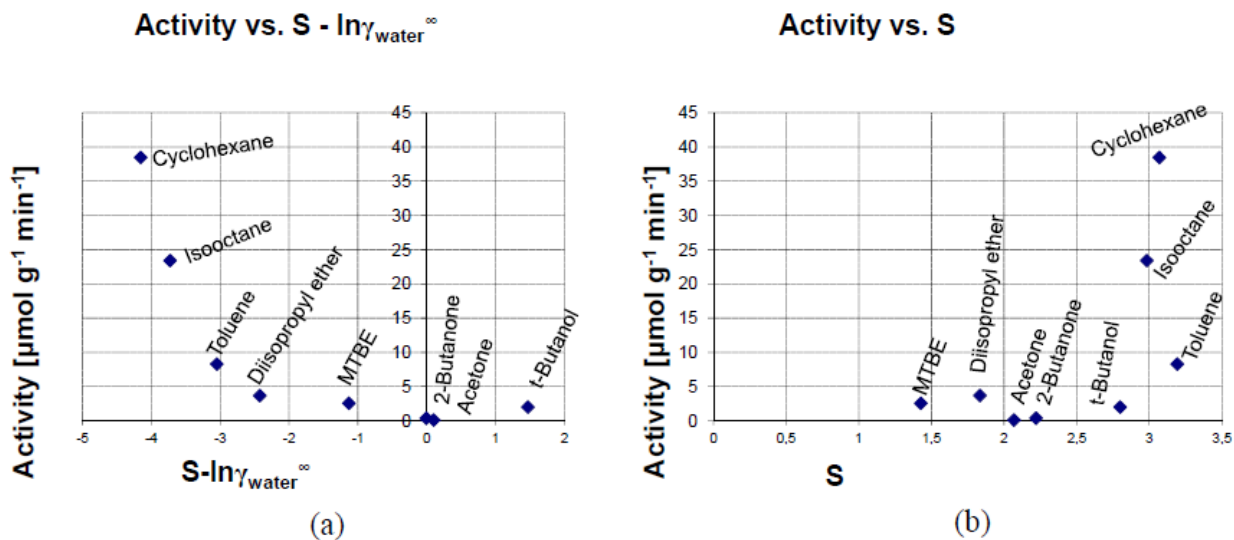
**Fig. 1.** Framework for enzymatic reactions in suitable solvents or mixtures.



**Fig. 2.** Transesterification of octanol with vinyl laurate in *tert*-butanol using lipase B from *Candida antarctica*. Symbols denotes measurements. The curve is from a fit to data.



**Fig. 3.** Transesterification of octanol with vinyl laurate using lipase B from *Candida antarctica*. %-conversion of vinyl laurate versus S calculated from eq. (8).



**Fig. 4.** Esterification activities of lipase from *Candida antarctica* using acryl acid/octanol in different media (with (a) and without (b) water removal). Activity data are from [44].  $S$  is calculated from eq. (8).

## References

---

- [1] a) A.M. Klivanov, *Biotechnol. Bioeng.* 19 (1977) 1351-1361; b) A.M. Klivanov, *Trends Biochem. Sci.* 14 (1989) 141; c) A.M. Klivanov, *Acc. Chem. Res.* 23 (1990) 114.
- [2] J.A. Akkara, M.S.R. Ayyagari, F.F. Bruno, *Trends Biotech.* 17 (1999) 67-73.
- [3] M. Perez, J.V. Sinisterra, M.J. Hernaiz, *Curr. Org. Chem.* 20 (2011) 2366-2383.
- [4] S.W. Chang, J.F. Shaw, *New Biotech.* 26 (2009) 109-116.
- [5] B.G. Davis, V. Borer, *Natural Product Reports* 18(6) 618-640.
- [6] F. Bordusa, *Chemical Rev.* 102 (2002) 4817-4867.
- [7] C.R. Wescott, A.M. Klivanov, *Biochim. Biophys. Acta* 1206 (1994) 1-9.
- [8] P.J. Halling, *Enzyme Microb. Technol.*, 16 (1994) 178-206.
- [9] C. Laane, S. Boeren, K. Vos, C. Veeger, *Biotech. Bioeng.* 30 (1987) 81-87.
- [10] F.J. Plou, M.A. Cruces, M. Ferrer, G. Fuentes, E. Pastor, M. Bernabe, M. Christensen, F. Comelles, J.L. Parra, A. Ballesteros, *J. Biotech.* 96 (2002) 55-66.
- [11] a) C.R. Wescott, A.M. Klivanov, *J. Am. Chem. Soc.*, 115 (1993) 10362-10363. ; b) T. Ke, C.R. Wescott, A.M. Klivanov, *J. Am. Chem. Soc.*, 118 (1996) 3366-3374.
- [12] P.J. Halling, *Biotechnol. Bioeng.*, 35 (1990) 691-701.
- [13] B. Zeuner, G.M. Kontogeorgis, A. Riisager, A.S. Meyer, *New Biotechnology*, 29 (2012) 255-270.
- [14] M.L. Damstrup, J. Abildskov, S. Kiil, A.D. Jensen, F.V. Sparsø, X. Xu, *J. Agric. Food Chem.*, 54 (2006) 7113-7119.
- [15] G. Carrea, S. Riva, *Angewandte Chemie-International Edition* 39(2000) 2226-2254.
- [16] U.T. Bornscheuer, R.J. Kazlauskas, 2005. 'Hydrolases in Organic Synthesis, Regio- and stereoselective biotransformations', 2nd ed., Wiley-VCH.
- [17] T.H. Rod, U. Ryde, Quantum mechanical free energy barrier for an enzymatic reaction. *Phys. Rev. Lett.* 2005; 94:138302-1-138302-4.
- [18] a) K.C. Satyanarayana, R. Gani, J. Abildskov, *Comp. Chem. Eng.*, 33 (2009) 1004; b) K.C. Satyanarayana, R. Gani, J. Abildskov, *Fluid Phase Equilibria*, 261 (2007): 58-63; c) K.C. Satyanarayana, J. Abildskov, R. Gani, G. Tsolou, V.G. Mavrantzas, *Brazilian J. Chem. Eng.*, 27 (2010): 369-380.
- [19] R. Vaidyanathan, M. El-Halwagi, *Ind. Eng. Chem. Res.* 35 (1996) 627.
- [20] C.D. Maranas, *Ind. Eng. Chem. Res.* 35 (1996) 3403.
- [21] K.G. Joback, G. Stephanopoulos, *Adv. Chem. Eng.*, 21 (1995) 257.
- [22] a) N. Churi, L.E.K. Achenie, *Ind. Eng. Chem. Res.* 35 (1996) 3788; b) A.P. Duvedi, L.E.K. Achenie, *Chem. Eng. Sci.* 51 (1996) 3727.
- [23] a) J. Abildskov, J.P. O'Connell, *J. Chem. Eng. Data*, 56 (2011): 1229-1237; b) I. Kouskoumvekaki, J. Abildskov, *Chem. Eng. Research & Design (Trans. IChemE)*, 84 (2006): 652-663.
- [24] R. Gani, B. Nielsen, A. Fredenslund, *AIChE J.* 37 (1991) 1318.
- [25] R. Gani, P.A. Gómez, M. Folić, C. Jiménez-González, D.J.C. Constable, *Comp. Chem. Eng.*, 32 (2008) 2420-2444.
- [26] P.-Y. Kim, D.J. Pollard, J.M. Woodley, *Biotechnol. Prog.* 23 (2007) 74-82.
- [27] L. Cao, L. van Langen, R.A. Sheldon, *Curr. Opin. Biotech.*, 14 (2003) 387-394.
- [28] R. Ter Haar, H.A. Schols, L.A.M. Van den Broek, D. Sağlam, A.E. Frissen, C.G. Boeriu, H. Gruppen, *J Mol Catal B: Enzym* 62 (2010) 183-189.
- [29] L.E.S. Brink, J. Tramper, *Biotechnol. Bioeng.*, 27 (1985) 1258.

- 
- [30] a) L.J. Bruce, A.J. Daugulis, *Biotechnol. Prog.*, 7 (1991) 116; b) T.B. Janikowski, D. Velicogna, M. Punt, A.J. Daugulis. *Appl. Microbiol. Biotechnol.*, 59 (2002) 368; c) C.T. MacLeod, A.J. Daugulis, *Appl. Microbiol. Biotechnol.*, 62 (2003) 291; d) G.P. Prpich, A.J. Daugulis, *Biotechnol. Bioeng.*, 97 (2007) 536.
- [31] W Y. Wang, L.E.K. Achenie, *Fluid Phase Equilibria*, 201 (2002) 1–18.
- [32] N.V. Sahinidis, M. Tawarmalani, M. Yu, *AIChE J.*, 49(7) (2003) 1761-1775.
- [33] D.K. Eggers, H.W. Blanch, J.M. Prausnitz, *Enzyme and Microbial Technology*, 11(2) (1989) 84-89.
- [34] R.H. Valivety, G.A. Johnston, C.J. Suckling, P.J. Halling, *Biotechnol. Bioeng.*, 38(10) (1991) 1137-1143.
- [35] J.M. Woodley, M. Bisschops, A.J.J. Straathof, M. Ottens, *J. Chem. Tech. Biotechnol.*, 83 (2008) 121-123.
- [36] A. Zaks, A.M. Klivanov, *Proc. Natl. Acad. Sci., USA*, 82 (1985) 3192-3196.
- [37] a) R.Gani, E.A.Brignole, *Fluid Phase Equilibria* 13 (1983) 331; b) E.A.Brignole, S.Bottini, R.Gani, *Fluid Phase Equilibria* 29 (1986) 125.
- [38] J.M. Smith, M.M. Abbott, H.C. Van Ness, 1996. "Introduction to Chemical Engineering Thermodynamics", 4<sup>th</sup> ed., McGraw-Hill.
- [39] H. Stamatis, E.C. Voutsas, C. Delimitsou, F.N. Kolisis, D. Tassios, *Biocatal. Biotransfor.*, 18 (2000) 259-269.
- [40] X.M. Shen, T.W. de Loos, J. de Swaan Arons, 1992. "Enzymatic Reaction in Organic Solvents and Supercritical Gases", *Biocatalysis in Non-conventional Media* (eds. Tramper et al.), 8: 417-423.
- [41] L.M.C. Sagis, C.G. Boeriu, G.E. Frissen, H.A. Schols, P.A. Wierenga, *Langmuir*, 24 (2008) 359-361.
- [42] A. Zaks, A.M. Klivanov, *J. Biol. Chem.*, 263(17) (1988) 8017-8021.
- [43] D.A. Robb, Z. Yang, P.J. Halling, *Biocat. Biotransf.*, 9(1) (1994) 277-283.
- [44] M. Nordblad, P. Adlercreutz, *J. Biotechnol.*, 133 (2008) 127–133.
- [45] T.M. Martin, D.M. Young, *Chem. Res. Toxicol.*, 14 (2001) 1378-1385.
- [46] R. Butler, G.M. Cooke, G.G. Lukk, B.G. Jameson, *Ind. Eng. Chem.*, 48 (1956) 808.
- [47] O. Odele, S. Macchietto, *Fluid Phase Equilibria*, 82 (1993) 47-54.
- [48] H.K. Hansen, P. Rasmussen, A. Fredenslund, M. Schiller, J. Gmehling, *Ind. Eng. Chem. Res.*, 30 (1991) 2352.
- [49] H.V. Kehiaian, *Fluid Phase Equilibria*, 13 (1983) 243.
- [50] J. Abildskov, R. Gani, P. Rasmussen, J.P. O'Connell, *Fluid Phase Equilibria*, 158 (1999) 349.
- [51] a) H.S. Wu, S.I. Sandler, *Ind. Eng. Chem. Res.*, 30 (1991) 881; b) H.S. Wu, S.I. Sandler, *Ind. Eng. Chem. Res.*, 30 (1991) 889.
- [52] R.P. Currier, J.P. O'Connell, *Fluid Phase Eq.*, 33 (1987) 245-265.
- [53] K. Balslev, J. Abildskov, *Ind. Eng. Chem. Res.*, 41 (2002) 2047.
- [54] Wittig, R.; Lohmann, J.; Gmehling, J. Vapor-liquid equilibria by UNIFAC group contribution. 6. Revision and extension, *Ind. Eng. Chem. Res.*, 42 (2003) 183.
- [55] L. Constantinou, R. Gani, *AIChE J.*, 40 (10) (1994) 1697.
- [56] P. Adlercreutz, *Eur. J. Biochem.*, 199 (1991) 609.
- [57] N.M. Micaêlo, C.M. Soares, *FEBS J.*, 274 (2007) 2424-2436.
- [58] R. Wedberg, J. Abildskov, G.H. Peters, *J. Phys. Chem. B*, 116 (2012) 2575–2585.

- 
- [59] R.V. Ulijn, L. De Martin, L. Gardossi, A.E.M. Janssen, B.D. Moore, P.J. Halling. *Biotechnol. Bioeng.*, 80(5) (2002) 509-515.
- [60] M.D. Ellegaard, J. Abildskov, J.P. O'Connell, *Ind. Eng. Chem. Res.*, 49 (2010) 11620–11632.
- [61] C. Blecker, J.-P. Chevalier, C. Fognies, J.-C. Van Herck, C. Deroann, M. Paquot, J. *Therm. Anal. Calorimetry*, 71 (2003) 215–224.
- [62] C.L.M. Hébette, J.A. Delcour, M.H.J. Koch, K. Booten, R. Kleppinger, N. Mischenkod, H. Reynaers, *Carbohydr. Res.*, 310 (1998) 65-75.
- [63] a) J.S. Rowlinson, F.L. Swinton, 1982. *Liquids and Liquid Mixtures*, 3rd edn., Butterworth Scientific, London, Chap. 5.; b) M.M. Abbott, M.V. Ariyapadi, N. Balsara, S. Dasgupta, J.S. Furno, P. Futerko, D.P. Gapinski, T.A. Grocela, R.D. Kaminsky, S.G. Karlsruhe, E.W. Kiewra, A.S. Narayan, K.K. Nass, J.P. O'Connell, C.J. Parks, D.F. Rogowski, G.S. Roth, M.B. Sarsfield, K.M. Smith, M. Sujanani, J.J. Tee, N. Tzouvaras, *Chemical Engineering Education*, Winter 1994, 18-24.
- [64] a) W. Rupp, S. Hetzel, I. Ojini, D. Tassios, *Ind. Eng. Chem. Process. Des. Dev.*, 23 (1984): 391; b) P.J. Stathis, D.P. Tassios, *Ind. Eng. Chem. Process Des. Dev.*, 24 (1985): 701.
- [65] A. Bondi, *Physical Properties of Molecular Crystals, Liquids and Glasses*, John Wiley & Sons, New York, 1968.
- [66] a) U. Weidlich, J. Gmehling, *Ind. Eng. Chem. Res.* 26 (1987) 1372–1381; b) J. Gmehling, J. Li, M. Schiller, *Ind. Eng. Chem. Res.* 32 (1993) 178–193.
- [67] J. Abildskov, R. Gani, P. Rasmussen, J.P. O'Connell, *Fluid Phase Equilibria*, 181 (2001): 163-86.
- [68] a) S. Christensen, G.H. Peters, F.Y. Hansen, J.P. O'Connell, J. Abildskov, *Molecular Simulation*, 33 (2007) : 449-457; b) S. Christensen, G.H. Peters, F.Y. Hansen, J.P. O'Connell, J. Abildskov, *Fluid Phase Equilibria*, 260 (2007): 169-176; c) S. Christensen, G.H. Peters, F.Y. Hansen, J. Abildskov, *Fluid Phase Equilibria*, 261 (2007): 185-190.
- [69] A. Jakob, H. Gremse, J. Lohmann, J. Gmehling, *Ind. Eng. Chem. Res.*, 45 (2006) 7924-7933.
- [70] a) H.E. Gonzales, B. Le Bert, J. Abildskov, R. Gani, *AIChE Journal*, 53 (2007): 1620-1632; b) H.E. Gonzales, J. Abildskov, R. Gani, *Fluid Phase Equilibria*, 261 (2007): 199-204.