Predicting when Precipitation-Driven Synthesis Is Feasible: Application to Biocatalysis

Rein V. Ulijn,[a] Anja E. M. Janssen,[b] Barry D. Moore,*[a] and Peter J. Halling[a]

Abstract: Precipitation-driven synthesis offers the possibility of obtaining high reaction yields using very low volume reactors and is finding increasing applications in biocatalysis. Here, a model that allows straightforward prediction of when such a precipitation-driven reaction will be thermodynamically feasible is presented. This requires comparison of the equilibrium constant, K_{eq}, with the saturated mass action ratio, Z_{sat}, defined as the ratio of product solubilities to reactant solubilities. A hypothetical thermodynamic cycle that can be used to accurately predict Z_{sat} in water is described. The cycle involves three main processes: fusion of a solid to a supercooled liquid, ideal mixing of the liquid with octanol, and partitioning from octanol to water. To obtain the saturated mass action ratio using this cycle, only the melting points of the reactants and products, and in certain cases the pK_a of ionisable groups, are required as input parameters. The model was tested on a range of enzyme-catalysed peptide syntheses from the literature and found to predict accurately when precipitation-driven reaction was possible. The methodology employed is quite general and the model is therefore expected to be applicable to a wide range of other (bio)-catalysed reactions.

Keywords: enzyme catalysis · peptides · solubility · thermodynamics

Introduction

In vivo and, commonly, in vitro enzyme catalysis involves dilute (mmolar) reactant concentrations. Indeed, during biocatalysed syntheses of pharmaceutical intermediates this may be unavoidable due to the poor aqueous solubility of the molecular building blocks. However, for biocatalysis to compete commercially with conventional chemistry, higher volumetric productivity is generally desirable. One common means of achieving this is to use enzymes in non-aqueous media, in which reactant solubilities are often higher.[1, 2] A more recent development is the application of biocatalysis in systems in which the initial reaction mixture consists mainly of solid reactant suspended in a low-volume, saturated liquid phase(aqueous or organic solvent) and in which the final mixture consists of mainly solid product.[3, 4] Such "solid to solid" reactions can be catalysed surprisingly efficiently by enzymes, resulting in very high volumetric productivity. Examples of biocatalysis in substrate suspensions include a range of reverse hydrolysis reactions (condensations) leading to amide,[5, 6, 7] glycoside[8] and ester[9, 10] formation.

A key feature of these predominantly solid reactant systems is that the reaction thermodynamics exhibit switch-like behaviour, giving either very low or very high yields.[11] Thus, if the product concentration required to attain equilibrium with saturated reactants is lower than the solubility limit, it simply accumulates in the low-volume liquid phase up to the equilibrium level, no further reaction occurs, and the overall yield is low. Alternatively, if the (theoretical) product concentration required to attain equilibrium with saturated reactants is higher than the solubility limit, the product will reach saturation before equilibrium is reached and start to precipitate out of the reaction mixture. This process will continue until the concentration of one of the reactant falls below saturation. Only then can equilibrium in the liquid phase be established. With only a small liquid phase present, the yield will necessarily be very high: the solid reactants in the initial reaction mixture are nearly all converted to solid product.

In a recent comparative study of enzyme-catalysed peptide synthesis, either in organic solvents or in solid to solid aqueous suspensions,[12] it was shown that the latter system gave both higher reaction rates and higher conversions. In particular, volumetric yields were very much better, reaching 0.8 g product per mL reaction mixture.[5] Given the rapid improvement in screening and evolution of new enzyme activities,

[a] B. D. Moore, R. V. Ulijn, P. J. Halling
Department of Pure and Applied Chemistry
Thomas Graham Building, 295 Cathedral Street
Glasgow G1 1XL (UK)
Fax: (+44) 141-548-4822
E-mail: b.d.moore@strath.ac.uk

[b] A. E. M. Janssen
Department of Food Technology and Nutritional Sciences
Food and Bioprocess Engineering Group
Wageningen University, P.O. Box 8129
6700 EV Wageningen (The Netherlands)

  0947-6539/01/0710-2089 $ 17.50+.50/0

product precipitation-driven biocatalysis offers a highly promising route for commercial synthetic chemistry.

A significant problem has been that until now it has not been possible to predict which type of "switch behaviour", low yield or high yield, a particular reaction would exhibit. While in simple cases it is possible to find this out experimentally, it is often the case that one of the reaction components (including the relevant enzyme) is not readily available. Before extensive efforts are expended on, for example, synthesising reactants or screening and isolating enzymes, it would be extremely useful to be able to predict whether or not the process will work. Herein we present a theoretical model that enables the thermodynamic feasibility of product precipitation-driven synthesis to be predicted reliably. In systems in which the aqueous reaction equilibrium constant is known or can be estimated, the only input parameters required are the melting points of the reactants and products and, in certain cases, their pK_a values.

The initial part of the following discussion deals with development of the general theoretical model. Application of the model to the specific example of precipitation-driven amide synthesis is then considered. Finally, we present an extensive comparison of theoretical predictions with the experimental results obtained from the biocatalysis literature.

Theoretical Model

The theory is presented in two parts; firstly we consider the important general features of a product precipitation-driven reaction, secondly we explain in detail how the feasibility of such reactions may be predicted using a hypothetical thermodynamic cycle.

When is a product precipitation-driven reaction favoured?

If we consider a simple thermodynamically controlled reaction such as that given in Equation (1) the thermodynamic equilibrium constant K_{th} for this reaction can be expressed in terms of thermodynamic activities a_x as given in Equation (2).

$$A+B \rightleftharpoons C \quad (1)$$

$$K_{th}=\frac{a_c}{a_A a_B} \quad (2)$$

If the reactants are present at dilute concentrations (even at saturation), then the equilibrium constant K_{eq} can be reasonably accurately formulated as given in Equation (3).

$$K_{eq}=\frac{[C]}{[A][B]} \quad (3)$$

Let us now consider the case in which initially the liquid phase of the reaction mixture is saturated with both reactants and, for each, a large excess of solid is also present. We see that, as the reaction proceeds, both the concentrations of A and B will remain constant at their limiting solubility values S_A and S_B, while product C will begin to accumulate in the liquid phase. Two different situations can then arise.

1) If [C] is able to reach its equilibrium value as defined by Equation (3), the reaction will proceed no further and most of A and B will still be present as a solid in the reaction mixture. Thus, in this case the overall reaction yield will be very low.
2) Alternatively, if before reaching the equilibrium concentration defined by Equation (3), C reaches its solubility limit, it will begin to precipitate from solution. Solid C will then continue to precipitate until all of the solid form of reactant A or B has dissolved. The concentration of this reactant will then fall below the solubility limit until the equilibrium constant defined by Equation (3) can be reached. In this case the overall yield is necessarily very high, since the reaction proceeds until all the excess solid of one (or both) of the reactants is used up.

To predict which of these two situations will arise, it is necessary to know the value of the saturated mass action ratio Z_{sat}. This is obtained by substituting the limiting solubility values of each of the reactants and products into the equilibrium constant expression (3) to give Equation (4).

$$Z_{sat}=\frac{S_c}{S_A S_B} \quad (4)$$

Comparing Equations (3) and (4), it can be seen that when $Z_{sat} > K_{eq}$, product C is able to reach its equilibrium value in the liquid phase corresponding to case 1) above. In contrast, when $Z_{sat} < K_{eq}$, the solubility limit of C is reached before equilibrium is established and a solid to solid reaction proceeds as described in 2). Figure 1 illustrates this latter situation diagrammatically.

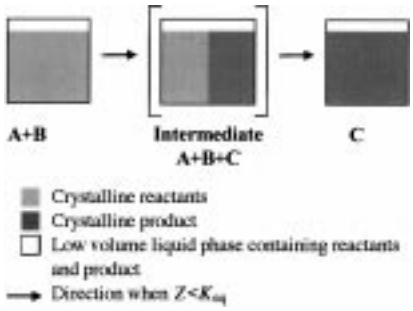

Figure 1. Schematic representation of a biocatalytic reaction in suspension. Arrows indicate the direction of product precipitation-driven catalysis when Z_{sat} is smaller than K_{eq}. If Z_{sat} were greater than K_{eq}, the arrows would be reversed. The intermediate state will always be thermodynamically unstable with respect to near complete conversion in one direction or other.

To predict whether or not a particular reaction can be precipitation-driven, it is therefore necessary to determine values for K_{eq} and Z_{sat}. Experimental or theoretical data allowing K_{eq} to be estimated often exist, but theoretical methods for calculation of Z_{sat} have not hitherto been reported.

Predicting the feasibility of product precipitation reactions

The thermodynamic feasibility of a product precipitation-driven reaction is expected to be largely independent of the solvent the reaction is carried out in.[3, 11] Therefore, since equilibrium constants for biocatalysed reactions are most often measured in water, our aim was to express the mass

0947-6539/01/0710-2090 $ 17.50+.50/0 *Chem. Eur. J.* **2001**, *7*, No. 10

action ratio (Z_{sat}) in terms of aqueous limiting solubility values. The approach taken was to relate solubilities to melting points by means of a hypothetical thermodynamic cycle.

The relationship between melting points and solubilities: The starting point for the theoretical model is the work of Grant and Higuchi[13] and Yalkowsky.[14] For further details of this type of treatment the reader should consult their work.

When a compound is at its solubility limit in a solvent, the process of dissolution has reached equilibrium. This means that the Gibbs free energy of the molecules in the liquid phase is equal to that of molecules in the solid phase. The free energy of the solid phase largely depends on the packing of individual molecules within the crystal lattice, strong intermolecular forces giving high crystal energies. Since for melting to occur the crystal energy of a solid must be overcome, high crystal energies also result in high melting points. This gives rise to the useful, but not widely recognised, rule that the melting point of a compound provides a good qualitative indication of its solubility.

To obtain more quantitative estimations of aqueous solubility, using melting points, the hypothetical thermodynamic cycle summarised in Figure 2 was devised. Here, the hypothetical indirect route to a saturated aqueous solution involves:

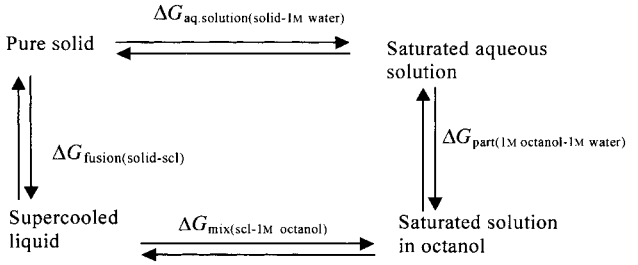

Figure 2. Hypothetical thermodynamic cycle for predicting aqueous solubilities.

1) fusion of a solid compound to a supercooled liquid,
2) ideal mixing of it with octanol to form a saturated solution, and
3) partitioning of it between octanol and water.

To understand how this alternate cycle can be employed for solubility predictions, the concept of different standard states needs to be introduced. It can be seen that the hypothetical route involves a series of reversible molecular processes. Such equilibrium processes are normally associated with a zero change in free Gibbs energy ΔG, because the reference point or standard state is chosen to be the same on both sides of the equilibrium. In this work we have instead deliberately chosen different standard states for the initial and final states of a number of reversible processes. As will be seen, this thermodynamic trick provides a method for quantifying certain useful parameters (such as ideal solubility) that arise within the cycle. It should be noted, however, that if the complete hypothetical thermodynamic cycle is considered, all the changes in standard state (indicated by the subscripts) cancel out and the free energy change is zero as required.

Ideal solubility: free energies of fusion and mixing: Let us begin by considering the first two steps in the above process separately, as shown in Figure 3.

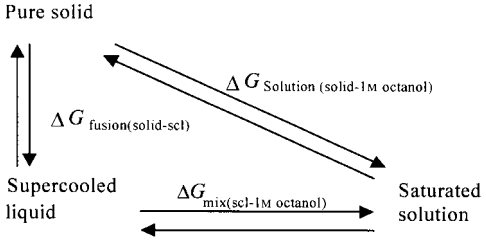

Figure 3. Thermodynamic route for prediction of ideal solubility of organic compounds in octanol.

Together, these result in formation of a saturated ideal solution in octanol. Since the saturated solution will be at equilibrium with the pure solid, the free energy change for the combined processes is zero (same standard state). However, we can see that the two contributing processes involve changes in free energy. These consist of, firstly, fusion of the pure solid to a "supercooled liquid" (scl) (that is, melting taking place below the melting point of the solid) and, secondly, mixing of the liquid melt with the solvent.

This can be expressed in terms of Equation (5) which can be reformulated to give Equation (6), whch in turn provides a method for obtaining the limiting solubility of the solid in octanol.

$$\Delta G_{\text{solution(solid}-1\text{M octanol)}} = \Delta G_{\text{fusion(solid}-\text{scl)}} + \Delta G_{\text{mix(scl}-1\text{M octanol)}} \quad (5)$$

$$\Delta G_{\text{fusion(solid-scl)}} = -\Delta G_{\text{mix(scl}-1\text{M octanol)}} \quad (6)$$

Estimation of Gibbs free energy of fusion: The Gibbs free energy of fusion can be written as given in Equation (7).

$$\Delta G_{\text{fusion}} = \Delta H_{\text{fusion}} - T\Delta S_{\text{fusion}} \quad (7)$$

If molecular melting takes place at the melting point temperature of a compound, the process of melting will be reversible, and hence Equation (8), in which T_m is the melting point, will become valid.

$$\Delta H_{\text{fusion}} = T_m \Delta S_{\text{fusion}} \quad (8)$$

If we assume that ΔS_{fusion} and ΔH_{fusion} are independent of temperature, then at a temperature T, below the melting point, the free energy of fusion from the pure solid to the pure supercooled liquid will be positive, and may be obtained by combining Equations (7) and (8) to give Equation (9).

$$\Delta G_{\text{fusion(solid}-\text{scl)}} = \Delta H_{\text{fusion}} - T\Delta S_{\text{fusion}} = (T_m - T)\Delta S_{\text{fusion}} \quad (9)$$

A method for estimating entropies of fusion, based on molecular flexibility, has previously been developed by

Dannenfelser and Yalkowsky.[15] For an asymmetrical organic molecule, Equation (10) becomes valid.

$$\Delta S_{\text{fusion}} = 50 + R\ln\phi \quad (10)$$

Here ϕ is an empirical molecular flexibility number, related to the number of torsion angles τ within the molecule by the relationship given in Equation (11).

$$\phi = 2.85^{\tau} \quad (11)$$

Evaluation of τ from the molecular structure is quite straightforward and involves a simple count of non-hydrogen atoms. Starting with the structure, the first step is to identify the first heavy atom such as C,N,O situated at the end of every branch. These are not included in the count. Atoms within terminal ring systems are a special case and are included as described below. A count is then made of the total remaining number of sp, sp^2 and sp^3 atoms (all elements), excepting those that are situated in rings. These totals are substituted into Equation (12), together with the total number of ring systems within the molecule. Where rings are fused together, these are all counted together as one ring. As is to be expected, sp atoms do not appear in Equation (12) since they do not contribute any new torsion angles.

$$\tau = \text{SP3} + 0.5\text{SP2} + 0.5\text{RING} - 1 \quad (12)$$

Here SP3 = total number of sp^3 atoms, SP2 = total number of sp^2 atoms and RING = total number of rings.

It should be noted that since an amide bond has pseudo-double bond character, we take the nitrogen to be sp^2-hybridised, as well as the carbonyl group.

As an example, let us consider N-acetylphenylalanine. Firstly, the N-terminal carbon atom of the methyl group and the C-terminal oxygen atom of the hydroxy group are eliminated. This leaves two sp^3 carbons (α, benzyl), two sp^2 carbons (carbonyls), one sp^2 nitrogen (amide) and one ring (phenyl). Substitution of these totals into Equation (12) gives $\tau = 3$. For more examples of the torsion angle numbers of a variety of molecules (see below) and the work by Dannenfelser and Yalkowsky.[15]

Gibbs free energy change of mixing: Organic molecules (particularly those containing hydrophobic groups) will tend to dissolve much more ideally in octanol than water. We have exploited this in the second step of our hypothetical thermodynamic cycle. To obtain an ideal saturated solution, the supercooled liquid is mixed with a solvent in which solubility is ideal. For such ideal mixing, $\Delta H_{\text{mix}} = 0$, while ΔS_{mix} is described for a two-component system by $\Delta S_{\text{mix}} = -R(n\ln\chi_1 + n\ln\chi_2)$, where n represents moles of species and χ represents mole fractions. Because the resulting solution tends to be dilute (mmolar range), the mole fraction of solvent will be close to 1. The Gibbs free energy change of mixing of the pure supercooled liquid to the ideally mixed state on a mole fraction basis is given by Equation (13).

$$\Delta G_{\text{mix(scl-1M octanol)}} = RT\ln\chi = 2.3\text{RT}\log\chi \quad (13)$$

Limiting solubility in octanol: Substituting Equation (9) and Equation (13) into Equation (6) and rearranging, we obtain the following expression [Eq. (14)] for the limiting mole fraction solubility in octanol at reaction temperature T.[16]

$$\log\chi_{\text{ideal}} = \frac{-\Delta S_{\text{fusion}}(T_{\text{m}} - T)}{2.3RT} \quad (14)$$

Since solubilities are generally expressed in terms of molarity rather than mole fraction, it is convenient to make a conversion using Equation (15) (assuming a dilute solution), where V_{solvent} is the molar volume of solvent and S_{ideal} is molar solubility.

$$\chi_{\text{ideal}} = S_{\text{ideal}} V_{\text{solvent}} \quad (15)$$

Altogether, the ideal molar solubility in octanol is then described by Equation (16).

$$\log S_{\text{ideal}} = \frac{-\Delta S_{\text{fusion}}(T_{\text{m}} - T)}{2.3RT} - \log V_{\text{octanol}} \quad (16)$$

From ideal solubility in octanol to aqueous solubility: log *P* values: From the hypothetical thermodynamic cycle shown in Figure 2, we can see that the link between ideal solubility and aqueous solubility can now be made by taking account of partitioning of the compound between octanol and water. This type of octanol to water partitioning is well described by the log P group contribution model. Tabulated log P values have been compiled for many molecular groups, for example by Rekker and De Kort.[17]

Here, partitioning is taken as being from an ideal saturated solution of known concentration, S_{ideal}, in octanol (1 M standard state) to a saturated solution of unknown concentration in water (1 M standard state). Figure 4 shows the thermodynamic cycle.

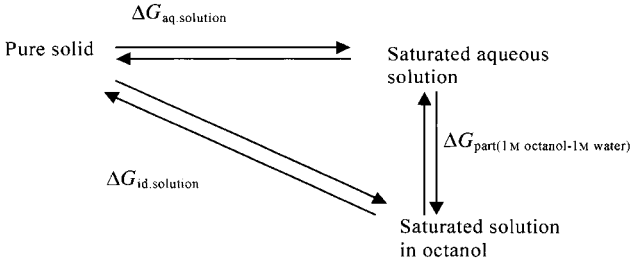

Figure 4. Thermodynamic route for prediction of aqueous solubility.

The free energy change of aqueous solution can now be described by Equation (17).

$$\Delta G_{\text{aq}} = \Delta G_{\text{ideal}} + \Delta G_{\text{part(1M octanol-1M water)}} \quad (17)$$

The standard states of ΔG_{aq} and ΔG_{ideal} are not explicitly defined here, since in practice we need only consider the final partitioning term.

The partitioning term is given by Equation (18), where $\log P = \log(C_{\text{octanol}}/C_{\text{water}})$.

$$\Delta G_{\text{part(1M octanol-1M water)}} = 2.3RT\log P \quad (18)$$

0947-6539/01/0710-2092 $ 17.50+.50/0 *Chem. Eur. J.* **2001**, *7*, No. 10

Importantly, since the solutions are ideally dilute, this ratio will also hold at the solubility limit so that Equation (19) is valid.

$$\log P = \log (S_{ideal}/S_{aq}) \quad (19)$$

Substituting Equation (16) into Equation (19) and rearranging then gives Equation (20).

$$\log S_{aq} = -\log P - \frac{\Delta S_{fusion}(T_m - T)}{2.3RT} - \log V_{octanol} \quad (20)$$

Equation (20) can be used to estimate the aqueous solubility limit of a compound, starting from the melting point, the molecular flexibility number and tabulated log P values. The individual solubilities predicted are usually accurate within one and always within two orders of magnitude of experimental values (see below). When ratios of solubilities are considered, these inaccuracies largely cancel out, as explained below. Predictions of Z_{sat}, using the model, therefore tend to be quite accurate.

From solubilities to mass action ratios and feasibility predictions: Prediction of the mass action ratio involves straightforward substitution of the reactant and product solubilities into the equilibrium constant expression. Interestingly, since the numerator (product) and denominator (reactants) necessarily exhibit many molecular similarities, some of the small errors introduced by simplifying the above thermodynamic cycle tend to cancel out. For example, with the log P group contributions the only residual difference between the reactants and product will be in the small region in which the covalent bond is formed. Also, the molecular flexibility number of the product will simply be the sum of the molecular flexibility numbers of the substrates with an extra SP2 for the terminal free amino group that is converted into an internal peptide nitrogen atom.

After Z_{sat} has been obtained, the final step simply involves comparison of its magnitude with the equilibrium constant K_{eq} as described above. If $Z_{sat} < K_{eq}$, then precipitation-driven synthesis is thermodynamically possible and should generate high yields. The reaction feasibility then reduces to one of kinetics and consideration of whether a suitable (bio)catalyst is available.

Complicating factors: ionisation effects: In the above treatment, we have assumed that the species A, B and C do not ionise on dissolution in water. In such cases it is safe to assume that the species present in the crystalline solid are the same as those present in solution. In many real systems, however, one or more of the species can ionise in water. In this case the treatment is more complex, because these ionisation processes must be taken into account. For example, a carboxylic acid may exist as a neutral species in the solid phase and be partially or nearly completely ionised in aqueous solution, depending upon the pH. To allow for this it needs to be ensured that Z_{sat} and K_{eq} are self-consistent and calculated using concentrations of species in the same ionisation state. For example, in the case of dipeptide formation from protected neutral amino acids (vide infra), the above thermodynamic cycle would predict the limiting aqueous solubility of the neutral forms of acid and amine, since these occur in the crystalline state. The value of Z_{sat} must then be compared with the equilibrium constant for neutral species: obtained from the overall equilibrium constant using simple ionisation calculations.

Application of the model to amide synthesis

Synthesis of an amide from an amine and an unactivated carboxylic acid is a simple and potentially commercially important reaction. However, in aqueous solution reaction yields are generally expected to be poor unless precipitation-driven synthesis is favoured. The model provides a simple way of assessing when this option is thermodynamically viable.

For the reaction of interest equilibrium (21) applies and the true thermodynamic equilibrium constant K_{th} will be given by Equation (22).

$$RCO_2H + R'NH_2 \rightleftharpoons H_2O + RCONHR' \quad (21)$$

$$K_{th} = \frac{a_{RCONHR'} a_{H_2O}}{a_{RCO_2H} a_{R'NH_2}} \quad (22)$$

However, since in aqueous solution the reactants and amide will tend to be present at dilute concentration, we can assume a_{H_2O} equals 1 and use the concentration-based equilibrium constant K_{eq} derived from Equation (23).

$$K_{eq} = \frac{[RCONHR']}{[RCO_2H]\,[R'NH_2]} \quad (23)$$

Precipitation-driven amide synthesis will be feasible when this is greater than the saturated mass action ratio Z_{sat} given by Equation (24).

$$Z_{sat} = \frac{S_{RCONHR'}}{S_{RCO_2H} S_{R'NH_2}} \quad (24)$$

Solubility predictions: assessing the model: Figure 5 illustrates how aqueous solubilities, predicted using the model,

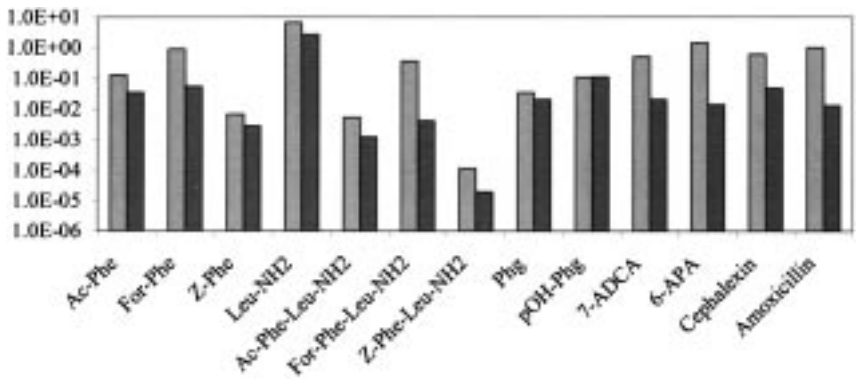

Figure 5. Predicted and measured solubilities on a logarithmic scale. Light bars are predicted values, dark bars are experimental values (see footnote [a] in Table 1 for abbreviations).

compare with data from the literature for a range of neutral dipeptides, ionisable amino acid derivatives and zwitterionic species (beta lactam antibiotics and their building blocks). The experimental solubility data have been reported in terms of non-ionised molecules[11] for the dipeptides and protected amino acids and in terms of the zwitterionic species for antibiotics and their building blocks.[18, 19]

Details of the values used in the calculations are given in Table 1. The melting points required for the solubility predictions were mainly obtained from the Beilstein Crossfire database. In cases where more than one melting point was available the highest (indicating high purity) or most recently reported value was chosen. When selecting melting points for ionisable species, it is important to ensure that values for the free base or free acid are selected and not those of complexes, hydrochlorides or other salts. The number of torsion angles (τ), as introduced in Equation (10), can be calculated using Equations (11) and (12). Finally, log P values were determined by summation of the relevant group contribution values listed by Rekker et al.[11]

From Figure 5 it can be seen that predicted aqueous solubilities are usually underestimated, but show a good correlation with experimental values for the different types of ionisable molecules. This confirms the supposition that the model predicts the aqueous solubility of protected amino acids and dipeptides in their neutral state. It is interesting that for zwitterions the predicted values much more closely resemble the reported solubilities of the zwitterionic species (according to Equation (26) the solubility concentration of the neutral species will be lower by a factor of 10 EΔpK_a). This is consistent with these molecules remaining as zwitterions within the solid crystal lattice.

Calculating the saturated mass action ratio for amide synthesis: We could use Equation (20) to predict the solubility of each component separately prior to substitution in the mass action ratio. However, the process can often be simplified, since on substitution into the mass action ratio many of the log P group contribution terms cancel out. In this case, the residual log P terms will simply be those relating to formation of the amide bond: log P(CONH), log P(CO_2H) and log P(NH_2). These combine to give an overall log P change for amide bond formation which is independent of the reactants used.

Hence, $\Delta\log P_{\text{amide formation}} = 2.446 - 0.938 - 1.420 = 0.088$.

Substituting into Equation (20) and combining all the terms then gives Eqaution (25) , which describes log Z_{sat} for amide synthesis, where $\Delta\log P_{\text{amide formation}}$ and log V_{octanol} are constants with values 0.088 and -0.81.

$$\log Z_{\text{sat}} = \Delta\log P_{\text{amide formation}} + \log V_{\text{octanol}} + \log \chi_{\text{RCONHR'}} - \log \chi_{\text{RCO}_2\text{H}} - \log \chi_{\text{R'NH}_2} \quad (25)$$

Equation (25) will apply, provided that none of the species are zwitterionic. It predicts the saturated mass action ratio of neutral reactants and products and so must be compared to the pH-independent equilibrium constant based on these species.

Dealing with zwitterions: When a zwitterionic species R'' $\pm$ is involved, the model predicts the solubility of the zwitterionic form. To be able to still make comparisons with the pH-independent equilibrium constant, this must be adjusted to give the solubility of the neutral form, R'', by applying Equation (26).

$$[\text{R}''] = \frac{K_a(\text{basic group})}{K_a(\text{acidic group})}[\text{R}''\pm] \quad (26)$$

This gives Equation (27).

$$\log \chi_{\text{R}''} = \log \chi_{\text{R}''\pm} - \Delta\text{p}K_a \quad (27)$$

Hence for each zwitterion in the reaction mixture we need to introduce an additional ΔpK_a term into Equation (25).

Comparison of predicted and experimental values of log Z_{sat}: Table 2 shows the predicted log Z_{sat} values for some systems for which experimentally determined values are available for comparison. As discussed previously, it is expected that some of the inaccuracies introduced in solubility predictions will cancel out in the mass action ratio. This is found to be the case (note the good predictions for amoxicillin and cephalexin, for which solubility predictions were very poor) and, gratifyingly, log Z_{sat} predictions are correct to within an order of magnitude. This will generally be good enough to predict accurately when product precipitation synthesis is feasible. The exception, of course, will be when Z_{sat} and K_{eq} are very close.

Table 1. Data required for solubility predictions. Entries 1 – 4 contain one ionisable group, 5 – 7 are neutral, 8 – 13 are zwitterions.[a]

		T_m	T	τ	log P	Solubility [M] predicted	Solubility [M] measured
1	Ac-Phe	172	40	3	0.013	1.3 E – 01	3.5 E – 02[11]
2	For-Phe	169	40	2.5	– 0.688	8.8 E – 01	5.3 E – 02[11]
3	Z-Phe	88	40	5.5	2.193	6.7 E – 03	2.8 E – 03[11]
4	Leu-NH_2	106	40	2.5	– 0.8	6.5 E + 00	2.7 E + 00[11]
5	Ac-Phe-Leu-NH_2	253	40	7	– 0.875	5.3 E – 03	1.2 E – 03[11]
6	For-Phe-Leu-NH_2	198	40	6.5	– 1.576	3.7 E – 01	4.2 E – 03[11]
7	Z-Phe-Leu-NH_2	195	40	9.5	1.305	1.1 E – 04	1.9 E – 05[b]
8	Phg	263	25	1	– 0.181	3.4 E – 02	2.0 E – 02[21]
9	pOH-Phg	245	25	1	– 0.495	1.1 E – 01	1.1 E – 01[18]
10	7-ADCA	240	25	0	– 0.782	5.0 E – 01	2.0 E – 02[21]
11	6-APA	209	25	0	– 0.964	1.4 E + 00	1.4 E – 02[18]
12	cephalexin	190	25	2.5	– 1.051	6.0 E – 01	4.7 E – 02[19]
13	amoxicillin	212	25	2	– 1.547	9.9 E – 01	1.3 E – 02[19]

[a] Abbreviations: Ac: acetyl; For: formyl; Z: carbobenzyloxy; 7-ADCA: 7-aminodeacetoxy cephalosporinic acid; 6-APA: 6-aminopenicillanic acid; Phg: phenylglycine; p-OH-Phg: *para*-hydroxyphenylglycine. [b] The literature solubility value for Z-Phe-Leu-NH_2 was 4.4 E – 4 M[11]. Our experiments gave 1.9 E – 5 M and this was used in the calculation.

0947-6539/01/0710-2094 $ 17.50+.50/0 *Chem. Eur. J.* **2001**, *7*, No. 10

Table 2. Comparison of theoretical and measured log Z_{sat} values. Measured Z_{sat} values were obtained from aqueous solubility data reported in the literature (references in table). All solubilities were recalculated in terms of the neutral species.

	Log of the saturated mass action ratio (log Z_{sat})	
	predicted	measured
Ac-Phe-Leu-NH_2	−2.2	−1.9[11]
For-Phe-Leu-NH_2	−1.2	−1.5[11]
Z-Phe-Leu-NH_2	−2.6	−2.6[11]
amoxicillin	4.2	4.3[18, 19]
cephalexin	5.9	6.5[19, 21]

Equilibrium constants for amide synthesis: Values for K_{eq} can be obtained from literature values, measured experimentally or else estimated. In general, equilibrium constants reported in the literature are "apparent" values (K_{eq} values), measured under defined conditions of pH and temperature. These cannot be compared directly to Z_{sat} because the model predicts solubilities of neutral species. Instead, the equilibrium constants need first to be converted to pH-independent equilibrium constants K_{ref}, based on non-ionised COOH and NH_2 forms, at a given temperature. This conversion is straightforward, provided ionisation constants (K_a values) for the reactants are available.

Specifically for amide synthesis, the concentration of uncharged RCO_2H can be calculated from the total (measured) concentration of acid, $[RCO_2H]_{total}$, using Equation (28).

$$[RCO_2H] = \frac{[RCO_2H]_{total}}{\left[1 + \frac{K_a(RCO_2H)}{H^+}\right]} \quad (28)$$

The concentration of uncharged $R'NH_2$ can be calculated using Equation (29).

$$[R'NH_2] = \frac{[R'NH_2]total}{\left[1 + \frac{H^+}{K_a(R'NH_2)}\right]} \quad (29)$$

The concentration of a zwitterion, $R''\pm$, in its neutral form, R'', can be obtained using Equations (26) and (30).

$$[R''\pm] = \frac{[R''\pm]_{total}}{1 + \frac{K_a(\text{basic group})}{H^+} + \frac{H^+}{K_a(\text{acidic group})}} \quad (30)$$

Literature values of equilibrium constants: The US National Institute of Standards and Technology[20] provides an electronically accessible database in which equilibrium constants from the literature are listed along with the conditions under which these were measured. Values for amide synthesis were processed by using Equations (26) and (28) to (30) in order to calculate the pH-independent equilibrium constants (K_{ref}) in terms of non-ionised species. The resulting values are summarised in Table 3.

Significantly, it can be seen that K_{ref} values differ by less than an order of magnitude over a wide range of substrates. This holds across many different reactions from synthesis of protected dipeptides to the synthesis of beta lactam antibiotics. Similarly, it was found that equilibrium constants show only slight temperature variations in the range 25 °C to 30 °C. This is because the enthalpy change associated with amide bond formation is relatively low. For example, $\Delta H_r = -5\ kJ\,mol^{-1}$ for the synthesis of Bz-L-Tyr-Gly,[20] corresponding to a change of 0.01 in log K_{ref} per Kelvin. Because the range of K_{ref} values is so small, when evaluating the feasibility of product precipitation-driven amide synthesis we simply used an average value, log $K_{ref} = 3.5$.

Results and Discussion

Selected examples of feasibility predictions

To illustrate practically how the model is used to predict the feasibility of product precipitation-driven reactions, we consider three typical examples of amide syntheses. As discussed, such reactions will be favourable when log Z_{sat} is less than 3.5.

Firstly, let us consider the thermolysin-catalysed synthesis of dipeptide Z-Phe-Leu-NH_2 (entry 20 in Table 4). The melting points were found in the Beilstein database (Z-Phe

Table 3. Log K_{ref} values calculated for selected amide syntheses.

AOH	pH	K_{eq}	pK_a (AOH)		BH	pK_a (BH)		pK_a (AB)		log K_{ref}	T [K]
			COOH	NH_2		COOH	NH_2	COOH	NH_2		
Phg	5.5	0.4[a]	2	9	7-ADCA	2.5	4.8	2.3	7.2	2.8[b]	various
p-OH-Phg	5.5	0.16[a]	2.5	9.2	6-APA	2.5	4.5	2.3	7.5	2.3	298
Ac-Phe	5.5	0.5[20]	3.6		Gly-NH_2		7.5			3.6	293
	7.3	0.4[20]								3.7	293
	8.2	0.2[20]								3.9	293
Bz-Tyr	8	0.2[20]	3.6[c]		Gly-NH_2		7.5			3.8	298
Bz-Tyr	6.5	0.1[20]	3.6[c]		Gly-anilide		7.5[c]			2.9	296
Ac-Phe	4.1	0.4[20]	3.6		di-Br-Tyr-OEt[c]		7.3			3.4	293
Ac-Phe	4.6	0.8[20]	3.6		Phe-Gly-OMe		7.3			3.6	293
Z-Gly	7.2	0.6[20]	3.7		Phe-NH_2		7.3			3.6	various
	7.1	0.7[20]								3.7	various
	7	1.0[20]								3.8	various
Ac-Phe	4.6	0.2[20]	3.6		Phe-Gly	3.7	7.7	3.7		3.4	293
average										3.5	

[a] The NIST database [20] also lists K_{eq} values for a range of antibiotics calculated by Svedas et al.[22] Log K_{ref} values calculated from these values were often an order of magnitude higher than the values in Table 2. The more recently published values[18, 21] are much closer to the values reported for peptides. [b] Average of the available K_{eq} values. [c] pK_a values were, when available, taken from the Beilstein Crossfire database. If they were not available, values were taken from similar compounds, as in Bz-Tyr (value of Bz-Phe), di-Br-Tyr-OEt (value of Tyr-OEt) and glycylanilide (GlyNH_2).

Table 4. Experimental and predicted mass action ratios for a variety of amide syntheses. Experimental log Z_{sat} values were calculated from measured solubilities.

Entry	AOH	BH	T_m			τ[a]		T	log Z_{sat}	Product precipitation?	
			AOH	BH	AB	AOH	BH			predicted	experimental
1	Z-Ala	Leu-NH_2	83	106	189	4	2.5	40	−2.3	yes	yes[11]
2	Z-Ala	Phe-NH_2	83	119	215	4	2	40	−2.6	yes	yes[23]
3	Z-Ala	Phe-Leu-NH_2	83	126	226	4	6	25	−3.3	yes	yes[23]
4	Fmoc-Asp	Tyr-OMe	191	135	172	5.5	3	35	0.0	yes	yes[b]
5	Z-Asp	Leu-NH_2	165	106	210	5.5	2.5	40	−1.6	yes	yes[24]
6	Z-Asp	Phe-NH_2	165	119	203	5.5	2	40	−1.3	yes	yes[24]
7	Z-Asp	Phe-OMe	165	liq.	135	5.5	3	37	−0.9	yes	yes[c][7]
8	Z-Gln	Leu-NH_2	136	105	232	6.5	2.5	40	−2.8	yes	yes[5]
9	Bz-Gly	Tyr[d]	185	325	200	2	2	40	7.3	no	no[e]
10	Bz-Gly	Phe[d]	185	319	177	2	2	40	7.2	no	no[e]
11	Z-Gly	Leu-NH_2	121	106	123	4	2.5	40	−0.4	yes	yes[24]
12	Z-Gly	Phe-Leu-NH_2	121	126	210	4	6	37	−3.9	yes	yes[f][7]
13	Z-Gly	Phe-OMe	121	Liq.	158	4	3	40	−2.0	yes	yes[24]
14	Z-His[d]	Leu-NH_2	162	106	200	5.5	2.5	40	0.5	yes	yes[11]
15	Z-His[d]	Phe-NH_2	162	119	177	5.5	2	40	1.2	yes	yes[25]
16	Ac-Phe	Leu-NH_2	172	106	253	3	2.5	40	−2.2	yes	yes[c][7,11]
17	Ac-Phe	Val-NH_2	172	269	275	3	1.5	25	−0.6	yes	yes[c, f][26]
18	For-Phe	Leu-NH_2	169	106	198	2.5	2.5	40	−1.2	yes	yes[11]
19	Z-Phe	Ala-NH_2	88	75	213	5.5	0.5	40	−3.0	yes	yes[24]
20	Z-Phe	Leu-NH_2	88	106	195	5.5	2.5	40	−2.6	yes	yes[11]
21	Z-Phe	Met-NH_2	88	51	220	5.5	3.5	40	−4.0	yes	yes[24]
22	Z-Phe	Phe-NH_2	88	119	230	5.5	2	40	−3.1	yes	yes[24]
23	Z-Ser	Leu-NH_2	118	106	182	5	2.5	40	−1.8	yes	yes[11]
24	Z-Ser	Phe-NH_2	118	119	186	5	2	40	−1.6	yes	yes[24]
25	Mand[g]	7-ADCA[d]	118	240	186	1	0	25	2.6	yes	no[21],[h]
26	Phac[i]	7-ADCA[d]	76	240	194	1	0	25	2.1	yes	no[21],[h]
27	Phg[d]	7-ADCA[d]	263	240	190	1	0	25	5.9	no	no[21]
28	pOH-Phg[d]	6-APA[d]	245	209	212	1	0	25	4.2	no	no[18]

[a] Note that τ for the product is simply the sum of the τ values for the substrate plus 1.5. [b] Unpublished result, P. Clapes. [c] Liquid phase was (in part) non-aqueous. [d] These are zwitterionic species. [e] Unpublished results, M. Erbeldinger, for method see ref. [5]. [f] Neither substrate saturated; product precipitation was observed. [g] Mand: mandelic acid. [h] Acyl substrate was not saturated. [i] Phac: phenylacetic acid.

88 °C; Leu-NH_2 106 °C; Z-Phe-Leu-NH_2 195 °C). Taking a reaction temperature, $T_r = 40$ °C, $R = 8.314\ J\,mol^{-1}K^{-1}$ and molecular flexibility values of $\tau = 5.5$ and 2.5 for the reactants and $\tau = 9.5$ for the product, these data can be inserted into Equations (10), (11) and (14) and the resultant log χ values used in Equation (25). Since none of the reactants are zwitterions, ΔpK_a values are not needed for the prediction. The calculated value of log $Z_{sat} = -2.6$. This value is much lower than the log K_{ref} value of 3.5, so product precipitation-driven synthesis is expected to be feasible. Halling et al[11] have confirmed this experimentally.

Secondly, let us take the synthesis of Z-His-Leu-NH_2 (entry 14). Here, one of the reactants, Z-His, is zwitterionic, so we now also need to include $\Delta pK_{a(Z-His)}$ in the calculation. For the Z-His side chain, $pK_a = 6$ and for the α-carboxyl $pK_a = 4$, giving $\Delta pK_{a(Z-His)} = 2$. With the aid of substitution into Equation (27) and subsequently into Equation (25), we see that this will change log Z_{sat} by +2. Inserting the melting points (162, 106, 200 °C) and flexibility numbers of $\tau = 5.5$, 2.5 (reactants) and 9.5 (product) into Equations (10), (11) and (14), and substituting into Equation (25), gives an overall value of log $Z_{sat} = 0.5$. This is again well under the value of 3.5, suggesting that product precipitation can be expected. This was confirmed experimentally.[11]

Finally in the case of amoxicillin synthesis (entry 28, Table 4), both the reactants and product are zwitterions so, besides the melting points, three ΔpK_a values are needed (see Table 3) (2 for 6-APA, 6.7 for pOH–Phg, and 5.3 for amoxicillin). Overall, this changes log Z_{sat} by $2+6.7-5.3 = +3.4$, and produces an overall predicted value of 4.2. Hence, "solid-to-solid" synthesis is predicted to be unfavourable, as confirmed experimentally[18]. Diender et al.[18] also attempted the synthesis in the presence of organic solvent, which again did not lead to precipitation of the product.

When product precipitation-driven synthesis is to be expected in peptide synthesis: Table 4 summarises the predicted and experimental results for 28 enzyme-catalysed peptide syntheses in which product precipitation-driven reaction has been studied. It can be seen that the feasibility of such a reaction is generally predicted correctly. For dipeptides formed between non-zwitterionic N-protected and C-protected amino acids, log Z_{sat} was found to be always much smaller in value than log K_{ref}. This was similar for tripeptide examples (Table 4, entries 3 and 12). For an additional 66 predictions of di- and tripeptide formation, for which the experimental outcome was not available, we found successful product precipitation-driven reactions were predicted. Since Z_{sat} was always many times smaller than the K_{ref}, it is probable that product precipitation-driven reactions are always feasible for the synthesis of protected di-, tri- and possibly oligopeptide synthesis, provided that the reactants are not zwitterionic

0947-6539/01/0710-2096 $ 17.50+.50/0 *Chem. Eur. J.* **2001**, *7*, No. 10

(N-protected basic amino acids or C-protected acidic residues).

In cases in which one of the substrates is zwitterionic (Table 4, entries 9, 10, 14, 15, 25–28) the feasibility depends largely on the ΔpK_a value of the zwitterion. We have previously considered Z-His, which has a ΔpK_a of 2. Although small, this is sufficient to put the feasibility of product precipitation-driven dipeptide synthesis in the balance. With both Phe-NH_2 and Leu-NH_2, successful synthesis is predicted to occur and has been observed experimentally (Table 4, entries 14 and 15). When the nucleophile is an oil, such as Phe-OMe, Ala-OEt and Gly-Oet, synthesis is also predicted to occur but has not as yet been reported. With higher melting point nucleophiles such as Tyr-OEt (T_m 109 °C) and His-OMe (T_m 213 °C), synthesis is predicted to be unfavourable. For most other cases of zwitterionic amino acids, ΔpK_a is found to be much greater than 2. For example, free amino acids such as Tyr and Phe have ΔpK_a values of around 6 (Table 4, entries 9 and 10). Precipitation-driven amide synthesis using zwitterionic reactants with such large ΔpK_a values is predicted always to be highly unfavourable. A possible exception to this will be where the product is also zwitterionic, in which case synthesis may still be possible.

There is much current interest in developing efficient enzyme-catalysed routes to antibiotics made from cephalosporin derivatives. The cephalosporin nuclei 6-APA and 7-ADCA are zwitterions with a ΔpK_a of only around 2.[18] Many of the beta lactam antibiotics, such as cephalexin, cephadroxin, amoxicillin and ampicillin, require acylation of this nucleus with unprotected amino acids with ΔpK_a values of around 7. Since the resultant product antibiotics generally have a ΔpK_a around 5,[18, 22] the overall effect on $\log Z_{sat} = 2 + 7 - 5 = +4$. This means product precipitation-driven synthesis is 10^4 times less favourable, simply because of ionisation effects. With non-zwitterionic acylating agents such as phenylacetic acid (Table 4, entry 26) and mandelic acid (Table 4, entry 25) (note that the product will not be zwitterionic in these cases), it is predicted that product precipitation may occur ($\log Z_{sat}$ just below $\log K_{ref}$). Schroën et al.[21] studied biocatalysis of these reactions in acid solution, starting with excess 7-ADCA present as a suspension. No product precipitation was observed. The prediction suggests the situation might change if the other reactant were to be presented as a solid as well.

We now consider cases in which only the product is zwitterionic. An example would be a dipeptide formed from a doubly acidic acyl donor such as Z-Asp and a doubly basic nucleophile such as Arg-OMe. Since Z-Asp-Arg-OMe has a ΔpK_{aAB} of around 8, $\log Z_{sat}$ will be reduced by this amount, and product precipitation-driven synthesis will become extremely favourable. This effect should be quite general and it may well be possible to design precipitation-driven coupling reactions in which ionisation effects are the predominate driving force.

Since mass action ratios are based on the crystal lattice energy of the reactants, they are theoretically expected to hold for syntheses carried out in any solvent. This leads to a valuable rule of thumb for optimising yields in a batch type reaction. If precipitation-driven synthesis is predicted to be favourable, the highest yields in terms of solid product will be reached in a solvent in which both reactants and products show poor solubility. This is because the minimum amount of reactant will then be left in solution when equilibrium is reached. Conversely, if product precipitation cannot be achieved, then the best possible solvent for maximising yield will be the one in which the product is most soluble.

When utilising the model developed here it should be noted that it is only the thermodynamically favoured direction of a precipitation that is predicted. Practical feasibility will depends on factors such as finding a suitable (bio-)catalyst and rates of dissolution and precipitation. It is, however, worth noting that if product precipitation-driven reaction is not favourable in the forward direction then it must, by default, be favourable in the reverse direction.

The model presented here can easily be applied to other condensation reactions, provided the change in $\log P$ upon reaction and the reference equilibrium constant are known. We are currently studying feasibility predictions with a variety of other reactions, including esterifications, glycosidations and isomerisations.

Conclusion

A general model that predicts the feasibility of product precipitation-driven reaction was developed. It gives values for saturated mass action ratios that are well within an order of magnitude of those obtained from experimental data. The model provides a general tool for exploring the scope of product precipitation-driven reactions and a simple method for preexperimental planning. Importantly, in addition to solubilities, we have found that the ionisation behaviour of the reaction components must be taken into account. For example, although precipitation-driven amide formation was found to be favourable for most neutral reactants, with zwitterionic species the feasibility depended strongly on ionisation constants.

Acknowledgements

This work was financially supported by the Biotechnology and Biological Sciences Research Council and Chemferm DSM, The Netherlands.

[1] P. J. Halling, *Curr. Opin. Chem. Biol.* **2000**, *4*, 74–80.
[2] Y. L. Khmelnitsky, O. J. Rich, *Curr. Opin. Chem. Biol.* **1999**, *3*, 47–53.
[3] M. Erbeldinger, X. Ni., P. J. Halling, *Enz. Microb. Technol.* **1998**, *23*, 141–148.
[4] A. J. J. Straathof, M. J. J. Litjens, J. J. Heijnen in *Methods in Biotechnology*: *Enzymes in non-aqueous solvents* (Eds.: E. N. Vulfson, P. J. Halling, H. L. Holland), Humana Press, in press.
[5] M. Erbeldinger, X. Ni, P. J. Halling, *Biotechnol. Bioeng.* **1998**, *59*, 68–72.
[6] V. Kasche, B. Galunsky, *Biotechnol. Bioeng.* **1995**, *45*, 261–267.
[7] R. López-Fandiño, I. Gill, E. N. Vulfson, *Biotechnol. Bioeng.* **1994**, *43*, 1016–1030.
[8] B. M. De Roode, M. C. R. Franssen, A. Van Der Padt, A. E. De Groot, *Biocat. Biotrans.* **1999**, *17*, 225–240.
[9] Y. Yan, U. T. Bornscheuer, L. Cao, R. D. Schmid, *Enz. Microb. Technol.* **1999**, *25*, 725–728.

[10] Y. Yan, U. T. Bornscheuer, R. D. Schmid, *Biotechnol. Lett.* **1999**, *21*, 1051 – 1054.
[11] P. J. Halling, U. E. Eichhorn, P. Kuhl, H. D. Jakubke, *Enz. Microb. Technol.* **1995**, *17*, 601 – 606.
[12] R. V. Ulijn, M. Erbeldinger, P. J. Halling, *Biotechnol. Bioeng.* **2000**, *69*, 633 – 638.
[13] D. J. W. Grant, T. Higuchi *Techniques in Chemistry: Solubility of Organic Compounds*, Wiley **1990**.
[14] S. H. Yalkowsky, *Solubility and Solubilization of Aqueous Media* ACS/Oxford, **2000**.
[15] R. Dannenfelser, S. H. Yalkowsky, *Ind. Eng. Chem. Res.* **1996**, *35*, 1483 – 1486.
[16] S. H. Yalkowsky, *J. Pharm. Sci.* **1981**, *70*, 971 – 972.
[17] R. F. Rekker, H. M. De Kort, *Eur. J. Med. Chem.* **1979**,*14*,479 – 488.
[18] M. B. Diender, A. J. J. Straathof, L. A. M. Van der Wielen, C. Ras, J. J. Heijnen, *J.Mol.Cat.B: Enz.* **1998**, *5*, 249 – 253.
[19] A. Tsuji, E. Nakashima, T. Yamana, *J.Pharm. Sci.* **1979**, 68, 308 – 311.
[20] NIST database *Thermodynamics of Enzyme Catalysed Reactions* **2000** to be found at http://wwwbmcd.nist.gov:8080/enzyme/enzyme.html
[21] C. G. P. H. Schroën, V. A. Nierstrasz, P. J. Kroon, R. Bosma, A. E. M. Janssen, H. H. Beeftink, J. Tramper *Enz. Microb. Technol.* **1999** , *24*, 489 – 506.
[22] V. K. Svedas, A. L. Margolin, I. V. Berezin, *Enz. Microb. Technol.* **1980**, *2*, 138 – 144
[23] P. Kuhl, P. J. Halling, H. D. Jakubke *Tetrahedron Lett.* **1990**, *31*, 5213 – 5216.
[24] U. Eichhorn, PhD thesis **1995**, University of Leipzig, Germany.
[25] U. Eichhorn, A. S. Bommarius, K. Drauz, H. D. Jakubke *J. Peptide Sci.* **1997**, *3*, 245 – 251.
[26] A. S. Feliciano, J. M. S. Cabral, D. M. F. Prazeres, *Enz. Microb. Technol.***1997**, *4*, 284 – 290.

Received: September 4, 2000 [F2708]

  0947-6539/01/0710-2098 $ 17.50+.50/0