ELSEVIER

Journal of Molecular Catalysis B: Enzymatic 7 (1999) 273–282

JOURNAL OF MOLECULAR CATALYSIS B: ENZYMATIC

www.elsevier.com/locate/molcatb

Study of enzyme-catalyzed reactions in organic solvents using multiple linear regression

David Nurok [a,*], Robert M. Kleyle [b], Barry B. Muhoberac [a], Megan C. Frost [a], Paul Hajdu [a], Daniel H. Robertson [a], Sanjay V. Kamat [c], Alan J. Russell [c]

[a] *Department of Chemistry, School of Science, Indiana University–Purdue University Indianapolis, 402 N. Blackford Street, Indianapolis, IN 46202-3274, USA*

[b] *Department of Mathematical Sciences, School of Science, Indiana University–Purdue University Indianapolis, 402 N. Blackford Street, Indianapolis, IN 46202-3274, USA*

[c] *Department of Chemical Engineering and Center for Biotechnology and Bioengineering, University of Pittsburgh, Pittsburgh, PA 15261, USA*

Received 23 June 1998; received in revised form 23 October 1998; accepted 19 February 1999

Abstract

We have used multiple linear regression to predict either initial rate, log initial rate or specificity for enzyme-catalyzed reactions performed in non-aqueous solvents. The Subtilisin Carlsberg catalyzed transesterification of *N*-acetyl-L-phenylalanine ethyl ester by methanol, 1-propanol, and 1-butanol was assayed in 30 non-aqueous solvents, and the lipase catalyzed transesterification of methyl methacrylate in 23 non-aqueous solvents. Both sets of reactions were performed at fixed thermodynamic water activity. The lipase catalyzed reactions were also performed in water saturated solvents and in dry solvents. The report illustrates that regression analysis may provide insight into how solvents can alter the activity and specificity of enzymes suspended therein. A regression model for the subtilisin catalyzed reaction suggests that solvents which have a flat hydrophobic region inhibit by competing with the substrate for an enzyme cleft. In the lipase catalyzed reaction, tetrachloroethylene is an outlier (i.e., behaves differently to other solvents) for all the regression models. This deviation, together with an element of structural similarity to the substrate, suggests that tetrachloroethylene acts as a competitive inhibitor. Log *P* is an important descriptor and it, or an expression containing log *P*, appears in all the regression equations. Log initial rate is predicted by a two-descriptor model for either enzyme system in solvents of high log *P* at fixed thermodynamic water activity. Regression models with the same two descriptors predict initial rate for the lipase system over the entire log *P* range for solvents maintained at fixed thermodynamic water activity and for dry solvents, but not for water saturated solvents.

Keywords: Enzyme-catalyzed reactions; Muliple linear regression; Solvents

1. Introduction

For over a decade there has been a concerted attempt to correlate the properties of organic solvents to the function of biocatalysts suspended therein [1–3]. This tour de force of research in non-aqueous enzymology has, however, yet to yield a predictive equation which correlates the variation of activity or specificity of multiple enzyme–substrate pairs to solvent properties. Indeed, it is rare that a prediction

* Corresponding author. Tel.: +1-317-274-6886; Fax: +1-317-274-4701; E-mail: nurok@chem.iupui.edu

1381-1177/99/$ - see front matter
PII: S1381-1177(99)00047-8

holds when even only the substrate is changed in a system [4,5].

Physical properties from solvent dipole moment [6,7] to dielectric constant [8–10] and log *P* [11–13] (the logarithm of the octanol/water partition coefficient for a solvent) have all been suggested as parameters which should be studied to guide the choice of a solvent for use with anhydrous biocatalysts, but no single parameter provides the elixir for those wishing to predict enzyme activity or specificity.

It has recently been demonstrated [14] that the initial rate for the subtilisin catalyzed transesterification of *N*-acetyl-L-phenylalanine ethyl ester by either methanol, 1-propanol, or 1-butanol can be modeled with a high level of predictability by a single equation in which log *P* and the non-polar unsaturated area, for each of the organic solvents, are the independent variables. This equation emerged from a study of the above reaction in 30 non-aqueous solvents. The equation is highly predictive only for those solvents that have log *P* values greater than 2.0. The same report illustrated that substrate specificity can be modeled with a moderate level of predictability. It is our belief that similar predictive equations can be found for many enzyme-catalyzed reactions in organic solvents, provided that a sufficiently large set of solvents is studied to confer statistical validity to these equations.

In the current report we discuss the data for three substrates of the commonly studied enzyme subtilisin, and one substrate with a lipase in order to determine if regression equations can be modeled for each of the systems using solvents with a wide range of log *P* values. The latter enzyme was studied with water saturated solvents, with anhydrous solvents, and with solvents of constant water activity.

Table 1
Lipase activity at different levels of water saturation

Solvent	Initial rate (mM/h)		
	Water saturated	Dry	Constant activity
1,1,1-Trichloroethane	0.186	0.05	0.167
2-Chlorotoluene	0.285	0.112	0.26
Bromobenzene	0.247	0.0592	0.32
Butyl acetate	0.00283	0.0046	0.006
Butyl ether	0.318	0.066	0.25
Benzene	0.183	0.06	0.19
Carbon tetrachloride	0.17	0.13	0.325
Dichloromethane	0.00486	0.00031	0.0054
Chloroform	0.0184	0.00025	0.000226
1-Chlorobutane	0.203	0.0728	0.21
Chlorobenzene	0.197	0.105	0.247
Cyclohexane	0.309	0.17	0.538
1,4-Dioxane	0.00881	0.00038	0.008
Ethyl acetate	0.00343	0.00029	0.00716
Ethylbenzene	0.314	0.12	0.316
Heptane	0.3	0.17	0.572
Hexane	0.359	0.19	0.481
Nonane	0.342	0.246	0.73
Octane	0.315	0.235	0.626
Propyl acetate	0.00228	0.00015	0.00471
Toluene	0.357	0.147	0.37
Fluorobenzene	0.162	0.059	0.161

2. Materials and methods

The description of the Subtilisin catalyzed reaction, and the experimental data, has been published elsewhere [14].

Lipase from *Candida rugosa* (860 units/mg solid) was purchased from Sigma. (A unit of activity was defined as that which would hydrolyze 1.0 microequivalent of fatty acid from olive oil in 1 h at pH 7.2 at 37°C.) All organic solvents were purchased from Aldrich Chemical (Milwaukee, WI) and were of the highest purity available. All solvents were dried using molecular sieves (3 Å). The water content of each solvent was determined by Karl Fischer titration in a Fisher Coulomatic apparatus. The rates for the lipase catalyzed reactions are listed in Table 1.

3. Activity studies in organic solvents

The following is typical of all the enzyme-catalyzed reactions that were studied. Enzyme powder (20 mg/ml) was added to 2 ml of

organic solvents containing 100 mM 2-ethylhexanol and 100 mM methyl methacrylate in a 4-ml Wheaton vial. Thereafter, the enzyme suspension was sonicated for 20 s before placing it inside a shaker (300 rmp) (New Brunswick, G-24 incubator/shaker) which was kept at 40°C. At regular intervals, 0.5 μl samples were taken from the reaction mixture and injected into the gas chromatograph for analysis. Reaction rates were determined by following the formation of 2-ethylhexyl methacrylate using a gas chromatograph (HP series II, equipped with cross linked methyl silicone capillary column and an FID detector). Initial rates were determined from the slopes of the linear regression plots of 2-ethylhexyl methacrylate formation vs. time, using Sigma plot (Jandel Scientific).

A constant water activity of 0.59 at 40°C was achieved by adding 0.2 g/ml $Na_4P_2O_7 \cdot 10H_2O$ [2,15] to each of the organic solvents.

4. The physical properties of the solvents

The following physical properties were considered as potential solvent descriptors: solvent-accessible non-polar saturated area (NPSA), solvent-accessible non-polar unsaturated area (NPUA), solvent-accessible polar surface area (PSA), polarizability, dipole moment, log P, density, molecular volume, dipolarity/polarizability (π^*), hydrogen bond donor ability (α), and hydrogen bond acceptor ability (β). The descriptors in Table 2 are those included in one or more of the regression models in this report.

The values of log P were literature shake-flask values [16], values calculated by the CLOGP methods as described by Leo [17] with Biobyte's MacLog P 1.0, or Ghose–Crippen values as calculated by Spartan 4.0. Log P is a particularly important descriptor in the regression models discussed in this report. In order to evaluate which of the above values were the best for this study, shake-flask and computer generated (both Maclog P and Ghose–Crippen) values were used to build regression models for all the systems discussed below. Both the Shake–flask and Maclog P values yielded very similar models with only small differences in regression coefficients and values of R^2. The latter values were generally higher than the corresponding values for models with the Ghose–Crippen log P values. Shake-flask log P values were used for the regression models in this report.

The methods for computing dipole moment, polarizability and the solvent-accessible surface

Table 2
Physical properties of organic solvents

Solvent	NPUA[a] (Å^2)	Dipole moment ($\times 10^{-30}$ cm)	log P	π^*
1,1,1-Trichloroethane	0.0	1.76	2.49	0.49
2-Chlorotoluene	38.9	1.56	3.42	–
Acetone[b]	0.0	2.92	−0.24	0.71
Bromobenzene	45.1	1.70	2.99	0.79
Butyl acetate	0.0	1.60	1.78	0.50
Butyl ether	0.0	1.17	3.21	0.24
Benzene	45.9	0.00	2.13	0.59
Tetrachloroethylene	15.1	0.00	3.40	0.28
Carbon tetrachloride	0.0	0.00	2.83	0.28
Dichloromethane	0.0	1.60	1.25	0.82
Acetonitrile[b]	16.7	2.89	−0.34	0.75
Nitromethane[b]	0.0	4.17	−0.35	0.85
Chloroform	0.0	1.01	1.97	0.58
1-Chlorobutane	0.0	2.05	2.64	0.39
Chlorobenzene	42.1	1.69	2.89	0.71
Cyclohexane	0.0	0.00	3.44	0.00
N,N-dimethylformamide	0.0	3.55	−1.01	0.88
1,4-Dioxane	0.0	0.00	−0.27	0.55
Ethyl acetate	0.0	1.78	0.73	0.55
Ethylbenzene	45.5	0.59	3.15	–
Heptane	0.0	0.00	4.66	−0.08
Hexane	0.0	0.00	3.90	−0.08
Triethylamine[b]	0.0	0.96	1.45	0.14
tert-Butylamine[b]	0.0	1.46	0.40	–
Nonane	0.0	0.00	5.81	−0.08
Octane	0.0	0.00	5.18	−0.08
Propyl acetate	0.0	1.61	1.24	–
Pyridine[b]	41.7	1.97	0.65	0.87
Tetrahydrofuran[b]	0.0	1.92	0.46	0.58
Toluene	43.7	0.36	2.73	0.54
Fluorobenzene[c]	44.5	1.57	2.27	–

[a] Non-polar unsaturated area.
[b] Not used with lipase.
[c] Not used with subtilisin.

areas were described in a previous report [14]. The solvent-accessible area was first defined by Lee and Richards [18,19] as the locus of the center of a solvent 'sphere' which is rolled over the van der Waals surface of the solute. This area may be partitioned into polar (the PSA) and non-polar components, and the latter may be further divided into the accessible surface area contributed by saturated atoms and that contributed by unsaturated atoms such as the carbons in an alkene or aromatic compound (the NPUA). Monovalent atoms attached to the unsaturated atoms of benzene are classified as saturated in the algorithm used [18]. As an example, toluene has contributions from both a saturated region (the methyl group and ring hydrogens) and an unsaturated region (the ring carbons) contributing to the total non-polar surface area.

5. Data analysis

Ordinary (forwards) stepwise regression, whereby the model is built one term at a time, was performed using SAS for Windows, version 6.11, (SAS Institute). The descriptor that correlates most highly with the dependent variable is the first to be inserted in the linear regression equation (hereafter referred to as the model), provided that it is significant at a specified level. In this study the 90% level was used. For

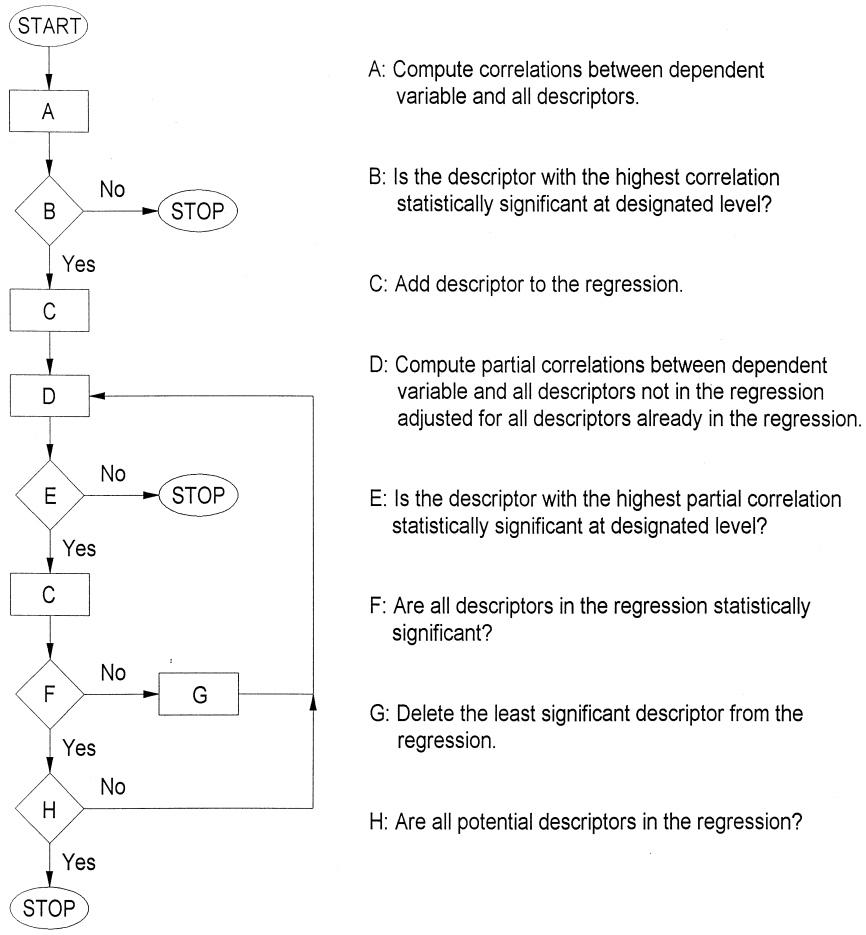

Fig. 1. Flow chart illustrating the stepwise mode of multiple linear regression.

most models, however, all terms were significant at a much higher level. The procedure then finds the descriptor that has the highest *partial* correlation with the dependent variable, adjusted for the presence of the first descriptor. If this partial correlation is significant at the specified level it is included in the model, otherwise the stepwise procedure stops. The procedure checks whether the significance of the term initially inserted is affected by the additional term. If the initial term fails the significance test it is eliminated from the model. The procedure is then continued in a similar manner for the remaining descriptors, until no term to be added is significant at the desired level. The overall procedure is illustrated in Fig. 1.

The regression models described in this report (with two exceptions noted in the text) all have values of the mean squared error (MSE) in the range 0.0007–0.06. These low values of the MSE indicate that it is reasonable to expect a good predictability from these models.

The MSE is defined as:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n} e_i^2}{n - p - 1} \quad (1)$$

where e_i is the ith residual (a residual is the difference between the observed and predicted value of the dependent variable), n is the number of observations (i.e., the number of solvents), and p is the number of descriptors.

6. Results and discussion

Transesterification was studied using both subtilisin and lipase, in a series of organic solvents that have a wide range of physical properties. The effect that these solvents have on initial rate was studied for both enzymes, while the effect on specificity was studied only for subtilisin.

The amount of water present in these systems substantially influences the kinetics of the reaction and for this reason was carefully controlled. In both the lipase and the subtilisin systems this was achieved by adding $Na_4P_2O_7 \cdot 10H_2O$. In addition the lipase system was studied both in a series of anhydrous and a series of water saturated solvents. While there are some similarities in the regression models describing the results in the two enzyme systems, there are also significant differences, and for this reason the systems are discussed separately.

7. Subtilisin

The following Subtilisin catalyzed transesterification was studied in the series of 30 solvents listed in Table 2, using methanol, 1-propanol or 1-butanol as the alcohol.

N-acetyl-L-phenylalanine ethyl ester + alcohol

→ *N*-acetyl − L − phenylalanine alkyl ester + ethanol

The activity of an enzyme in a non-aqueous solvent depends on the amount of bound water. Hydrophilic solvents (i.e., solvents of low log *P*) can distort the way this water is associated with the enzyme and can adversely affect the activity of an enzyme. Laane has suggested that this distortion of the way that water is bound accounts for the diminished activity of enzymes in non-aqueous solvents with log *P* values less than 2.0 [20].

Only poor regression models, for predicting log initial rate for the above transesterification, are obtained for the complete set of 30 solvents. As previously demonstrated [14], a good model is obtained for each of the three substrates when the reaction is performed in 16 solvents with log *P* values greater than 2.0. This value of log P was selected based on the observation of Laane noted in the previous paragraph. The model is:

$$\log \nu = -a + b \log P - c\,\mathrm{NPUA} \quad (2)$$

where NPUA is the non-polar unsaturated area associated with the aromatic solvents and with

tetrachloroethylene. The values of R^2 were in the range 0.8944–0.9426 with methanol having the lowest and 1-butanol the highest value. R^2 measures the percentage (expressed as a proportion) of the total variability in the dependent variable that can be accounted for by its linear relationship with the descriptors. Thus for 1-butanol the model explains 94% of the variation (with change of solvent) of log initial rate about the mean. [The R^2 values are slightly different to those reported previously. That report used computed (Mac log P) values, and was for a set of 15, rather than 16, solvents.]

A cleft is present in subtilisin, and the substrate slots into this cleft during reaction [21]. Other molecules with a flat hydrophobic region should also be capable of fitting into this cleft. The only solvents in this study which have such a flat region are the benzene derivatives, pyridine and tetrachloroethylene. These are also the only molecules (of log P greater than 2.0) that have a value for the NPUA. The negative value of NPUA in Eq. (1) is apparently indicative of these solvents competing with the phenylalanine substrate for a position in the cleft. This illustrates the manner by which a model obtained by multiple linear regression can provide insight into how a solvent can modulate enzyme activity. [This argument holds only for common solvents. It would not hold for compounds such as buckminsterfullerene which has a non-planar NPUA.]

This interpretation is concordant with data for subtilisin obtained by a completely different approach. A survey of the Brookhaven X-ray crystallographic database found that five different inhibitors with a flat hydrophobic region slot into the same cleft on the surface of subtilisin. This is illustrated in Fig. 2 for L-napthyl-1-acetamido boronic acid where the naphthyl group is partially embedded in a cleft on the surface of the enzyme. The latter is depicted using its solvent-accessible surface, with the inhibitor drawn as a CPK model.

We suggest that the rate of transesterification is affected in at least two different ways when the enzyme is suspended in each of a series of solvents of increasing log P. The more substantial of these is a large increase in rate with increasing log P. The second is a relatively small diminution of rate superimposed on the increase noted above, and may be related to decreasing flexibility of the enzyme in solvents of high log P. This latter effect is discussed in the following paragraphs.

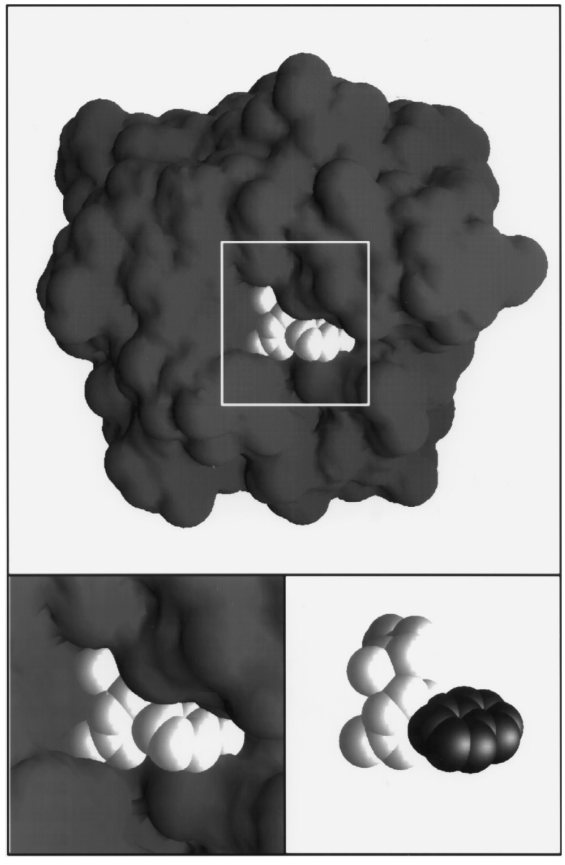

Fig. 2. The top panel shows the Subtilisin Carlsberg L-napthyl-1-acetamido boronic acid inhibitor complex. The lower left panel is a close-up showing the inhibitor fitting into the cleft of the protein. The lower right panel shows the inhibitor with the planar naphthyl group highlighted in black.

Water can be more highly ordered when in contact with a non-polar molecule such as an alkane [22], than when in contact with a polar molecule. We suggest that the same ordering effect holds for the enzyme bound water when the enzyme is suspended in a non-polar solvent

(of high log P). We also suggest that the greater ordering of the enzyme-bound water molecules confers a degree of structural inflexibility to the enzyme. These assumptions lead to the following interpretation of the above regression model.

Fig. 3 is a plot of the residuals (i.e., the differences between experimental and predicted values) for log initial rate using 1-butanol. The figure was constructed using Eq. (2) to predict the log rate in all 30 solvents, even though only 16 of log P greater than 2.0 were used in generating the equation. There is a clear trend of increasing residuals with decreasing log P, which indicates that this enzyme exhibits a *greater* activity than predicted by the regression equation in the solvents of lowest log P (even though the *absolute* activity is lowest in these solvents). If it is assumed that this change is not associated with the amount of protein-bound water (since we are at constant water activity), it may be related to a different ordering of this water depending on the polarity of the solvent in which the enzyme is suspended. We suggest that, in the more hydrophobic solvents, the water assumes a relatively ordered structure in which there is minimal hydrogen bonding between water and solvent molecules. Such a structure would in turn confer a degree of rigidity to the enzyme which would be the same for all solvents of sufficient hydrophobicity. Fig. 3 suggests that this would occur for solvents with log P greater than 2.0. This diminished enzyme flexibility would result in a lower activity than in water.

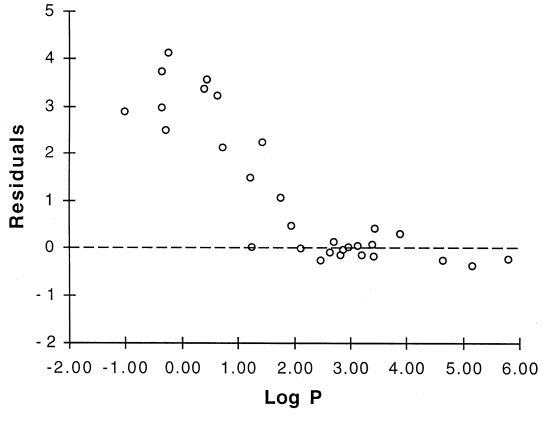

Fig. 3. Residuals for Eq. (2) plotted against log P, for butanol as substrate. The plot is for all solvents, but Eq. (2) was derived for solvents with log P greater than 2.0.

Consider now what happens when the enzyme is suspended in each of a series of solvents of successively *lower* log P, i.e., in a series of solvents of increasing hydrophilicity. Below a log P value of approximately 2.0, there will be an increasing interaction between solvent and enzyme-bound water molecules with each successive solvent of lower log P. This increase will result in decreasing order in these water molecules which will result in greater flexibility of the enzyme. This in turn will result in greater activity than predicted by Eq. (2), as shown by the positive residuals in Fig. 3. It must be emphasized that, although the residuals are positive for solvents of low log P, the *absolute* activity of the enzyme is lowest in these solvents.

Models for predicting initial rate were of poorer fit than those for predicting log initial rate. Even with solvents of log P greater than 2.0, the values of R^2 are in the range 0.66–0.88, and log P is the only descriptor that enters the model.

In our earlier report [14], it was shown that a plot of specificity ($\nu_{\text{methanol}}/\nu_{\text{butanol}}$) against log P was approximately bell shaped. The following descriptor models a symmetrical bell shaped curve centered at m and having inflection points at $m \pm s$.

$$\chi = e^{-\left(\frac{\log P - m}{s}\right)^2} \tag{3}$$

In addition to modeling specificity, χ is also a useful descriptor for modeling log rate. m and s are constants that are determined by a least squares optimization of the non-linear relationship between each specificity (or initial rate or log initial rate) and χ. Thus the values of m and s are specific for a given system. The value of the above descriptor is increasing in log P, for log $P < m$, and decreasing for log $P > m$.

Log initial rate is predicted for each of the three substrates by the following model, when the transesterification is performed in 16 solvents with log P greater than 2.0:

$$\log \nu = a - b\chi - c\,\mathrm{NPUA} \quad (4)$$

The values of the coefficients and R^2 are: Methanol—$a = 0.62$, $b = 4.02$, $c = 0.02$, $R^2 = 0.9449$; 1-Propanol—$a = 1.27$, $b = 4.62$, $c = 0.02$, $R^2 = 0.9421$; 1-Butanol, $a = -0.05$, $b = 3.08$, $c = 0.01$, $R^2 = 0.9528$.

In these models the χ term is significant at the 99.99% level, and the NPUA term is significant at a level greater than 99.6%. The intercept is highly significant for methanol and 1-propanol, but is insignificant for 1-butanol. A comparison of Eqs. (2) and (4) indicates that χ acts as a surrogate for log P in the prediction of initial rate. The regression fits with Eq. (4) are better than those with Eq. (2). The latter two equations may be reconciled by noting that for log $P > m$, χ decreases with increasing log P. This is the case for all models for solvents with log $P > 2.0$, because the values of $m < 2.0$ for each of the three substrates. The term $-\chi$ then becomes less negative (i.e., more positive) as log P increases.

8. Specificity

We define specificity as the ratio:

$$\text{Specificity} = \frac{[k_{cat}/K_m]_1}{[k_{cat}/K_m]_2} \quad (5)$$

where 1 and 2 are the members of a substrate pair. Because we determine initial rate at $[S] \ll K_m$, and because the term for enzyme concentration is the same in both numerator and denominator [22], the ratio of initial rates simplifies to the ratios of k_{cat}/K_m.

One of the descriptors used for predicting specificity is π^*, which is a solvatochromic descriptor for the dipolarity/polarizability of a molecule [23]. This spectroscopically derived descriptor has been widely used [24–26] in studies ranging from modeling reaction kinetics [27,28] to predicting chromatographic retention [29–31]. π^* values were available for only 14 of the solvents with log P greater than 2.0. The regression model below applies to both $\nu_{methanol}/\nu_{propanol}$ and $\nu_{methanol}/\nu_{butanol}$.

$$\text{Specificity} = a + b\chi + c\pi^* \quad (6)$$

where a, b and c are regression coefficients which are different for each of the two specificities. The model for $\nu_{methanol}/\nu_{propanol}$ has a value of R^2 of 0.9434, and for the $\nu_{methanol}/\nu_{butanol}$ model the value of R^2 is 0.8365. The MSE (0.11) for the former model is not as good as the corresponding values for the other regression equations in this report, but is still considered satisfactory. While the corresponding value for the $\nu_{methanol}/\nu_{butanol}$ model is too high (3.2) to allow confident prediction, it is nevertheless satisfying that the two regression equations for specificity have identical descriptors.

9. Lipase

The following Lipase catalyzed transesterification was studied in a series of 23 solvents.

Methyl methacrylate + 2-ethyl-1-hexanol

→ methanol + 2-ethylhexyl methacrylate

The solvents were either dry, water saturated or maintained at a constant water activity by incorporating $Na_4P_2O_7 \cdot 10H_2O$ in the reaction mixture. These are referred to as the Dry, the Water Saturated, and the Constant Activity series, respectively.

The initial attempt to construct regression models predicting either initial rate or log initial rate resulted in only mediocre values of R^2 for each of the systems. Inspection of the residuals for these models showed that the data point for tetrachloroethylene is an outlier. It is unlikely that this is due to experimental error as tetrachloroethylene was used in three independent experiments. This solvent was dropped from the study which is restricted to the 22 other sol-

vents. Tetrachloroethylene was not an outlier in the subtilisin study, and in fact was a solvent with a low residual for the regression Eq. (2). The atypical behavior of this solvent may be due to competitive inhibition resulting from an element of structural similarity to the substrate. Both tetrachloroethylene and methyl methacrylate are 1,1-disubstituted ethylene compounds.

As is the case for the subtilisin catalyzed transesterification, the best regression model for predicting log initial rate in the constant water activity series is obtained for the more hydrophobic solvents. The following model is for the 15 solvents with log P greater than 2.2:

$$\log \nu = -0.74 + 1.07 \log P - 0.13\mu$$

$$R^2 = 0.8982 \qquad (7)$$

The models predicting log initial rate in either the dry or the water saturated series of solvents were not satisfactory irrespective of the range of hydrophobicities considered.

In contrast to the model predicting initial rate in the subtilisin series, the corresponding model in either the Dry or the Constant Water Activity series has a relatively high R^2 and applies to the complete set of 22 solvents considered (log P range $= -0.27$ to 5.81). The model is:

$$\nu = a + b \log P - c\mu \qquad (8)$$

where μ is the dipole moment. The value of the intercept is statistically insignificant in both series, and is essentially zero. The other two terms in the models are significant at the 99.92%, or better, level. The values of the coefficients are: Dry series—$b = 0.043$, $c = 0.031$, $R^2 = 0.9033$; Constant activity series—$b = 0.126$, $c = 0.079$, $R^2 = 0.9237$.

The values of R^2 indicate that in either the Dry or the Constant Water Activity series, more than 90% of the total variation of initial rate results from its relationship with log P and dipole moment. The initial rate is more sensitive to the values of the descriptors in the Constant Water Activity than in the Dry series, as indicated by coefficients that are almost three times larger in the former series.

There is an apparent anomaly in the application of Eq. (8) to dioxane and ethyl acetate, where very small negative reaction rates are predicted. The agreement between predicted and experimental rates for these solvents is, however, very good because the experimental values are very low (less than 0.008 mM/h). These small negative predicted values are within the 95% prediction intervals for these two solvents.

The model in Eq. (8) applies also when the solvents are restricted to the 16 solvents with a log P value greater than 2.0, in which case the values of R^2 for the Constant Water Activity and the Dry series are 0.9224 and 0.8704, respectively. It is not surprising that log P effects are more noticeable in the Dry than in the Constant Water Activity series of solvents, if changes in the models are related to stripping of bound water in hydrophilic solvents. The presence of the salt hydrate should eliminate the ability of the hydrophilic solvents to strip bound water from the enzyme. In contrast to the Constant Water Activity and Dry series, the Water Saturated series provides unsatisfactory models for both initial rate and log initial rate. This suggests that water saturation allows the amount of bound water to vary with the identity of solvent, which does not allow for a predictive regression model.

χ (Eq. (3)), which is a very useful descriptor in the Subtilisin series, does not enter any of the models predicting either initial rate or log initial rate in the Lipase series.

10. Conclusions

This study demonstrates that it is possible to build good regression models for transesterification in non-aqueous solvents catalyzed by subtilisin or lipase, respectively. Log P is an important descriptor, and all models in this report include either log P or a descriptor containing log P. It is likely that good regression models

will be found also for other enzyme-catalyzed reactions in non-aqueous solvents.

This report also demonstrates that regression model can provide insight into solvent–enzyme interactions. An example is given by the regression model (Eq. (2)) that gives a possible answer to the question as to why there is not a uniform increase in reaction rate with increasing solvent lipophilicity. As interpreted in the text, the model suggests that solvents having a planar unsaturated region are able to occupy a cleft in subtilisin that is adjacent to the active center, thus hindering access of substrate to the center. A consideration of the residuals for the above regression model suggests that there is a diminution of flexibility of the enzyme when suspended in solvents of high hydrophobicity. Another example of such an insight is that the only solvent that does not fit the regression models (i.e., is an outlier) in the lipase series is tetrachloroethylene, and it is suggested that the elements of structural similarity of this solvent to the substrate (methyl methacrylate) results in its acting as a competitive inhibitor.

There are important differences between the models for the subtilisin and the lipase series. In the subtilisin series good models for predicting log initial rate (and mediocre models for predicting rate) apply only to solvents with log P values greater than 2.0, whereas in the lipase series good models for predicting initial rate apply over the complete log P range (-0.27 to 5.81). A good model for predicting log initial rate in the lipase constant activity series, however, applies only to solvents with log $P > 2.2$. It is intriguing that for this enzyme, initial rate can be predicted with a high R^2 for all 22 solvents, but log initial rate only for 15 solvents.

The fact that either rate, log rate, or specificity may be predicted with high confidence by simple regression equations containing, in most cases, only one or two descriptors, suggests that these equations have the potential to furnish insight into the role of the non-aqueous solvents in these enzyme-catalyzed reactions.

References

[1] A.J. Russell, S. Chatterjee, I. Rapanovich, J.G. Goodwin, Biomolecules in Organic Solvents, CRC Press, London, 1995, pp. 91–111.
[2] P.J. Halling, Enzyme Microb. Technol. 16 (1994) 178–206.
[3] A.M. Blinkovsky, B.D. Martin, J.S. Dordick, Curr. Opin. Biotechnol. 3 (1992) 124–129.
[4] S. Tawaki, A.M. Klibanov, J. Am. Chem. Soc. 114 (1992) 1882–1884.
[5] C.R. Wescott, A.M. Klibanov, J. Am. Chem. Soc. 115 (1993) 1629–1631.
[6] P.A. Fitzpatrick, A.M. Klibanov, J. Am.Chem. Soc. 113 (1991) 3166–3177.
[7] Z. Yang, A.J. Russell, in: A.M.P. Koskinen, A.M. Klibanov (Eds.), Enzymatic Reactions in Organic Media, Blackie Academic and Professional, Glasgow, UK, 1995, pp. 43–69.
[8] R. Affleck, Z. Xu, V. Suzawa, K. Focht, D.S Clark, J.S. Dordick, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 110–1104.
[9] Z.-F. Xu, R. Affleck, P. Wangikar, V. Suzawa, J.S. Dordick, D.S. Clark, Biotechnol. Bioeng. 43 (1994) 515–520.
[10] R.M. Guinn, P.S. Skerker, P. Kavanaugh, D.S. Clark, Biotechol. Bioeng. 37 (1991) 303–308.
[11] M.D. MacNaughtan, A.J. Daugulis, Enzyme Microb. Technol. 15 (1993) 114–119.
[12] V. Laroute, R. Willemot, Enzyme Microb. Technol. 14 (1992) 528–534.
[13] Z. Yang, Q.A. Robb, Biotechnol. Bioeng. 43 (1994) 365–370.
[14] A.K. Chaudhary, S.V. Kamat, E.J. Beckman, D. Nurok, R.M. Kleyle, P. Hadju, A.J. Russell, J. Am. Chem. Soc. 118 (1996) 12891–12901.
[15] P.J. Halling, Biotechnol 6 (1992) 271–276.
[16] C. Hansch, A. Leo, D. Hoekman, Exploring QSAR, American Chemical Society, Washington, DC, 1995.
[17] A.J. Leo, Chem. Rev. 93 (1993) 1281–1306.
[18] B. Lee, F.M. Richards, J. Mol. Biol. 55 (1971) 379–400.
[19] F.M. Richards, Methods Enzymol. 115 (1985) 440–464.
[20] C. Laane, S. Boeren, K. Vos, C. Veerer, Biotechnol. Bioeng. 30 (1987) 81–87.
[21] D.A. Estell, T.P. Graycar, J.V. Miller, D.B. Powers, J.P. Burnier, P.G. Ng, J.A. Wells, Science 233 (1986) 659–663.
[22] R. Chang, Physical Chemistry with Applications to Biological Systems, MacMillan, New York, 1981.
[23] M.J. Kamlet, J.-L.M. Abboud, M.H. Abraham, R.W. Taft, J. Org. Chem. 48 (1983) 2877–2887.
[24] M.J. Kamlet, R.M. Doherty, M.H. Abraham, Y. Marcus, R.W. Taft, J. Phys. Chem 92 (1988) 5244–5255.
[25] J.E. Brady, P.W.J. Carr, Phys. Chem. 89 (1985) 1813–1822.
[26] D.C. Leggett, Anal. Chem. 65 (1993) 2907–2909.
[27] M.H. Abraham, R.M. Dorherty, M.J. Kamlet, J.M. Harris, R.W. Taft, J. Chem. Soc., Perkin Trans. 2 (1987) 913–920.
[28] P. Aparna, S. Kothari, K. Banerji, Proc. Indian Acad. Sci. 107 (1995) 213–220.
[29] L.C. Tan, P.W. Carr, Anal. Chem. 66 (1994) 450–457.
[30] M.H. Abraham, C.F. Poole, S. Poole, J. Chromatogr. A 749 (1996) 201–209.
[31] P.G. Muijsellar, H.A. Claessens, C.A. Cramers, Anal. Chem. 69 (1997) 1184–1191.