Predicting the Solvent Dependence of Enzymatic Substrate Specificity Using Semiempirical Thermodynamic Calculations

Charles R. Wescott[†] and Alexander M. Klibanov*

Department of Chemistry
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

The discovery that substrate specificity of enzymes is markedly dependent on the solvent[1,2] has prompted a search for a mechanistic explanation of this phenomenon. Recently, we have succeeded[2] in rationalizing the observed solvent effect on the transesterification reactions of N-Ac-L-Ser-OEt (Ser) and N-Ac-L-Phe-OEt (Phe) with 1-propanol catalyzed by the protease subtilisin Carlsberg.[3] A thermodynamic model has been elaborated which explains a strong preference of the enzyme for Ser in some anhydrous solvents and for Phe in others on the basis of substrate solvation/desolvation differences. Furthermore, a mathematical equation has been derived which relates the substrate specificity in an organic solvent to that in water and to the solvent-to-water partition coefficients (P) of the substrates:

$$\log\left[\frac{(k_{cat}/K_M)_{Ser}}{(k_{cat}/K_M)_{Phe}}\right]_{solvent} = \log\left[\frac{(k_{cat}/K_M)_{Ser}}{(k_{cat}/K_M)_{Phe}}\right]_{water} + \log\frac{P_{Phe}}{P_{Ser}} \quad (1)$$

where k_{cat} and K_M are the turnover number and the Michaelis constant for the ester substrate, respectively, in the medium indicated. When we measured and plotted the substrate specificities[4] of subtilisin *vs* the experimentally determined ratio of the partition coefficients, the expected linear dependence indeed ensued.[2]

A fundamental limitation of this approach, however, is that it can be used only with *water-immiscible* solvents, because direct measurements of partition coefficients between water and *water-miscible* solvents are not feasible. In addition, even the data obtained for water-immiscible solvents are quite imperfect. First, due to mixing of the solvent and water during the measurement of partition coefficients, the actual two phases in contact are unavoidably water-saturated solvent and solvent-saturated water, instead of the anhydrous solvent and pure aqueous solution in which the substrate specificities are measured. Second, 1 M propanol is present in the nonaqueous reaction medium as a nucleophile during the specificity measurements. The partition coefficients, however, cannot be determined in its presence because the propanol will partition into the aqueous phase. Finally, the measurement of partition coefficients is rather laborious and time-consuming.

In the present work, we eliminate the aforementioned problems and broaden the utility of our approach by using a computer program to calculate the P ratio of the substrates for different solvents. It can be shown[6] that

$$\frac{P_{Phe}}{P_{Ser}} = \left(\frac{\gamma_{Ser}}{\gamma_{Phe}}\right)_{solvent} \left(\frac{\gamma_{Phe}}{\gamma_{Ser}}\right)_{water} \quad (2)$$

where γ is the thermodynamic activity coefficient of the substrate indicated. The γ values for a given molecule in a solvent can be calculated on the basis of the van der Waals volumes and surface areas of the constituent groups of that molecule and of those of the solvent and empirically determined interaction parameters between these groups.[7] Such calculations can be carried out using the UNIFAC group contribution algorithm.[7,8] Unfortunately, insufficient UNIFAC interaction parameters are available in the literature[7,8] to calculate the activity coefficients for N-Ac-Phe-OEt and N-Ac-Ser-OEt. However, log P of a molecule is an additive function of its component groups.[9] Thus, when the P ratios are calculated for two similar molecules, the contributions of identical groups which exist in both will cancel out. Because our two substrates differ only in that the hydroxyl group in Ser is replaced by a phenyl group in Phe, $P_{Phe}/P_{Ser} = P_{tol}/P_{MeOH}$, where P_{tol} and P_{MeOH} are the corresponding partition coefficients for toluene and methanol, respectively. Consequently, we have written a computer program which implements UNIFAC to calculate γ_{tol} and γ_{MeOH} in organic solvents containing 1 M propanol; the γ ratio in water has been determined experimentally.[10] Equation 2 was then employed to calculate the P_{Phe}/P_{Ser} ratios.

Table I contains the calculated γ values and the resultant P ratios of the substrates for eight water-miscible, as well as 11 water-immiscible, solvents. According to equation (1), a double-logarithmic plot of substrate specificity *vs* P_{Phe}/P_{Ser} should yield a straight line with a slope of 1.0 and an intercept equal to the logarithm of the substrate specificity in water (−1.7).[2] When such a plot is produced using the calculated values of P_{Phe}/P_{Ser} (Figure 1), linear regression yields a slope of 0.89 and an intercept of −1.7, with a correlation coefficient of 0.96. The model on which eq 1 is based assumes that the substrates are fully removed from the solvent in the transition state.[2] If one or both of the

[†] An NIH Biotechnology Predoctoral Trainee.

(1) Zaks, A.; Klibanov, A. M. *J. Am. Chem. Soc.* **1986**, *108*, 2767–2768. Gaertner, H.; Puigserver, A. *Eur. J. Biochem.* **1989**, *181*, 207–213. Clapés, P.; Adlercreutz, P. *Biochim. Biophys. Acta* **1991**, *1118*, 70–76. Gololobov, M. Y.; Voyushina, T. L.; Stepanov, V. M.; Adlercreutz, P. *FEBS Lett.* **1992**, *307*, 309–312. Ryu, K.; Dordick, J. S. *Biochemistry* **1992**, *31*, 2588–2598.

(2) Wescott, C. R.; Klibanov, A. M. *J. Am. Chem. Soc.* **1993**, *115*, 1629–1631.

(3) Alkaline serine protease from *Bacillus licheniformis* (EC 3.4.21.14).

(4) Defined as the ratio of the specificity constants,[5] k_{cat}/K_M, for the substrates.

(5) Fersht, A. *Enzyme Structure and Mechanism*, 2nd ed.; Freeman: New York, 1985.

(6) The partition coefficient of Phe (P_{Phe}) is defined by the expression: P_{Phe} = [Phe]$_{solvent}$ / [Phe]$_{water}$, where the brackets represent the molar concentration in the indicated phase at equilibrium. The thermodynamic activity (a) is related to the activity coefficient (γ) and the mole fraction (x) by: a_{Phe} = $\gamma_{Phe}x_{Phe}$. Since at equilibrium a_{Phe} is the same in both phases, $(\gamma_{Phe})_{water}/(\gamma_{Phe})_{solvent}$ = $(x_{Phe})_{solvent}/(x_{Phe})_{water}$. For dilute solutions, $n_{Phe} \ll n_{solvent}$, where n is the number of moles of the designated component, and thus, $(x_{Phe})_{solvent} \approx n_{Phe}/n_{solvent}$ = [Phe]$_{solvent}$/[solvent]. Likewise, $(x_{Phe})_{water} \approx$ [Phe]$_{water}$/[water]. Substitution of these expressions for x into the expression for the γ ratio yields $(\gamma_{Phe})_{water}/(\gamma_{Phe})_{solvent}$ = ([Phe]$_{solvent}$/[Phe]$_{water}$)([water]/[solvent]). Solving this equation for [Phe]$_{solvent}$/[Phe]$_{water}$ and substituting it into the expression for P_{Phe} yields P_{Phe} = {$(\gamma_{Phe})_{water}/(\gamma_{Phe})_{solvent}$}([solvent]/[water]). When the same procedure is repeated for N-Ac-L-Ser-OEt and the ratio of P_{Phe}/P_{Ser} is expressed, the concentrations of solvent and water cancel out, yielding eq 2.

(7) Fredenslund, A.; Gmehling, J.; Rasmussen, P. *Vapor–Liquid Equilibria Using UNIFAC*; Elsevier: New York, 1977.

(8) Steen, S.-J.; Bärbel, K.; Gmehling, J.; Rasmussen, P. *Ind. Eng. Chem. Process Des. Dev.* **1979**, *18*, 714–722. Rasmussen, P.; Fredenslund, A. *Ind. Eng. Chem. Process Des. Dev.***1982**, *21*, 118–127. Macedo, E. A.; Weidlich, U.; Gmehling, J.; Rasmussen, P. *Ind. Eng. Chem. Process Des. Dev.***1983**, *22*, 678–681. Teigs, D.; Gmehling, J.; Rasmussen, P.; Fredenslund, A. *Ind. Eng. Chem. Res.* **1987**, *26*, 159–161. Hansen, H. K.; Rasmussen, P.; Schiller, M.; Gmehling, J. *Ind. Eng. Chem. Res.* **1991**, *30*, 2355–2358.

(9) Rekker, R. F. *The Hydrophobic Fragmental Constant*; Elsevier: New York, 1977.

(10) Since 10 mM toluene is insoluble in water under our conditions, we could not use UNIFAC to calculate $(\gamma_{tol})_{water}$. Instead, we rewrote eq 2 for methanol and toluene and used it to calculate $(\gamma_{tol}/\gamma_{MeOH})_{water}$ from the experimentally determined values of the P ratios and the calculated values of the γ ratios in water-immiscible solvents. The average of the resultant values was 88 ± 17.

Communications to the Editor

J. Am. Chem. Soc., Vol. 115, No. 22, 1993 10363

Table I. Calculated Activity Coefficients for Methanol and Toluene and Calculated Partition Coefficient Ratios for N-Ac-L-Phe-OEt and N-Ac-L-Ser-OEt[a]

solvent	γ_{MeOH}	γ_{tol}	$\gamma_{MeOH}/\gamma_{tol}$	P_{Phe}/P_{Ser}[b]
water	c	c	0.011[c]	1
Water-Miscible				
tert-butyl alcohol	1.1	3.5	0.32	28
acetonitrile	2.1	3.8	0.55	48
dioxane	2.0	3.2	0.63	56
pyridine	1.1	1.7	0.64	56
acetone	2.3	1.8	1.3	120
2-butanone	2.4	1.3	1.8	160
methyl acetate	2.8	1.6	1.8	160
tetrahydrofuran	2.9	1.5	2.0	170
Water-Immiscible				
tert-amyl alcohol	1.1	1.29	0.39	34
ethyl acetate	2.8	1.4	2.1	180
isopropyl acetate	2.9	1.2	2.4	210
tert-butyl acetate	2.9	1.2	2.4	210
diethyl ether[d]	3.8	1.3	2.8	250
chloroform	4.1	0.74	5.5	480
octane	7.9	1.4	5.6	490
dichloromethane	4.7	0.76	6.2	540
toluene	6.4	1.0	6.2	550
benzene	7.2	0.97	7.5	660
carbon tetrachloride	8.5	0.99	8.6	750

[a] Activity coefficients (γ) for 10 mM each methanol (MeOH) and toluene (tol) in the indicated organic solvents containing 1 M 1-propanol were calculated using the UNIFAC group contribution method. [b] Partition coefficient (solvent-to-water) ratios were calculated from activity coefficients as described in footnote 6. Note that the calculated partition coefficient ratios are different from those measured experimentally.[2] This is because the latter involve partitioning between water-saturated solvents and solvent-saturated water and do not include 1 M propanol. In contrast, the calculated values were for the pure phases where the organic phase contained 1 M propanol. Separately, it is worth mentioning that the use of partition coefficients of reactants is becoming increasingly useful in describing enzyme action in organic solvents: Yang, Z.; Robb, D. A.; Halling, P. A. In *Biocatalysis in Non-Conventional Media*; Tramper, J., Vermüe, M. H., Beeftink, H. H., von Stackar, U., Eds.; Progress in Biotechnology 8; Elsevier: New York, 1992; pp 585–592. [c] Because γ_{tol} could not be calculated in water, $\gamma_{MeOH}/\gamma_{tol}$ in water was calculated as described in footnote 10. [d] Another ether, tert-butyl methyl ether, which was used by us previously,[2] is not included here because insufficient UNIFAC groups exist in the literature[8] to model tertiary ethers.

substrates are partially solvated in the transition state,[11] then only a fraction of the free energy of solvation of the substrates

(11) Our preliminary molecular modeling estimates show that in the ground state of the acyl-subtilisins, approximately one third of the substrate moiety is exposed to the solvent. This analysis is based on the assumption that the structures of subtilisin Carlsberg in water and in anhydrous solvents are the same; that assumption is supported by our recent X-ray crystallographic data: Fitzpatrick, P. A.; Steinmetz, A. C. U.; Ringe, D.; Klibanov, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8653–8657.
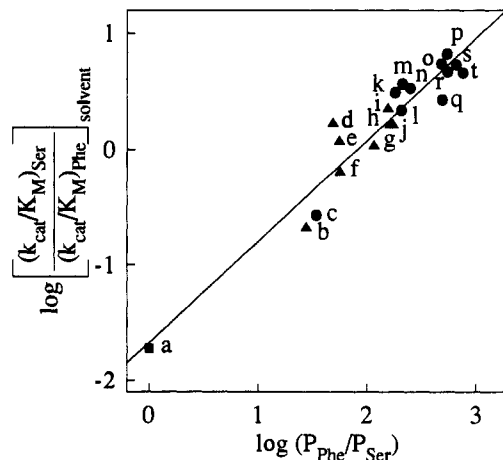

Figure 1. Dependence of the substrate specificity of subtilisin Carlsberg in water-miscible (▲) and water-immiscible (●) solvents on the ratio of the calculated solvent-to-water partition coefficients of N-Ac-L-Phe-OEt and N-Ac-L-Ser-OEt. Solvents: (a) water, (b) tert-butyl alcohol, (c) tert-amyl alcohol, (d) acetonitrile, (e) dioxane, (f) pyridine, (g) acetone, (h) 2-butanone, (i) methyl acetate, (j) tetrahydrofuran, (k) ethyl acetate, (l) isopropyl acetate, (m) tert-butyl acetate, (n) diethyl ether, (o) chloroform, (p) dichloromethane, (q) octane, (r) toluene, (s) benzene, and (t) carbon tetrachloride. The k_{cat}/K_M values were measured as previously described.[2] For 2-butanone, methyl acetate, and tetrahydrofuran, which had not been used previously,[2] the substrate specificity values were 1.7, 2.3, and 1.7, respectively. For experimental details, see footnotes to Table I. The use of the double-logarithmic (instead of linear) plot is necessary to give the same weighting to all of the data points when analyzed by linear regression. For instance, in logarithmic, but not in linear, coordinates, specificity values below 1 (the Phe substrate is favored) contribute to the regression equally to the specificity values exceeding 1 (the Ser substrate is preferred).

is available to influence the enzymatic specificity. This may be the reason why the observed slope (Figure 1) is slightly less than the expected value of 1.

In principle, our approach is independent of the enzyme[2] and thus should be of general significance. This fact, combined with the ability to calculate the P ratios of substrates by computer, allows the quantitative prediction of enzymatic specificity for various enzyme/substrate systems in any solvent, given the substrate specificity in a single reference solvent (e.g., in water, as herein). The applicability of this concept to other systems is currently being verified.