Rationalization of the Enantioselectivity of Subtilisin in DMF

Giorgio Colombo,[‡] Samuel Toba,[†] and Kenneth M. Merz, Jr.*,[†]

*Contribution from the Department of Chemistry, The Pennsylvania State University,
152 Davey Laboratory, University Park, Pennsylvania 16802, and Istituto di Biocatalisi e Riconoscimento
Molecolare, CNR, via Mario Bianco 9, 20131 Milano, Italy*

Received November 11, 1998

Abstract: Herein we examine the origin of enantioselectivity in the serine protease subtilisin in DMF through the use of molecular dynamics (MD) and free energy perturbation (FEP) simulations. In particular, we are interested in the resolution of a racemic mixture of *sec*-phenethyl alcohol by a transesterification reaction with the acylating agent vinyl acetate, catalyzed by subtilisin in anhydrous dimethylformamide (DMF). To study the enantioselectivity in this case, we examined the tetrahedral intermediate as a model of the enzyme transition state (as has been done in the past). A critical aspect of this study was the determination of the charge distribution of the two (*R* and *S*) tetrahedral intermediates through the use of a combined quantum mechanical/molecular mechanical electrostatic potential fitting methodology. In designing the active site charge model, we found that the *R* and *S* tetrahedral intermediates have significantly different charge distributions due to the presence of the stereodifferentiating environment presented by the enzyme. In contrast the charge distribution obtained for models of the tetrahedral intermediate in the gas phase have similar charge distributions. From MD simulations we find that both steric and electrostatic complimentarity plays a role in the enantioselectivity of this enzyme-catalyzed reaction. Through the use of FEP simulations we obtained a free energy difference that is in good accord with experiment, which quantitatively supports the accuracy of our model and suggests that all-atom molecular simulations are capable of providing accurate qualitative and quantitative insights into enzyme catalysis in nonaqueous environments.

Introduction

The study of enzyme selectivity in nonaqueous solvents is a particularly intriguing field of research. In these media, the absence of a continuous aqueous layer around an enzyme makes it possible for it to interact directly with the nonaqueous solvent, which results in modifications of the properties of the enzyme. Specifically the biocatalyst can attain new properties in terms of stability, activity, and specificity/selectivity.[1] More importantly, in organic solvents, enzymes such as hydrolases and proteases can catalyze esterification and transesterification reactions readily with high product yields.[2]

However, a complete understanding of the enzyme–substrate–solvent interaction is necessary to increase their utility to synthetic chemists.[3,4] Several theories have been proposed to rationalize the mechanism by which organic media influence enzymatic reactivity. One theory proposes that the selectivity is altered by solvent molecules bound within the active site where they modify the interactions between the enzyme and its substrate.[5–7] Alternatively the solvent could alter the conformation of the enzyme, thereby affecting the molecular recognition process.[8,9] A third model suggests that the selectivity dependence arises due to differences in the thermodynamics of substrate solvation, and because this model has a basis in thermodynamics, it can make some quantitative predictions.[10,11] Haeffner et al. suggested that enantioselectivity could be expressed as a function of the energy difference between two diastereoisomeric enzyme/substrate complexes. This energy difference evaluation was performed by defining subsets within the enzyme structure using molecular modeling procedures. Two different strategies were used: The first used predefined parts of the enzyme and the substrate as subsets. The second approach formed energy-based subsets. The selection of residues to be included was based on the energy of the interaction between the specific residue and the transition state analogue.[12] Importantly, this approach was able to qualitatively predict which enantiomer was the fast-reacting one.

Another recent computational model used to study enzyme enantioselectivity (α-chymotrypsin was the model enzyme) was described by Ke, Tidor, and Klibanov.[13] In this study they determined the substrate charge distribution (determined via STO-3G electrostatic potential fits) for the *R* and *S* substrates using an ensemble of energy-minimized structures (10 in all) obtained from gas-phase molecular dynamics (MD) simulations.

[†] The Pennsylvania State University.
[‡] CNR.
(1) Zacks, A.; A. M., K. *Science* **1984**, *224*, 1249–1251.
(2) Carrea, G.; Ottolina, G.; Riva, S. *Trends Biotechnol.* **1995**, *13*, 63–70.
(3) Klibanov, A. M. *Trends Biotechnol.* **1997**, *15*, 97–101.
(4) Tramper, J.; Vermue, M. H.; Beeftink, H. H.; Von Stockar, U. **1992**.
(5) Secundo, F.; Riva, S.; Carrea, G. *Tetrahedron: Asymmetry* **1992**, *3*, 267–280.
(6) Nakamura, K.; Takebe, J.; Kitayama, T.; Ohno, A. *Tetrahedron Lett.* **1991**, *32*, 4941–4944.
(7) Hirose, Y.; Kariya, K.; Sasaky, J.; Kurono, Y.; Ebicke, H.; Achiwa, K. *Tetrahedron Lett.* **1992**, *33*, 7157–7160.

---

(8) Fitzpatrick, P. A.; Klibanov, A. M. *J. Am. Chem. Soc.* **1991**, *113*, 3166–3171.
(9) Wu, H. S.; Chu, F. Y.; Wang, K. *Bioorg. Med. Chem. Lett.* **1991**, *1*, 399.
(10) Wescott, C. R.; Noritomi, H.; Klibanov, A. M. *J. Am. Chem. Soc.* **1996**, *118*, 10365–10370.
(11) Ke, T.; Wescott, C. R.; Klibanov, A. M. *J. Am. Chem. Soc.* **1996**, *118*, 3366–3374.
(12) Haeffner, F.; Norin, T.; Hult, K. *Biophys. J.* **1998**, *74*, 1251–1262.
(13) Ke, T.; Tidor, B.; Klibanov, A. M. *Biotechnol. Bioeng.* **1998**, *57*, 741–745.

Scheme 1. Transesterification Reaction Catalyzed by Subtilisin in DMF

(I)
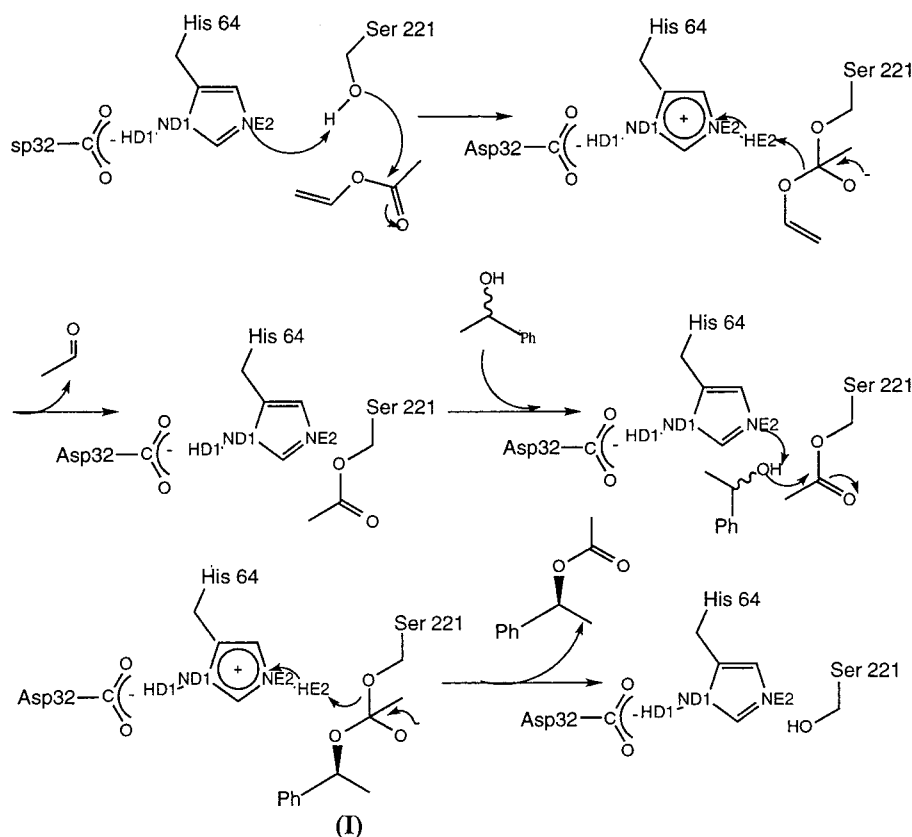
Then they docked the substrates into the enzyme active site and used vacuum (modeled using a distance-dependent dielectric) MD simulations followed by energy minimization to obtain suitable structures of the enzyme/substrate complexes. To mimic the presence of solvent, these authors used continuum electrostatic models (on the enzyme/substrate structures) and were able to obtain good estimates of enzyme enantioselectivity in water. While this study was a significant advance over earlier efforts, which were essentially vacuum studies,[14–17] we felt that still further improvements could be made especially through the use of explicit solvent and the determination of atomic point charges that included the effect of the enzyme environment. This has been done in the past on ester cleavage by a β-cyclodextrin,[18] but, to our knowledge, has not been applied to an enzyme system.

Previous work from our laboratory has focused on the effect of nonaqueous solvents on protein structure and dynamics.[19,20] To extend this work to how enzyme function is affected, we became interested in enantioselectivity in organic solvents (as well as in water). Enantioselectivity is an appealing characteristic of enzyme catalysts, and we were interested in investigating the mechanism by which the solvent environment might influence enzyme enantioselectivity. The ultimate goal of this work is to aid in the rational design of biocatalytic systems suitable for organic synthesis.

In particular, we were interested in garnering a deeper understanding of the enantioselectivity of the serine protease

subtilisin in the organic solvent dimethylformamide (DMF). The reaction of interest is shown in Scheme 1. This is the resolution of a racemic mixture of *sec*-phenethyl alcohol by a transesterification reaction with the acylating agent vinyl acetate, catalyzed by subtilisin in anhydrous DMF. The enantioselectivity of an enzyme following Michaelis–Menten behavior can be expressed in terms of the parameter E (which depends on the $(k_{cat}/K_M)_S/(k_{cat}/K_M)_R$ ratio),[21] which, in turn, can be related to the free energy $\Delta\Delta G^*$ by the expression[22]

$$\Delta\Delta G^* = -RT \ln E$$

Since two enantiomeric substrates in an achiral environment (DMF) have the same ground-state free energy, $\Delta\Delta G^*$ can be expressed as the free energy difference between the rate-determining transition states leading to the S and R products:

$$\Delta\Delta G^* = \Delta G_{(S-R)}$$

The experimental value for enantioselectivity (E), of the reaction shown in Scheme 1, was found by chiral GC to be 2.2 ($\Delta\Delta G^* = \Delta G_{(S-R)} = 0.4$ kcal/mol), with the S enantiomer reacting faster than the R enantiomer.[23]

The formation of the tetrahedral intermediate is thought to be the rate-determining step in catalysis of serine proteases.[24] Moreover, it is thought that the structure of the transition state for formation of the tetrahedral intermediate closely resembles

(14) Wipff, G.; Dearing, A.; Weiner, P. K.; Blaney, J. M.; Kollman, P. A. *J. Am. Chem. Soc.* **1983**, *105*, 997−1005.

(15) Norrin, M.; Hult, K.; Mattson, A.; Norrin, T. *Biocatalysis* **1993**, *7*, 131−147.

(16) DeTar, D. F. *Biochemistry* **1981**, *20*, 1730−1743.

(17) DeTar, D. F. *J. Am. Chem. Soc.* **1981**, *103*, 107−110.

(18) Luzhkov, V.; Åqvist, J. *J. Am. Chem. Soc.* **1998**, *120*, 6131−6137.

(19) Toba, S.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1997**, *119*, 9939−0048.

(20) Toba, S.; Hartsough, D. S.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1996**, *118*, 6490−6498.

(21) Chen, C. S.; Fujimoto, Y.; Girdaukas, G.; Sih, C. J. *J. Am. Chem. Soc* **1982**, *104*, 7294−7299.

(22) Fersht, A. R. *Enzyme Structure and Mechanism*; W. H. Freeman and Co.: New York, 1985.
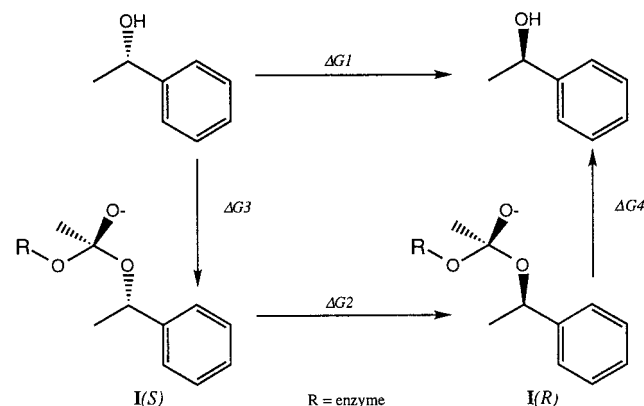
(23) Ottolina, G.

Scheme 2. Thermodynamic Cycle Used for the Free Energy Perturbation Calculations

Figure 1. Structure of the tetrahedral intermediates **I** (enzyme bound) and **II** (in vacuo calculation).

the structure of the tetrahedral intermediate itself.[24] Thus, in the past workers interested in studying serine proteases have used the structure of the tetrahedral intermediate to approximate the transition state structure.[12,13] In this paper we determine the free energy difference between the two tetrahedral intermediates **I**(R) and **I**(S) given in Scheme 2, which have been used to represent the R and S transition states for the transesterification of *sec*-phenethyl alcohol catalyzed by subtilisin (see Scheme 1). We also report studies that address the role that structural, energetic, and electrostatic factors play in influencing the enantioselectivity of this reaction when carried out by subtilisin in DMF.

In particular, the importance of the electrostatic interactions in the differentiation between two enantiomers, to our knowledge, has not been analyzed in detail. Clearly, two enantiomeric molecules in an achiral environment must have the same charge distribution on the corresponding atoms. The same enantiomeric substrates, complexed or bound to an enzyme, however, experience a chiral environment, which gives rise to two diastereoisomeric complexes. In this case analogous atoms of the two substrates will be perturbed by different electrostatic fields, generated by different electrostatic environments. This "electrostatic stereodifferentiation" is certainly a factor in the determination of the free energy difference defining enzyme enantioselectivity. The availability of quantum mechanical/molecular mechanical electrostatic potential (QM/MM ESP) fitting methods[25−28] has allowed us, for the first time, to carefully take into consideration this aspect of enzymatic reactivity, by explicitly considering polarization and charge-transfer effects. In particular, we show differences in the charge distributions on the atoms of the substrate and on the atoms of the catalytically important residues and differential fluctuations of these charges arising during MD simulations. Most critically, the accurate treatment of the electrostatic problem has allowed us to evaluate a $\Delta\Delta G^*$ value which is in good accord with the experimentally determined enantioselectivity value.

Computational Details

The starting enzyme coordinates were obtained from the crystal structure (PDB entry 1scn).[29] The intermediates, **I**(S) and **I**(R), were linked to the Oγ (OG in Figure 1) of Ser-221. The O$^-$ atom of the

intermediate was placed in the oxyanion hole defined by the side chain CONH$_2$ group of Asn-155 and the backbone NH group of Ser-221. Three chloride atoms were added to neutralize the excess charge on the enzyme/substrate complex. The 50 most strongly bound water molecules (as determined from experimental *B*-factors) were retained. This number of water molecules has been found to be approximately the amount of water required for the protein to retain its catalytic activity in nonaqueous environments.[30,31] The enzyme/substrate complex was then solvated in a cubic box of 1515 DMF molecules. The solvent and the enzyme were modeled using the AMBER force field.[32] All-atom AMBER parameters were used for the enzyme/substrate complex,[32] TIP3P for the water molecules,[33] and OPLS[34] united-atom parameters for the DMF. This system was then AMBER[32] minimized using ROAR 1.0.[35]

The AMBER-minimized system was then QM/MM minimized using ROAR 1.0.[35] The QM region included the substrate and the residues involved in catalysis (Ser-221, Asp-32, His-64, Asn-155), and the PM3 Hamiltonian[36] was used. The total charge of the QM region was set to −1. The solvent and the rest of the enzyme were treated at the MM level, using the potential functions described above.

The structures of the complexes obtained after the first QM/MM minimization were used to evaluate the atomic charges of the intermediate and of the active site residues in the enzyme-bound state, with QM/MM (MNDO Hamiltonian)[28,37] methods coupled with ESP fitting (the total charge of the QM system fixed to −1). We used the MNDO Hamiltonian for ESP fitting because MNDO has been shown to give ESP fitted charges that are well correlated to HF/6-31G* ESP derived charges, while PM3 does not.[28,37] It is important to notice that the set of charges obtained from these calculations (see Table 1) include the influence of the enzyme solvent environment on the substrate. Using the charges obtained in this way, we ran a MD equilibration run of the enzyme/substrate complex in a DMF solvent box.

The MD simulations were performed using the SANDER module of AMBER.[38] The temperature of the system was slowly raised from 0 to 313 K (the experimental reaction temperature) over 9 ps, followed by equilibration for 120 ps, at 313 K at a constant pressure of 1 atm.
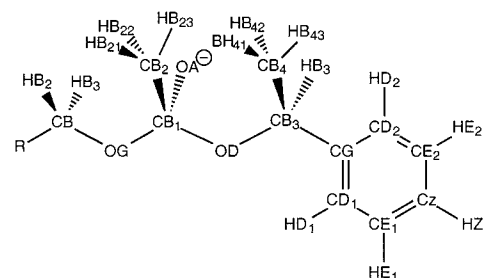
(24) Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J.-K. *Biochemistry* **1989**, *28*, 3629−3637.

(25) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.

(26) Stanton, R. V.; Hartsough, D. S.; Merz, K. M., Jr. *J. Comput. Chem.* **1994**, *16*, 113−128.

(27) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(28) Besler, B. H.; Merz, K. M. J., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431−439.

(29) Steinmetz, A. U. C.; Demuth, H. U.; Ringe, D. *Biochemistry* **1994**, *33*, 10535−10544.

(30) Zacks, A.; Klibanov, A. M. *J.Biol. Chem* **1988**, *263*, 3194−3201.

(31) Klibanov, A. M. *TIBS* **1989**, *14*, 141−144.

(32) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Phys. Chem.* **1983**, *79*, 926−935.

(34) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(35) Cheng, A.; Stanton, R. S.; Vincent, J. J.; Damodaran, K. V.; Dixon, S. L.; Hartsough, D. S.; Best, S. A.; Merz, K. M., Jr. ROAR 1.0, The Pennsylvania State University, 1997.

(36) Stewart, J. J. P. *J. Comput. Chem.* **1991**, *12*, 320−341.

(37) Hoops, S. C.; Anderson, K. W.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 8262−8270.

(38) Case, D. A.; Pearlman, D. A.; Caldwell, J. C.; Cheatham, T. E. I.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M., Jr.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. AMBER 5.0, University of California, San Francisco, 1997.

Table 1. Calculated Charges for Model Compound **I** Using QM/MM ESP Calculations[a]

atom	first QM/MM min	120 ps	atom	first QM/MM min	120 ps
CB	(S) −0.1397	(S) −0.0234	HB41	(S) 0.0024	(S) 0.0779
	(R) −0.0864	(R) −0.1238		(R) 0.0543	(R) 0.0665
HB2	(S) 0.0933	(S) 0.0557	HB42	(S) 0.0171	(S) 0.0179
	(R) 0.0739	(R) 0.1018		(R) 0.0658	(R) 0.0457
HB3	(S) 0.0622	(S) 0.0387	HB43	(S) 0.0830	(S) 0.0339
	(R) 0.0429	(R) 0.0512		(R) 0.0008	(R) 0.0955
OG	(S) −0.2999	(S) −0.3458	CG	(S) −0.0486	(S) −0.1219
	(R) −0.3471	(R) −0.2220		(R) −0.0939	(R) 0.1432
CB1	(S) 0.6157	(S) 0.6437	CD1	(S) −0.0770	(S) −0.0681
	(R) 0.6456	(R) 0.5995		(R) −0.0330	(R) −0.1933
CB2	(S) −0.3827	(S) −0.3297	HD1	(S) 0.1148	(S) 0.1173
	(R) −0.3581	(R) −0.2984		(R) 0.0503	(R) 0.1065
HB21	(S) 0.0759	(S) 0.0684	CE1	(S) −0.0994	(S) −0.1369
	(R) 0.1202	(R) 0.0308		(R) −0.1265	(R) −0.0777
HB22	(S) 0.0652	(S) 0.0311	HE1	(S) 0.1295	(S) 0.1175
	(R) 0.0571	(R) 0.0745		(R) 0.1040	(R) 0.1161
HB23	(S) 0.0940	(S) 0.0805	CZ	(S) −0.1491	(S) −0.1205
	(R) 0.0874	(R) 0.0975		(R) −0.1047	(R) −0.1510
OA	(S) −0.8003	(S) −0.9310	HZ	(S) 0.0990	(S) 0.0790
	(R) −0.8109	(R) −0.8860		(R) 0.1025	(R) 0.0796
OD	(S) −0.5965	(S) −0.4533	CE2	(S) −0.0582	(S) −0.1419
	(R) −0.5650	(R) −0.4672		(R) −0.0813	(R) −0.0802
CB3	(S) 0.5276	(S) 0.5004	HE2	(S) 0.0889	(S) 0.1084
	(R) 0.5150	(R) 0.0856		(R) 0.0980	(R) 0.0899
HB3	(S) −0.0813	(S) −0.0558	CD2	(S) −0.2211	(S) −0.0646
	(R) −0.0808	(R) 0.1012		(R) −0.1654	(R) −0.1627
CB4	(S) −0.2076	(S) −0.2316	HD2	(S) 0.1136	(S) 0.0798
	(R) −0.2267	(R) −0.2678		(R) 0.1149	(R) 0.1283

[a] See Figure 1 for atom labels.

Table 2. Calculated Charges for Asp-32 Using QM/MM ESP Calculations[a]

atom	AMBER	first QM/MM min	120 ps
CB	−0.0303	(S) −0.3453	(S) −0.3534
		(R) −0.3523	(R) −0.3274
HB2	−0.0122	(S) 0.1034	(S) 0.1268
		(R) 0.1034	(R) 0.0737
HB3	−0.0122	(S) 0.0915	(S) 0.0979
		(R) 0.0915	(R) 0.0922
CG	0.7994	(S) 0.7103	(S) 0.7035
		(R) 0.7150	(R) 0.6865
OD1	−0.8014	(S) −0.7494	(S) −0.7810
		(R) −0.7587	(R) −0.7467
OD2	−0.8014	(S) −0.7056	(S) −0.7552
		(R) −0.7116	(R) −0.7117

[a] AMBER charges are given for reference purposes.

Table 3. Calculated Charges for His-64 Using QM/MM ESP Calculations[a]

atom	AMBER	first QM/MM min	120 ps
CG	−0.0012	(S) 0.0974	(S) 0.1274
		(R) 0.0955	(R) 0.1569
ND1	−0.1513	(S) −0.0430	(S) −0.1812
		(R) −0.0679	(R) −0.2278
HD1	0.3866	(S) 0.2939	(S) 0.3700
		(R) 0.3025	(R) 0.4014
CE1	−0.0170	(S) 0.0369	(S) −0.0071
		(R) −0.0161	(R) 0.0150
HE1	0.2681	(S) 0.2827	(S) 0.2963
		(R) 0.2961	(R) 0.2538
NE2	−0.1718	(S) −0.1093	(S) 0.1853
		(R) 0.0366	(R) 0.1038
HE2	0.3911	(S) 0.3042	(S) 0.1605
		(R) 0.2155	(R) 0.1910
CD2	−0.1141	(S) −0.2033	(S) −0.2759
		(R) −0.2219	(R) −0.2660
HD2	0.2317	(S) 0.2461	(S) 0.2528
		(R) 0.2572	(R) 0.2667

[a] AMBER charges are given for reference purposes.

Table 4. Calculated Charges for Asn-155 Using QM/MM ESP Calculations[a]

atom	AMBER	first QM/MM min	120 ps
CG	0.7130	(S) 0.6157	(S) 0.5491
		(R) 0.6223	(R) 0.5861
OD1	−0.5931	(S) −0.5902	(S) −0.5719
		(R) −0.6140	(R) −0.6207
ND2	−0.9191	(S) −0.7413	(S) −0.6941
		(R) −0.7488	(R) −0.6787
HD21	0.4196	(S) 0.4197	(S) 0.4036
		(R) 0.4317	(R) 0.4041
HD22	0.4196	(S) 0.2637	(S) 0.3206
		(R) 0.2761	(R) 0.2645

[a] AMBER charges are given for reference purposes.

Scheme 3. Free Energy Perturbation of the Tetrahedral Complex Carried out in Vacuo

The temperature and pressure were controlled using the methods of Berendsen and co-workers.[39] Periodic boundary conditions and a time step of 1.5 fs were used in all the simulations. In all cases, the bond lengths were constrained using the SHAKE algorithm with a tolerance of 0.0005 Å.[40] The structure obtained at 120 ps was used to recalculate the charges on the groups in the QM region as described previously. The final results are summarized in Tables 1−4. The final charge models were then used for all subsequent MD (one run of 300 ps total) and free energy perturbation (FEP) simulations.

The free energy calculations were carried out using the GIBBS module of AMBER 5.[38] We used the slow growth method, coupled with the dual topology representation for the groups undergoing changes during the FEP simulation.[41] Thus, the topologies for both the *R* and *S* enantiomers were simultaneously defined, and as the slow-growth FEP simulation proceeds the phenyl group of the *S* enantiomer of *sec*-phenethyl alcohol slowly disappears, while the corresponding group

(39) Berendsen, H. J. C.; Potsma, J. P. M.; van Gunsteren, W. F.; DiNola, A. D.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(40) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.
(41) Miller, J. L.; Kollman, P. A. *J. Phys. Chem.* **1996**, *100*, 8587−8594.

of the *R* enantiomer slowly appears. A similar run on intermediate **II** (Scheme 3) was carried out in the gas phase (using gas-phase MNDO-calculated charges) and in DMF to better quantitate the effect the enzyme environment has on the atomic point charges relative to those obtained in the gas phase.

The FEP simulations were carried out over three different time scales (120, 450, and 750 ps) starting from the structure obtained after another 90 ps of equilibration using the charges obtained after 120 ps of MD simulation (see above for details). The reason for utilizing three different FEP simulation lengths was to determine whether the simulations were converged. We also tested convergence by starting the FEP simulations from a different starting structure obtained after another 30 ps of equilibration (120 ps total) on the structure and charges obtained from the 120 ps MD simulation described above. In this case, the calculations were only carried out over two different time scales (450 and 750 ps). Periodic boundary conditions, a 1.5 fs time step, a constant *T* of 313 K, and a constant *P* of 1 atm were used in all FEP simulations.[39] The SHAKE algorithm was used to constrain all bonds with a tolerance of 0.0005 Å.[40]

Table 5. Select Average Distances (Å) and RMS Deviations (Å) for Atoms within Catalytically Important Residues

Ser-221@OG−His-64@HE2		His-64@HD2−Asp-32@0D2	
(*S*) 2.2835	0.3262	(*S*) 2.2300	0.2590
(*R*) 2.3883	0.2732	(*R*) 2.2632	0.2406
Ser-221@OD−His-64@HE2		Asn-155@HD21−Ser-221@OA	
(*S*) 2.6951	0.6366	(*S*) 1.8296	0.0941
(*R*) 3.5654	0.3030	(*R*) 1.8244	0.0992
His-64@HD1−Asp-32@OD1		Asn-155@HD22−Ser-221@OA	
(*S*) 1.9620	0.2384	(*S*) 3.3777	0.1454
(*R*) 1.8126	0.1235	(*R*) 3.4033	0.1351

Discussion

MD Simulations. A primary aim of this study was to examine the structural factors that influence the stereoselectivity of subtilisin in DMF. One we can focus our attention on is the hydrogen-bonding pattern between the substrate and the active site of the enzyme. We determined the average values of the parameters of interest over the last 150 ps of a 300 ps MD simulation. As can be seen from Table 5 and from Figure 2, in the case of the *S* complex, the HE2 of His-64 forms hydrogen bonds with both the OD and OG oxygens of the enzyme/substrate complex. In the case of the *R* complex the hydrogen bond between OD of the substrate and HE2 of His-64 is weakened as indicated in Table 5 and Figure 3. Hence, in the case of the *R* complex this interaction is completely missing during the MD simulation while it is present in the *S* complex. The distance analysis also shows that the interaction between HD1 of His-64 and the two oxygens of Asp-32 are equally favorable for the *S* and *R* complexes.

From the structural analysis we also noticed that once the complex was formed and equilibrated, the phenyl group of the *S* substrate fit very nicely into a hydrophobic pocket defined by residues Leu-126, Gly-127, Gly-128, and Asn-155. In the case of the *R* enantiomer, this region is occupied by a DMF molecule and the phenyl ring is oriented toward the surrounding solvent. The methyl group of the alcoholic moiety of the *R* complex is now pointing into this hydrophobic pocket, and in total, these arrangements contribute to disrupt the catalytically essential hydrogen bond (OD to HE2) to His-64. These observations suggest that, in the transesterification reaction, the *R* alcohol cannot readily donate its proton to the catalytic histidine residue (His-64) which is essential for the catalysis.[42] This, in itself, favors the reactivity of the (*S*)-*sec*-phenylethyl alcohol over that of the corresponding *R* enantiomer.

This hypothesis is further confirmed by an analysis of the RMS fluctuation per residue applied to the active site amino acids. Table 6 shows that Asp-32, His-64, and Ser-221 in the *S* complex were more flexible than the corresponding amino acids in the *R* complex. This higher flexibility may favor reactivity by allowing the catalytic residues to adopt a more favorable orientation for reaction 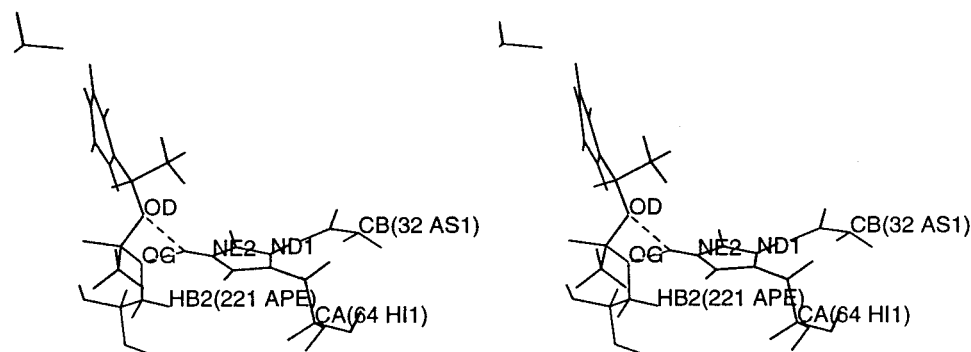in the case of the *S* complex. An examination of the solvent-accessible surface areas (SASA) for the catalytically important amino acid residues (in particular, Ser-221 bound to the intermediate) suggests that, in the case of the *R* complex, the substrate is more exposed to solvent (Table 7). This SASA calculation supports the previous observation of a differential penetration of the solvent into the active site region. In particular, in the case of the *R* complex the solvent occupies the hydrophobic pocket occupied by the substrate phenyl ring in the *S* complex case.

In Table 1 the various ESP charges used in this study for the *R* and *S* substrates are given. We observed significant differences between the QM/MM ESP charges for enzyme-bound *R* and *S* substrates. Upon further reflection we determined that the differences observed in the *R* and *S* substrate charges arise due to the presence of a chiral environment that allows each diastereomeric enzyme/substrate complex to experience a differing electrostatic environment. Thus, the solvent and enzyme atoms can have an "electrostatic influence" on the enzyme enantioselectivity by presenting an electrostatic field that alters the charge distribution of the incoming *R* and *S* substrate molecules through polarization and charge-transfer effects.[43] To further examine this effect, we examined the charge distribution (through QM/MM ESP fitting) in the *R* and *S* substrate/enzyme complexes at a number of points (12 total, every 15 ps) during the last 180 ps of the corresponding MD simulations. From these calculations we determined the average and standard deviations of the charges on each of the QM atoms.

The ESP derived atomic charges show rather substantial variations during the MD trajectories. The instantaneous fluctuations of the partial charges were examined for all the atoms in the QM region. Examples of these variations are given in Figures 4 and 5, where we report the variation versus time of the charges on the oxygen atoms of the tetrahedral intermediates and on the N atoms of the catalytically essential His-64. From the observations of these figures and of the values reported in Tables 8 and 9 we find that the largest differences and deviations are on the atoms which are "involved" in the reactive process and in the H-bond network. For example, the NE2 atom on His-64 has a higher positive charge for the *S* complex case. As we have pointed out in the structural analysis, this atom is bound to HE2, which is involved in H-bond formation with the OG and OD oxygen atoms on the substrate. Interestingly, HE2 itself does not vary too much between the *R* and *S* complexes. The charges on the oxygen atoms (OA, OD, and OG) also appear to be important. From Figure 5 and Table 9 we find that these atoms have a higher concentration of negative charge, in the case of the S complex. This factor helps in the stabilization of the H-bonding pattern in the active site, thereby lowering the energy of the transition state leading to preferential reactivity of the *S* enantiomer. Significant charge variation was also observed for the stereogenic carbon (CB3) along with the bound hydrogen (HB3), methyl carbon (CB4), and aromatic carbon (CG). This reflects the fact that this part of the molecule is placed in differing environments in the *R* (the methyl group is in the hydrophobic pocket, and the phenyl ring is solvent exposed) and *S* (the methyl group is solvent exposed, and the phenyl is in the hydrophobic pocket) cases.

These results also suggest that environmental influences observed in our MD simulations give rise to changes and fluctuations in the calculated partial charges. One very important factor we noticed was the asymmetric charge distribution on the substrate, reflecting the fact that the two enantiomers of *sec*-phenethyl alcohol in the transition state are in a stereodifferentiating environment not only from the steric but also from the electrostatic point of view. These results confirm the suggestion that environmental changes during the course of the simulation play a significant role in causing charge changes and reorganization in the model reacting system.[44]

To investigate this point further, we calculated the atomic point charges of the intermediate model compound **II** in the

(42) Uppenberg, J.; Ohrner, N.; Norin, M.; Hult, K.; Kleywegt, G. J.; Patkar, S.; Waagen, V.; Thorleif, A.; Jones, T. A. *Biochemistry* **1995**, *34*, 16838−16851.

(43) Nadig, G.; Van Zant, L. C.; Dixon, S. L.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1998**, *120*, 5593−5594.

(44) Hartsough, D. S.; Merz, K. M., Jr. *J. Phys. Chem.* **1995**, *99*, 11266−11275.

Figure 2. Active site of the enzyme with the *S* enantiomer bound. The phenyl ring is directed into a hydrophobic pocket.
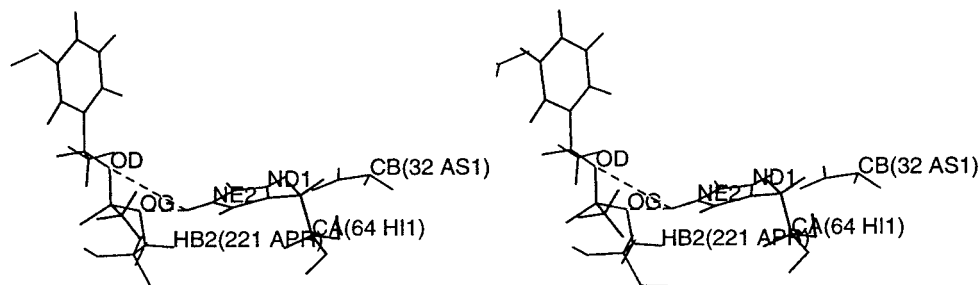

Figure 3. Active site of the enzyme with the *R* enantiomer bound. The phenyl ring is now directed outside, and the hydrophobic pocket is occupied by a DMF molecule.

Table 6. RMS Fluctuation per Residue of the Active Site (Å)

residue	S complex	R complex	residue	S complex	R complex
Asp-32	1.4485	0.9351	Asn-155	1.5107	0.8533
His-64	1.5076	1.0451	Ser-221	1.5016	1.0316

Table 7. SASA for Active Site Residues (Å²)

residue	S complex	R complex	residue	S complex	R complex
Asp-32	2.746	1.427	Asn-155	48.136	50.785
His-64	27.719	33.608	Ser-221	96.137	135.707
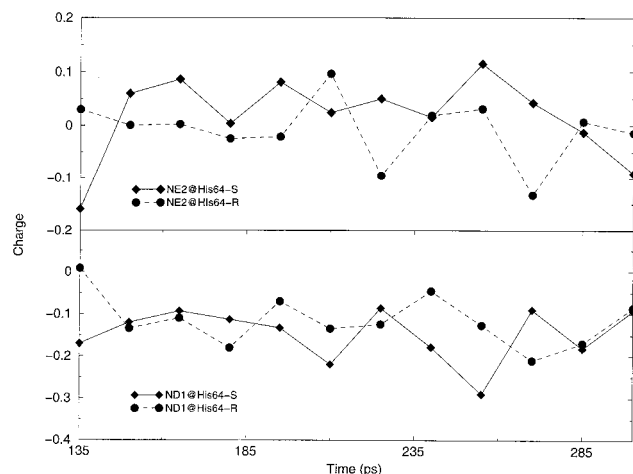
Figure 4. Charge fluctuations versus time for ND1 (bottom graph) and NE2 (top graph).

Figure 5. Charge fluctuations versus time for the oxygen atoms on the substrate. From the bottom to the top we report OA, OD, and OG, respectively. The solid line refers to the *S* complex, and the dashed line refers to the *R* complex.

gas phase (see Scheme 3 and Figure 1), in which the carbon CB1, which has three oxygens (OG, OA, and OD) bound to it, is forced to have a an *S* configuration as is required by the location of the oxyanion hole. As can be seen from Table 10, the differences between the charges calculated in this way are smaller than in the enzyme-bound solvent case. This is particularly true for the atoms highlighted above. Thus, the charge variation for the oxygen atoms OG, OA, and OD is less
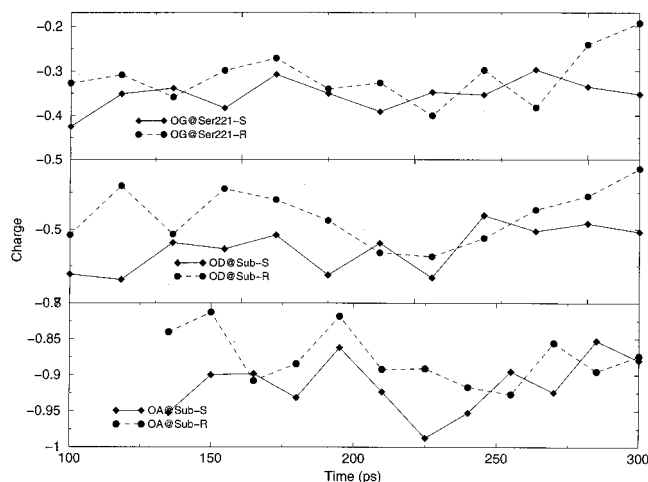
for the gas-phase model complex relative to the enzyme-bound charge set. The most dramatic difference is observed around the stereogenic carbon (CB3), where in the gas phase there is only a modest variation of the charges around this center, while in the enzyme-bound case this region is strongly affected by the surrounding environment. Overall, this clearly indicates that the presence of the enzyme/solvent environment strongly determines the asymmetric distribution of the atomic charges on the substrate.

Free Energy Perturbation Simulations. To place our qualitative observations from the MD simulations on a more quantitative footing, we performed free energy perturbation simulations for the conversion of the substrate **II***(S)* to **II***(R)* in vacuo (with gas-phase calculated charges which are reported in Table 10*)*, and the free energy difference was, as expected, 0.0 kcal/mol. The same perturbation was performed for compound **II** in DMF (again using the gas-phase charges in Table

Table 8.   Averaged QM/MM ESP Charges Calculated for Asp-32, Asn-155, and His-64

	R Asp av	std dev	S Asp av	std dev	charge diff (Δ)
CB	−0.331	0.023	−0.343	0.0.20	0.012
HB1	0.090	0.014	0.100	0.013	−0.010
HB2	0.075	0.019	0.088	0.014	−0.013
CG	0.691	0.018	0.682	0.032	0.008
OD1	−0.735	0.018	−0.745	0.034	0.010
OD2	−0.727	0.015	−0.723	0.029	−0.003

	R Asn av	std dev	S Asn av	std dev	charge diff (Δ)
CG	0.586	0.031	0.568	0.021	0.018
OD	−0.605	0.016	−0.576	0.016	−0.025
ND	−0.698	0.043	−0.682	0.055	−0.017
HD21	0.382	0.032	0.396	0.030	−0.013
HD22	0.299	0.021	0.276	0.031	0.024

	His R chg av	std dev	His S chg av	std dev	charge diff (Δ)
CG	0.082	0.037	0.107	0.042	−0.025
ND1	−0.123	0.065	−0.149	0.061	0.025
HD1	0.346	0.042	0.343	0.035	0.002
CE1	−0.001	0.051	0.029	0.060	−0.031
HE1	0.269	0.016	0.262	0.021	0.007
NE2	0.001	0.064	0.031	0.087	−0.030
HE2	0.259	0.038	0.254	0.054	0.005
CD2	−0.166	0.058	−0.204	0.057	0.038
HD2	0.253	0.016	0.247	0.013	0.005

Table 9.   Averaged ESP Charges Calculated for the R and S Substrates[a]

	sub R chg av	std dev	sub S chg av	std dev	charge diff (Δ)
CB	−0.056	0.062	−0.029	0.063	−0.027
HB2	0.081	0.027	0.064	0.029	0.017
HB3	0.035	0.023	0.042	0.036	−0.007
OG	−0.305	0.061	−0.351	0.033	0.047
CB1	0.588	0.106	0.697	0.101	−0.108
CB2	−0.331	0.065	−0.297	0.067	−0.034
HB21	0.075	0.038	0.045	0.033	0.030
HB22	0.092	0.027	0.046	0.026	0.046
HB23	0.060	0.027	0.063	0.034	−0.003
OA	−0.876	0.036	−0.914	0.038	0.037
OD	−0.462	0.073	−0.543	0.065	0.081
CB3	0.205	0.118	0.464	0.118	−0.258
HB3	0.045	0.059	−0.044	0.038	0.089
CB4	−0.277	0.051	−0.222	0.059	−0.054
HB41	0.052	0.022	0.042	0.027	0.010
HB42	0.065	0.024	0.033	0.031	0.032
HB43	0.095	0.027	0.050	0.017	0.045
CG	0.055	0.071	−0.047	0.075	0.103
CD1	−0.154	0.048	−0.069	0.053	−0.085
HD1	0.095	0.020	0.126	0.019	−0.031
CE1	−0.109	0.054	−0.142	0.049	0.033
HE1	0.101	0.021	0.126	0.021	0.025
CZ	−0.114	0.052	−0.116	0.042	0.001
HZ	0.097	0.020	0.093	0.023	0.004
CE2	−0.107	0.037	−0.131	0.034	0.024
HE2	0.111	0.014	0.100	0.011	0.011
CD2	−0.125	0.051	−0.139	0.045	0.014
HD2	0.117	0.013	0.099	0.015	0.018

[a] See Figure 1 for atom labels.

Table 10.   Charges for Model Compound **II** in Vacuo[a]

atom	S chg	R chg	charge diff (Δ)	atom	S chg	R chg	charge diff (Δ)
CB	0.1575	0.0145	0.143	HB41	−0.0032	0.0494	−0.0526
HB1	−0.0354	−0.0138	−0.0216	HB42	0.0304	0.0034	0.027
HB2	−0.0386	0.0141	−0.0527	HB43	0.0452	0.0659	−0.0207
HB3	−0.0189	0.0372	−0.0561	CG	0.059	−0.0459	0.1049
OG	−0.4392	−0.4056	−0.0336	CD1	−0.1038	−0.0519	−0.0519
CB1	0.8315	0.7976	0.0339	HD1	0.1434	0.2046	−0.0612
CB2	−0.3664	−0.2269	−0.1395	CE1	−0.0912	−0.1546	0.0634
HB21	0.0579	0.0326	0.0253	HE1	0.0804	0.0849	−0.0045
HB22	0.0554	0.0179	0.0375	CZ	−0.1718	−0.116	−0.0558
HB23	0.0594	0.0235	0.0359	HZ	0.0861	0.0738	0.0123
OA	−0.8059	−0.8129	0.007	CE2	−0.0715	−0.1526	0.0811
OD	−0.5847	−0.5965	0.0118	HE2	0.0697	0.0781	−0.0084
CB3	0.4117	0.3895	0.0222	CD2	−0.2351	−0.1457	−0.0894
HB3	−0.0443	−0.0335	−0.0108	HD2	0.1164	0.084	0.0324
CB4	−0.1941	−0.2151	0.021				

[a] See Figure 1 for atom labels.

Table 11.   Calculated Free Energy Difference (kcal/mol) for the Thermodynamic Cycle Given in Scheme 2

starting structure	FEP time (ps)	free energy (ΔΔG*)	starting structure	FEP time (ps)	free energy (ΔΔG*)
90 ps	150	1.5 ± 0.5	120 ps	450	0.8 ± 0.4
90 ps	450	0.6 ± 0.3	120 ps	750	1.3 ± 0.5
90 ps	750	1.2 ± 0.5			

[a] Experimental value: ΔΔG* = 0.4 kcal/mol.

these simulations we obtained $\Delta\Delta G^*$ values of 11 ± 3 kcal/mol (150 ps), 9.2 ± 2.2 kcal/mol (450 ps), and 9.6 ± 2.5 kcal/mol (750 ps), which are all in poor agreement with the experimental value of 0.4 kcal/mol. This is in contrast to the QM/MM derived charges which give $\Delta\Delta G^*$ values that are in much better agreement with experiment (see below). Thus, the importance of taking electrostatic differentiation into account becomes evident in these examples: charge values have a fundamental influence on the determination of $\Delta\Delta G^*$, and using the same charge set for chiral molecules in a stereodifferentiating environment may not give an accurate electrostatic representation.[13]

The thermodynamic cycle used for our FEP simulations was presented in Scheme 2 above. The two enantiomers in the achiral solvent DMF have the same ground-state free energy; thus, ΔG_1 is equal to zero. Hence, the problem of defining enantioselectivity is reduced to the evaluation of the free energy difference between the two diastereoisomeric complexes **I**(S) and **I**(R), in which the alcohols are part of the acyl intermediate.

The results obtained from a series of FEP simulations for this interconversion are given in Table 11. To assess convergence, we ran FEP simulations ranging from 150 to 750 ps (forward and backward), and we find that in all cases the calculated free energy is on the order of 1 kcal/mol. The best estimate comes from the 750 ps runs (from two temporally separated starting structures), which is 1.25 ± 0.5 kcal/mol. The experimental value for this change has been determined to be 0.4 kcal/mol. Thus, our value is in the right direction, but too strongly favors the *S* enantiomer. Nonetheless, the results are in good accord with experiment[23] and lend credence to the qualitative MD results discussed above.

Conclusions

Through the use of FEP calculations and MD simulations we have been able to rationalize the observed enantioselectivity of subtilisin in DMF. A number of qualitative factors come into play in determining enantioselectivity, and through the use of an all-atom modeling approach, we have garnered insights into

10), which resulted in a free energy difference of −0.08 kcal/mol. A third FEP calculation in DMF, but using the 120 ps calculated charges for each of the two complexes (see charges in Table 1; these were slightly modified to ensure that each substrate had a net −1 charge), resulted in a free energy difference of −5.74 kcal/mol, almost entirely due to the differences in the electrostatic contribution to the free energy. Finally, we carried out FEP simulations using the gas-phase charges given in Table 10 in the presence of the enzyme (90 ps equilibration and 150, 450, and 750 ps FEP simulations). From

some of these factors. Steric fit factors play a role in governing enatioselectivity, and in this case the more reactive *S* enantiomer is able to place its phenyl ring into an active site pocket, while the less reactive *R* enantiomer is unable to do so. Because of the steric complimentarity of the *S* enantiomer with the active site of subtilisin, we also observe better hydrogen-bonding complimentarity especially for the atoms involved in catalysis (OA, OG, and OD) and their hydrogen-bonding interactions with His-64.

We also found that electrostatic complimentarity is critical in determining which enantiomer is favored. Generally, it is assumed that the charge distribution of enantiomers is identical in the gas phase or a homogeneous aqueous phase. However, in a stereodifferentiating environment like an enzyme active site this is not necessarily the case. Indeed, in the present example, the stereodifferentiating nature of the active site alters the charges on the *R* and *S* enantiomers to such an extent that their solvation free energy difference in a homogeneous (i.e., DMF) environment is quite large (almost 6 kcal/mol). Clearly, this effect needs to be accounted for when enzyme enantiselectivity is modeled.

We have described one of the first quantitative attempts to evaluate enantioselectivity through molecular modeling methods. Our model and approach have shown promise in predicting the reactivity of biocatalytic systems in nonaqueous environments, and indeed, we can evaluate the major factors (structural and electrostatic) that contribute to determination of enzyme enantioselectivity.