VOLUME 119, NUMBER 42
OCTOBER 22, 1997
© Copyright 1997 by the
American Chemical Society

JOURNAL
OF THE
AMERICAN
CHEMICAL
SOCIETY

The Concept of Solvent Compatibility and Its Impact on Protein Stability and Activity Enhancement in Nonaqueous Solvents

Samuel Toba and Kenneth M. Merz, Jr.*

Contribution from the Department of Chemistry, The Pennsylvania State University, 152 Davey Laboratory, University Park, Pennsylvania 16802

Received February 26, 1997[⊗]

Abstract: Chen and Arnold (*Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5618−5622) have generated a 10 amino acid mutant (PC3) of subtilisin E (SE) that has enhanced activity in mixed DMF/water solvent systems. Through the use of molecular dynamics simulations on both SE and PC3 in water, DMF, and DMF/water (PC3 only) solvent systems, we have provided insights into how nonaqueous solvents affect protein structure and dynamics. On the basis of the observations reported herein, we propose that the PC3 mutant protein is more compatible with DMF as solvent than is the native SE protein. The concept of solvent compatibility embodies the ideas that in order for a protein to be active in organic solvents it must be able to retain its overall shape and that it must not become too rigid such that catalytic activity is compromised. Moreover, neither should the active site region be obstructed by conformational changes that block or structurally alter the active site nor should the active site binding pocket be blocked by solvent molecules (*i.e.*, solvent inhibition). Attempts to determine how each individual amino acid substitution might cause these effects met with mixed results. Clearly, in the present case the individual mutations synergistically lead to the alteration in function through numerous subtle local changes in the structure and dynamics of the protein. Nonetheless, from the simulations, we were able to make some predictions regarding how a protein might be stabilized in a nonaqueous solvent environment.

Introduction

In addition to the important advantage of enhancing the solubilities of organic substrates, organic solvents can lead to novel chemistries in enzyme systems. Interestingly, enzyme-catalyzed reactions that are not possible in an aqueous medium due to kinetic or thermodynamic constraints can become favored in a nonaqueous one. An example of this is the transesterification and selective peptide syntheses catalyzed by serine proteases in organic solvents.[1] Serine proteases have been widely used to carry out specific peptide syntheses via their amidase activities (the thermodynamic reverse of the normal hydrolytic reaction) or, alternatively, via the esterase activity in a kinetically controlled process.[2] With either method, peptide yields are significantly enhanced by carrying out the reaction in the presence of organic solvents because water limits yields through hydrolysis of the acyl enzyme intermediate or the peptide product. Thus, in organic solvents, serine proteases have become promising catalysts for organic synthesis and the preparation of unusual polymers.[3−7]

We can divide organic solvents into two categories, the nonpolar organic solvents and polar organic solvents. The use

⊗ Abstract published in *Advance ACS Abstracts,* August 15, 1997.
(1) Wong, C.-H. *Science* **1989**, *244*, 1145−1152.
(2) Barbas, C. F., III.; Matos, J. R.; West, J. B.; Wong, C. H. *J. Am. Chem. Soc.* **1988**, *110*, 5162−5166.
(3) Margolin, A. L.; Fitzpatrick, P. A.; Dubin, P. L.; Klibanov, A. M. *J. Am. Chem. Soc.* **1991**, *113*, 4693−4694.
(4) Zhong, Z.; Bibbs, J. A.; Yuan, W.; Wong, C.-H. *J. Am. Chem. Soc.* **1991**, *113*, 2259−2263.
(5) Zhong, Z.; Liu, J. L.-C.; Dinterman, L. M.; Finkelman, M. A. J.; Mueller, W. T.; Rollence, M. L.; Whitlow, M.; Wong, C.-H. *J. Am. Chem. Soc.* **1991**, *113*, 683−684.
(6) Patil, D. R.; Dordick, J. S. *Macromolecules* **1991**, *24*, 3462−3463.
(7) Riva, S.; Chopineau, J.; Kieboom, A. P. G.; Klibanov, A. M. *J. Am. Chem. Soc.* **1988**, *110*, 584−589.

S0002-7863(97)00629-X CCC: $14.00 © 1997 American Chemical Society

of very nonpolar solvents for some enzymatic syntheses have resulted in low solubilities for some important reactants and problems associated with enzyme aggregation.[8] On the other hand, polar organic solvents are excellent media for many chemical transformations, but these solvents often drastically reduce both the stability and activity of enzymes.[9,10] This is because polar organic solvents can compete with and disrupt many of the competing noncovalent interactions (*e.g.*, hydrogen bonding, salt-bridge formation, van der Waals contacts) that both stabilize the folded enzyme and determine its ability to stabilize a reaction transition state. Solvent can also adversely affect the partitioning of a substrate into the enzyme's substrate binding pocket by competing with critical noncovalent interactions between the enzyme and its substrate.

Recently, molecular dynamics (MD) simulations have been used to address the structure and dynamics of proteins in organic solvents.[11−16] Among the best-studied enzymes in organic solvents are the serine proteases chymotrypsin and subtilisin. In this work, we have chosen to work with the serine protease subtilisin E from *Bacillus subtilis*. This enzyme has potential applications in peptide synthesis and for use in transesterification reactions in organic solvents.[2,7] The structure and function of the subtilisins has also been well characterized.[17−19] Despite considerable differences in amino acid sequences, the secondary structure elements of the subtilisins and overall folding are very similar. Furthermore, subtilisins have been extensively studied in the past decade to provide insights into both their stability and the mechanism and specificity of their catalysis.[19−21] These studies have helped engineer this protein to obtain properties such as enhanced stability and specificity.[22]

The active site catalytic triad of the subtilisin family is nearly identical to that of the mammalian serine protease chymotrypsin, but their fold is totally different. Subtilisin E is expressed as a pre-pro enzyme, with a presequence of 29 residues responsible for secretion and a prosequence of 77 residues needed for the proper folding of subtilisin E. The mature enzyme is a single polypeptide chain containing 275 amino acids and no disulfide bridges. Once exported and correctly folded, the pre and pro portions are cleaved by a series of autocatalytic reaction. Without the prosequence, subtilisin E will not fold to form a catalytically competent enzyme.[23] The X-ray crystallographic structure of subtilisin E was not available at the start of this project,[24] but the structure of the highly homologous (99%) subtilisin DY[25] from *Bacillus mesentericus* was available and was used to generate a high-quality model of the structure of subtilisin E.

Because little is known regarding the mechanism(s) by which polar solvents reduce the activity of soluble enzymes, random mutagenesis approaches have often been used to improve the catalytic performance of proteins.[26,27] Random mutagenesis and screening has been used to enhance or alter various enzyme features, including enhanced thermal stability,[28−32] alkaline stability,[33] substrate specificity,[34] higher activity in organic solvents,[35] and recovering the catalytic activity of an enzyme damaged by site-directed mutagenesis.[36] Although proteins in organic solvents have been extensively studied,[37] their amino acid sequences and even their three-dimensional structures do not provide us clear insights into how to confer desirable properties on these systems (*e.g.*, enhanced activity or stability in a variety of solvent systems). Recently, random mutagenesis has been used to engineer subtilisin E to function in high concentrations of the polar organic solvent dimethylformamide (DMF). Through sequential rounds of mutagenesis and screening, Arnold *et al*. generated a mutant subtilisin E (identified as PC3) that hydrolyzes a peptide substrate 256 times more efficiently than the wild-type protein in 60% DMF.[35] The 10 amino acids substitutions in PC3 that result in enhanced activity are clustered on one face of the enzyme, near the active site and substrate binding pocket, and all are located in loops that connect the core secondary structure elements together. Moreover, these substitutions exhibit considerable sequence variability in subtilisins from different sources.[20]

From the simulations described in detail below on subtilisin E and its mutant PC3, we hope to provide insights into the mechanism(s) by which activity in polar organic solvents is lost and subsequently recovered in the engineered enzyme. Through an enhanced understanding of these processes, we can begin to rationally formulate design rules for engineering enzymes to function in polar organic media.

Computational Methods

Molecular dynamics (MD) simulations were carried out on subtilisin E and its mutant derivative PC3 beginning from five different starting configurations, which we have labeled SWAT, SDMF, PWAT, PDMF, and PMIX. The details regarding the starting conditions for the five simulations presented herein are given in Table 1. All simulations were

(8) Arnold, F. H. *Protein Eng.* **1988**, *2*, 21−25.
(9) Butler, L. G. *Enzyme Microb. Technol.* **1979**, *1*.
(10) Arnold, F. H. *Current Opin. Biotech.* **1993**, *4*, 450−455.
(11) Toba, S.; Hartsough, D. S.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1996**, *118*, 6490−6498.
(12) Hartsough, D. S.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1992**, *114*, 10113−10116.
(13) Hartsough, D. S.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1993**, *115*, 6529−6537.
(14) Zheng, Y.-J.; Ornstein, R. *J. Am. Chem. Soc.* **1996**, *118*, 4175−4180.
(15) Zheng, Y.-J.; Ornstein, R. L. *Biopolymers* **1996**, *38*, 791−799.
(16) Zheng, Y.-J.; Ornstein, R. L. *Protein Eng.* **1996**, *9*, 485−492.
(17) Markland, J. F. S.; Smith, E. L. *Subtilisins: Primary Structure, Chemical and Physical Properties*; Markland, J. F. S., Smith, E. L., Eds.; Academic Press: New York, 1971; Vol. 3, pp 561−608.
(18) Krait, J. *Subtilisin: X-ray Structure*; Krait, J., Ed.; Academic Press: New York, 1971; Vol. 3, pp 547−560.
(19) Wells, J. A.; Estell, D. A. *Trends Biochem. Sci.* **1988**, *13*, 291−297.
(20) Siezen, R. J.; de Vos, W. M.; Leunissen, J. A. M.; Dijkstra, B. W. *Protein Eng.* **1991**, *4*, 719−737.
(21) Takagi, H.; Matsuzawa, H.; Ohta, T.; Yamasaki, M.; Inouye, M. *Ann. NY Acad. Sci.* **1992**, *672*, 52−59.
(22) Sears, P.; Schuster, M.; Wang, P.; Witte, K.; Wong, C.-H. *J. Am. Chem. Soc.* **1994**, *116*, 6521.
(23) Ikemura, H.; Takagi, H.; Inouye, M. *J. Biol. Chem.* **1987**, *262*, 7859−7864.
(24) Chu, N.-M.; Chao, Y.; Bi, R.-C. *Protein Eng.* **1995**, *8*, 211−215.
(25) Dauter, Z.; Betzel, C.; Genov, N.; Lamzin, V.; Navaza, J.; Schnebli, H. P.; Visanji, M.; Wilson, K. S. *FEBS* **1993**, *317*, 185−188.
(26) Chen, K.; Robinson, A. C.; Van Dam, M. E.; Martinez, P.; Economou, C.; Arnold, F. H. *Biotechnol. Prog.* **1991**, *7*, 125−129.
(27) Chen, K.; Arnold, F. H. *Biotechology* **1991**, *9*, 1073−1077.
(28) Bryan, P. N.; Rollence, M. L.; Pantoliano, M. W.; Wood, J.; Finzel, B. C.; Gilliland, G. L.; Howard, A. J.; Poulas, T. L. *Proteins: Struct., Funct., Genet.* **1986**, *1*, 326−334.
(29) Liao, H.; McKenzie, T.; Hageman, R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 576−580.
(30) Joyet, P.; Declerck, N.; Gaillardrin, C. *Biotechnology* **1992**, *10*, 1579−1583.
(31) Bryan, P. N. *Engineering Dramatic Increases in the Stability of Subtilisin*; Bryan, P. N., Ed.; Plenum Press: New York, 1992; pp 147−181.
(32) Risse, B.; Stempfer, G.; Rudolph, R.; Schumacher, G.; Jaenicke, R. *Protein Sci.* **1992**, *1*, 1710−1718.
(33) Cunningham, B. C.; Wells, J. A. *Protein Eng.* **1987**, *1*, 319−325.
(34) Oliphant, A. R.; Struhl, K. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9094−9098.
(35) Chen, K.; Arnold, F. H. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5618−5622.
(36) Hermes, J. D.; Blacklow, S. C.; Knowles, J. R. *Proc. Natl. Acad. U.S.A.* **1990**, *87*, 696−700.
(37) Wescott, C. R.; Klibanov, A. M. *Biochim. Biophys. Acta* **1994**, *1206*, 1−9.

Table 1. Summary of the Setup of the Starting Configurations

molecule type	SWAT	SDMF	PWAT	PDMF	PMIX
X-ray water	312	50	312	50	312
added water	8917	0	8962	0	4596
added DMF	0	1865	0	1909	1718
added chloride ions	2	2	5	5	5

performed using a parallel version[38] of the SANDER module of AMBER 4.1[39] simulation package in conjunction with the OPLS force field,[40] TIP3P[41] water, and an OPLS model for dimethylformamide (DMF).[42]

The starting coordinates for all heavy atoms were based on the crystal structure of subtilisin DY (Protein Data Bank reference 1MEE).[25] Subtilisin E differs from subtilisin DY only at three amino acid positions (Ala 85 → Ser, Ala 183 → Ser, Ser 259 → Asn) and subtilisin E differs from PC3 by 10 residues (Asp 60 → Asn, Asp 97 → Gly, Gln 103 → Arg, Gly 131 → Asp, Glu 156 → Gly, Asn 181 → Ser, Ser 182 → Gly, Ser 188 → Pro, Asn 218 → Ser, Thr 255 → Ala). The amino acid substitutions from subtilisin DY to subtilisin E and to PC3 were carried out using MIDAS[43] and then minimized using the SANDER module of AMBER, while the remaining protein residues were fixed. After 4000 steps of minimization for both the subtilisin E and PC3 structures, Ramachandran plots[44] that were generated by PROCHECK[45] showed only one (Asp 131) residue in the "disallowed" region in PC3 and no residues in the disallowed region for the subtilisin E structure. The observation of a few residues with ϕ,ψ angles outside of the "allowed" conformational region have been previously observed in a subtilisin crystal.[24] Thus, we decided that our structures were within acceptable limits to begin our simulations. From hereon, the minimized structures of subtilisin E and PC3 were used as the reference crystal structures for our simulations and subsequent analyses.

In the SDMF and PDMF starting configurations, the 50 most tightly bound water molecules (as determined from experimental *B*-values) were retained. This amount of water approximately corresponds to that required for hydration of charged groups and has been found to be roughly the right amount of water required to retain catalytic activity in anhydrous solvent systems.[46,47] In the SWAT, PWAT, and PMIX systems, all of the 312 water molecules observed in the crystal structure were included in the simulation. For the subtilisin E and PC3 systems, three and five chloride ions were introduced, respectively, near the positively charged groups as counterions to neutralize the protein charge. For the PMIX simulation, the initial configuration was prepared using the following procedure. A layer of DMF molecules was placed around the protein before solvating it in water, creating a DMF/water mixture of ~60:40 (volume:volume) respectively. Keeping the protein structure fixed, the resulting DMF/water mixture was optimized and equilibrated for 60 ps at 300 K to insure a random mix. The solvent molecules were subsequently kept fixed, and the protein was minimized for 1000 steps to adjust to the solvent environment. The five resulting starting configurations of SWAT, SDMF, PWAT, PDMF, and PMIX were then subjected to steepest descent energy minimization for 2000 steps with periodic boundary conditions. In all cases, the bond lengths were constrained to their equilibrium values using the SHAKE[48] algorithm with a tolerance of 0.0005 Å.

The minimized starting configurations were then equilibrated by gradually increasing the temperature from 0 to 300 K over a period of ~30 ps by coupling to a temperature bath.[49] A time step of 1.5 fs as well as a constant pressure of 1 atm[49] was used in all the simulations with a residue-based cutoff distance of 12 Å. Both the SWAT and SDMF simulations covered 615 ps, PDMF covered 495 ps, and PWAT covered 705 ps, and the PMIX simulation was run for over 1 ns (1005 ps). Coordinates were saved every 50 steps for all systems. All analyses were done using the last 210 ps of the trajectories.

Results and Discussion

RMS Deviation. The root-mean-squared (RMS) deviations of the instantaneous structures from the starting "crystal" structures are plotted as a function of time in Figure 1 for both the entire protein and for the backbone atoms. The RMS deviations for all of the systems are typical of those seen in protein simulations using the OPLS parameter set.[50] Both the SDMF and SWAT simulations appear to have reached an equilibrium state with respect to RMS deviation by *ca.* 405 ps for both backbone and total RMS deviations. The PDMF and PWAT systems appear to have reached an equilibrium state with respect to RMS deviation by *ca.* 285 and 495 ps, respectively, and PMIX reached equilibrium by *ca.* 795 ps. Why the PMIX simulation required a longer simulation time scale to reach equilibrium with respect to RMS is not clear; however, with the two solvents in the PMIX system, not only is the protein trying to find the most favorable solvent environment but at the same time another competing process exists in which the solvent molecules are attempting to optimize their intermolecular interactions. As the solvent molecules begin to form their most-favored orientations, the positioning of the protein atoms may no longer be in the most energetically favorable position. This results in the protein relaxing further and continuing its search for a new low-energy position. Given the complicated behavior of this system, it is not surprising that it required a longer time scale to equilibrate the PMIX system. The average RMS deviations for the five systems after equilibrium has been reached (*i.e.*, the instantaneous RMS averaged over the last 210 ps of the simulations) are reported in Table 2.

Comparing the five simulations, we find that the RMS deviations values for the protein backbone are lowest for the SWAT system (1.2), followed by identical values for the SDMF and PDMF systems (1.6) and slightly higher values for PWAT (1.8) and PMIX (2.0). The protein in the SWAT simulation had a lower total RMS deviation when compared to the SDMF simulation. This indicates that the SWAT case had a closer resemblance to the X-ray structure which was expected since the starting structure was obtained from a "native-like" X-ray crystal structure. For the three PC3 systems, the protein molecule in the PDMF simulation had the lowest total RMS deviation (1.9), followed by PWAT (2.1) and PMIX (2.3). The higher RMS deviation for the DMF/water (PMIX) system compared to the water (PWAT) or the DMF (PDMF) simulations is intriguing. Recently, Griebenow and Klibanov[51] have shown that proteins have a greater tendency to denature in aqueous organic mixtures than in the corresponding aqueous or pure organic solvents. Our RMS deviation results for the

(38) Vincent, J. J.; Merz, K. M., Jr. *J. Comput. Chem.* **1995**, *16*, 1420–1427.
(39) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, C.; Weiner, P.; Kollman, P. A. *AMBER 4.1*, 4.1 ed.; Pearlman, D. A., Case, D. A., Caldwell, J. C., Seibel, G. L., Singh, C., Weiner, P., Kollman, P. A., Eds.; University of California: San Francisco, CA, 1993.
(40) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(42) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 569–578.
(43) Ferrin, T. E. *J. Mol. Graphics* **1988**, *6*, 13–27.
(44) Ramakrishnan, C.; Ramachandran, G. N. *Biochem. J.* **1965**, *5*, 909–933.
(45) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
(46) Zaks, A.; Klibanov, A. M. *J. Biol. Chem.* **1988**, *263*, 3194–3201.
(47) Klibanov, A. M. *TIBS* **1989**, *14*, 141–144.
(48) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.
(49) 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.
(50) Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1990**, *112*, 2773–2781.
(51) Griebenow, K.; Klibanov, A. M. *J. Am. Chem. Soc.* **1996**, *118*, 11695–11700.

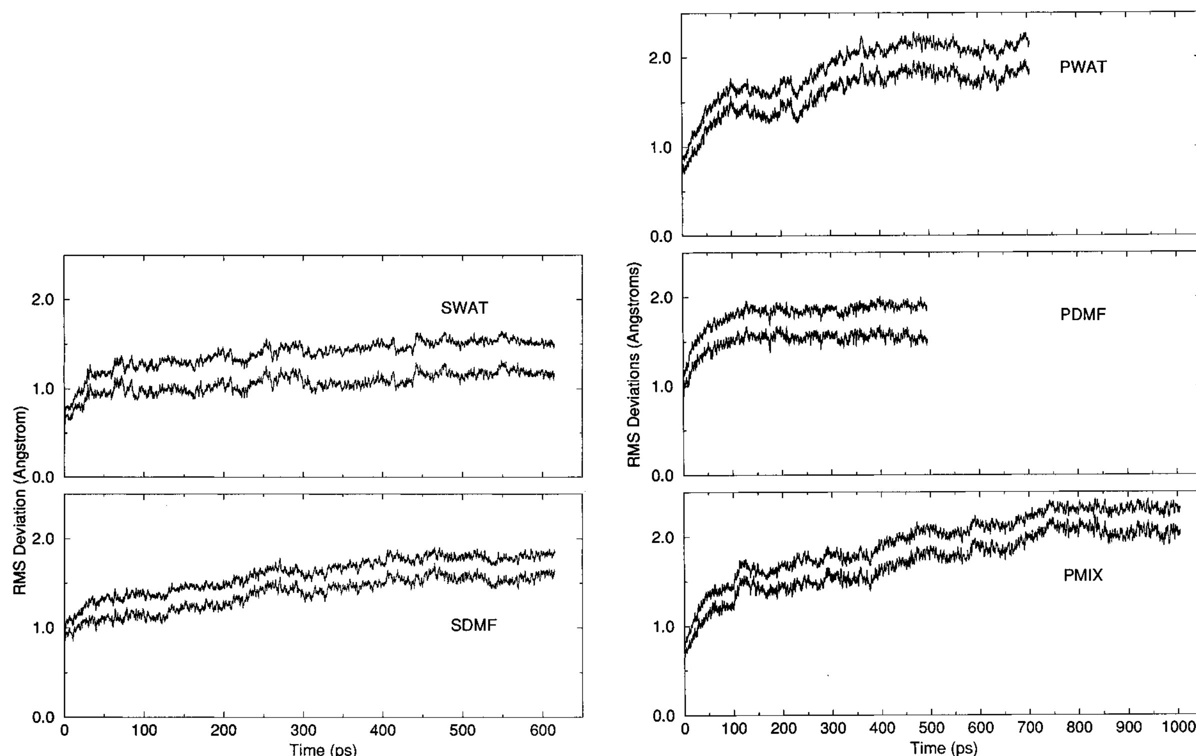

Figure 1. RMS deviation between the instantaneous computed structures and the crystal structure for all residues as a function of time. In all cases, the upper curve is for all atoms, while the lower curve is for backbone atoms only. For average values of the RMS deviations, see Table 1.

Table 2. Summary of the Radius of Gyration, RMS Deviation, RMS Fluctuations, SASA, Solvent Diffusion, and Hydrogen-Bonding Data from the Simulations

	SWAT	SDMF	PWAT	PDMF	PMIX
	Radius of Gyration (Å)				
X-ray	16.79	16.79	16.79	16.79	16.79
average	16.81	16.72	16.81	16.87	16.86
	RMS Deviation (Å)				
total	1.5	1.8	2.1	1.9	2.3
backbone	1.2	1.6	1.8	1.6	2.0
	RMS Fluctuation (Å)				
total	0.70	0.61	0.75	0.67	0.67
backbone	0.59	0.53	0.65	0.56	0.57
	SASA ($Å^2$)				
Ser 221	19.1	2.9	3.2	1.0	0.2
Asp 32	0.6	1.3	4.4	1.6	3.1
His 64	40.4	23.8	41.6	50.5	58.8
10-point mutations[c]	600	578	466	450	432
hydrophobic (%)	26.81	28.23	30.30	31.15	31.23
hydrophilic (%)	73.19	71.77	69.70	68.85	68.77
total ($\times 10^3$)	10.15	9.89	10.24	10.06	10.57
	Diffusion (10^{-5} cm^2/s)				
WAT	4.28	0.58	4.24	0.64	2.44
DMF	–	1.58	–	1.72	1.52
	Hydrogen Bonding				
total	548	569	582	616	532
>90%[a]	128	159	106	140	124
<10%[b]	221	196	244	239	204

[a] Number of hydrogen bonds present >90% of the time. [b] Number of hydrogen bonds present <10% of the time. [c] The SASA in the gas phase for the starting structure is 628.1 for native subtilisin and 584.1 for PC3.

PC3 systems is in agreement with their observation. Nevertheless, the overall average structures obtained from all five of the simulations were not far away from the initial X-ray structure.

Analyzing the average RMS deviations on a per residue basis (see Figure 2), we observe very distinct regions of high RMS and low RMS deviations for all five systems. The high RMS regions share the same trait, that is, they are all located in the loop regions and are positioned on the surface of the protein. Higher deviations in the surface loop regions were also observed in our previous simulation of chymotrypsin in hexane.[11] Surface loops in protein, including loop length and loop sequences, have been known to have little effect on overall protein stability which may afford the high RMS deviations. Surface loop regions have been shown to be able to tolerate significant sequence variability or mutations without impairing the overall fold and functional properties of proteins.[52,53] Chu *et al.* have also observed that the surface loops or turns have higher temperature factors in the crystal structure of subtilisin E.[24] With the exception of Glu 54, the protein loop regions have undergone an increase in RMS deviations on going from the SWAT to the SDMF systems (Figure 2). It is not surprising that Glu 54 is conformationally flexible, since it is a charged surface residue which when placed into the relatively less-polar DMF environment experiences an inward "folding" of the side chain into the protein's interior. This folding behavior was also observed in PDMF and PMIX, where the Glu 54 side chain was oriented inward but, as expected, was oriented outward in PWAT. This suggests that, by mutating Glu 54 to a nonpolar residue, large conformational changes can be avoided in this region of the protein. Selective removal of charged surface residues for proteins in organic solvents may help the protein avoid unfavorable interactions[8,54] in the nonaqueous environment. However, whether this will stabilize or destabilize the protein in nonaqueous environments is unclear, and indeed, conflicting results have been obtained.[22,55] For example, one can argue that this inward folding may also contribute to higher number of intraprotein hydrogen bonding and thus increase protein stability. Glu 54 provides an interest-

(52) Toma, S.; Campagnoli, S.; Margarit, I.; Gianna, R.; Grandi, G.; Bolognesi, M.; DeFilippis, V.; Fontana, A. *Biochemistry* **1991**, *30*.
(53) Hardy, F.; Vriend, G.; van der Vinnie, B.; Frigerio, F.; Grandi, G.; Venema, G.; Eijsink, V. G. H. *Protein Eng.* **1994**, *7*, 425–430.
(54) Arnold, F. H. *Trends Biotechol.* **1990**, *8*, 244–249.
(55) Martinez, P.; Arnold, F. H. *J. Am. Chem. Soc.* **1991**, *113*, 6336–6337.

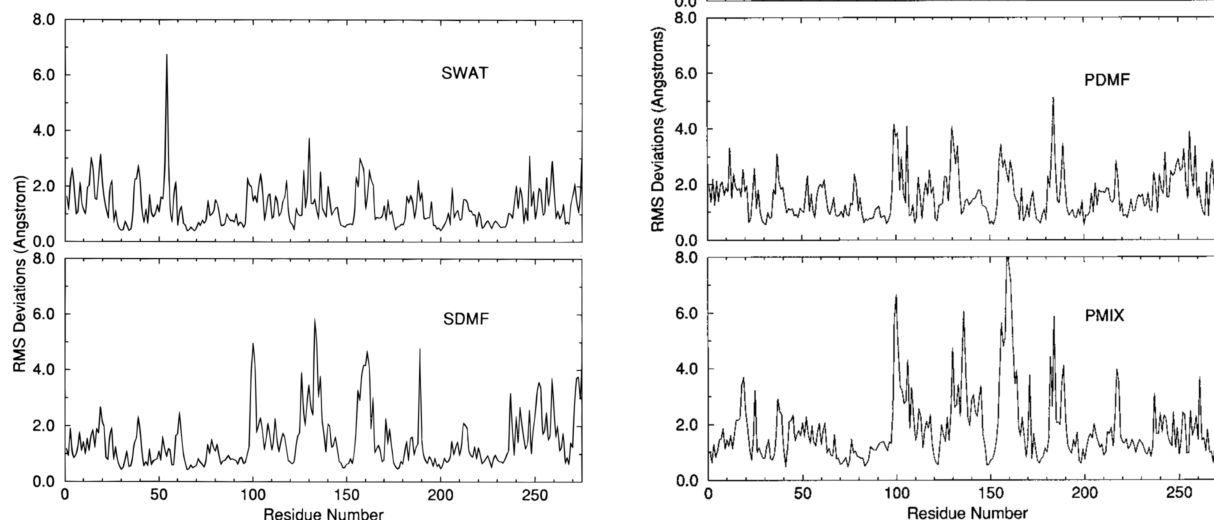

Figure 2. RMS deviations per residue. Averaged over the last 210 ps of each simulation relative to the starting X-ray structure.

Table 3. RMS Deviations and Flexibility of Active Site Residues and RMS Flexibility of Calcium Sites

	SWAT	SDMF	PWAT	PDMF	PMIX
RMS Deviations of Active Site Residues (Å)					
Ser 221	0.93	1.01	1.19	0.85	1.31
Asp 32	0.76	0.76	1.42	0.99	1.45
His 64	0.60	0.74	1.20	1.06	1.15
RMS Flexibility of Active Site Residues (Å)					
Ser 221	0.37	0.45	0.57	0.49	0.47
Asp 32	0.44	0.39	0.49	0.41	0.42
His 64	0.38	0.39	0.49	0.47	0.50
RMS Flexibility of Calcium Sites (Å)					
site A	0.51	0.53	0.64	0.52	0.47
site B	0.44	0.42	0.48	0.46	0.41

ing situation where the modeling gives a clear prediction of a large conformational change which can then be studied experimentally to determine what the outcome of altering the characteristic of this residue will be on protein activity and stability.

Our simulations also show that the RMS deviations for the active site residues are lower than the protein's average (Table 3). This indicates that the active site of the protein in our simulations is very similar to the protein's X-ray structure. This is not surprising since it has been shown that subtilisin can be catalytically active in both water and DMF.[35] The ability of the protein to conserve its active site structure in dissimilar solvents like water and DMF suggests that this ability probably accounts for its catalytic activity in other nonaqueous solvents.[56]

Protein Flexibility. The RMS fluctuations from average coordinates, in any time window, is a good measure of the flexibility of the system during that time period. Thus, we did a structural averaging and then calculated the RMS fluctuations about the time-averaged structure for the last 210 ps of all the simulations. We observe that the total average flexibility of the protein is higher than the average flexibility of the backbone atoms for all five simulations (see Table 2). Both subtilisin and PC3 were more flexible in water than in DMF. This is consistent with previous MD simulations results where several proteins were found to be less flexible in organic solvents.[11−13,16] The identical RMS fluctuation values for PDMF and PMIX imply that the PC3 protein experienced similar flexibility in both DMF and DMF/water mixtures. This is, however, in contrast to earlier reports where an increase in enzyme hydration was accompanied by an increase in enzyme mobility.[57,58] Although, one can use the calculated average RMS fluctuation values to assess the total flexibility of the protein, caution should be taken in interpreting the results. The average results do not distinguish between regions in a protein where flexibility is increased in water while they are decreased in organic solvents (or *vice versa*). The most flexible amino acid group in the SDMF, PDMF, and PMIX is the hydrophobic group of a Phe residue, yet the most flexible group of amino acid in SWAT and PWAT is the hydrophilic group of a Glu residue.

Comparing the RMS fluctuations per residue in all five systems, we observe that the regions of high flexibility within the protein were again mainly located on or adjacent to surface loops (see Figure 3). If we consider a region in a protein with a RMS fluctuation of greater than 1.0 Å as having significant flexibility,[16] we find that there are no long regions in the primary sequence in any of the simulations that meet this criteria. Indeed, if we examine the secondary structural elements (see Table 4), individually none of these have an average RMS fluctuation greater than 1 Å. However, since RMS fluctuations are calculated from the time-averaged equilibrated structure, low RMS fluctuations do not necessarily mean the conservation of a specific secondary structure in a protein. It only provides us with an insight into the flexibility of these elements.

There are more regions of high flexibility (*i.e.*, larger than 1.0 Å RMS fluctuations) in the aqueous systems (SWAT and

(56) Fitzpatrick, P. A.; Steinmetz, A. C.; Ringe, D.; Klibanov, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8653−8657.

(57) Affleck, R.; Xu, Z. F.; Suzuwa, V.; Focht, K.; Clark, D. S.; Dordick, J. S. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 1100−1104.

(58) Burke, P. A.; Griffin, R. G.; Klibanov, A. M. *Biotechnol. Bioeng.* **1993**, *42*, 87−94.

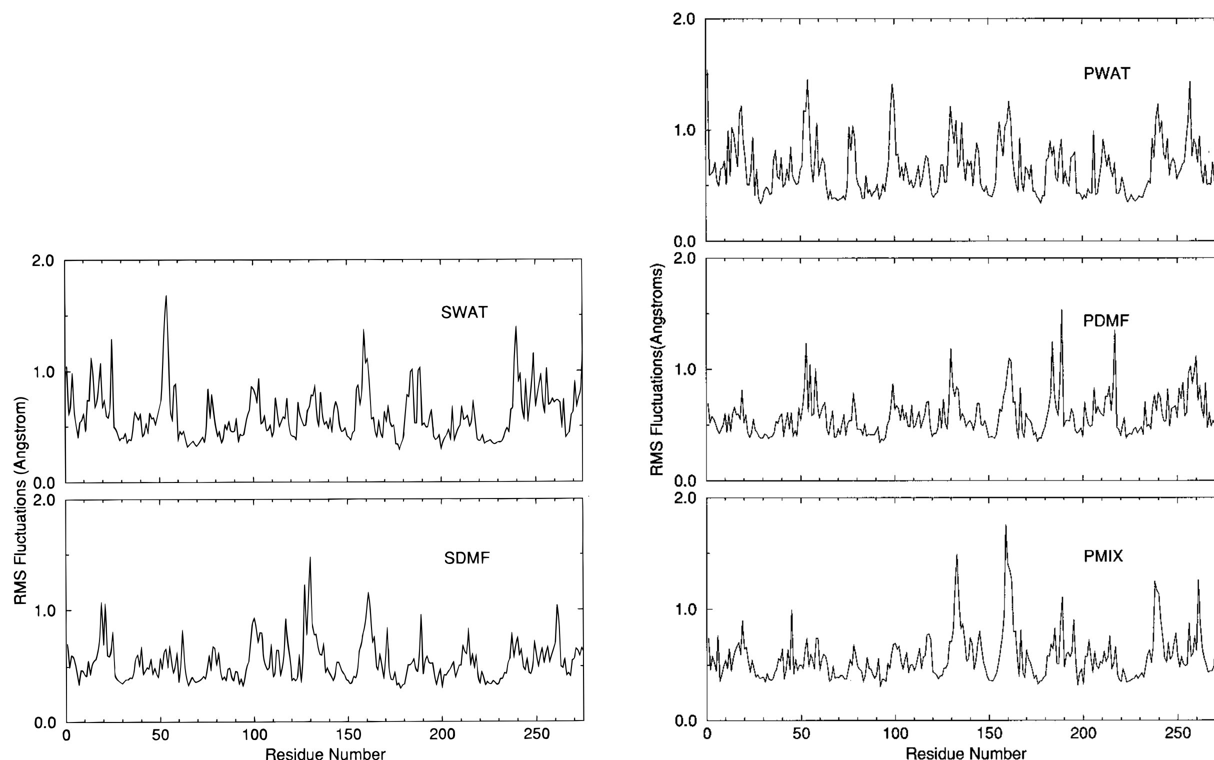

Figure 3. RMS fluctuations per residue from the time-averaged structure. Values averaged over the last 210 ps of each simulation relative to the starting X-ray structure.

Table 4. RMS Fluctuation of Secondary Structures[a]

	RMS fluctuations (Å)				
	SWAT	SDMF	PWAT	PDMF	PMIX
hA 6−11	0.52	0.43	0.60	0.48	0.51
hB 13−17	0.84	0.55	0.81	0.58	0.60
hC 63−73	0.35	0.38	0.40	0.48	0.40
hD 104−116	0.53	0.55	0.58	0.56	0.53
hE 133−144	0.60	0.54	0.75	0.59	0.77
hF 220−236	0.38	0.40	0.43	0.47	0.43
hG 243−253	0.79	0.54	0.70	0.67	0.59
hH 270−274	0.76	0.62	0.75	0.76	0.78
sheet	0.46	0.43	0.49	0.51	0.45

[a] hA to hH are α-helices. The β-sheet regions are comprised of residues 27−32, 43−49, 89−94, 120−124, 148−152, 175−180, 198−201, 206−209, and 213−217.

PWAT) than the DMF and DMF/water systems. In addition, much higher flexibility is observed for the peptide termini in the aqueous systems, while the organic systems experience less. The N-terminus of the protein in the DMF environment is hydrogen bonded to several DMF molecules, which encapsulate and immobilize this region of the protein. This has to do with the fact that DMF can only act as a hydrogen-bond acceptor, which readily allows it to hydrogen bond to the charged N-terminus. This, however, is not found for water in the aqueous phase simulations. The water molecules hydrogen bonded to the N-terminus, having the ability to be both a hydrogen donor and acceptor, can readily exchange with bulk water. The DMF molecules also cannot serve as a proton donor to the C-terminus of the protein. The PWAT, SWAT, and PMIX systems have enough water molecules to solvate the protein C-terminus; thus, a relatively high flexibility in this region is observed. However, with the lack of large quantities of water in the PDMF and SDMF systems the C-terminus folds back and fulfills its hydrogen bonds through intramolecular interactions. This stabilizes the protein C-terminus in DMF and reduces its flexibility. These observations, taken as a whole, suggest that it is unlikely that proteins unravel from the terminal regions in nonaqueous environments. It has been proposed that the presence of excess water in organic solvents acts as a molecular lubricant which results in a higher conformational flexibility of the enzyme resulting in enhanced catalysis.[59,60] From our simulations, we have shed some light on the role of water as well as DMF molecules on the protein. Water is considered as a "molecular lubricant" due to the ability of water to be both hydrogen acceptor and donor and thus allowing the protein to extensively sample conformational space. On the other hand, the ability of DMF molecules only to be hydrogen-bond acceptors may have helped to prevent the onset of unraveling at the protein termini by promoting intramolecular interactions and, thereby, stabilizing the protein in nonaqueous solvents.

The RMS fluctuations for the catalytic triad active site residues in all the five systems are consistently lower than the protein's average (see Table 3). This was also observed in MD simulations of chymotrypsin in water[61] and in hexane.[11] The reduced flexibility of the active site residues further shows the ability of the protein to retain its active site structure and catalytic ability in both aqueous and nonaqueous solvents.

Radius of Gyration. The radius of gyration (R_{gyr}) is defined as the mass-weighted RMS distance of a collection of atoms from their common center of gravity. This provides insight into the relative size of a protein molecule. The radius of gyration values for the five systems were calculated and averaged over the sampling phase of the MD simulations, and the results are given in Table 2.

The results show that the R_{gyr} for subtilisin E in water is similar to the R_{gyr} from the X-ray structure. However, when subtilisin E is placed in DMF (SDMF), it experienced a decrease

(59) Schmitke, J. L.; Wescott, C. R.; Klibanov, A. M. *J. Am. Chem. Soc.* **1996**, *118*, 3360−3365.
(60) Poole, P. L.; Finney, J. L. *Int. J. Biol. Macromol.* **1983**, *5*, 308−310.
(61) Yu, H.-A.; Karplus, M.; Nakagawa, S.; Umeyama, H. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 172−194.

in the R_{gyr}. This decrease in the R_{gyr} is expected for a protein when placed in a nonaqueous solvent. We observed similar behavior in our earlier BPTI simulations where a general reduction in the R_{gyr} of the protein was observed in chloroform relative to that in water.[12,13] The R_{gyr} for PWAT is also similar to the X-ray structure and is identical to the SWAT value. However, both in PDMF and PMIX the protein experienced a slight increase in the R_{gyr} instead of undergoing a reduction in the R_{gyr}. It appears that PC3 has been well designed through selective mutagenesis and screening to take on a different globular size to effectively function in DMF, while in water retaining a size similar to the native subtilisin in water. This is a manifestation of the idea that solvent compatibility is critical to the correct functioning of an enzyme exposed to non-native solvent environments. Thus, in this instance PC3 is able to retain a degree of compactness in DMF that is more reminiscent of the native protein in water. This change is mainly observed in the surface loops of PC3 where the loops are slightly more solvent exposed (extended into the solvent) in DMF and in the DMF/water mixture than in water. This observation along with the increase in the RMS fluctuation on going from SDMF to PDMF and PMIX suggests that PC3 is more flexible than the native subtilisin in DMF. This could be the origin of the increased activity of PC3 relative to the native protein in DMF.

It is however, important that the amino acid changes that are made in a protein do not induce unexpected large changes in the overall structure. A prerequisite for a redesigned active enzyme function is maintenance of its folded globular structure. Although we observed a change in R_{gyr} from native subtilisin to its mutant PC3, the change is very small and is within the reasonable fluctuations of the protein in both water and the nonaqueous environment. The protein remains spherical, and no unwinding or denaturation was observed.

Solvent Accessible Surface Area. Solvent accessible surface area (SASA) is another indicator of how the surrounding medium affects protein structure and dynamics. In our simulations, the calculated SASA values were determined using an algorithm developed in our laboratory[62] with a probe radius of 1.4 Å. The total SASA for the native subtilisin is lower in the SDMF simulation than in the SWAT system. This is in accord with our R_{gyr} calculation where we observed the native subtilisin to be more compact in DMF than in water. The total SASA for PC3 is lowest for PDMF followed by PWAT and is highest for the PMIX system (see Table 2). To further analyze these systems, we separated the hydrophilic (polar) amino acids into three groups: the neutral groups, the basic (hydrogen donor) groups, and the acidic (hydrogen acceptor) groups. In DMF, the acidic polar surface residues of the protein tend to be folded inward due to the repulsive interaction as DMF can only act as hydrogen-bond acceptor. The SASA for the hydrogen-bond acceptor residues Asp and Glu accounts for 7% (666 $Å^2$) of the total SASA in the SWAT system but only accounts for 4% (396 $Å^2$) in the SDMF system. Similarly, Asp and Glu residues account for 6% (605 $Å^2$) in the PWAT system but only 4% (405 $Å^2$) in both the PDMF and PMIX systems (Figure 4). The basic and the neutral groups for all the five systems remained relatively the same.

By separating the total SASA into hydrophilic and hydrophobic categories, we find that the SASA value for hydrophobic residues in both native subtilisin and PC3 in DMF have experienced a marked increase in SASA when compared to the hydrophilic residues (see Table 2). Many water-soluble proteins, such as subtilisin, have highly hydrophilic surfaces with

(62) LeGrand, S. M.; Merz, K. M., Jr. *J. Comput. Chem.* **1993**, *14*, 349–352.

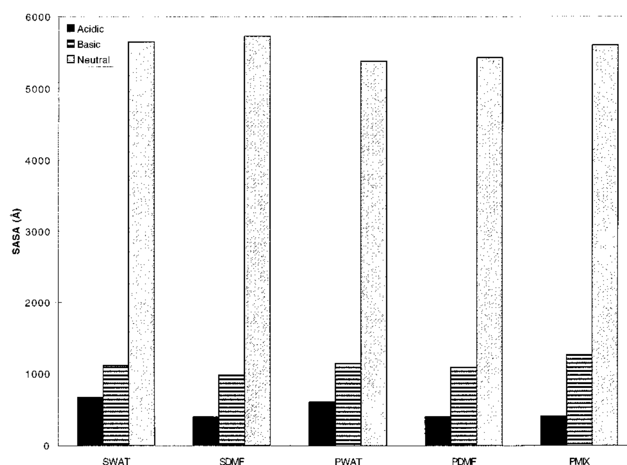

Figure 4. SASA for acidic residues (Asp and Glu) and basic residues (Lys, Arg, His). There is a decrease in SASA for acidic residues in the DMF environment but the SASA for the basic residues remained relatively constant in DMF.

numerous charged residues that are solvated by water. Transfer to a medium that is less effective in solvating charges could lead to unfavorable interactions as these charges want to become neutralized[63] or search for better solvation through protein aggregation or misfolding. Thus, it is not surprising when we consider that subtilisin E has 129 hydrophilic residues (46.9%) and the PC3 mutant has 125 (45.5%) hydrophilic residues. Hence, when either subtilisin or PC3 was placed in DMF, the inward orientation for hydrophilic residues and the outward folding (solvent exposed) hydrophobic residues would change the hydrophobic to hydrophilic SASA ratio. Moreover, PC3 (net charge +1) has two less negative charges than does the native protein (net charge −2). It has been proposed that the replacement of noncritical charged surface residues by uncharged amino acids would stabilize a protein for use in nonaqueous solvents by rendering the protein surface less hydrophilic and presumably less dependent on solvation by water for proper folding.[8,54,55]

Analyzing the SASA for the catalytic triad (see Table 2) indicated that Asp 32 remained buried in all five simulations. For Ser 221, there was a significant decrease in the SASA from SWAT to SDMF, and it remained relatively buried in all the PC3 systems. The change from the SWAT system to the PWAT system resulted in the burial of Ser 221. However, the burial of Ser 221 does not greatly affect catalysis since both these enzymes have been shown to be catalytically active in these two solvents.[35] The burial of Ser 221 can not be attributed to a single or directly observable change in the neighboring residues. Rather, many small contributions from many neighboring residues make Ser 221 less solvent accessible.

On the other hand, the change in SASA for His 64 is due to several large contributors. His 64 has a SASA of 40.4 $Å^2$ in the SWAT simulation, but is reduced by 40% in the SDMF simulation. In the PC3 mutant this reduction in the SASA is not observed on going from the aqueous phase simulation (PWAT) to the DMF and DMF/water simulations (PDMF and PMIX, respectively). Amino acid substitutions near or in the enzyme's active site and substrate binding pocket have been demonstrated to affect catalysis.[27] One of the amino acid mutations is at position 60 which is located near to His 64 of the catalytic triad. The native subtilisin to PC3 mutation at position 60 involved the replacement of a negatively charged residue (Asp) by a neutral one (Asn), which altered the charge

(63) Zheng, Y.-J.; Ornstein, R. L. *J. Am. Chem. Soc.* **1996**, *118*, 11237.

environment and conformational tendencies of the active site near His 64 in particular. Asp 60 is near the surface and is adjacent to Tyr 59, a surface residue. In the SWAT simulation, Tyr 59 is extended into the solvent and the side chain oxygens of Asp 60 form hydrogen bonds with the amide hydrogen on the His 64 backbone (1.88 and 2.19 Å, respectively). In the SDMF simulation, to avoid the unfavorable DMF solvent environment, Tyr 59 buries itself into the protein. This burial of Tyr 59 (as well as the change in the solvent dielectric) indirectly affects Asp 60 by shortening the hydrogen bonds between Asp 60 and His 64 (1.59 and 1.65 Å, respectively). Another surface residue, Tyr 217, which is also situated near His 64, is extended into solvent in SWAT but is buried in SDMF, which further obstructs the solvent accessibility of His 64. Both Asp 60 and Tyr 217 affect the burial of His 64 in the SDMF system which renders it less accessible to potential substrate molecules and, thus, partially accounts for the significant reduction in catalytic activity of subtilisin E when placed in DMF. However, this was not observed in the PC3 system due to the local change resulting from the Asp 60 → Asn mutation. This mutation eliminates the hydrogen bonds between Asp 60 and His 64, which in turn allows His 64 to better interact with the hydroxyl oxygen of Ser 221. Thus, this is another example of how the mutations in PC3 generate an enzyme that is more compatible with a solvent system that contains DMF. From these observations, we predict that the maintenance of the solvent exposure of His 64 is necessary to retain the catalytic activity of subtilisin.

The total SASA for the 10 mutations in the native subtilisin and PC3 systems have decreased significantly when compared to their respective starting structures (see Table 2). Interestingly, for the native subtilisin the percent reduction in the SASA was 4.5 and 8.0% for SWAT and SDMF, respectively. On the other hand, for PC3 it was 20.3, 23.0, and 26.0% for PWAT, PDMF, and PMIX, respectively. Thus, on a percentage basis, relative to the starting structure in the gas phase, the PC3 mutant experiences a larger reduction once placed in any solvent system. It has been suggested that reduction in the SASA of solvent-exposed hydrophobic residues in the active site region results in increased thermal stability of proteins in water.[64,65] Arnold *et al.* have shown that the improved stability of these proteins appears to be connected with enhanced hydrophobicity, reduced surface area, and increased packing density of the interior core.[10] Similarly, the mutations from subtilisin E to PC3 are all located on the variable loops near the active site and substrate binding pocket. Our study shows that the net effect of the mutations is the reduction of SASA near the active site and on the protein surface. This reduction in SASA near the active site makes it less solvent exposed and possibly enhanced the packing of the active site region and reduced substrate−solvent competition (as described below). This reduction in SASA and enhanced packing may in turn help maintain protein stability when placed in an organic solvent such as DMF.

Active Site Analysis. The active site of subtilisin contains the classic catalytic triad of a serine protease, and in this case, it consists of Ser 221, His 64, and Asp 32. In addition to the effect of water molecules with direct contact with the protein surface or with a particular part of the protein (such as the substrate binding cleft), we were also interested in the extent of solvent penetration into the protein interior, especially into the active site "pocket". Polar organic solvents have been shown to be capable of not only interacting with water molecules at the protein surface but also in the active site.[66] Solvent binding in the active site region could possibly compete with the substrate and, thus, become a competitive inhibitor of subtilisin. This has been observed in the cross-linked crystal of subtilisin where four organic solvent molecules were found to be bound in the active site region.[56,67]

To effectively sample the active site pocket, we analyzed the solvent penetration within a 5 Å cutoff radius from the oxygen at Ser 221. There were no water molecules present in the active site pocket in both the crystal structure and our starting configurations. However, during the SWAT and the PWAT simulations the active sites were solvated by an average of seven and five water molecules, respectively. In the SDMF and the PDMF simulations, DMF molecules also diffused into the active site. There is at least one DMF molecule residing in the protein active site at any given time in the SDMF and PDMF trajectories. In the PMIX simulation, we observed an interesting solvation pattern. With the solvent competition between DMF and water in the PMIX simulation, we noted that the water molecules (on the average three molecules) seem to have preferentially diffused into the active site. There was only one DMF molecule present in the active site of the protein in the PMIX simulation, and it was present for <30 ps total. This was also observed in our previous simulation of chymotrypsin in hexane, where no hexane was observed in the active site, but instead between two and four water molecules resided in the pocket throughout the simulation.[11]

We further observed that the water molecules within the active site of the proteins in the PMIX, PWAT, and SWAT simulations were able to readily exchange with the water molecules in the bulk solvent. For both SDMF and PDMF, the DMF solvent molecules were "bound" to the active site region and, therefore, could compete with the substrate and possibly inhibit subtilisin. In PDMF, a DMF solvent molecule (DMF-1656) was present in the active site for the entire trajectory and occasionally another DMF solvent molecule (DMF-1450) would also be found (<30 ps) in the active site. However, in SDMF only one DMF solvent molecule (DMF-1488) was present for the entire length of the trajectory. This binding of a protein active site by an organic solvent has also been observed experimentally. Sampaio *et al.* have suggested the possibility of competitive inhibition by an organic solvent (toluene in this case) for subtilisin.[68] Another advantage of having water molecules in the active site of the proteins in the SWAT, PWAT, and PMIX simulations is the ability of water to both act as a hydrogen-bond acceptor and donor. This ability of water molecules to act as a "hydrogen-bond bridging group" helps fulfill intraprotein hydrogen bonding in the active site, thereby stabilizing the protein as well as making it unfavorable for DMF molecules to be present and compete for these interactions. Moreover, water molecules in the active site may allow substrate molecules to readily access this region because of their ability to both accept and donate hydrogen bonds to an incoming substrate molecule. DMF, on the other hand, with its ability to only accept hydrogen bonds may more effectively inhibit access to active site pockets by an incoming substrate molecule.

Our average distance calculations for the hydrogen bonds of Asp 32, His 64, and Ser 221 indicate that the interaction between Asp 32 and His 64 is stronger than that between His 64 and

(64) Wigley, D. B.; Clarke, A. R.; Dunn, C. R.; Barstow, D. A.; Atkinson, T.; Chia, W. N.; Muirhead, H.; Holbrook, J. J. *Biochim. Biophys. Acta* **1987**, *916*, 145−148.

(65) Meng, M.; Bagdasarian, M.; Zeikus, J. G. *Biotechnology* **1993**, *11*, 1157−1161.

(66) Economou, C.; Chen, K.; Arnold, F. H. *Biotechnol. Bioeng.* **1992**, *39*, 658−662.

(67) Fitzpatrick, P. A.; Ringe, D.; Klibanov, A. M. *Biochem. Biophys. Res. Commun.* **1994**, *198*, 675−681.

(68) Sampaio, T. C. d.; Melo, R. B.; Moura, T. F.; Michel, S.; Barreiros, S. *Biotechnol. Bioeng.* **1996**, *50*, 257−264.

Table 5. Calculated Average Distances

	hydrogen bonding distances between active site residues Ser-His-Asp (Å)				
	SWAT	SDMF	PWAT	PDMF	PMIX
His 64 ND1−Asp 32 OD1	2.74	2.77	3.83	2.88	3.06
His 64 ND1−Asp 32 OD2	3.34	3.27	2.68	3.02	2.93
His 64 NE2−Ser 221 OG	4.16	4.06	3.69	3.23	3.64

distance (Å)	SWAT	SDMF
218@OD1-204@HN	1.64	2.78
219@HN-202@O	1.70	2.10

distance (Å)	PWAT	PDMF	PMIX
218@OG-204@HN	2.81	3.05	3.72
218@OG-205@HN	2.18	1.85	1.91
219@HN-202@O	2.48	1.73	1.75

Ser 221 (see Table 5). This was also observed experimentally and indicates that the hydrogen-bonding distances between the active site residues for subtilisin and PC3 are within the expected range. The X-ray crystallographic study of Matthews *et al.* reveals[69] that there is a hydrogen bond between Asp 32 and His 64, whereas there are none between the side chains of Ser 221 and His 64, or it is severely distorted in native subtilisin.

In our simulations, we further found a hydrogen bond that is present in the active sites of SDMF and PDMF only. In SDMF, the side chain amide hydrogen of Asn 155 is oriented toward the active site forming a hydrogen bond with the hydroxyl oxygen of Ser 221. This close proximity of Asn 155 to Ser 221 was not observed in the SWAT, PWAT, and PMIX systems. The residue Asn 155 is adjacent to the solvent-exposed residue Glu 156. In SWAT, the Glu 156 side chain is extended into the solvent but in SDMF the side chain folds inward into the protein. This inward folding of Glu 156 in SDMF pushed Asn 155 into close proximity with Ser 221 and, hence, the formation of the hydrogen bond. With the mutation of Glu 156 to Gly in PC3, one would expect that there would no longer be any unfavorable solvent interaction and no inward folding. However, in PDMF, the presence of the carbonyl oxygen of a DMF molecule (DMF 1450) may have helped to stabilize the amide hydrogen of Asn 155 and allow it to form a hydrogen bond with Ser 221. This hydrogen bond formed only in the neat DMF simulations, and it can be imagined that it further reduces the catalytic activity of Ser 221 in these environments. Hence, the presence of a DMF molecule or the presence of Glu 156 can cause a blocking of the active site of the enzyme further diminishing catalytic activity.

Hydrogen-Bonding Interactions. In the hydrogen-bonding analysis, the criteria used to determine the presence or absence of a hydrogen bond were solely geometric. All polar hydrogens were identified, and distances to potential hydrogen bond acceptors were calculated. A distance cutoff of 2.6 Å between the donor and the acceptor and an angle cutoff between 120° and 180° at the hydrogen atom were applied. The criteria used is the same as that previously reported by Hartsough and Merz.[12,13]

Experimentally, it has been shown that mutations enhancing catalytic efficiency can occur without corresponding changes in stability, which may indicate that activity and stability are not coupled traits.[70] Hydrogen bonds can be considered as one of the main contributing factors to the stability of the enzyme in different solvation environments. Thus, examining these interactions in detail can give us insight into the stability of a given protein in a given solvent environment.

In a study of a single Asn 218 to Ser mutation in subtilisin BPN′, the mutant variant has been shown to exhibit enhanced thermal stability.[28] Asn 218 to Ser is also one of the 10 mutations from subtilisin E to PC3. In our simulation, we observed that the strong hydrogen bonds formed by the side chain of Asn 218 (204@HN-218@OD1) and its neighboring residue (219@HN-202@O) in SWAT was significantly weaker in SDMF (see Table 5). This loss of hydrogen-bond interactions may partially account for the loss of catalytic activity and destabilization of subtilisin E in organic solvent. However, the Asn 218 → Ser mutation in PC3 has allowed the protein to recover the lost hydrogen bonds. The hydrogen bond found in SWAT between the side chain of Asn 218 and the amide hydrogen of Ser 204 has been replaced by Ser 218 and Ile 205 in the PC3 systems. On the basis of the distance, the hydrogen bonds between 218@OG-205@HN and 219@HN-202@O were especially strong in the PDMF and PMIX system (see Table 5). This illustrates one of the many subtle ways in which the loss of catalytic activity of subtilisin E in DMF could possibly be regained by the mutant PC3 in DMF and DMF/water mixtures.

We observe that, in general, the total number of hydrogen bonds is greater for native subtilisin E in a DMF environment versus an aqueous one. This is more clearly seen when we separate the strong (present >90% of the time) from the weak (present <10% of the time) hydrogen bonds. It can be clearly seen that, for both subtilisin E and PC3, as the protein is moved from an aqueous to a DMF environment the number of intramolecular hydrogen bonds increases (see Table 2). Interestingly, but not unexpectedly, the PMIX simulation gave a value that was intermediate between those of the PWAT and the PDMF simulations. These data, taken together, clearly indicate the preference for formation of more intramolecular hydrogen bonds within a protein when it is placed in an organic solvent. The increase in intramolecular hydrogen bonds was also observed in our previous simulations of BPTI in chloroform[12,13] and chymotrypsin in hexane.[11] However, for PC3 in water the number of long-lived hydrogen bonds is less than that of the native protein in water (128 *versus* 106) making it a more flexible system, as evidenced by the total RMS fluctuation of this system (see Table 2). Addition of neat DMF increases the number of long-lived hydrogen bonds above that observed for the native protein in water, but the PMIX simulation gives a value close to the native protein (128 *versus* 124). This again highlights the concept of solvent compatibility because, while it is advantageous in some regards to have an increase in the number of strong hydrogen bonds, it is important to not have so many that the structure and the dynamics are altered too excessively when the protein is placed in an organic solvent. The PC3 mutant appears to satisfy this criteria. Overall, we conclude that the formation of more long-lived intraprotein hydrogen bonds is true for a protein in nonaqueous solvents whether they are polar or nonpolar, but that the formation of too many of these types of interactions may be detrimental to the structure, function, and dynamics of a protein in an organic solvent. Thus, a delicate balance between stability and activity must be reached in order to obtain an optimally active protein in organic solvents.

If we consider the kinetic stability of a protein in a nonaqueous solvent as being correlated to its activity in the same solvent, Chen and Arnold[35] have found that 43% of the activity of PC3 remains after 460 h in 70% DMF *versus* 49% of the activity left for the subtilisin E. This indicates that the wild-

(69) Matthews, D. A.; Alden, R. A.; Birktoft, J. J.; Freer, S. T.; Kraut, J. *J. Biol. Chem.* **1977**, *252*, 8875−8883.
(70) You, L.; Arnold, F. H. *Protein Eng.* **1996**, *9*, 77−83.

type protein is more stable than PC3. Our hydrogen bond analysis results are in agreement with these experimental findings, assuming that we can correlate stability with the number of hydrogen bonds present (we suspect that this correlation is likely to be better in a nonaqueous solvent than in an aqueous one). Thus, SWAT gave 128 strong hydrogen bonds (>90%) *versus* 106 for PWAT, while SDMF gave 159 *versus* 140 for PDMF. Hence, the native protein seems to have an inherent tendency to form more long-lived strong hydrogen-bond interactions in water and DMF. Even though we have not carried out simulations on both proteins in 70% DMF, it is clear that PC3 will more than likely have a reduced number of strong hydrogen bonds when compared to the native protein.

Conclusions

A protein in water has several important characteristics that allow for catalytic activity. It must first of all fold into an active structure that does not alter significantly with time under normal conditions.[71,72] The enzyme must also not be too rigid such that it becomes difficult to bind and turnover substrate molecules.[71,72] Moreover, the active site should not become obstructed during the course of substrate binding, catalysis, and product release.[71,72] From previous work on proteins in nonaqueous solvents, we observe that proteins under these conditions retain a structure that is related to the folded structure in aqueous solution.[11−13] At the same time, however, the protein molecules undergo significant structural alterations that include a reduction in size (as estimated by R_{gyr}), undergo numerous subtle local structural distortions (due to the interaction of polar and charged solvent exposed side chains with the protein matrix), and become more rigid due to an increase in long-lived hydrogen-bonding interactions.[11−13] These effects are not necessarily compatible with an highly active protein. Since these deleterious effects arise from a change in solvent environment, what is needed is to alter the properties of a protein such that it becomes more compatible with the solvent environment of choice.

From our simulated results, we observe that PC3 in a number of ways has been modified such that it has become more compatible with the polar nonaqueous solvent DMF. Firstly, amino acid side chains that are incompatible with DMF have been removed. Thus, three carboxylate bearing side chains have been removed, while only one carboxylate group has been introduced. As noted above DMF can not stabilize negatively charged side chains, which causes them to become embedded into the protein matrix in order to form stabilizing hydrogen-bond interactions. When, at the same time, some of the solvent incompatible residues are eliminated, new solvent compatible groups are introduced. For example, an Arg residue was introduced at position 103, and this residue, because of its ability to donate hydrogen bonds, is relatively (*e.g.*, to a Glu residue) compatible with the hydrogen-bond-accepting capabilities of DMF. Indeed, in a newer mutant generated in the Arnold group, another Arg residue has been introduced (Ala 48 → Arg) as well as more hydrophobic residues (Ile 107 → Val and Gln 206 → Leu).[70] Because negatively charged groups are DMF incompatible, this can cause major unfavorable structural changes. Take, for example, the case of Glu 156 which is mutated to Gly in PC3. The rearrangement of Glu 156 in DMF due to its solvent incompatibility causes distortions within the active site of subtilisin E, which are eliminated in PC3.

Native subtilisin E in water and DMF has several defining features that we have quantified in this study (see Table 2). For example, the R_{gyr} on going from water to DMF decreases (16.81 → 16.72), the total RMS fluctuation decreases (0.70 → 0.61), and the number of strong hydrogen bonds increases (128 → 159). This is consistent with what we have observed before.[11−13] PC3, in comparison to subtilisin E in water behaves quite differently. Thus, while the R_{gyr} is the same (16.81), the total RMS fluctuation increases (0.7 → 0.75) and the total number of long-lived hydrogen bonds decreases (128 → 106). This suggests that PC3, in water, is a more flexible and dynamic protein than is its native precursor. Nonetheless, it retains good activity in water.[35] The interesting comparison arises when we look at PC3 in the three solvent systems. The R_{gyr} increases when placed into DMF and DMF/water relative to the native and mutant protein in water. The RMS fluctuations decrease relative to PC3 in water, but the values obtained in DMF and DMF/water (0.67) are close to the value obtained for the native protein in water (0.70). Moreover, the number of hydrogen bonds increases on going from water to DMF and DMF/water, but for the DMF/water simulation the number of strong hydrogen bonds (124) is similar to that observed for the native protein in water (128). Taken together, it appears that PC3 has recovered many characteristics of the native protein in aqueous solution in a mixed DMF/water environment. Hence, the concept of solvent compatibility.

We have examined each point mutation site individually in an attempt to establish basic protein design principles in each case. As shown in the text, we were successful in several instances, but in general we have found it difficult to make individual predictions. While some design principles are obvious in general (*e.g.*, remove negatively charged side chains when DMF is used as a solvent), it becomes difficult to decide how to implement them in practice. Thus, which of the Glu and Asp residues to remove becomes a difficult question that potentially can only be adequately addressed through alteration of all residues of this type. Thus, at this point in time, we can say that we have made some observations which form the basis of a general prediction (*i.e.*, solvent compatibility), but how to implement this in a specific sense (*i.e.*, which positions to mutate) is not clear.

Acknowledgment. We thank the Office of Naval Research for support of this research through grant no. N00014-90-3-4002. The Pittsburgh Supercomputer Center and the Cornell Theory Center are acknowledged for generous allocations of supercomputer time through a MetaCenter grant. We thank Dr. Greg Farber and Dr. Frances Arnold for helpful discussions.

Supporting Information Available: Analyses of solvent diffusion, secondary structure, and calcium binding (11 pages). See any current masthead page for ordering and Internet access instructions.

JA970629A

(71) Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5250−5254.
(72) Shoichet, B. K.; Baase, W. A.; Kuroki, R.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 452−456.