Reprinted from the Journal of the American Chemical Society, 1995, *117*.

Molecular Dynamics Simulations of $H_2NSO_2C_6H_4CONH(Gly)_3OBn$ Bound to the Active Site of Human Carbonic Anhydrase II

Donovan N. Chin and George M. Whitesides*

Contribution from the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

Received November 7, 1994[⊗]

Abstract: This paper describes stochastic boundary molecular dynamics simulations of a benzyl-terminated oligoglycine inhibitor, $H_2NSO_2C_6H_4CO(NHCH_2CO)_3R$ (SG_3Bn), bound to the active site of human carbonic anhydrase II (HCAII, EC 4.2.1.1) in the presence of water. The position of the terminal benzyloxy group (R = $OCH_2C_6H_5$) was not defined in a crystal structure of this complex. The simulation suggested that the benzyl group associated with the hydrophobic residues Phe-20 and Pro-202 and that this association accounted for the decrease in the value of K_d observed for this benzyl-terminated inhibitor relative to other inhibitors with hydrophilic terminal groups. The average conformation observed for the oligoglycines in the inhibitor from the simulation was significantly different from the conformation inferred in the crystal structure. For example, the simulation indicated that the Ψ angle of Gly-1 and the Φ angle of Gly-2 shifted by about 50° from their values in the crystal structure within the first 2.5 ps of the 150 ps simulation. The simulation suggested that the calculated average conformation of the oligoglycines in the inhibitor—in its bound state with the protein—was a result of (1) the formation of hydrogen bonds with the surrounding molecules of water (these were included in the simulation) and (2) an accompanying improvement of π-face associations with the hydrophobic wall of the protein. These results, and their implications for the design of new inhibitors, are discussed.

Introduction

The objective of this study was to simulate the motion of an inhibitor of human carbonic anhydrase II (HCAII)—*N*-[*N*-[*N*-(4-sulfamylbenzoyl)glycyl]glycyl]glycine benzyl ester (SG_3Bn)—bound at the active site of the enzyme in water. We are using the binding of arylsulfonamide-derived ligands to carbonic anhydrase (CA) as a model system to examine the physical–organic chemistry of protein–ligand interactions.[1-4] CA has a well-defined active site cavity with one hydrophobic and one hydrophilic "wall" and is a structurally well-defined model system for examining hydrophobic interactions. It accepts, as ligands, a wide variety of *p*-substituted benzene sulfonamides; the geometry of the binding of the benzene sulfonamide group is invariant with changes in the *para* substituent. Christianson and co-workers have defined the structure of SG_3Bn bound to HCAII crystallographically (resolution and crystallographic *R* factor were 2.4 Å and 0.2, respectively),[5] and we have determined the dissociation constant of this ligand and a number of related inhibitors. Table 1 summarizes the values of the dissociation constants of several

[⊗] Abstract published in *Advance ACS Abstracts*, May 1, 1995.

(1) Chu, Y. H.; Chen, J. M.; Whitesides, G. M. *Anal. Chem.* **1993**, *65*, 1314.

(2) Avila, L. Z.; Chu, Y. H.; Blossey, E. C.; Whitesides, G. M. *J. Med. Chem.* **1993**, *36*, 126.

(3) Jain, A.; Huang, S. G.; Whitesides, G. M. *J. Am. Chem. Soc.* **1994**, *116*, 5057.

(4) Jain, A.; Whitesides, G. M.; Alexander, R. S.; Christianson, D. W. *J. Med. Chem.* **1994**, *37*, 2100.

(5) Cappalonga, A. M.; Alexander, R. S.; Christianson, D. W. *J. Am. Chem. Soc.* **1994**, *116*, 5063.

0002-7863/95/1517-6153$09.00/0

Table 1. Selected Dissociation Constants, K_d for *p*-Substituted Benzene Sulfonamides $H_2NSO_2C_6H_4CO-R$

R	abbrev	K_d (nM)	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})
$-NH_2$	SNH_2	120[a]		
$-NHCH_2CO_2^-$	SG	313[c]	3.8×10^5	0.12
$-NHCH_2CO_2CH_3$	SGMe	63[b]		
$-NHCH_2CONHCH_2CO_2^-$	SG_2	292[c]		
$-NHCH_2CONHCH_2CONHCH_2CO_2^-$	SG_3	234[c]		
$-NHCH_2CONHCH_2CONHCH_2COCH_2Ph$	SG_3Bn	75[b]		
$-NHCH_2CONHCH_2CONHCH_2CONHCH(CH_2Ph)CO_2^-$	SG_3Phe	196[c]		
$-NHCH_2CONHCH_2CONHCH_2CONHCH_2CO_2^-$	SG_4	356[c]	2.5×10^5	0.09

[a] Taken from ref 4. [b] Taken from ref 5. [c] Sigal, G.; Whitesides, G. M. Unpublished results.

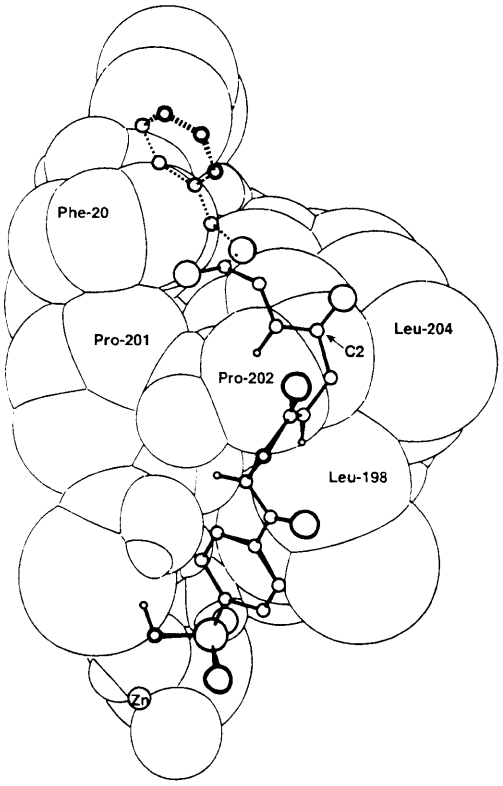


Figure 1. Partial view of the crystal structure of SG_3Bn bound to HCAII (from ref 5). The SG_3Bn molecule (ball and stick) and the hydrophobic wall of the HCAII active site (van der Waals spheres) are shown. Atoms in SG_3Bn that were crystallographically defined are connected by solid bonds; ill-defined atoms are connected by dashed bonds. The benzyloxy group is assigned using standard valence geometries. The carbonyl carbon of the second glycine, C2, is labeled for reference in the stochastic boundary molecular dynamics methodologies.

arylsulfonamides relevant to the present study. Crystallography shows that the $(Gly)_3$ group binds to the active site by hydrophobic interactions of the π-faces of the amide groups with the hydrophobic surface of the enzyme.[5] The binding constants in Table 1 are insensitive to the length of the oligoglycine chain. We have suggested that this insensitivity reflects the cancellation of the favorable increase in enthalpy with the unfavorable decrease in conformational entropy on binding.[3]

The benzyloxy-terminated triglycine tail of SG_3Bn was designed to associate with the hydrophobic wall of the protein.[4-5] The geometry of the G_3Bn group suggested that it might associate with hydrophobic residues located near the lip of the active site of the protein (Figure 1). The association of SG_3Bn with HCAII is approximately five times tighter than that of SG_3. The benzyloxy group of SG_3Bn bound to HCAII was not, however, observed in the crystal structure.[5] In other benzene sulfonamides with oligoglycine tails that are not terminated by a benzyl group, a similar disorder is observed for tails longer than three glycines.[3]

We wished to reconcile the observations that, although the benzyl group appeared to contribute approximately a factor of 5 to protein−inhibitor affinity, its binding was so ill-defined geometrically that it was not observed in the crystal structure. We have focused on two issues while using molecular dynamics simulations to investigate the complex between SG_3Bn and HCAII in water. First, we wished to know if the simulations would replicate, and thereby help to rationalize, the crystallographic structure. Second, we wished to determine if simulations could suggest a family of conformations that both rationalized the contributions to binding from the terminal benzyl group and were compatible with the crystallographically observed disorder in this group. We also hoped that studying the interactions between SG_3Bn−HCAII and water using simulations would suggest plausible structures for amino acids and terminal groups in oligopeptides related to $(Gly)_3OBn$ that would exploit hydrophobic associations and lead to compounds with tighter binding.

Modeling Methodology

The crystal structure of the SG_3Bn−HCAII complex (2.4 Å resolution) at 298 K was kindly provided to us by Professor David Christianson and served as the initial structure for the simulations.[5] The CHARMm 22 molecular mechanics program and Quanta 3.3 parameter set from MSI were used for the computations in this study.[6,7] Additional parameters were used to describe the $-SO_2NH-Zn-(His)_3-$binding site and are described in the following sections. The list of nonbonded interactions was cut off at 15 Å, and a switching potential between 11 and 14 Å was used for van der Waals and electrostatic terms (with the dielectric constant equal to 1).[8] Polar hydrogens were added to the crystal structure using HBUILD,[7] and an extended atom representation was used for all other nonpolar hydrogenated atoms.[6] The TIP3P model of water was used.[9] The benzyloxy group, ill-defined in the crystal complex, was initially added to the SG_3Bn inhibitor in the crystal structure using standard valence geometries.

Solvation of the Active Site. A pre-equilibrated sphere (15 Å radius) of bulk water[9] was centered on the carbonyl carbon of the second glycine in SG_3Bn (identified in the figures as C2). Molecules of bulk water that overlapped atoms in the crystallographically defined aggregate of SG_3Bn, HCAII, and water were deleted. The minimum distance allowed between the added molecules of water and the crystal structure (which we will call the "SG_3Bn−HCAII−crystal water" structure), without deletion of these waters, was 2.8 Å. We repeated this

(6) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(7) *QUANTA 3.3 Parameter Handbook*; MSI: Burlington, MA, 1992.
(8) Brunger, A. T.; Karplus, M. *Proteins: Struct. Funct. Genet.* **1988**, *4*, 148.
(9) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 927.

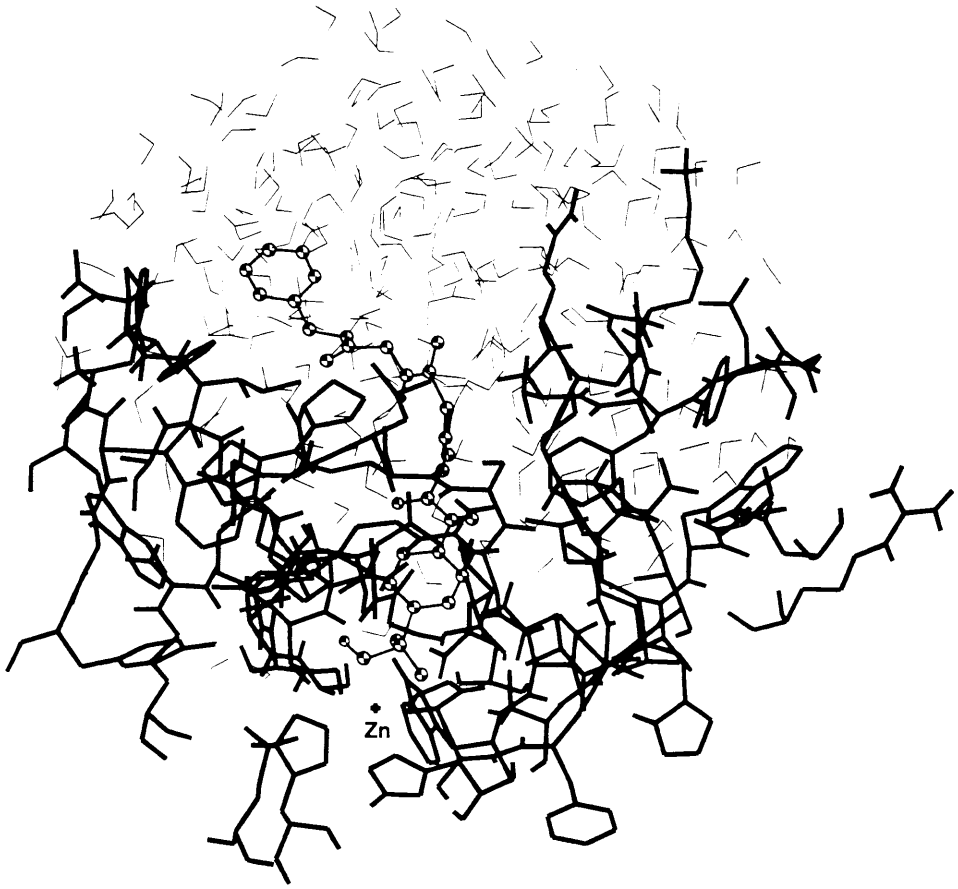


Figure 2. Initial system used in the SBMD simulation. Heavy lines represent HCAII; light lines represent the molecules of water; SG_3Bn is shown in ball and stick representation.

procedure of adding and deleting molecules of bulk water three successive times, each time applying a different angle of rotation to the sphere of added waters. Adding spheres of waters in this way places molecules of water in any artificial "holes" that may have been created in the deletion step. This procedure resulted in the inclusion of 211 bulk waters (adding to the 13 crystallographic waters already present) into and around the active site of HCAII containing SG_3Bn. The incorporation of solvent in the active site region provided a more realistic model of the SG_3Bn–HCAII interaction than simulations without water in this region.

Adjustment of the System Due to Added Hydrogens and Solvent. The new SG_3Bn–HCAII–water structure was allowed to relax, energetically, in the presence of the added hydrogens and molecules of water by the following prescription: All non-hydrogen atoms in the SG_3Bn–HCAII–water structure were constrained using a harmonic potential with initial force constant of 20 kcal/(mol·Å^2), and the energy of this system was minimized by the Fletcher–Powell method for 500 steps.[6] This minimization sequence was repeated six times, each time *decreasing* the harmonic force constant by 1.5 kcal/(mol·Å^2). After the sixth sequence, the harmonic constraints were removed and the coordinates of this SG_3Bn–HCAII–water structure used as the starting structure for the simulation.

Stochastic Boundary Molecular Dynamics. For computational efficiency, we performed a stochastic boundary molecular dynamics (SBMD) simulation starting from the minimized coordinates of the SG_3Bn–HCAII–water structure. Detailed discussions of SBMD exist in the literature,[10-12] and we will describe only the pertinent details in this section. In the SBMD simulation, only the spatial region of interest is included in the calculations; the rest is excluded. Starting from the energy-minimized SG_3Bn–HCAII–water structure, residues in the protein and/or crystallographic waters that lay outside a sphere of radius 15 Å centered on the C2 atom in SG_3Bn (Figure 1) were deleted from the structure. This reduced-particle complex (Figure 2) was then partitioned into a reaction zone and a reservoir region (Scheme 1). The atoms in the reservoir region were excluded from the calculation, but their effects on the remaining atoms were described by appropriate mean and stochastic forces. The reaction zone was further divided into a reaction region radius of 13 Å surrounded by a 2 Å buffer region. These reaction and buffer regions were centered on the C2 atom of the SG_3Bn inhibitor in the crystal structure (Scheme 1). Mean and stochastic boundary forces were applied to the atoms in the buffer region. The mean boundary forces were of two kinds: those on the protein and those on the solvent.

Boundary Forces on the Protein were described by a harmonic restoring term that kept the non-hydrogen atoms of the protein in the buffer region near the positions defined for them in the crystal structure (eq 1). F_i^{protein} was the effective

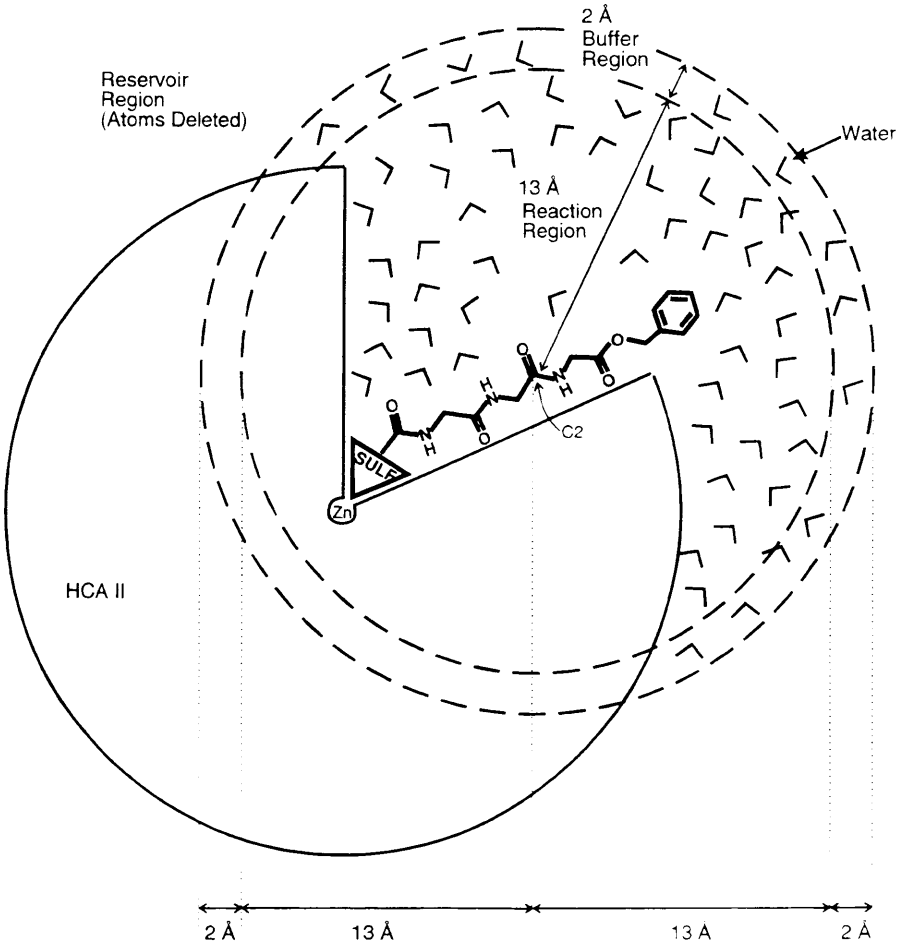
$$F_i^{\text{protein}} = \frac{-3k_B TS(r_i)}{\langle \Delta u_i^2 \rangle} \tag{1}$$

$$B_i = \frac{8\pi^2}{3\langle \Delta u_i^2 \rangle} \tag{2}$$

force constant, k_B was the Boltzmann constant, T was the

(10) Brooks, C. L.; Brunger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843.
(11) Brooks, C. L.; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159.
(12) Nakagawa, S.; Yu, H.; Karplus, M. *Proteins: Struct. Funct. Genet.* **1993**, *16*, 172.

Scheme 1. Diagram Showing the Partitioning of the SG_3Bn−HCAII−Water System for the SBMD Simulation[a]



[a] Residues in the protein and crystallographic waters that were in the reservoir region were deleted from the system.

temperature, and $\langle \Delta u_i^2 \rangle$ was the mean-square displacement for atom i in the protein and was calculated from the isotropic Debye−Waller factors, B_i, shown in eq 2.[13] The force constants in the buffer region were mediated by a switching function,[6] $S(r_i)$, that varied from 0 at the reaction/buffer region boundary to 0.5 at the buffer/reservoir region boundary. Averaged values of B were used in determining F_i^{protein} for atoms in the buffer region (averaged over side-chain and main-chain atoms in different residue types in the crystal structure): 0.5 Å^2 for the Zn ion, 1 Å^2 for Thr-199 and -200 residues, 5 Å^2 for sulfur atoms, and 9 Å^2 for the side-chain and main-chain atoms in the rest of the residues.

Boundary Forces on the Solvent were described by an effective potential having a radial distribution (centered on the C2 atom) and were applied to the oxygen atoms of the molecules of water. This solvent boundary potential prevented molecules of water at the buffer/reservoir border from evaporating and kept the structure of the water in the buffer region close to that of bulk water. While solvent boundary forces were only applied to molecules of water in the buffer region, these waters were allowed to diffuse across the buffer/reaction boundary during the simulation.

Stochastic Forces resulting from thermal motion were described by the Langevin equation of motion and stimulated fluctuations about the mean positions of atoms. The frequency of these fluctuations generated by the Langevin equation were governed by the collisional frequency, γ, and had the values 100 and 62 ps^{-1} for non-hydrogen atoms of the protein and water oxygens in the buffer region, respectively.[11] The buffer region, set up with these mean and stochastic forces, acted as a heat sink and source for atoms in the reaction region and also served to maintain the correct structural features of the water, protein, and inhibitor atoms in the reaction zone (as will be shown). The Verlet algorithm, with a time step of 1 fs, was utilized to describe the molecular dynamics of the atoms in the reaction region.[14]

The collision frequencies, solvent, and protein boundary force parameters were applied to their respective atoms in the buffer zone; covalent bonds to hydrogens were constrained using SHAKE,[15] and the SBMD simulation was started by allowing the system to equilibrate at 298 K for 10 ps. This equilibration period was followed by a simulation for 150 ps in which the coordinates were saved every 0.5 ps. Comparisons with experiment and analyses were performed over the 150 ps simulation period.

Correction of the Geometry of the $-SO_2NH-Zn-(His)_3$ Interaction. The initial parameter set did not distribute partial charges for the $-SO_2NH-Zn-(His)_3$ complex adequately (as judged from a short SBMD simulation); the geometry of this complex deviated from its geometry in the crystal structure and became hexacoordinated at the Zn ion. A new partial charge distribution for the $SO_2NH-Zn-(His)_3$ interaction (Figure 3), based on previous quantum mechanical studies on similar sulfonamide inhibitors bound to HCAII by Merz, Murko, and

(13) Willis, B. T. M.; Pryor, A. W. *Thermal Vibrations in Crystallography*; Cambridge University Press: Cambridge, U.K., 1975.

(14) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, U.K., 1987.

(15) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 4484.

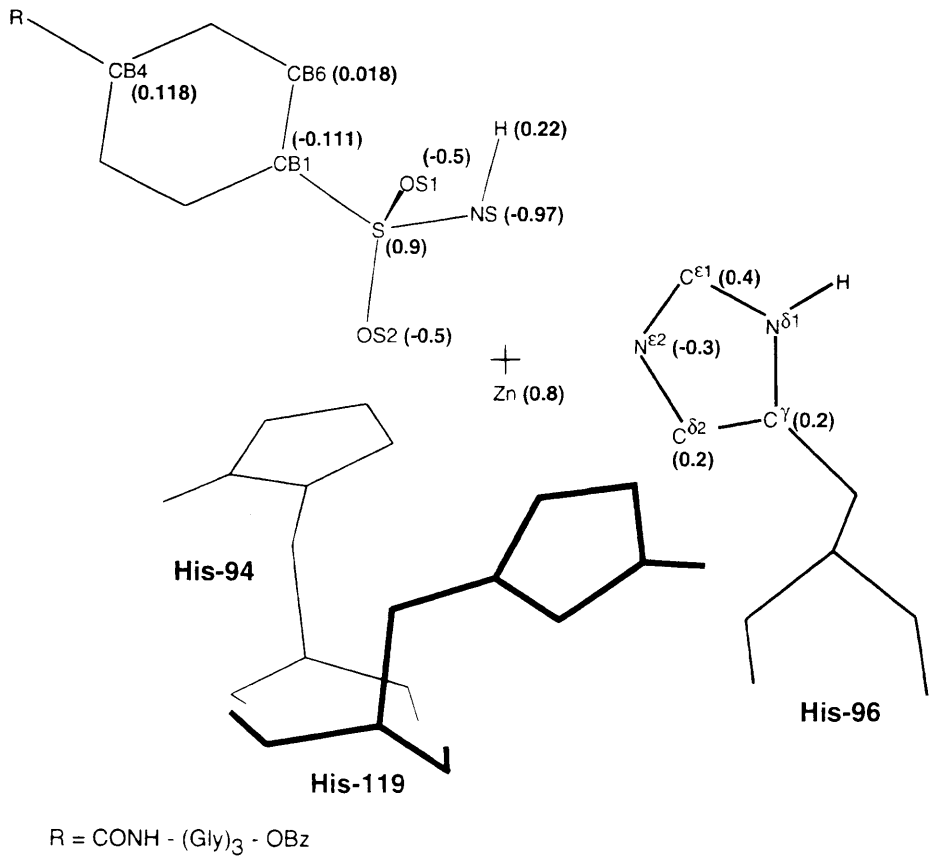


Figure 3. Partial charges are shown in parentheses for the $-ArSO_2NH$, N^{ϵ}-protonated His-96, and the Zn groups. The partial charges shown for His-96 reflect that a value of 0.1 e was added to the initial partial charges assigned by QUANTA. Analogous partial charges were on the other N^{δ}-protonated His-94 and the N^{ϵ}-protonated His-119.

Kollman,[16] was implemented for this study. This charge distribution reduced the +2 charge on the Zn ion to +0.8 and dispersed the remaining +1.2 units of charge evenly on the three chelating His residues (Figure 3). Merz et al. reported that this charge distribution still led to a $-SO_2NH-Zn-(His)_3$ complex with hexacoordinated character; this undesired geometry led the authors to apply explicit bond and angle terms to the atoms of $SO_2NH-Zn-(His)_3$, in addition to the new charge distribution, in order to maintain the correct pentacoordinated Zn geometry. We found that the use of a flat-bottom potential with quadratic walls at the $-SO_2NH-Zn-(His)_3$ site was sufficient to maintain the correct geometry (Table 2). The combination of these constraints and the modified partial charge distribution maintained the correct pentacoordination of the sulfonamide group with the Zn ion and sustained a geometry for $-SO_2NH-Zn-(His)_3$ that was similar to its geometry in the crystal structure.

Results and Analysis

Comparisons between the Simulated HCAII–SG_3Bn Structure in Solution and the Crystalline State. To judge the quality of the SBMD parameters of the simulation, we compared the spatial root-mean-square (rms) difference and fluctuations between the simulated structure and the crystal. These comparisons were made using only the non-hydrogen atoms. The rms fluctuations for the crystal were derived from Debye–Waller factors (from eq 2; rms = $\langle u^2 \rangle^{1/2}$); for the simulation, fluctuations relative to the average structure were used.[17]

Table 2. Minimum (R_{min}) and Maximum (R_{max}) Distances Allowed between Atoms Interacting at the $-SO_2NH-Zn-His$ Site before Applying a Constraining Energy[a] on This Geometry during the SBMD Simulation

interaction	R_{min}[a] (Å)	R_{max}[a] (Å)	X ray[b] (Å)
SG_3Bn-OS_2- - -Zn	2.5	2.7	2.6
SG_3Bn-OS_2- - -$N^{\epsilon 2}$-His-94	3.0	3.1	3.0
$SG_3Bn-OS2$- - -$N^{\delta 1}$-His-119	3.1	3.3	3.2
SG_3Bn-NS- - -Zn	2.0	2.2	2.1
SG_3Bn-NS- - -$N^{\epsilon 2}$-His-96	3.4	3.6	3.5
SG_3Bn-NS- - -$N^{\delta 1}$-His-119	3.7	3.9	3.8
SG_3Bn-NH- - -$N^{\epsilon 2}$-His-94	3.4	3.6	3.5

[a] The energies of these constraints depended on the following relationships: $E(R) = 0.5K(R - R_{min})^2$ for $R < R_{min}$; $E(R) = 0$ for $R_{min} \leq R \leq R_{max}$, and $E(R) = 0.5\ K(R - R_{max})^2$ for $R_{max} < R$, where K = 70.0 kcal/mol·Å^2 and R is the distance between the constituent atoms at any time during the simulation. The values of R_{min}, R_{max}, and K were chosen as those that reproduced the geometry of this interaction in the crystal structure. [b] These were the distances observed in the crystal structure.

Isotropic rms Difference between Components in the Reaction Region of HCAII for the Initial Crystal Structure and the Minimized System Used for the SBMD Simulation. Table 3 lists the rms differences between the initial and minimized structure for residues in HCAII, the SG_3Bn molecule, bulk and crystallographic waters in the reaction region. These differences were small for most of the residues in HCAII and

(16) Merz, K. M.; Murcko, M. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 4484.

(17) Findsen, L. A.; Subramanian, S.; Lounnas, V.; Pettitt, B. M. In *Principles of Molecular Recognition*; Buckingham, A. D., Legon, A. C., Roberts, S. M., Eds.; Blackie Academic and Professional: London, 1993; Chapter 7.

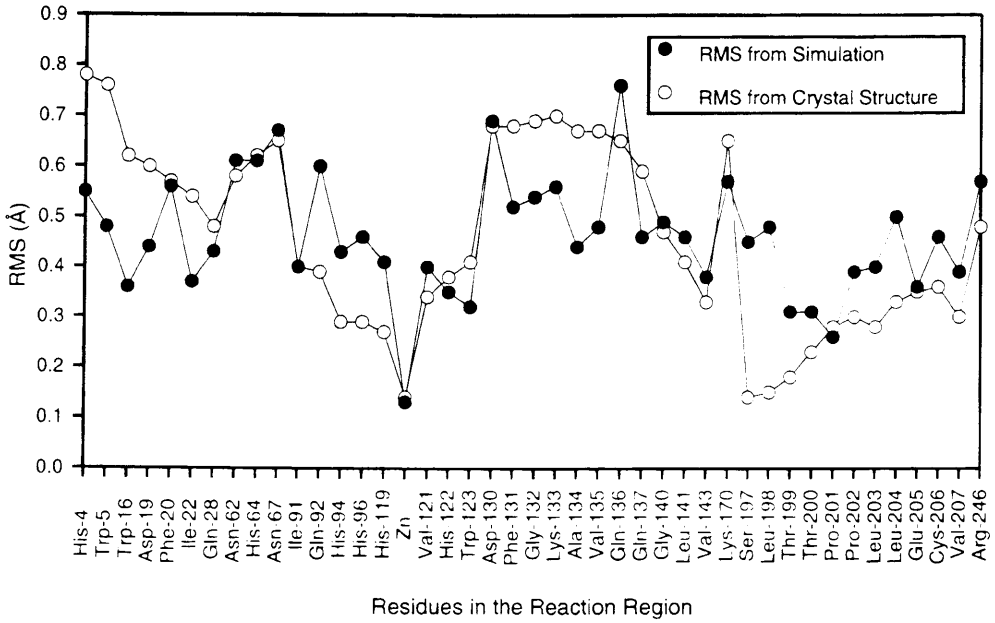


Figure 4. Isotropic rms fluctuations for residues that reside in the 13 Å reaction region (Scheme 1) of the active site of HCAII from both the SBMD simulation and the crystal structure. The crystallographic *R* factor, rms deviation from ideal bond lengths, and angles were 0.2, 0.011 Å, and 2.3°, respectively.

Table 3. Isotropic rms Difference between the Initial Crystal Structure[a] and the Minimized Structure for Residues of HCAII, the SG_3Bn Molecule, and the Molecules of Water in the Reaction Region

residue	rms (Å)	residue	rms (Å)	molecule	rms (Å)
Leu-141	0.40	**Pro-202**	**0.19**	**SG_3Bn**	**0.39**
Leu-24	**0.35**	Gln-92	0.19	bulk[b]	0.43
Lys-170	0.34	Cys-206	0.19	Cwat[b]	0.46
Trp-16	0.32	Gly-140	0.18		
Asn-67	0.31	Ile-22	0.18		
Asp-130	0.30	Glu-205	0.18		
Val-143	0.30	Asn-62	0.18		
His-4	0.28	Phe-131	0.17		
His-96[c]	0.26	Gln-28	0.17		
Ser-197	0.26	**Pro-201**	**0.17**		
Gln-137	0.25	**Phe-20**	**0.16**		
Asp-19	0.25	Val-207	0.16		
Thr-200	0.24	Gln-136	0.16		
His-64	0.23	His-122	0.14		
Lys-133	0.22	Leu-203	0.14		
Thr-199	0.22	Trp-5	0.12		
His-94[c]	0.22	Val-121	0.11		
Trp-123	0.21	Val-135	0.10		
Ile-91	0.21	Gly-132	0.10		
Leu-198	**0.20**	Ala-134	0.07		
His-119[c]	0.20	**Zn**	**0.00**		

[a] Residues of HCAII and the SG_3Bn molecule depicted in Figure 1 are shown in bold. [b] These values are for the bulk and crystallographic (Cwat) waters. [c] These histidines are bound to the Zn ion.

resulted from spatial readjustments to the added hydrogens, relaxation of any geometric artifacts due to the averaging effect in crystallography, and molecules of water that readjusted to the protein–inhibitor complex. The largest rms difference in HCAII came from Leu-141; the bulk of this difference was due to the carbonyl group adjusting its position in order to optimize a hydrogen bond with a NH group of Thr-208. The SG_3Bn molecule had a greater change in its coordinates than most of the residues of the HCAII in the reaction region.

Isotropic rms Fluctuations of Residues in the Reaction Region during the SBMD Simulation. Figure 4 shows a comparison between the isotropic rms fluctuations of residues in the reaction region from the SBMD simulation with the fluctuations from the crystal structure. The overall trends in

Table 4. Rms Fluctuations for Side-Chain Atoms[a] in the Protein for a Region around the Primary Binding Site of the Zn Ion (A) and the Reaction Region (B)

side-chain atoms	simulation (Å)	crystal structure (Å)
(A) 7 Å area around Zn	0.36	0.38
(B) 13 Å area around C2 on SG_3Bn (reaction region)	0.49	0.54
(A/B)	0.73	0.72

[a] That is, excluding the following main chain atoms: C, O, N, and C_α.

rms fluctuations from the simulation were in general agreement with those from the crystal structure. The largest discrepancies occurred in three regions: (1) Trp-16, Asp-19, Phe-20; (2) Gln-136; (3) Ser-197, Leu-198, Thr-199. Regions 1 and 3 were located at or near the reaction/buffer region boundary. Region 2 (Gln-136) protruded out of the surface of the protein into the solvent; the large amplitude of its fluctuations, relative to those of adjacent residues, reflected the fact than the simulation was for a single protein surrounded by water, rather than for a crystalline array of proteins.[18]

Mobilities of the Side Chains within the Active Site. Comparisons between rms fluctuations of side-chain atoms around the primary binding site at the Zn ion (defined as those atoms in the reaction region that were within a 7 Å sphere centered on the Zn atom), with those from the rest of the 13 Å reaction region, are shown in Table 4 for the simulation and the crystal structure. The magnitudes of these rms fluctuations from the simulation were slightly less than those from the crystal structure. The lower rms fluctuations from the simulation were not surprising since Debye–Waller factors (and hence the resulting rms fluctuations) contain a contribution from lattice disorder[19,20] (the amount of which was unknown, but its contribution was assumed to be constant) in addition to the dominant contribution from thermal averaging; rms fluctuations from the simulation, however, result only from thermal averag-

(18) As judged from viewing HCAII in its crystalline environment using QUANTA 3.3.

(19) Frauenfelder, H.; Petsko, G. A.; Tsernoglou, D. *Nature* **1979**, *280*, 558.

(20) See ref 14, pp 97–99.

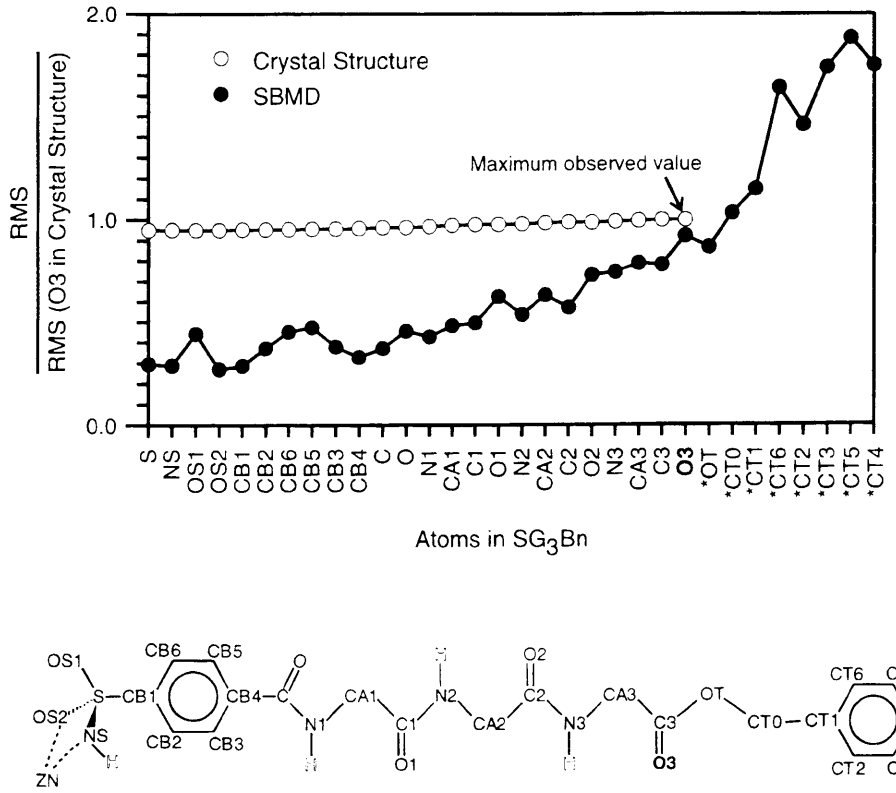


Figure 5. Rms fluctuations for atoms in SG_3Bn from the simulation and the crystal structure. The molecular structure of SG_3Bn with the corresponding atom labels in the graph is also shown. Atoms that did not appear in the crystal structure are labeled with an asterisk. All rms fluctuations were normalized with the value of the O3 atom in the crystal structure—the maximum observed value. SBMD = stochastic boundary molecular dynamics.

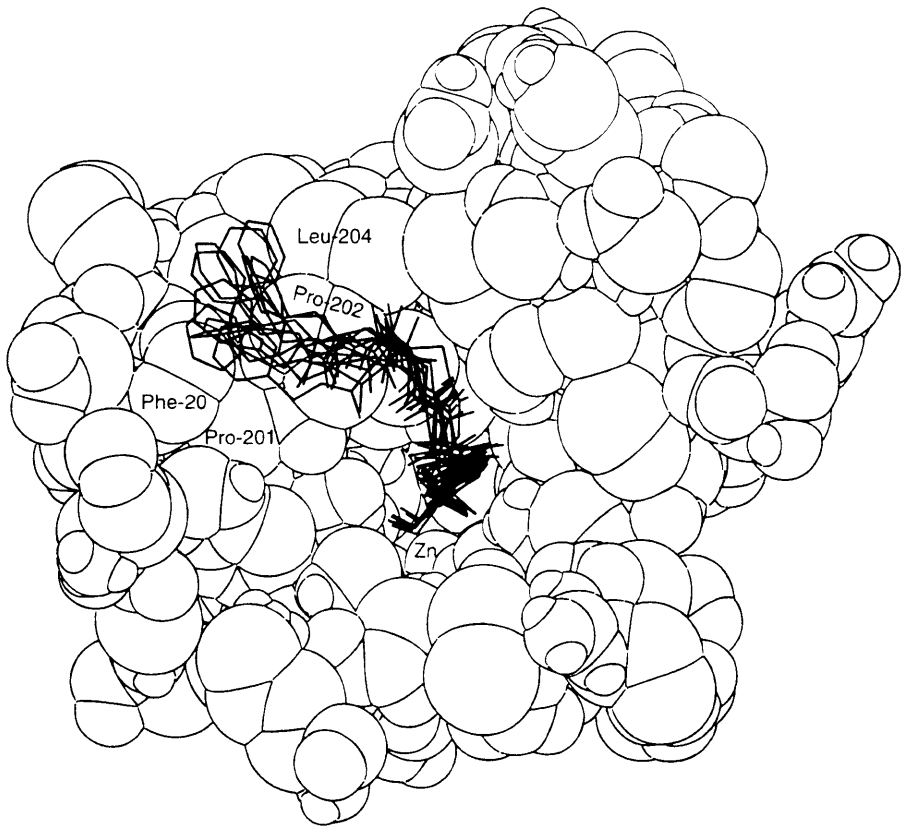


Figure 6. Superimposed structures of SG_3Bn (10 ps apart for 150 ps) shown in the active site of the averaged coordinates of HCAII (150 ps).

ing. The *ratios* of these rms fluctuations (that is, [catalytic site]/[reaction region]) were similar for the simulation (0.73) and the crystal structure (0.72).

Rms Fluctuations of SG_3Bn in the Active Site. We compared the normalized rms fluctuations of atoms in SG_3Bn from the simulation with those from the crystal structure (Figure 5). The SG_3Bn molecule was poorly resolved in the crystal structure; a detailed comparison of these atomic fluctuations with those from the crystal structure may therefore not be valid. The general trends of these fluctuations in the simulation and

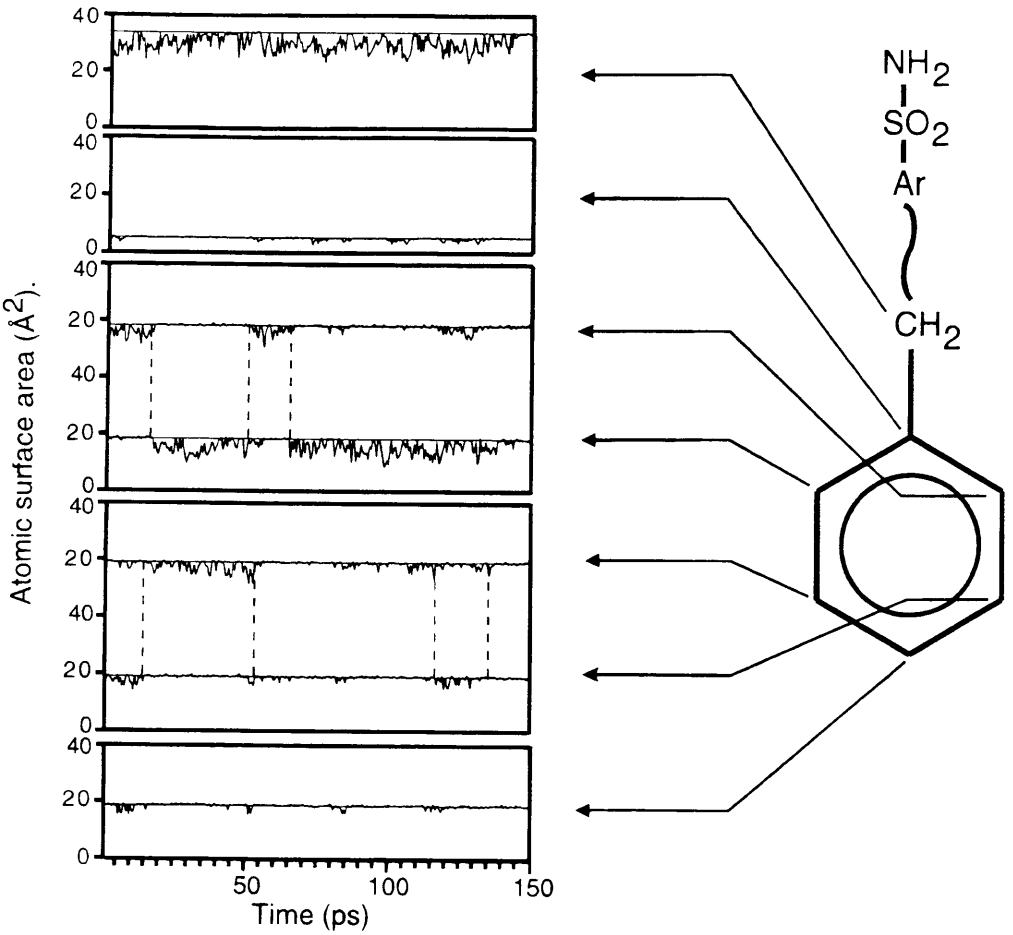


Figure 7. Molecular-surface area for atoms in the benzyl group of SG_3Bn in its association with the hydrophobic wall. This association was primarily with Phe-20 and Pro-202.

the crystal structure were, however, compatible. That is, the rms fluctuations from the simulation were greatest for atoms that did not appear in the difference electron density map and were therefore disordered in the crystal structure (Figure 5); this trend suggested that groups closer to the zinc end of the sulfonamide had stronger enthalpic interactions with HCAII than groups toward the benzyl group which was qualitatively consistent with trends in relaxation times, determined from ^1H NMR spectroscopy, for a series of similar oligoglycine inhibitors bound to carbonic anhydrase.[3] To illustrate this trend in the motions of the atoms of SG_3Bn, Figure 6 shows selected conformations of the inhibitor (10 ps apart) superimposed on the active site of the averaged coordinates of HCAII (150 ps).

The fluctuations inferred from the crystal structure were more uniform in magnitude than those from the simulation. The electron density map, however, was substantially more "noisy" for SG_3Bn than for the rest of the crystal structure.[5]

From these comparisons of the various rms data from the crystal structures and the simulations, we conclude that the SBMD parameters provide a reasonable model of the interactions of SG_3Bn with HCAII in the presence of water.

The Simulation Rationalizes the Enhanced Binding of SG_3Bn, Relative to SG_3, and the Absence of Localized Electron Density in the Crystal Structure. We computed the molecular-surface area[21] (MSA) of the benzyl group in SG_3Bn when free in solution and when associated with the underlying surface of HCAII as one metric of its hydrophobicity. The results of these values for the MSA, broken down by individual atomic contributions, are shown in Figure 7. An atom whose MSA *decreased* by more than 5%[22] from its value in the minimized *unassociated* structure in solution (shown by the horizontal reference line in each of the graphs in Figure 7) indicated that it formed a contact with the hydrophobic wall: the values of MSA that decreased the most indicated the strongest associations. The data for the values of MSA, and molecular graphics, suggested that the benzyl group associated predominately with Phe-20 and Pro-202. In this association: (1) The average orthogonal orientation of the benzyl group relative to Phe-20 (Figure 8a) was consistent with the free-energy favored "T-shaped" arrangement for a benzene dimer in dilute aqueous solution[23,24] and phenylalanine–phenylalanine interactions in proteins.[25] The average distance between the center of the phenyl ring of SG_3Bn and the center of the phenyl ring of Phe-20 was 5.9 ± 0.7 Å.[26] (2) The edge of the benzyl group resided in a groove defined by Phe-20 and Pro-202 (Figure 8b).

The Φ and Ψ Torsional Angles of the Three Glycines in SG_3Bn from the Simulation Indicated a Preferred Conformation Related to, but Not Identical to, That of the Crystal Structure. We wanted to know whether the conformation of SG_3Bn inferred from the crystal structure was at, or very close

(21) Using the Lee and Richards algorithm with probe radius set to 0: Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

(22) MSA values in this range were due to the fluctuations of bonds and angles.

(23) Jorgensen, W. L. *Chemtracts: Org. Chem.* **1991**, *4*, 91.

(24) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768.

(25) Hunter, C. A.; Singh, J.; Thorton, J. M. *J. Mol. Biol.* **1991**, *218*, 837.

(26) An average value of 5.5 Å was reported in ref 23; a value of 5.0 Å was reported in ref 24.

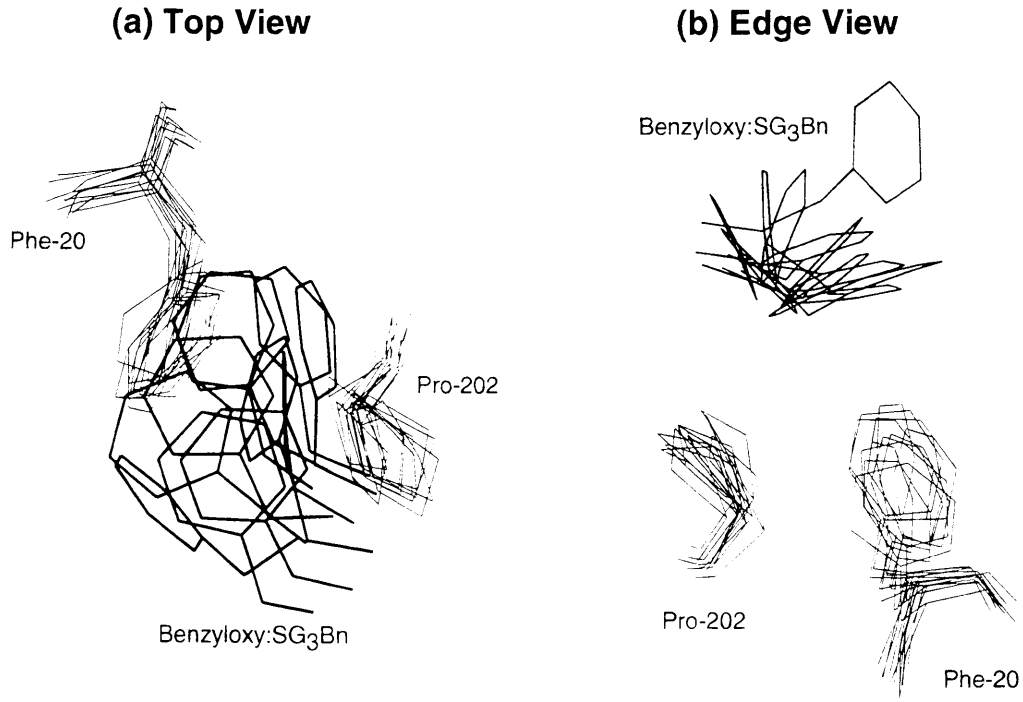


Figure 8. Different perspectives of the superimposed structures (10 ps apart for 150 ps) of the benzyl group (heavy dark lines) and Phe-20 and Pro-202 during the simulation: (a) top view of this association; (b) edge view of this association.

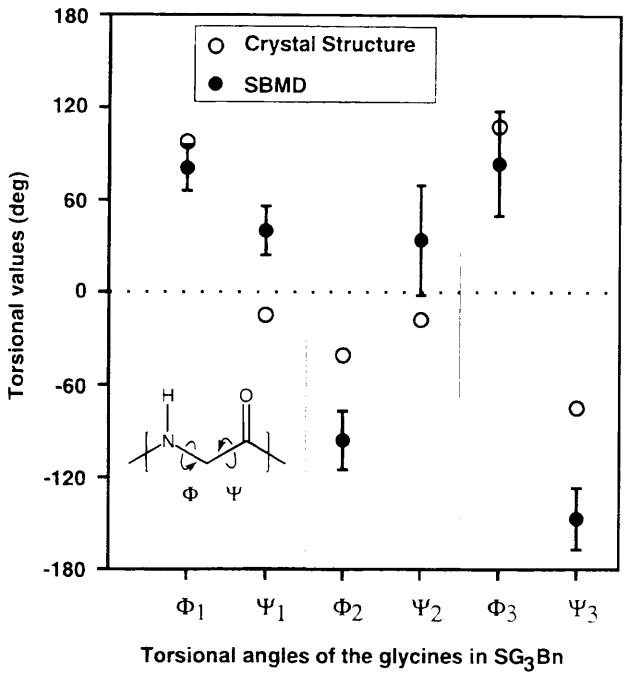


Figure 9.

to, its equilibrium geometry for the bound state or was an artifact of its poor resolution relative to the protein. Specifically, we wished to know whether the conformation of SG_3Bn in the crystal structure should be used as a template for designing other inhibitors of HCAII based on oligopeptides. We examined the conformations of SG_3Bn from the simulation by plotting the ranges of the Φ and Ψ torsional angles for the oligoglycine tail in the inhibitor (Figure 9); Figure 10 summarizes the difference between these two structures. These data showed that Φ for Gly-1 and Gly-3 were within 17° and 24°, respectively, of their values in the crystal structure while the other torsional values had larger differences. The bulk of this change in conformation for the oligoglycines occurred within the first 2.5 ps of the 10 ps equilibration segment of the simulation. The net effect of the change in the conformation of the triglycine that occurred in the simulation (relative to the crystal structure after minimization of the energy) was to (1) increase their association with the hydrophobic wall (the average enthalpy of interaction between SG_3Bn and HCAII from the simulation was approximately 10 kcal/mol lower than that from the crystal structure) and (2) optimize hydrogen bonding with the surrounding molecules of water (as we will show). As the electron-difference map was considerably noisier for SG_3Bn than for the protein in the crystal structure, the non-hydrogen atoms of SG_3Bn, up to Gly-1, were defined with the aid of the structurally well-defined SGMe molecule in its complex with HCAII.[5] Defining SG_3Bn in this way, however, oriented the CONH group between Gly-1 and Gly-2 perpendicular to the hydrophobic surface of HCAII just below it (see Figure 2); a consequence of this orientation was that the NH group of this amide bond points toward the hydrophobic surface. Christianson et al. noted that there was no electron density in the electron-difference map for the oxygen atom of the CONH group between Gly-1 and Gly-2 in SG_3Bn, at least in the conformation that was modeled with the aid of SGMe.[5] The simulation suggested that the CONH group between Gly-1 and Gly-2 preferred to orient itself such that it was approximately coplanar with the proximate hydrophobic surface of HCAII and that this conformation also promoted favorable enthalpic interactions between this group and surrounding molecules of water. We suggest that it is possible that the conformations of the oligoglycines in SG_3Bn observed during the simulation were representative of structures that were more favored in free energy than that reported for the crystal structure. We have, however, no unambiguous way of judging which of these two structures is the more reliable.

Manifestations of the Conformational Shift Involving the Oligoglycines. We analyzed the hydrogen bonds between molecules of water around the peptide groups in SG_3Bn to see whether these interactions had any effect on the conformation of the inhibitor. In the crystal structure, one oxygen in a

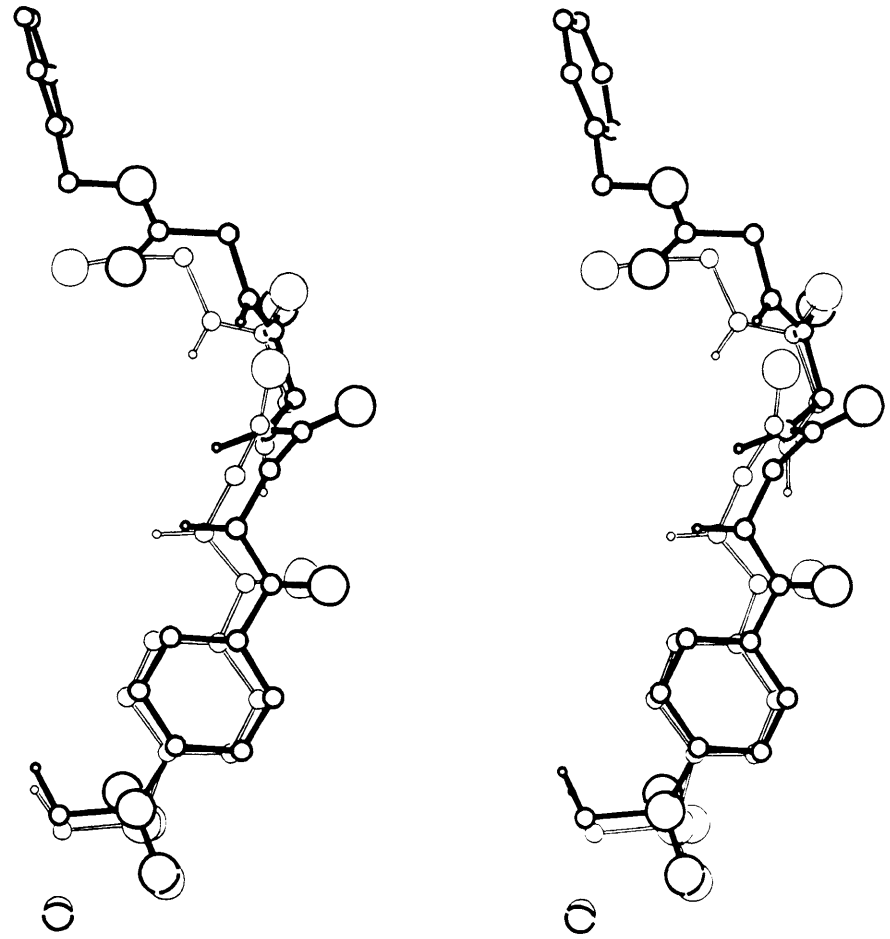


Figure 10. Crystal structure of SG_3Bn (thin lines), in its conformation associated with HCAII, shown in relation to the average structure of SG_3Bn (thick lines) derived from simulation.

molecule of water exhibited poor hydrogen-bond geometry[27] with the NH group of Gly-1 in SG_3Bn (we will refer to this molecule of water as Cwat-94). The absence of any significant amounts of crystallographically defined water in the active site, and of any apparent hydrogen bonding between the oligoglycines of SG_3Bn with either HCAII or molecules of water in the crystal complex, suggested that the role of solvent might be important above that which is observable in the crystal structure. The simulation suggested that six molecules of water (including Cwat-94) associated strongly with SG_3Bn (Table 5). These molecules of water interacted most strongly with the CO groups of the glycines and less strongly with the NH groups; Figure 11 shows representative "snapshots" of these interactions. The interactions between SG_3Bn and water influenced the conformation in the simulation in two ways that accounted for the difference between crystallographic and simulated conformation: (1) they facilitated a shift in each of the values of Ψ_1, Φ_2, Ψ_2, and Ψ_3 by approximately 50° or more from their values in the crystal structure (Figures 9 and 10) and (2) they were responsible for the favorable hydrophobic association between the π-face of Gly-2 and residues in HCAII (Pro-202 and Leu-198). The van der Waals component of the average energy of interaction between SG_3Bn and HCAII from the simulation was *more* favorable by approximately 20 kcal/mol than that calculated for the conformation of the crystal structure (the concomitant electrostatic component of this average energy of interaction from the simulation was *less* favorable by 10 kcal/mol than that from the crystal structure).

(27) Creighton, T. C. *Proteins Structures and Molecular Properties*; W. H. Freeman and Co.: New York, 1993; pp 147–148.

Table 5. Average Hydrogen-Bond Geometries for the Molecules of Water Interacting Most Strongly with the Peptide Groups of SG_3Bn during the Simulation

functional group[a,b]	water[c]	∠⟨C=O- -H⟩ ± rms (deg)	r⟨C=O- -H⟩ ± rms (Å)[d]
>C=O	167	148 ± 13	2.4 ± 0.6
		145 ± 11	3.1 ± 0.6
>C1=O1	82	151 ± 13	3.0 ± 0.7
		150 ± 13	2.8 ± 0.6
>C2=O2	148	135 ± 18	3.4 ± 1.0
		137 ± 18	3.1 ± 1.1
>C3=O3	72	134 ± 18	3.7 ± 1.0
		138 ± 19	3.3 ± 1.2
>C3=O3	44	131 ± 20	3.4 ± 1.2
		132 ± 21	3.8 ± 1.3
		∠⟨N−H- -O⟩ ± rms (deg)	r⟨N−H- -O⟩ ± rms (Å)[d]
>N1−H1	94[e]	129 ± 18[f]	3.5 ± 0.9[f]
>N1−H1	44	126 ± 14	3.7 ± 0.4
>N2−H2	94[e]	138 ± 22	3.7 ± 0.7
>N2−H2	44	139 ± 19	3.2 ± 0.3
>N3−H3	44	112 ± 28	3.9 ± 0.6

[a] Refer to the nomenclature used in Figure 5. [b] Each carbonyl group has two listings per molecule of water for interactions with each of the hydrogens in that water. [c] Each listing is the identification number for the molecule of water in the simulation. [d] These distances involve the following atoms: O_{SG3Bn} to H_{water} for >CO groups and N_{SG3Bn} to O_{water} for >NH groups. [e] This molecule of water is referred to as Cwat-94 in the text. [f] This interaction is observed in the crystal structure with a distance of 3.5 Å and an angle of 138°.

The Conformation and rms Atomic Fluctuations of $H_2NSO_2C_6H_4CONHCH_2CO_2CH_3$ (SGMe)—from a Simulation of the SGMe−HCAII−Water Complex—Were in Very

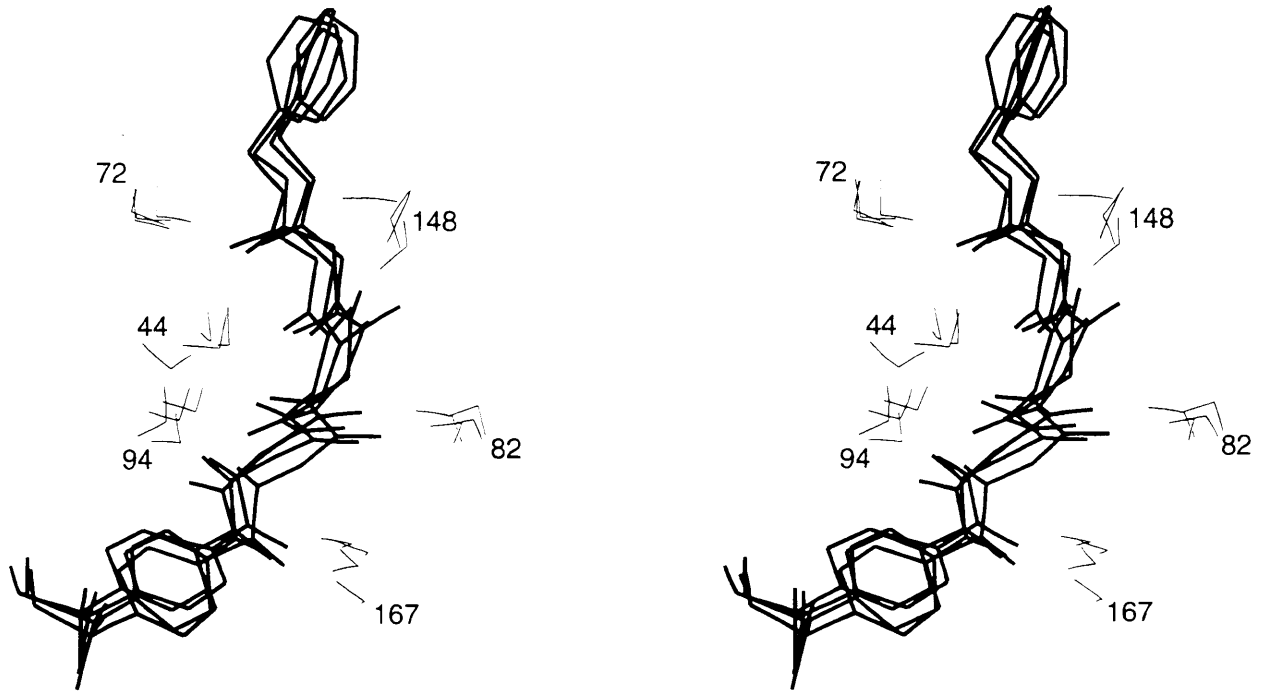


Figure 11. Four representative "snapshots", 0.5 ps apart (taken from time 100.0 to 101.5 ps), of the interactions between the oligoglycines in SG_3Bn (thick lines) and the six molecules of water that interacted most strongly (thin lines). These six waters consisted of bulk water labeled 167, 82, 148, 72, and 44 and a crystallographic water labeled 94. Table 5 lists some statistics for these interactions.

Good Agreement with Its Values from a Crystallographically Well-Defined Structure of SGMe–HCAII. We questioned whether the average conformation of SG_3Bn inferred from the simulation could be believed sufficiently to be used in ligand design, since it differed significantly from the conformation in the crystal structure (although this structure was not well defined). We therefore examined the conformation and rms atomic fluctuations from a second crystal complex for atoms in a structurally similar inhibitor—SGMe—bound to HCAII and whose electron density map was well defined for this inhibitor.[5] A SBMD simulation of the SGMe–HCAII complex suggested that the conformation of SGMe (Figure 12), and the trends in its atomic fluctuations (Figure 13), were in very good agreement with those from the crystal structure.[28] These favorable comparisons further suggested that the conformations observed from simulations of these inhibitors, in their bound states with HCAII, were believable.

Discussion

Molecular dynamics simulations suggest that the benzyl group of SG_3Bn associates with Phe-20 and Pro-202 of HCAII and that its edge predominately rests in a groove defined by these residues. We interpret the presence of the benzyl–HCAII interaction as important in the relatively tight binding of this benzyloxy-terminated oligoglycine inhibitor with HCAII. This benzyloxy group, however, is ill defined in the crystal structure—an observation that is consistent with results from the simulation, in that this group has values of rms atomic fluctuations that are larger than the other groups in SG_3Bn.

During the simulation, the glycines in SG_3Bn adopt conformations that are significantly different from those of the crystal structure. The range of conformations from the simulation, however, is wide and overlaps at the extreme the conformation of SG_3Bn suggested from the crystal structure. The flexibility of the glycines in SG_3Bn, and the absence of any hydrogen bonds between these glycines and HCAII, contribute to the noisy electron-difference map for this complex. The conformation of SG_3Bn in the crystal structure was modeled—up to Gly-1—from the complex of SGMe with HCAII; this empirically determined conformation placed the CONH group between Gly-1 and Gly-2 perpendicular to the hydrophobic wall (Figure 2). The molecules of water localized in the simulation appear to be energetically important in determining the conformations

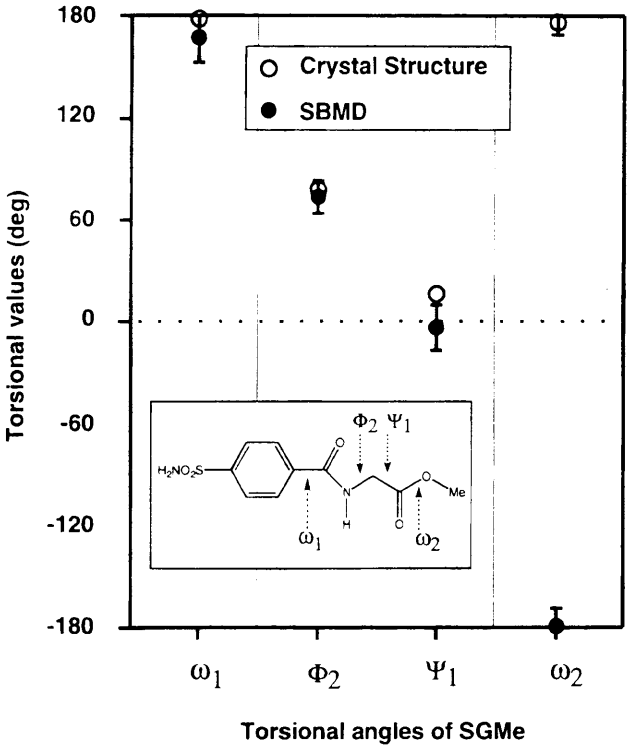


Figure 12. Average Φ, Ψ, and ω torsional values of SGMe from a simulation of the SGMe–HCAII complex (filled circles) and the corresponding values from the crystal structure (open circles).

(28) We used the crystal structure of the SGMe–HCAII complex as the starting point for this SBMD simulation (ref 5); the modeling methodology for this simulation was consistent with that in this paper, except that the analyses were performed over 100 ps.

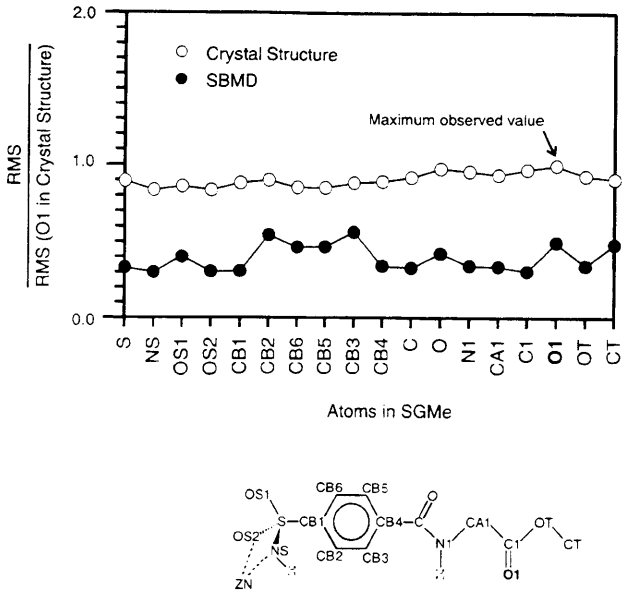


Figure 13. Rms fluctuations of atoms in SGMe from the simulation and the crystal structure. The molecular structure of SGMe shows the corresponding atom labels in the graph. All rms values were normalized with the value of the O1 atom in the crystal structure—the maximum observed value.

observed for SG_3Bn; the simulation also suggests that the differences between the conformation of the oligoglycines in the simulation and the crystal structure reflect enthalpic interactions with the molecules of water and π-face associations with the hydrophobic wall of HCAII. Comparing the average interaction energy between SG_3Bn and HCAII from the simulation with that from the crystal structure provides a quantitative description of the conformational change in the oligoglycines: the electrostatic component of the average interaction energy between SG_3Bn and HCAII from the simulation is 10 kcal/mol *less* favorable than that from the crystal structure, while the van der Waals component of this energy is 20 kcal/mol *more* favorable. We note explicitly two ambiguities in this discussion: first, the conformations of SG_3Bn observed in the simulation are not necessarily the true equilibrium geometry; second, the disorder in the conformation derived from the crystal structure, and the fact that this conformation shows surprisingly few of the expected hydrogen bonds, suggests that it may be misleading in some of its details. We have no present method of deciding which conformation is more accurate.

The orientation of the phenyl group within the groove defined by Phe-20 and Pro-202 and the results of previous physical organic studies on the relationships between structural orientations and stabilities[29] suggest that the binding affinity of these inhibitors for HCAII could be improved by replacing the terminal phenyl group with one that has a better complimentary fit into this groove. Properly designed, this new terminal group might possess the desired shape to fit in this groove and be more hydrophobic than the benzyl group.[30-32] A search of the Cambridge Crystallographic Database[33] (CCD) reveals a limited set of tripeptide sequences—those that do not contain glycine and are hydrophobic—that have successive Φ/Ψ angles within the ranges defined by the oligopeptides in SG_3Bn from the simulation. The identification of a hydrophobic tripeptide sequence for a sulfonamide inhibitor that can conform to the hydrophobic surface in the active site of HCAII, and is more conformationally restrictive than the oligoglycines, is a potentially useful outcome from the simulation. In this regard, we are pursuing the design of an inhibitor of HCAII based on the above inferences from the simulation and will report on these findings at a later date. While the results from the simulation suggest new, plausible, metrics with which to design sulfonamide inhibitors that bind tighter than SG_3Bn to HCAII, the extent to which this paradigm for rational drug design reflects reality remains to be determined experimentally. Molecular dynamics has extended our understanding of the SG_3Bn–HCAII complex from that of crystallographic studies and has provided us with a clearer, dynamic picture of this interaction—that involving a flexible inhibitor—within the active site of HCAII.

Acknowledgment. This study was supported by the NIH (NIH GM30367). We thank Professor David Christianson (U. Penn) and George Sigal for helpful discussions during the preparation of this manuscript.

Supplementary Material Available: Listings of parameters used (5 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, can be ordered from the ACS, and can be downloaded from the Internet; see any current masthead page for ordering information and Internet access instructions.

JA943608E

(29) Chin, D. N.; Gordon, D. M.; Whitesides, G. M. *J. Am. Chem. Soc.* **1994**, *116*, 12033.
(30) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.
(31) Dec, S. F.; Gill, S. J. *J. Solution Chem.* **1984**, *13*, 27.
(32) Gao, J. M.; Whitesides, G. M. Unpublished results.
(33) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. *Acta Crystallogr.* **1979**, *B35*, 2331.