proceeds via the reaction of a zinc-bicarbonate intermediate with CO_2 through a trigonal bipyramidal TS, and our calculations predict that water, apparently, inhibits this reaction. The hydrolysis of the subsequently formed zinc-bicarbonate intermediate, however, requires the presence of several water molecules in the CA active site. Once water replaces bicarbonate, it transfers a proton to His-64 via the surrounding waters (two-water-proton relay) and regenerates the zinc-hydroxide intermediate. His-64 subsequently transfers the proton to buffer external to the enzyme and the catalytic cycle continues. The second mechanism (see Scheme VI), which we will call the AM1 mechanism, is identical with the above one except for the involvement of a buffer molecule in the intramolecular PT. Finally, we note that our calculations have not unambiguously been able to eliminate the mechanism of Kannan et al.,[4] and therefore it is still a distinct mechanistic possibility.

It is useful at this point to differentiate our computational results from those of other researchers. Previous workers have found that the hydration of CO_2 proceeds via a nearly activationless inner-sphere mechanism,[15] in which CO_2 initially coordinates with the zinc ion, while our calculations suggest an inner-sphere mechanism where CO_2 does not initially coordinate with the zinc ion. In both cases, though, the zinc ion is found to serve as a template to bring CO_2 and a hydroxide ion into contact. We attempted to locate an outer-sphere TS, which has been espoused by Lipscomb,[13] but were able to locate only a hilltop that led to an E_a that was greater than that for the inner-sphere mechanism.

The two-water-proton relay mechanism has been studied by Liang and Lipscomb,[24] and they have calculated an E_a for this process of 34 kcal/mol (from four-coordinated zinc) and 54 kcal/mol (from five-coordinated zinc), while we have predicted the E_a to be 18 kcal/mol (from four-coordinated zinc). Both predictions are certainly too high, and the reason for the overestimation of this E_a is due in part to the theoretical model employed[24] and the neglect of the contributions from tunneling and the protein environment,[40] which tend to reduce the E_a of proton transfers.

Recently, the role of His-64 in the intramolecular PT has been studied with the aid of site-directed mutagenesis and has yielded the interesting conclusion that this group is *not* crucial to HCA activity.[45] These authors studied glutamine, glutamic acid, lysine, and alanine mutants and found that the k_{cat} was reduced only by a factor of 1.5-3.5 relative to the native enzyme. Our conclusions are consistent with these results in that they do suggest that other protein residues or molecules (e.g., bicarbonate) can serve as a proton-transfer group and thus keep k_{cat} close to that observed for the native enzyme.

Acknowledgment. Realistic modeling of the active site, with solvent, would not have been possible without the computation facilities of the Center for Theory and Simulation in Science and Engineering at Cornell University, which receives funding in part from the NSF, New York State, and IBM Corp. Our research at Cornell was generously supported by the NSF through research Grant CHE8406119. We are grateful to Jane Jorgensen and Elizabeth Fields for their expert drawings. We also thank G. G. Hammes and Peter Kollman for helpful discussions.

Registry No. CA, 9001-03-0; CO_2, 124-38-9; 14H, 121173-69-1; 15H, 28414-71-3.

(45) Forsman, C.; Behravan, G.; Jonsson, B.-H.; Liang, Z.; Lindskog, S.; Ren, X.; Sandström; J.; Wallgren, K. *FEBS Lett.* **1988**, *229*, 360.

Free Energy Perturbation Simulations of the Inhibition of Thermolysin: Prediction of the Free Energy of Binding of a New Inhibitor

Kenneth M. Merz, Jr., and Peter A. Kollman*

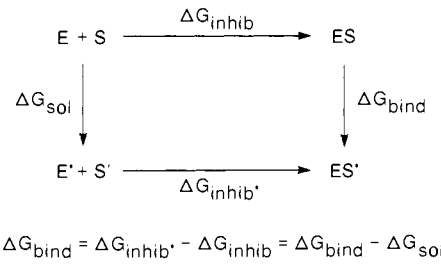
Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California 94143. *Received September 19, 1988*

Abstract: We report free energy perturbation simulations on a series of thermolysin inhibitors. The three compounds that were examined are transition-state analogues and have the general structure carbobenzoxy-Gly^P(X)-L-Leu-L-Leu, where X = NH, O, and CH_2. While the NH and O compounds have had their crystal structure solved as well as having their K_i's determined, this is not true for the CH_2 compound. Our simulations predict the CH_2 compound to have a K_i similar to that of the NH compound ($\Delta\Delta G_{bind}$ = -0.3 kcal/mol favoring the NH compound), while both have K_i's greater than the O compound. The reason why the CH_2 compound is a potent inhibitor can be traced to two factors: (1) the solvation free energy for the CH_2 is less than that for the NH and O compounds (hydrophobic effect) and (2) the CH_2 compound does not have a repulsive interaction between the main chain carbonyl of Ala 113 and the oxygen. We have also redone the NH → O simulations reported earlier (*Science* **1987**, *235*, 574) using a new set of charges based on a more complete quantum mechanical geometry optimization of models of the inhibitors. These new charges led to calculated $\Delta\Delta G$ for NH → O of 5.9 kcal/mol, in contrast to the 4.2 kcal/mol calculated earlier. By mutating the new charges into the old, we were able to more closely reproduce ($\Delta\Delta G_{calcd}$ = 4.05 kcal/mol) the calculated results of Bash et al., albeit with different relative contributions of solvation and binding free energies. We have also carried out simulations in which we modified the force field parameter used for the neutral carboxylic acid X-C-O-H torsion. This was done because Glu 143 interacts with the perturbation site, and hence we expect that the smaller torsion barrier used previously (3.6 kcal/mol) might bias the calculation considering that the experimental value for this torsion is approximately 10 kcal/mol. In these simulations, which also used the new charges, we find that the free energy value for the CH_2 → NH perturbation changes very slightly (-0.3 vs 0.0 kcal/mol), while for the NH → O perturbation there is a much larger change (5.9 vs 3.3 kcal/mol). Concomitantly, a significant change in the active site geometry between the two parameter sets is observed, with the newer set matching the experimental structure more closely.

Thermolysin is a zinc-requiring endopeptidase that has a specificity for a peptide bond in which the adjacent side chains are hydrophobic (e.g., leucine, isoleucine, and phenylalanine). The crystal structure has been solved to a resolution of 1.6 Å,[1] which

0002-7863/89/1511-5649$01.50/0

Scheme I



$$\Delta\Delta G_{bind} = \Delta G_{inhib'} - \Delta G_{inhib} = \Delta G_{bind} - \Delta G_{sol}$$

yields a tertiary structure that has two spherical domains separated by a deep groove with a zinc ion at the center of it which constitutes the active site region.

A large number of inhibitors for thermolysin have been studied crystallography by Matthews et al.,[2,3] which has resulted not only in a better understanding of the mode of inhibition of transition-state analogues but has also shed light on salient mechanistic features of the enzyme. We have chosen to continue our study on thermolysin inhibitors due to the initial successes we have had in the application of the free energy perturbation methodology to this problem.[4] The series of inhibitors that are being studied are the Cbz-GlyP-(X)-Leu-Leu, where X = NH, O, and CH_2. The K_i's for the NH and O compounds are known[5] (and thus the $\Delta\Delta G_{bind}$), but the corresponding data are not available for the CH_2 compound. We have already reported calculations on the NH and O compounds,[4] but in those calculations the relative binding free energy of the compounds was known prior to the theoretical calculations. For theoretical calculations to have the same credibility in the biomolecular area that they have for small molecule gas-phase problems,[6] it is important that one make predictions. Thus, it is the major aim of this paper to *predict* the $\Delta\Delta G_{bind}$ of the CH_2 compound relative to the former two compounds.

Computational Approach. The free energy perturbation method is a statistical mechanical approach that was first described by Zwanzig[7] that has recently found extensive application[8] due, in no small part, to the recent advances in computer technology.

The Gibbs free energy can be calculated according to eq 1[7]

$$G(\lambda + \Delta\lambda) - G(\lambda) = -k_B T \ln\langle\exp[(-H(\lambda + \Delta\lambda) - H(\lambda))/k_B T]\rangle_\lambda \quad (1)$$

where k_B is Boltzmann's constant, T is the absolute temperature, $H(\lambda + \Delta\lambda)$ and $H(\lambda)$ are the hamiltonians at the states λ and $\lambda + \Delta\lambda$, and the $\langle\ \rangle_\lambda$ indicates the ensemble average at the intermediate points along the conversion pathway defined by the coupling parameter λ. A simulation is then run between the states $\lambda = 1$ to $\lambda = 0$ (or the reverse) where the free energy is evaluated at each of the intermediate λ points. The total free energy for this change is then given by eq 2

$$\Delta G = \sum G(\lambda + \Delta\lambda) - G(\lambda) \quad (2)$$

In order to determine the $\Delta\Delta G_{bind}$ between one inhibitor and another we make use of the following thermodynamic cycle (Scheme I) where ΔG_{inhib} and $\Delta G_{inhib'}$ represent the free energy of binding of the inhibitor S and S', respectively, ΔG_{sol} represents the free energy of solvation difference between S and S', and, lastly, ΔG_{bind} represents the free energy of binding difference between S and S' in the enzyme active site. By using simulations we are currently unable to determine the ΔG_{inhib} terms; however, it is possible to mutate one inhibitor into another and thereby readily determining ΔG_{sol} or ΔG_{bind}. Since we are dealing with a state function (ΔG) the following relationship, based on the thermodynamic cycle, holds. Thus we have at our disposal all that is necessary to determine the $\Delta\Delta G_{bind}$ between two inhibitors.

$$\Delta\Delta G_{bind} = \Delta G_{inhib'} - \Delta G_{inhib} = \Delta G_{bind} - \Delta G_{sol} \quad (3)$$

Computational Procedure. The coordinates for the thermolysin Cbz-Gly^P-(NH)-Leu-Leu (ZG^P(NH)LL) complex were kindly made available to us by B. Matthews. This structure served as the starting point for all of the enzyme simulations outlined below. The charges for the three inhibitors were obtained in the following way: We first carried out 4-31G geometry optimizations on the model structures **1–3**. This was done in order to get a better representation of the geometry around the crucial region (the -G^P(X)- region) in which the perturbations will be carried out. The structures are given in Figure 1. One major geometric change that occurs on improving the basis set from STO-3G* to 4-31G is the opening up of the C–P–X bond angle. Thus, the STO-3G* bond angle ranged from 92° to 96°, while the 4-31G angles range from 97° to 101°. The value obtained crystallographically[3] for the NH inhibitor bound to thermolysin is 100°; thus, the 4-31G basis set is in much better agreement with experiment. Another large geometric difference is observed in the P–O bond lengths. For the smaller basis set the bond lengths are 1.45 Å, and with the larger one the bond lengths increase from 1.558 to 1.575 Å. Again the 4-31G basis set is in better agreement with the experimental value of 1.52 Å.[3] The requisite geometric variables were then transferred to a STO-3G* optimized structure (**4**). We

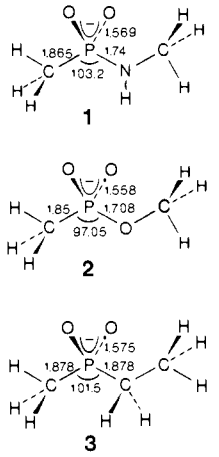
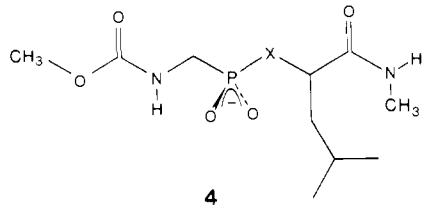


Figure 1. Structures for the 4-31G optimized inhibitor models. Bond lengths are in Å and angles are in deg.

Table I. New and Old Atomic Charges of Cbz-Gly^P(X)-L-Leu-L-Leu for X = NH, O, and CH_2

	new			old	
atom	X = NH	X = O	X = CH_2	X = NH	X = O
C	−0.16	−0.191	−0.2115	−0.246	−0.216
H	0.037	0.0545	0.051	0.065	0.06
H	0.037	0.0545	0.051	0.065	0.06
P	0.803	0.903	0.862	1.046	1.004
O	−0.617	−0.6365	−0.638	−0.667	−0.665
O	−0.617	−0.6365	−0.638	−0.667	−0.665
X	−0.605	−0.532	−0.634	−0.705	−0.451
X	0.237		0.1085	0.227	
X			0.1085		
C	−0.205	−0.056	−0.1135	−0.055	−0.132
H	0.092	0.042	0.056	0.061	0.074

(1) Holmes, M. A.; Matthews, B. W. *J. Mol. Biol.* **1982**, *160*, 623.
(2) For a detailed discussion of this work, see: Matthews, B. W. *Acc. Chem. Res.* **1988**, *21*, 333.
(3) Tronrud, D. E.; Holden, H. M.; Matthews, B. W. *Science* **1987**, *235*, 571.
(4) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* **1987**, *235*, 574.
(5) Bartlett, P. A.; Marlowe, C. K. *Biochemistry* **1987**, *26*, 8553. Bartlett, P. A.; Marlowe, C. K. *Science* **1987**, *235*, 569.
(6) See, for example: Schaefer, H. F. *Science* **1986**, *231*, 1100.
(7) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.
(8) See, for example: Potsma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. *Faraday Symp.* **1982**, *17*, 55. Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281. Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050. Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564. Singh, U. C.; Brown, F. K.; Bash, P. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1987**, *108*, 1607. Also see ref 4.



4

then carried out electrostatic potential calculations,[9] by using the STO-3G* basis set, on this structure. We, subsequently, took these charges and fixed the leucine portion(s) of the molecule with standard AMBER leucine charges, while we used the newly calculated charges for the $-CH_2-P(O)_2-X-CH-$ portion of the inhibitor. Thus, the only charges that change during the course of the free energy perturbation simulations are the ones for the $-CH_2-P(O)_2-X-CH-$ fragment. The resulting charges are given in Table I.

Having thus obtained the charges for the three inhibitors we assembled a force field, to describe the inhibitor from standard AMBER parameters.[10]

The free energy perturbation calculations were carried out using the GIBBS module of the AMBER[11] suite of programs. We opted to use the "windowing" approach to the determination of the free energy.[4] The protocol for these calculations is similar to that used in our previous effort.[4] In the solution-phase simulations we placed the inhibitor in question in a box of approximately 800 TIP3P[12] water molecules generated from a Monte Carlo simulation. We then partially minimized this system (200 steps of minimization) in order to remove any bad intermolecular contacts. Following minimization we then equilibrated for 4 ps at constant pressure and temperature under periodic boundary conditions.[13] SHAKE[14] was used to constrain all bond distances to their equilibrium values. The perturbations were carried out at constant temperature (300 K) and pressure (1 atm) over 20 windows with a $\Delta\lambda$ of 0.05, and at each window 500 steps of equilibration (0.002 ps timestep) and 300 steps of sampling (0.001 ps timestep) was employed. A constant dielectric of 1 was used throughout, and the nonbonded pairlist had a cutoff of 8.0 Å and was updated every 100 timesteps. While updating the nonbonded pairlist every 100 steps is longer than is commonly used in molecular dynamics simulations of aqueous systems, we have found that for free energy simulations of small solutes in water that updating the nonbonded pairlist "infrequently" has little or no effect on the computed free energies.[15] Furthermore, in our previous work on thermolysin[4] we used the same nonbonded pairlist update frequency and obtained excellent results. In order to assess the degree of hysteresis (i.e., the degree of thermodynamic reversibility) in our calculations we run simulations in both the forward ($\lambda = 1$ to $\lambda = 0$) and backward ($\lambda = 0$ to $\lambda = 1$). Thus, the simulations covered a total of 54.6 ps.

For the enzyme simulations the following protocol was followed: the coordinates supplied to us were first partially minimized (200 steps of minimization) to remove any bad contacts which were presented. A cap of approximately 200 TIP3P waters were placed 18 Å from the zinc ion at the center of the active site and were restrained at the 18 Å boundary only by a harmonic potential with a force constant of 0.5 kcal/Å. All residues which lie 15 Å or more from the zinc ion were fixed, while the residues within 15 Å and the water molecules were allowed to move during the course of the MD run. Glu 143 was treated as neutral, while His 231 was protonated on the basis of the crystallographic results, which strongly suggested that these residues have these charge states.[2] We found it necessary to restrain, using a harmonic restraining force of 50 kcal/Å, the following atoms to their original coordinates: His 142, NE2; His 146, NE2; Glu 155, OE1; the two oxygen atoms on the phosphorus of the inhibitor. The reason for doing this is to stop the distortion of the active site due to the presence of many anionic residues (e.g., glutamic acid) being too strongly attracted to the zinc ion. The zinc ion has a charge of +2.0 and therefore is subject to very strong electrostatic interactions from negatively charged residues, with the end result of this being hexacoordination of the zinc ion. In more recent studies on carbonic anhydrase[16] we have found that the presence of divalent ions will likely be a general problem in the simulation of metalloenzyme systems, in which a catalytically important zinc is present, given the current level of sophistication of force field methods. We are attempting to formulate a more realistic approach to the simulation of systems of this type.[16] The enzyme substrate complex was then equilibrated at 300 K for 4 ps followed by a 10 window perturbation run ($\Delta\lambda = 0.1$) with 500 steps equilibration (0.001 ps timestep) and 300 steps of sampling (0.001 ps timestep) being done at every window. Again, a constant dielectric of 1 was used throughout, and the nonbonded pairlist had a cutoff of 8.0 Å and was updated every 100 timesteps. SHAKE was again used,[14] and the simulations were run in the forward and reverse direction. Thus, the enzyme simulations were run over a total of 17.6 ps.

We also carried out simulations in which our "new" charges were perturbed into the "old" charges from our previous effort.[4] Since we were only dealing with electrostatic changes which only cause small changes in the system, we used the following simulation protocol: For both the solution and enzyme simulations we equilibrated the system for 4 ps, while we carried out the perturbation runs over four windows ($\Delta\lambda = 0.25$) for the enzyme runs and five windows ($\Delta\lambda = 0.2$) for the solution runs. All simulations used SHAKE[14] and 500 steps of equilibration (0.002 timestep) with 300 steps of sampling (0.001 timestep) at each window. The solution runs were carried out at constant temperature and pressure, while the enzyme runs were carried out at constant temperature.[13] The number of water molecules involved and the way in which they were used is identical with that described above.

Finally, we carried out simulations in which we modified our force field parametr for the X–C–O–H torsion contained in neutral carboxylic acids. In the original work[4] on thermolysin a value of 3.6 kcal/mol was chosen, which is significantly lower than the value of 10 kcal/mol used in other force fields.[17] We felt that this modification would have a significant effect on our results due to the interaction of Glu 143 with the perturbation site (see

Table II. Calculated Free Energies of NH → O in kcal/mol[b]

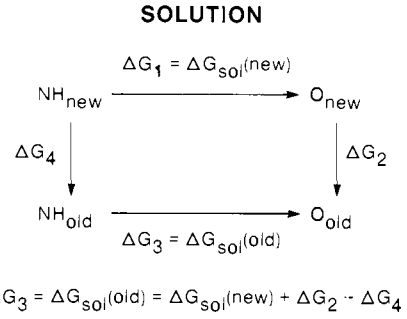
run		A → B		B → A	av
			NH → O		
ΔG_{sol}	F[a]	−0.617	F	+0.282	
					−0.464 ± 0.27
	R	+0.765	R	−0.190	
ΔG_{bind}	F	+6.03	F	−4.978	
		(+3.83)		(−1.881)	+5.507 ± 0.43
					(+2.86 ± 0.87)
	R	−5.465	R	+5.553	
		(−3.32)		(+2.43)	

$\Delta\Delta G_{bind} = \Delta G_{bind} - \Delta G_{sol} = 5.971 \pm 0.6$ kcal/mol
$(3.33 \pm 1.1$ kcal/mol)

[a] In this table and subsequent ones F and R represent the free energies computed in the $(\lambda + \Delta\lambda)$ and $(\lambda - \Delta\lambda)$ directions, respectively. See eq 1. [b] Set 2 values in parentheses.

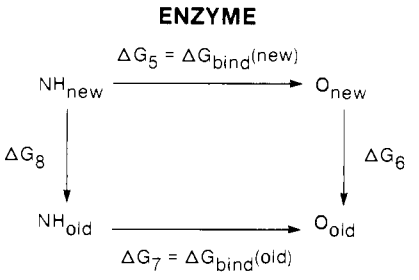
(9) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
(10) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(11) Singh, U. C.; Weiner, P. K.; Caldwell, J. W.; Kollman, P. A. AMBER (UCSF), Version 3.0.
(12) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(13) Berendsen, H. J. C.; Potsma, J. P. M.; van Gunsteren, W. F.; DiNola, A. D.; Haak, J. R. *J. Phys. Chem.* **1984**, *81*, 3684.
(14) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.
(15) Pearlman, D. A.; Kollman, P. A. Unpublished results.
(16) Merz, K. M., Jr.; Kollman, P. A.; Murcko, M.; Hoffmann, R. Manuscript in preparation.
(17) Allinger, N. L.; Chang, S. H.-M. *Tetrahedron* **1977**, *33*, 1561 and references cited therein.

Scheme II



$$\Delta G_3 = \Delta G_{sol}(old) = \Delta G_{sol}(new) + \Delta G_2 - \Delta G_4$$

Scheme III



$$\Delta G_7 = \Delta G_{bind}(old) = \Delta G_{bind}(new) + \Delta G_6 - \Delta G_8$$

below). Since the enzyme simulations were the only ones effected by this change we have redone them using the same protocol outlined above.

Results

We first determined the $\Delta\Delta G_{bind}$ between the NH and O compound and found a value of 5.97 kcal/mol. Previously,[4] our results (4.21 kcal/mol) closely agreed with the experimental (4.1 kcal/mol) $\Delta\Delta G_{bind}$; however, we find in this current study that our results are 1.9 kcal/mol (see Table II) larger than the experimental numbers. We suspected that this difference was wholly due to our choice of a new charge model (see below), which is a direct result of the better structure for the inhibitor we obtained using the 4-31G basis set (Table I gives our new and old charges for the critical atoms involved in the perturbation). While for the NH compound the nitrogen and hydrogen charge differences are small, we find that for the O case that the oxygen charge is more negative for the new set of charges, which will cause a greater repulsive interaction between the carbonyl oxygen on Ala 113 and the oxygen of the inhibitor.[4] There also is a large change in the charge on the phosphorus atom between the "old" and "new" charges of 0.1 (in the O case) and 0.243 (in the NH case). Note also that on going from NH → O with the new charges the difference in the charges on the phosphorous is 0.1, while for the old charges it was only 0.042. The increased repulsion between the P and the Zn atoms would tend to decrease the binding affinity of the O compound. Qualitatively we suspect these interactions to be the dominant reason for the differences in the $\Delta\Delta G_{bind}$ between the two simulations, but we decided to quantitate our results by carrying out simulations in which we perturb the new charges into the old ones. If we find that the changes are constant, some other effect is responsible for the difference in our results (e.g., our simulations are too short); if, however, they are not constant the charges contribute to the differences. The results are given in Table III. We find from these simulations that indeed there is a large difference between the two charge models. In the O case the changes are nearly constant (i.e., $\Delta G_2 \approx \Delta G_6$), while in the NH case there is a 1.68 kcal/mol difference between ΔG_4 and ΔG_8. From the numbers given in Table III and with the use of the thermodynamic cycles given above we can determine the $\Delta\Delta G_{bind}$ for the old charges (see Table III at bottom). It is interesting to note that the ΔG_{sol}(old) and ΔG_{bind}(old) values we calculate here (1.03 and 5.08 kcal/mol, respectively) are significantly different than the 3.44 and 7.64 kcal/mol values calculated previously.[4] The difference in these numbers, however, almost certainly results from the distinctly different ways in which they were arrived at. Thus, in the earlier work we directly evaluated ΔG_3 and ΔG_7 by using simulations covering a total of 80 ps, while in the present work we determined ΔG_3 and ΔG_7 by using shorter simulations lengths and the thermodynamic cycles given above. Given that these simulations are of limited length means sampling of conformational space is not complete. Hence, it is not surprising that these numbers differ. However, the $\Delta\Delta G_{bind}$ we obtain from this computation is 4.05 kcal/mol, which is in good agreement with the previous number of 4.21 kcal/mol.

Table III. Calculated Free Energies for NH(new) → NH(old) and O(new) → O(old) in kcal/mol

run		A → B		B → A	av
		NH(new) → NH(old)			
ΔG_4	F	+1.939	F	−2.824	+2.307 ± 0.58
	R	−1.692	R	+2.772	
ΔG_8	F	+4.185	F	−3.927	+3.981 ± 0.15
	R	−3.824	R	+3.989	
		O(new) → O(old)			
ΔG_2	F	+3.939	F	−3.727	+3.80 ± 0.18
	R	−3.584	R	+3.953	
ΔG_6	F	+3.999	F	−3.120	+3.559 ± 0.59
	R	−4.125	R	+2.990	

$\Delta G_3 = \Delta G_{sol}(old) = \Delta G_{sol}(new) + \Delta G_2 - \Delta G_4 = 1.03$ kcal/mol
$\Delta G_7 = \Delta G_{bind}(old) = \Delta G_{bind}(new) + \Delta G_6 - \Delta G_8 = 5.08$ kcal/mol
$\Delta\Delta G_{bind} = \Delta G_{bind}(old) - \Delta G_{sol}(old) = 4.05$ kcal/mol

Table IV. Calculated Free Energies of CH_2 → NH in kcal/mol[a]

run		A → B		B → A	av
		CH_2 → NH			
ΔG_{sol}	F	−2.217	F	+2.367	
					−2.395 ± 0.28
	R	+2.099	R	−2.627	
ΔG_{bind}	F	−1.852	F	+3.356	
		(−2.72)		(+29.14)	+2.722 ± 0.84
					(+2.40 ± 0.52)
	R	+2.16	R	−3.519	
		(+2.94)		(−1.798)	

$\Delta\Delta G_{bind} = \Delta G_{bind} - \Delta G_{sol} = -0.33 \pm 0.83$ kcal/mol
$(-0.003 \pm 0.80$ kcal/mol)

[a] Set 2 values in parentheses.

The results for the NH → O simulation that used the modified X-C-O-H torsion present in Glu 143 are given in Table II in parentheses. We find that a significant difference in the ΔG_{bind} is obtained when a more realistic torsion potential is used. With the new values for the ΔG_{bind} we calculate a $\Delta\Delta G_{bind}$ of 3.33 kcal/mol, which is in much better agreement with the experimental value of 4.1 kcal/mol.

We next went on to determine the $\Delta\Delta G_{bind}$ between the CH_2 and NH compounds. We expected that the CH_2 compound should have a K_i intermediate between the O and NH compounds because of the anticipated hydrophobic effect that would reduce the solubility of the CH_2 compound. We postulated also that the CH_2 compound would bind to the enzyme better than the O compound because of the reduction of the repulsive interaction between Ala 113 and the inhibitor. Our calculations do indeed suggest this to be the case (see Table IV). We observe, as expected, that the CH_2 is not as well solvated as the NH and O compounds and that it binds the enzyme better than the O compound. In order to make these comparisons we are assuming that the kinetic and intra-perturbed group contributions to the free energy are negligible for all of the inhibitors studied. For the former this has been demonstrated to be the case,[19] and for the latter this is generally so.[20] Although, one does not know that these are negligible in every case, in general we expect them to be small.

(18) Wolfenden, R.; Williams, R. *J. Am. Chem. Soc.* **1983**, *105*, 1028.
(19) van Gunsteren, W. F. *Protein Eng.* **1988**, *2*, 5.
(20) Singh, U. C. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4280.

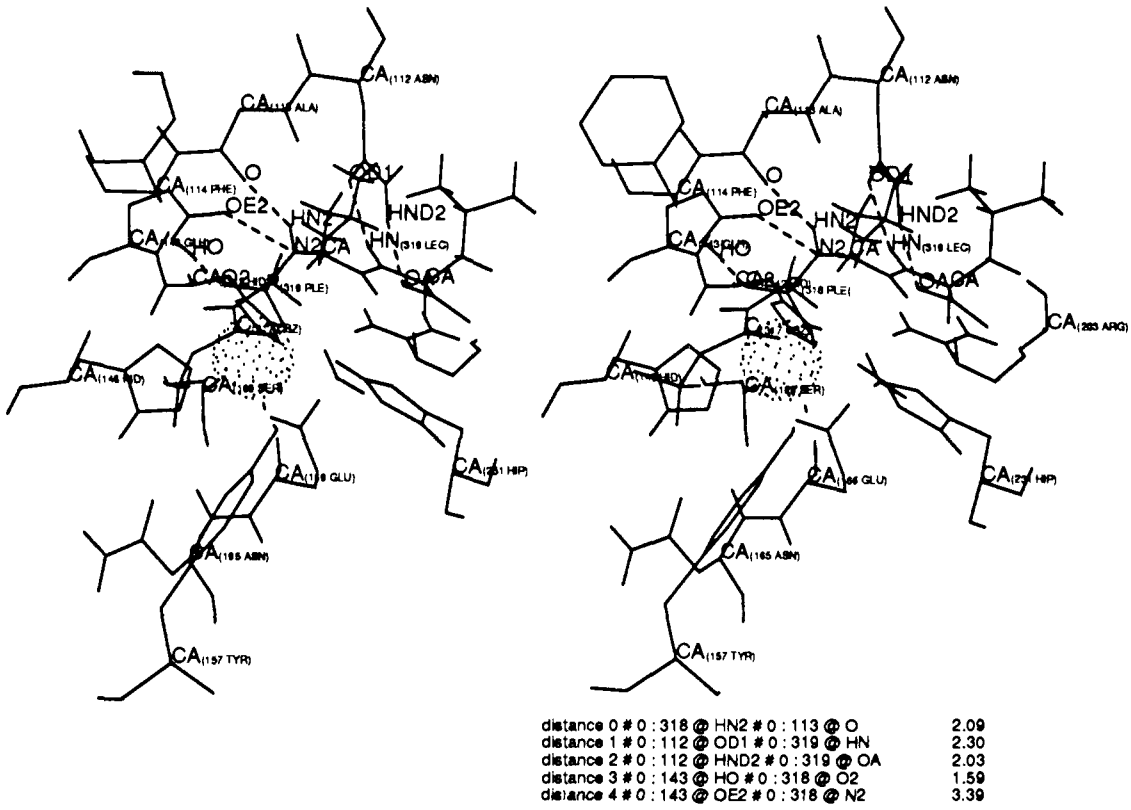


Figure 2. Structure of the thermolysin active site with the NH inhibitor bound.

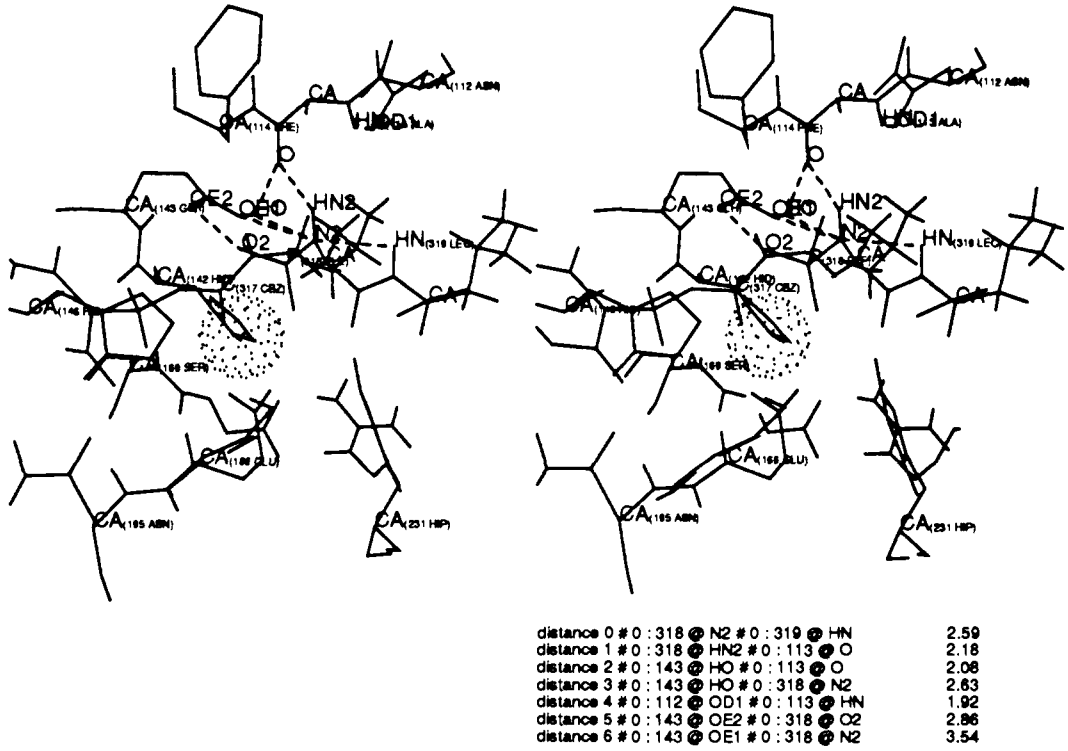


Figure 3. Structure for the set 1 simulations: NH inhibitor and thermolysin active site.

We have also reproduced our results using the modified force field and the results are given in Table IV in parentheses. In this particular situation the results using the modified force field are very similar to the original results. The new $\Delta\Delta G_{bind}$ is 0.0 kcal/mol, while the old value is −0.3 kcal/mol; thus, this simulation is not strongly affected by the change in force field parameter.

Figure 2 gives a stereo picture of the crystallographically determined structure of the enzyme-inhibitor interactions for the NH compound. The residue labeled Ple has the structure –NH-CH_2-$P(O)_2$-Leu, and the residue labeled Lec is a C-terminal leucine. In Figures 3, 4, and 5 we present stereo pictures that are representative of the computationally predicted interactions present between the enzyme and the inhibitor for the first set of simulations. Figures 6, 7, and 8 contain the corresponding stereo pictures for the simulations in which we modified the force field. In what follows we will refer to the first set of simulations (unmodified force field) as set 1 and the second set as set 2 (modified force field).

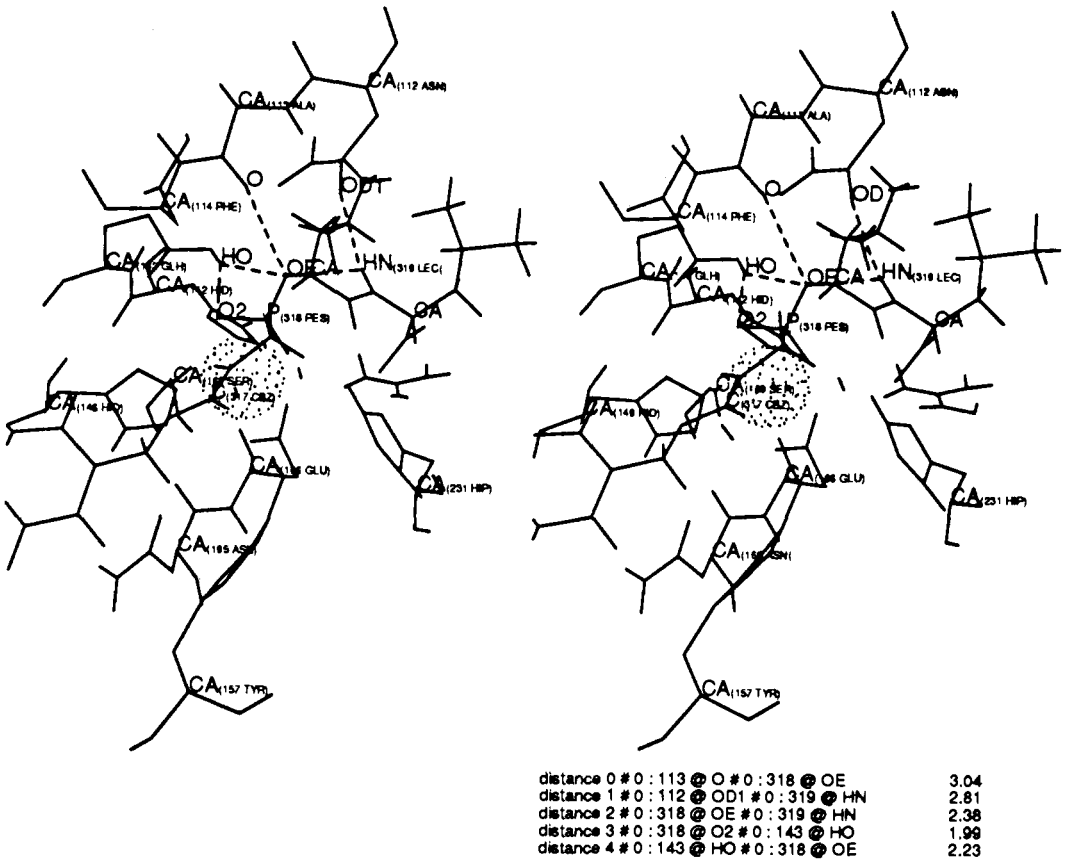


Figure 4. Structure for the set 1 simulations: O inhibitor and thermolysin active site.

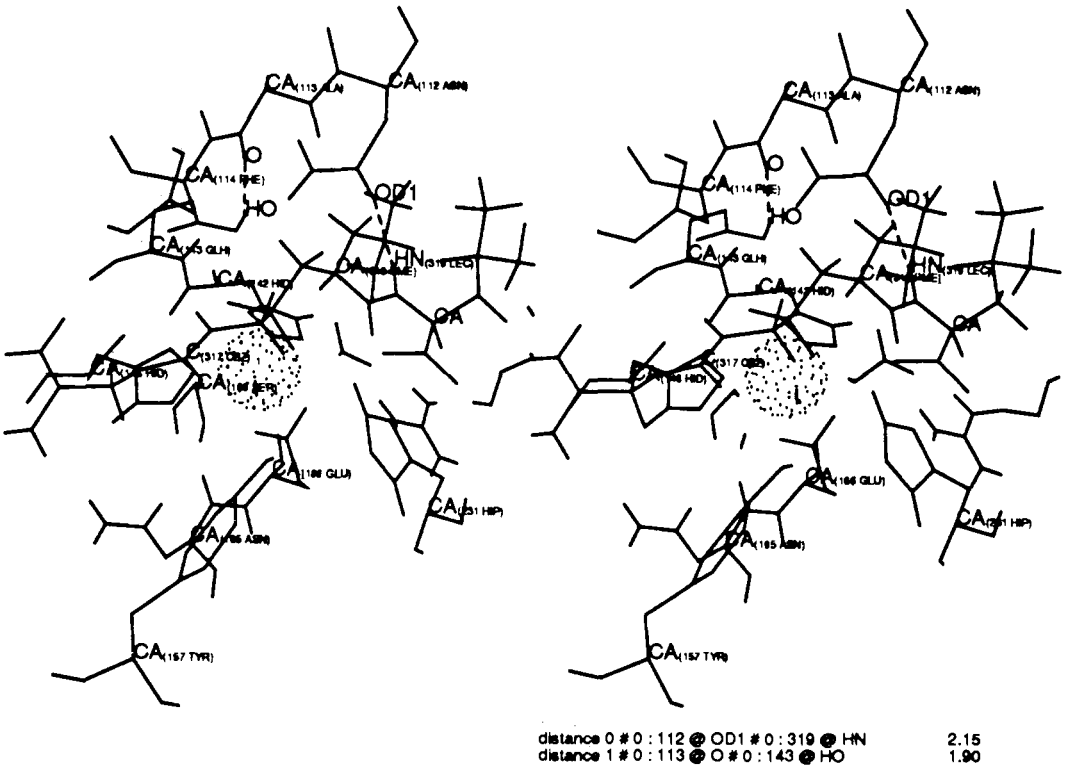


Figure 5. Structure for the set 1 simulations: CH_2 inhibitor and thermolysin active site.

The protein inhibitor hydrogen bonding interactions from the experimentally determined tertiary structure that we want to highlight in Figure 2 are the following: The Glh 143 (Glh is a protonated Glu residue) at HO to Ple 318 at O2, Ala 113 at O to Ple 318 at HN2, the Asn 112 at OD1 to Lec 319 at HN, and the Asn 112 at HND2 to Lec 319 at OA. The first interaction is between a proton bound to glutamic acid and one of the oxygens that is bound to the zinc ion. The bond distance for the hydrogen bonding interactions is $\approx$1.6 Å. The second hydrogen bonding interaction has been implicated in the difference between the NH and O compounds K_i's, while the latter two are conserved in both the NH and O compounds.[3-5] In our simulations we find that

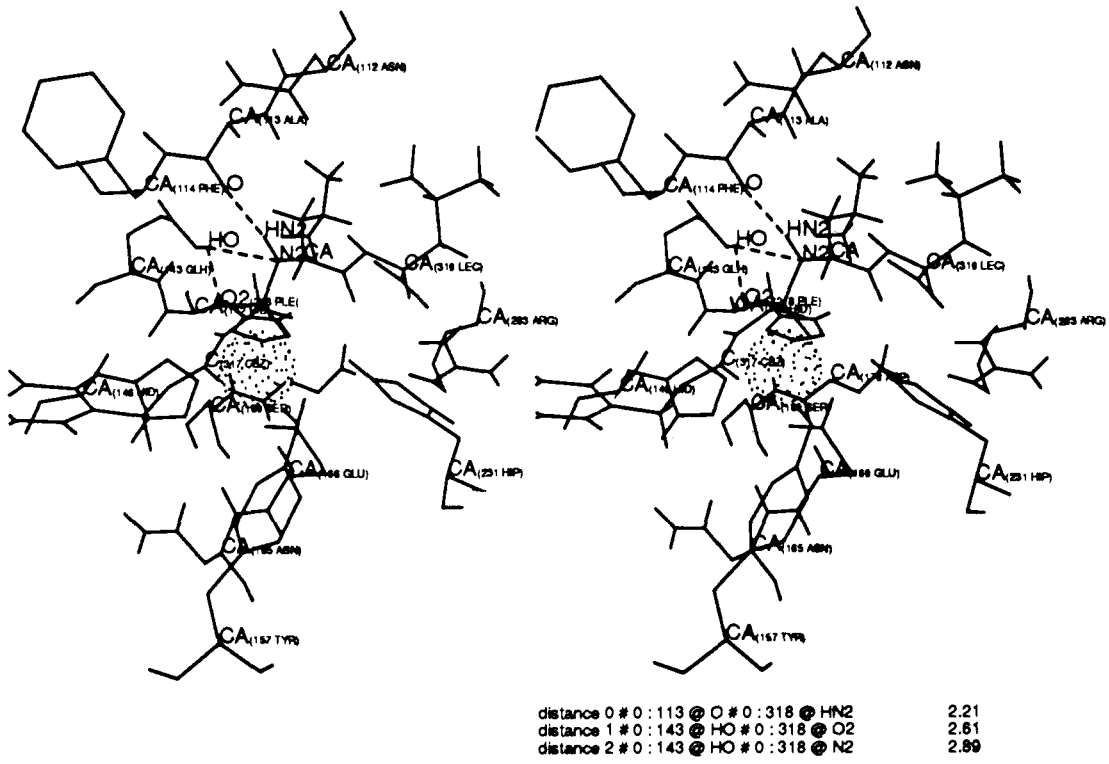


Figure 6. Structure for the set 2 simulations: NH inhibitor and thermolysin active site.

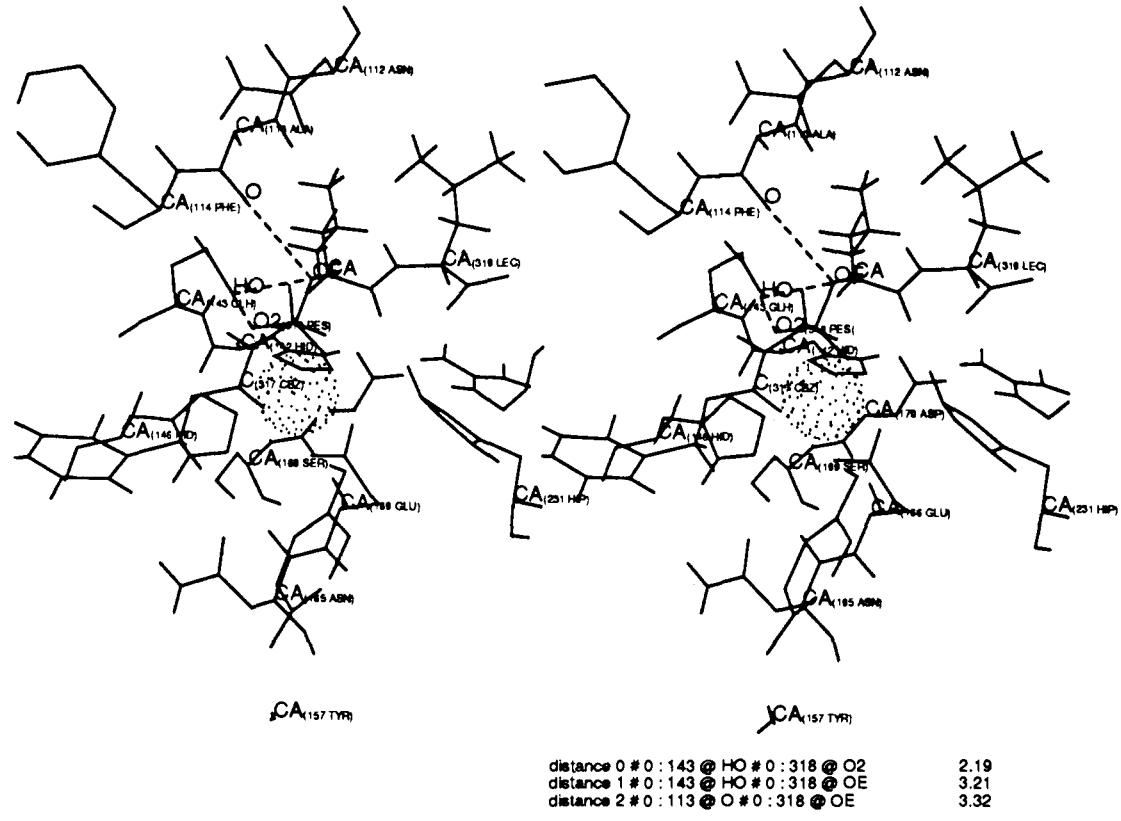


Figure 7. Structure for the set 2 simulations: O inhibitor and thermolysin active site.

these interactions vary considerably.

In both sets of simulation on the NH compound (see Figures 3 and 6) we observe the carbonyl oxygen of Ala 113 interacting with the NH hydrogen. For set 1 we note that the Glh 143 has flipped from the crystallographically inferred position where the proton hydrogen bonds exclusively with oxygen (O2) bound to the zinc (see Figure 2). For the set 1 NH structure the proton interacts strongly with the carbonyl oxygen (O) from Ala 113 (2.07 Å) and weakly with the nitrogen (N2, 2.63 Å). The proton from Glh 143 has also adopted a nonplanar conformation in order for the interactions with N2 and the carbonyl oxygen from Ala 113. However, in the set 2 NH structure, Glh 143 has not flipped completely away from the crystallographically determined position (see Figure 4). Thus, it would appear that the observation that Glh 143 rotates away from its crystallographic position is due, in part, to the choice of the force field parameter for the X–C–

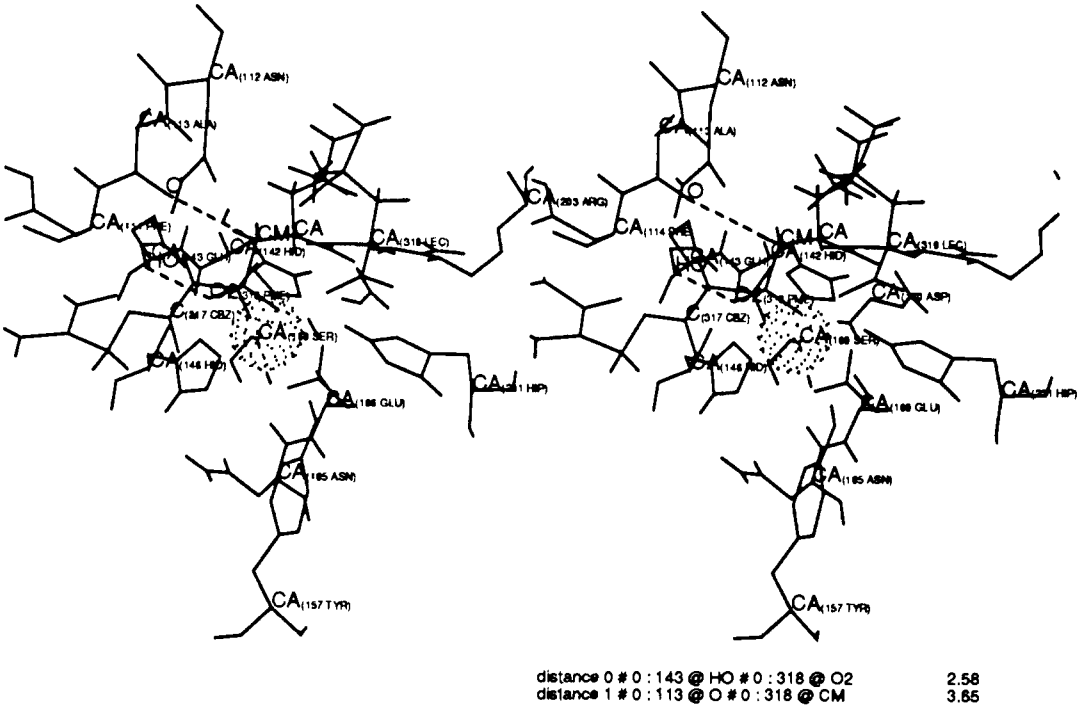


Figure 8. Structure for the set 2 simulations: CH_2 inhibitor and thermolysin active site.

O–H torsion. The interactions between Asn 112 and the terminal leucine (Lec 319) from the inhibitor are no longer present in Figures 3 or 6: these interactions have been disrupted by the formation of hydrogen bonds with the water molecules surrounding the active site.

For the O compound (see Figures 4 and 7) we find that the proton (HO) forms a bifurcated interaction between OE (set 1,2.21 Å; set 2, 3.21 Å) and O2 (set 1, 1.97 Å; set 2, 2.19 Å), while the O from Ala 113 is 3.04 Å (set 1; set 2, 3.36 Å) away from the oxygen. The interactions present between Asn 112 and the terminal Leu from the inhibitor are again absent. Thus, the difference in the ΔG_{bind} for the NH and O compound can be readily understood by the difference in the Ala 113 interaction between NH and O, as has been done previously,[3-5] while the differences between the set 1 and set 2 simulations is entirely due to the choice of force field parameter. A decomposition of what individual interactions cause these differences cannot be easily done here, but apparently the flexibility of the X–C–O–H torsion in the set 1 simulations allows the direct interaction of the hydrogen with the perturbation site (i.e., NH and O), which is then manifested in the computed free energy.

What is the situation in the CH_2 case? In Figure 5 we reproduce the active site region for set 1, from which we observe that the proton from Glh 143 is hydrogen bound to the carbonyl oxygen of Ala 113 (1.91 Å) exclusively. This last interaction might also stabilize the CH_2 inhibitor relative to the NH one, but since this hydrogen bond is not intimately involved in the perturbation, our method will not directly determine its contribution to the enzyme-inhibitor stability. The observation that the Glh 143 proton has rotated by about 90° to hydrogen bond with Ala 113 is an artifact of the parameter choice for the set 1 simulations. In Figure 8 the structure, for the set 2 runs, is more reasonable. The set 2 structure is very close to the NH and O crystal structure, in that Glh 143 has not completely rotated about is Cγ–Cδ bond.

We have also determined the rms difference between the final structures in our enzyme simulations versus the pertinent starting structure. The number of structures for each inhibitor type that were compared was two for the O and CH_2 compounds and four for the NH compound. In all cases the rms deviation is around 1.3 Å.

Finally, a note on these pictures. The structures are taken from the endpoints of the forward and reverse simulations and therefore do not represent the average structure that one would obtain from a "regular" MD simulation. Thus, these pictures just represent a snapshot along the trajectory and could possibly significantly differ from the average structure. For a qualitative discussion, though, these structures are useful, and it was these very snapshots that first helped us realize that the torsional parameter for Glh 143 was possibly too small.

Discussion

Having repeated our previous[4] free energy perturbation simulation for NH → O we have discovered, not unexpectedly, that these simulations can be sensitive to the choice of charge model (i.e., geometry and the basis set chosen for the ESP calculations[9]). In the present context, however, a strong dependence on the charges used might be expected because the perturbation involves an anionic group ($P(O)_2^-$) whose solvation free energy is expected to be of the order of 70 kcal/mol. Thus, the apparently small changes in the charges that occur during the course of the simulation could have a several kcal/mol effect because the overall binding energy of the *whole* inhibitor to the enzyme is large (this binding energy is large due to the strong coulombic attraction present between the zinc and phosphate). A similar argument can be made for the solution simulations because again the free energy of solvation for the whole inhibitor is probably large. It will be interesting to see if a corresponding change in charge model for a neutral inhibitor will behave similarly.

We also have observed that these simulations can be significantly effected by the choice of the force field parameter for the X–C–O–H torsion present in a neutral glutamic acid. It should be stressed, though, that the only reason that we have observed such a large effect is because the neutral glutamic acd in question (Glh 143) interacts with the perturbation site. If this group had been located far away from the perturbation site its effect on the calculated free energy would likely have been small.

In spite of these problems our calculations are still in qualitative agreement with the observation that there is a large preference for the binding of the NH compound. There is no explicit experimental data for the absolute values of ΔG_{sol} and ΔG_{bind} to compare our calculated numbers to; however, there is a ΔG_{sol} value for the conversion of trimethyl phosphate to the corresponding amide,[18] which effectively mimics the conversion of the NH inhibitor to the O one. Wolfenden and Williams[18] showed that the former conversion results in a relative free energy change of 0–1 kcal/mol, and our previous simulations on this model system gave

a value of 0.28 kcal/mol.[4] However, the ΔG_{sol} from our previous work on the conversion of the NH inhibitor into the O one was 3.44 kcal/mol, while our current value is −0.46 kcal/mol, which is in better agreement with the experimental number for the simpler model system. This improvement in the ΔG_{sol}, though, has not translated into a corresponding change in our $\Delta\Delta G_{bind}$.

The determination of the free energy differences using the free energy perturbation method in conjunction with molecular dynamics can be subject to a fair amount of hysteresis (i.e., $\Delta G_A \rightarrow B \neq \Delta G_B \rightarrow A$). In order to reduce the amount of hysteresis observed one has to be acutely aware of the relaxation time scale for the system under scrutiny. This is because the accuracy of a free energy perturbation simulation is directly related to the adequacy of the sampling of phase space. Since the free energy is integrated over the phase space available to the system, the accuracy of the results strongly depend on how effectively phase space is sampled by the simulation. If entropy is a strong contributor to the free energy, then it is even more important to have sampled phase space extensively. Thus, if the simulation is carried out over a time scale that is much greater than the relaxation time scale for the system, it can be expected that the observed hysteresis should be small (i.e., the simulation is reversible). If the simulation is run over a time scale similar to that for the relaxation process we expect that the free energy determined in the reverse direction will not be identical with the forward direction. A real pitfall occurs if the simulation is run over a shorter period of time than the relaxation process, because the determined free energies would appear to be very accurate (i.e., no hysteresis) when in fact they may be very inaccurate because of inadequate sampling of phase space. No firm guidlines are available to decide how long a simulation should be run to get converged results. It is our expectation, though, that the aqueous simulations were run for an adequate length of time (54.6 ps total),[19] while the enzyme simulations, it can be argued, were run for too short of a period of time (17.6 ps). However, in the latter case it is not clear that longer simulation times would necessarily give converged results. The crystallographic results show that no major conformational changes occur on going from NH → O,[2,3] but we observe that during the course of these simulations the active site structure is drifting away from the X-ray structure. Thus, shorter runs might give more reliable $\Delta\Delta G_{bind}$ values in the present situation because the active site distortions would be less pronounced at these time regimes than at longer ones.

The $CH_2 \rightarrow NH$ simulations and the resulting $\Delta\Delta G_{bind}$ make qualitative sense: the computed ΔG_{sol} has the NH inhibitor being better solvated than the CH_2 compound, and the NH compound is more tightly bound to the enzyme. We expected that this would be the case because the loss of a hydrogen bond and the hydrophobic effect should destabilize the CH_2 compound relative to the NH compound in solution, and the loss of the Ala 113 hydrogen bond should also reduce the effectiveness of the CH_2 compounds interaction with the enzyme.

Conclusions

We predict that the order of strength of the present series of thermolysin inhibitors is NH ≈ CH_2 > O. The NH inhibitor owes its stability to the, aforementioned, Ala 113 interaction, while the O inhibitor is weaker because it substitutes this stabilizing hydroben bonding interaction for a repulsive interaction. The CH_2 compound owes its stability to the reduction of the repulsive interaction present in the O compound, along with a reduction in the solvation free energy of this compound relative to the others. We look forward to an experimental study of the CH_2 system in order to either refute or verify our conclusions. Such a study is being carried out by Paul Bartlett and co-workers at UC Berkeley. It is important to carry out such *predictive* studies in order to establish credibility of the free energy perturbation approach. On the basis of realistic error bars for our calculations, it would be fair to characterize our prediction a success if the free energy of binding of the CH_2 compound turned out to be within 1 kcal/mol of that of the NH compound. Nonetheless, one also should not lose sight of the fact that the prediction is being carried out on a system that involves a small topological and structural change. As one validates the method for making predictions on systems with small changes being made, predictions on systems with larger changes can be attempted.

Our results also demonstrate that the choice of a charge model is crucial. What is the best approach for the determination of a charge model? There is no simple answer, but obviously we would prefer to determine charges, by using an electrostatic potential fit,[9] at a level of sophistication that best reproduces the electrostatic properties of the molecule. We further demonstrate that the choice of a force field parameters around the perturbation site is crucial, and the use of less accurate parameters can potentially have a significant effect on the results.

Table V. Inhibition of Thermolysin by Phosphinic Acid Tripeptide Analogues[a]

inhibitor X	K_i values (nM)[b] Y = CH_2[c]	Y = NH[c]	Y = O[c]
NH_2	1400	760	660000
Gly	300	270	230000
Phe	66	78	53000
Ala	18.4	16.5	13000
Leu	10.6	9.1	9000

[a] Determined at pH 7.0, 25 °C as described in ref 1. [b] K_i values for Y = NH taken from ref 1, for Y = O from ref 2. [c] Series.

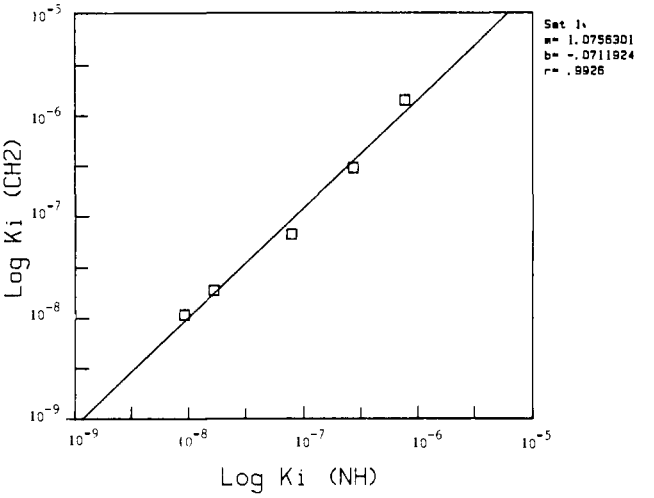
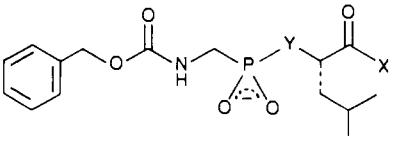


Figure 9. Linear free energy plot correlating equilibrium constants for phosphinates with those for phosphonamidates.

Addendum[21]

Since submission of this communication, we have been informed by Bartlett and Morgan of their synthesis and evaluation of the phosphinic acid tripeptide analogues **5** as inhibitors of thermolysin. These analogues were synthesized in parallel to the theoretical studies reported here, and the inhibition constants were determined in experiments which began in October 17, 1988 (Table I). Figure 9 reveals a strong correlation between the K_i values of the phosphinates **5** and the previously reported phosphonamidates **6**.[22]



5; Y = CH_2, **6**; Y = NH, **7**; Y = O

The correlation suggests that the phosphinates **5** are bound in the active site of thermolysin in a manner similar to that observed

(21) Personal communication from Paul A. Bartlett and Bradley P. Morgan, Department of Chemistry, University of California, Berkeley, CA 94720.

(22) Bartlett, P. A.; Marlowe, C. K. *Biochemistry* **1983**, *22*, 4618–4624.

crystallographically[23] for the phosphonamidates **6** and phosphonate esters **7**.[24] As the inhibition constants recorded in Table I indicate, the binding affinities of the phosphinates **5** are very similar to those of the corresponding phosphonamidates **6**. Across the series, the average difference in free energy of binding ($\Delta\Delta G_{bind}$) is -0.1 kcal/mol, favoring the NH compound), in close agreement with the theoretical predictions of this paper.

Acknowledgment. We are pleased to acknowledge research support from the NIH (GM-29072). The facilities of the UCSF computer graphics laboratory, supported by RR-1081, are gratefully acknowledged as is the supercomputer time from the San Diego Supercomputer Center, supported by the NSF. Finally, K.M. thanks Shashi Rao for many helpful discussions and Paul Bash for providing the coordinates for the thermolysin-inhibitor complex.

(23) Matthews, B. W. *Acc. Chem. Res.* **1988**, *21*, 333-340.
(24) Bartlett, P. A.; Marlowe, C. K. *Science* **1987**, *235*, 569-571.

An ab Initio Study of Crystal Field Effects: Solid-State and Gas-Phase Geometry of Acetamide

P. Popelier, A. T. H. Lenstra, C. Van Alsenoy, and H. J. Geise*

Contribution from the University of Antwerp (UIA), Department of Chemistry, Universiteitsplein 1, B-2610 Wilrijk, Belgium. *Received October 11, 1988*

Abstract: A model of the solid state is constructed by optimizing a central molecule in an electrostatic field of the proper symmetry, produced by Mulliken point charges. Attention is paid to ensure convergence, with respect to the number of contributing neighboring molecules as well as to self-consistency of the field. Despite the neglect of orbital overlap between neighboring molecules, the simple crystal field adaptation of standard ab initio methods is capable of reproducing all significant differences observed between the gas-phase and the solid-state geometry of acetamide. Specifically we note (i) the rotation of the methyl group from a form in which a CH eclipses the C=O to a form in which a CH is perpendicular to C=O, (ii) the increase of C=O and NH bond lengths, the decrease of the CN bond, the stability of the CH bonds, and the variation in valence angles, and (iii) the pyramidization of the amide group. Since the method does not increase the number of two-electron integrals, the crystal simulations do not seriously increase the demands on computer facilities.

There are few theoretical techniques that reliably calculate the geometrical and other properties of molecules in the gaseous *and* solid state. An attractive, computationally simple method has been known for some time,[1,2] but it has not yet been used extensively and its scope is yet unknown. After having proved its usefulness in calculating the librational movements of ethyne in the crystalline phase,[3] we report here on its potential in conformational analysis, taking acetamide as an example.

The molecular structure of acetamide has been studied by diffraction techniques both in the gaseous phase and in the solid state. Kitano and Kuchitsu have analyzed[4] the gas phase by electron diffraction, which shows a planar structure with two hydrogen atoms of the methyl group pointing symmetrically out of the molecular plane and the third C-H eclipsing the C=O bond (Figure 1). In addition to X-ray studies,[5,6] Jeffrey et al.[7] have reported an elastic neutron diffraction analysis at 23 K of the rhombohedral form of acetamide, locating the H atoms with great accuracy. These studies show a solid-state structure, which is significantly different from the gas phase on the following points:

(i) rotation of the methyl group over about 30°, such that one C-H is perpendicular to the C=O bond. In this way the molecule in the solid state has lost the C_s symmetry characteristic of the gas phase (Figure 1);

(ii) large shifts in bond lengths, notably an increase of C=O by 0.030 Å and a decrease of C-N by 0.043 Å, when going from the gaseous to the solid state. These differences are about 10 times their esd's;

(iii) pyramidization of the amide group, showing most clearly in the torsion angles H(4)-N-C(2)-C(1) and H(4)-N-C(2) = 0, which deviate from 180° and 0°, respectively. Consistent with the gas-phase electron diffraction studies,[4] ab initio calculations at the 3-21G level with complete geometry relaxation[7] showed the C_s form to be the lowest energy conformation for the isolated molecule. Attempts, however, were unsuccessful to reproduce by similar calculations the experimentally observed geometry changes when going to the solid. When the methyl group was constrained to the conformation found in the crystal, the other geometrical parameters remained virtually unaltered. We will demonstrate that the incorporation of a simple crystal field model of the solid state into standard ab initio calculations provides molecular models consistent with the available experimental evidence and allows one to reproduce conformational changes that occur when the molecule goes from the gaseous to the solid state. Standard ab initio methods are employed to calculate the geometry and atomic charges of the free (gaseous) molecule. Then, as proposed by Almlöf et al.[1] and later by Saebo et al.,[2] the solid-state model is constructed by surrounding a central molecule with these atomic point charges, put at atomic positions given by, e.g., an X-ray experiment. To be precise, the X-ray experiment provides an external geometry, i.e., the center of mass position and the Eulerian angles of the surrounding molecules with respect to the crystallographic system of axes. The external geometry is fixed, at least for the time being, whereas the internal geometry ($3N - 6$ parameters representing bond lengths, valence, and torsion angles) remains refinable. Subsequently, the influence of the electrostatic crystal field on the central molecule is calculated by ab initio methods, and the resulting new internal parameters of geometry and charges are fed back into the calculations until convergence is reached.

Calculations

Equilibrium geometries of acetamide were calculated using Pulay's gradient method, his computer program TEXAS,[8-10] the 4-21G basis set,[11]

(1) Almlöf, J.; Kvick, A.; Thomas, J. O. *J. Chem. Phys.* **1973**, *59*, 3901-3906.
(2) Saebo, S.; Klewe, B.; Samdal, S. *Chem. Phys. Lett.* **1983**, *97*, 499-502.
(3) Popelier, P.; Lenstra, A. T. H.; Van Alsenoy, C.; Geise, H. J. *Acta Chem. Scand., Ser. A* **1988**, *42*, 539-543.
(4) Kitano, M.; Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1973**, *46*, 3048-3051.
(5) Denne, W. A.; Small, R. W. H. *Acta Crystallogr.* **1971**, *B27*, 1094-1098.
(6) Ottersen, T. *Acta Chem. Scand., Ser. A* **1975**, *29*, 939-944.
(7) Jeffrey, G. A.; Ruble, J. R.; McMullan, R. K.; De Frees, D. J.; Binkley, J. S.; Pople, J. A. *Acta Crystallogr.* **1980**, *B36*, 2292-2299.
(8) Pulay, P. *Mol. Phys.* **1969**, *17*, 197.

0002-7863/89/1511-5658$01.50/0