Extending Crystallographic Information with Semiempirical Quantum Mechanics and Molecular Mechanics: A Case of Aspartic Proteinases

Amiram Goldblum,* Anwar Rayan, Amit Fliess, and Meir Glick

Department of Pharmaceutical Chemistry, School of Pharmacy, Hebrew University of Jerusalem, Jerusalem, Israel 91120

Received September 24, 1992

The results of crystallographic analysis of a complex between an aspartic proteinase, endothiapepsin, and an inhibitor have been extended through the assignment of protons in the active site, to study various steps in the reaction with a substrate. Mechanistic implications are suggested as a consequence of semiempirical quantum mechanical calculations, indicating that most of the activation energy is required to bring the substrate from an initial binding mode to close distance to a water molecule.

INTRODUCTION

X-ray crystallography is considered to be the best method for studying the structure of molecules. However, large molecules such as proteins and their complexes suffer from resolution limitations which result in the lack of information about the positions of their protons. In native enzymes, we are thus unable to determine unequivocally whether some basic residues are protonated or not, especially histidines, and similar difficulty is encountered with the state of carboxylic groups of aspartic and glutamic acids. In addition, the spatial direction of hydroxyl groups of Ser, Thr, and Tyr are not easily assigned. In enzyme–inhibitor complexes, if the inhibitors were designed to mimic intermediates of the enzymatic reactions with substrates, then the correct positioning of protons may lead to an improved understanding of the mechanism of action, even if kinetic studies are not available.

An example is the recently determined complexes of aspartic proteinases with difluoroketone inhibitors.[1,2] The mechanism of peptide cleavage by aspartic proteinases (AP) is not yet known, but the vast interest in this family (that also includes pepsins, chymosin, cathepsins, renin, retroviral proteinases)[3] promotes the design of efficient enzyme inhibitors, preferably based on the concept of mimicking the transition state (TS). This TS is not accessible by experiments, and thus other methods must be employed to study the structure of these short-lived species. The two difluoroketone inhibitors, both in their hydrated *gem*-diol form, are assumed to be "TS mimics" or tetrahedral intermediates that are close in structure to the initial product of an attack by a water molecule on peptide substrates.

Many of the crystals of native AP display electronic density at the center of the active site, attributed to the oxygen of a water molecule. It is nearly equidistant from the two equivalent pairs of oxygen atoms (the two closer ones, called "inner" oxygens, and the two "outer" ones) which are almost at the same plane (Figure 1). The two aspartates are the main machinery of AP for peptide cleavage, which probably involves at least one water molecule. They are believed to be in a monoanionic state due to the bell-shaped pH-profile that shows lower enzyme activity for the assumed neutral or dianionic states.[4] The positions of the three protons in the active site, two of water and one from the two aspartates, could shed some light on the sources of stabilization of the native enzymes' structure. But the oxygen density could also reflect an average position of the water molecule.[5] Another

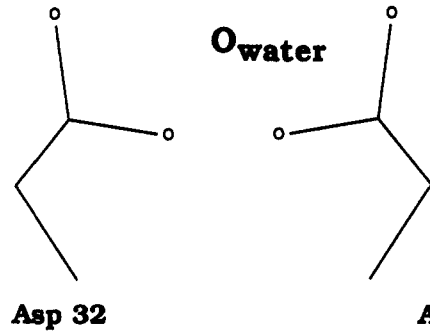


Figure 1. Schematic arrangement in space of two aspartic side chains and oxygen of a water molecule, as found in native aspartic proteinases. The numbers are according to pepsin.

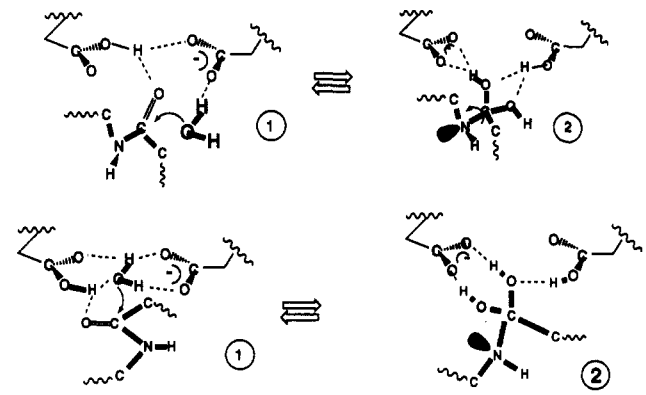


Figure 2. Two alternatives for binding and water attack (step 1) on a peptide substrate in the active site of aspartic proteinase and for the structure of an intermediate (step 2).

water molecule is found in some native AP near the active site, close to the ionized aspartate.[6] The direction of approach and the binding of a substrate in the active site can be affected by the positions of protons in the active site, and such an approach can also determine the spatial relationship to a water molecule and its attack on the peptide carbonyl. NMR evidence for the existence of a tetrahedral intermediate in pepsin's interaction with pepstatone, a noncleavable analog of the effective inhibitor pepstatin, was given by Rich et al.[7] This evidence can not, however, differentiate between alternatives for this "general acid–general base" reaction, such as those depicted schematically in Figure 2. In the upper part, the peptide's carbonyl is polarized by a proton that forms a hydrogen bond between the two inner oxygens of the aspartates, and an outer water molecule, polarized in its turn by its aspartic

0095-2338/93/1633-0270$04.00/0

	32	33	34	35		215	216	217	218
Pepsins:	ASP	THR	GLY	SER	----	ASP	THR	GLY	THR

	25	26	27	28		25'	26'	27'	28'
HIV-PR:	ASP	THR	GLY	ALA	----	ASP	THR	GLY	ALA

Figure 3. Residues of the active sites of AP that comprise our quantum mechanical models.

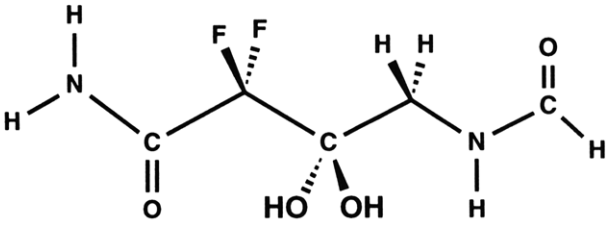


Figure 4. Model of the difluoroketone hydrate inhibitor in quantum mechanical calculations.

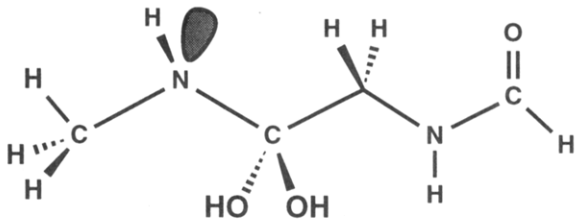


Figure 5. Model of the substrate *gem*-diol intermediate employed in the quantum mechanical calculations.

acid anionic neighbor, attacks the carbon and forms the *gem*-diol intermediate. This scheme requires that the inner water molecule should leave the active site prior to binding of the substrate. In the lower part of Figure 2, the substrate occupies a different position, and the polarizing proton is on an outer oxygen of the neutral aspartate. The water molecule in the center of the aspartates attacks the carbon, again leading to a *gem*-diol intermediate. This intermediate is expected to have a different hydrogen-bonding scheme with the active site.

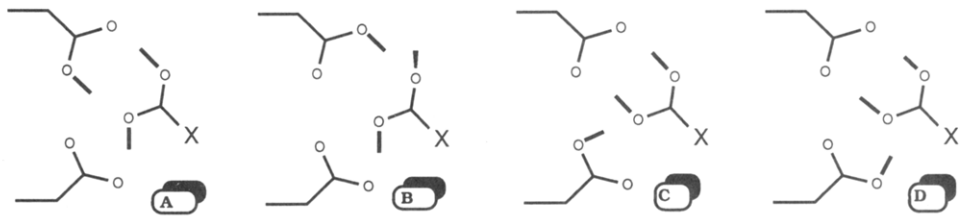
The interaction of the substrate with the active site is an interaction between neutral and negatively charged species, and the approach of a negatively charged oxygen into the anionic center should be repulsive. This consideration is not in favor of the upper scheme of Figure 2. For the lower scheme, the position of a single proton on an outer oxygen in the negatively charged active site does not seem to be an accord with intuition, which favors a hydrogen bond between the two closer oxygens of the two acids. The introduction of a substrate or an inhibitor could alter this intuitive preference, and thus two questions are raised: what is the position of this proton in the native active site, and is it altered by ligand binding. Charge interactions, in principle, should be less specific than dipole–charge, dipole–dipole, or higher multipole interactions. Thus, hydrogen bonding relations of enzymes with ligands could have a detrimental effect on the positioning of substrates and inhibitors in the active site.

Not all AP have uniform interactions with substrates and inhibitors. Even the "universal inhibitor" pepstatin that is employed to identify AP differs largely in its inhibitory potency of several AP. It is, thus, possible that the interactions with ligands are not similar for all AP, and there could be variations in their mechanism of action.

The main goal of this study is to advance from the assignment of active-site protons in the complex of endothiapepsin and a difluoroketone inhibitor to gain more insight to the mechanism of aspartic proteinases. The results for the positions of the three protons mentioned above[8] indicated a clear-cut preference for one arrangement compared to any other, and this structure favors one of a few mechanisms of action. Initial probing of a pathway for this mechanism indicates that activation energy is required to bring the substrate into close contact with a water molecule in the active site, from the substrate's initial binding position. Proton transfers during the attack of a water molecule require low activation energies and are probably not involved in transition states.

METHOD

Our study includes a few steps. First, we position protons in the active site by a semiempirical quantum mechanical method, MNDO/H,[9] that can properly simulate multiple hydrogen-bonding interactions for both geometries and energies. This experiment was done with a relatively small part of the enzyme–inhibitor complex that includes eight active-site residues, considered to be the "rigid core" of the active site. These eight residues include the "aspartic triad" (Asp-Thr-Gly) which appear twice along the main chain in the pepsins or once in each monomer of retroviral AP. In addition, a fourth residue was added to the computations (Figure 3) because in some AP it supports the hydrogen-bonding scheme and contributes to the rigidity of the active site. We have shown that variations in this fourth residue have an effect on local acidity of the active sites and predicted that hydrogen-bonding ligands would bind more strongly to some AP, based on this local acidity property.[10] Our model for quantum mechanical calculations of the difluoro inhibitor contains the most important groups that bind to the active site and to its close vicinity, i.e., to some of the other residues represented (Figure 4).



	A	B	C	D
ΔH_f (rel.)=	16.6	24.9	23.6	0.0
ΔE (rel.)=	21.4	13.6	16.6	0.0

Figure 6. Schemes for the relations between the inhibitor's *gem*-diol and the monoanionic active site of the aspartates. The schemes portray the minima found by MNDO/H. The two lines underneath display the relative energies of (upper line) MNDO/H calculations and (lower line) DISCOVER simulations of HIV-1 PR and the full tripeptide difluoro inhibitor. X denotes the rest of the difluoro-containing inhibitor.

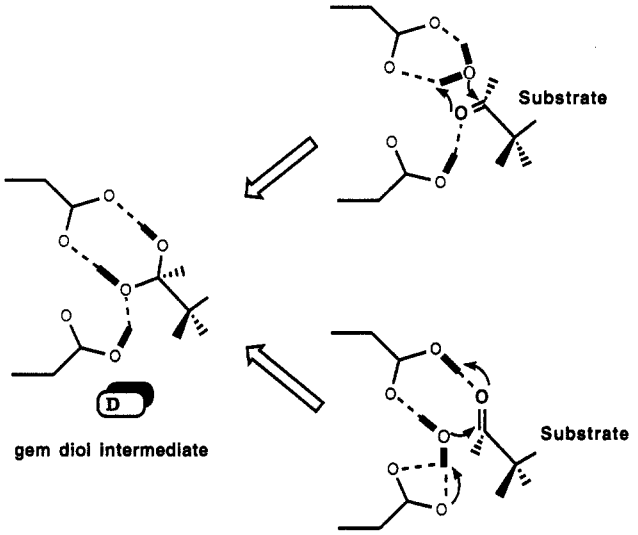


Figure 7. Most stable structure of a hydrated substrate (tetrahedral intermediate) can be produced mainly by the two alternative pathways shown on the right. In these schemes, Asp-32 is the "upper" in each pair.

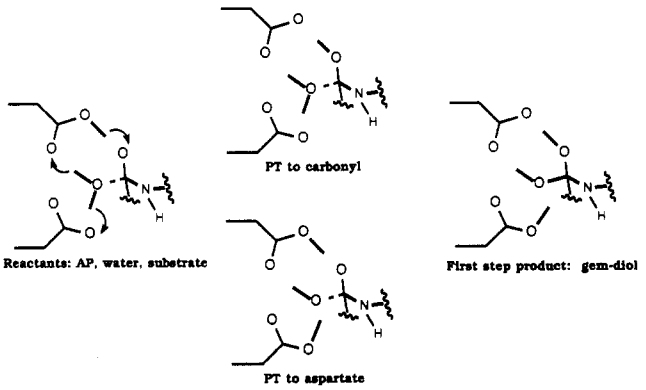


Figure 8. Scheme of species calculated by MNDO/H at several distances of O (water) to C (substrate carbonyl cation).

A full conformational search for all mutual positions of the *gem*-diol has been followed by MNDO/H for each of the four different positions of a proton on an aspartate oxygen. All four were tested since the active site of endothiapepsin is not symmetric, certainly so in the presence of an inhibitor, but also in the native structure, even in the relatively small part of the active site that was represented in the quantum mechanical calculations.

The same study of positioning the three protons was followed with the full native structure of HIV-1 proteinase. The inhibitor from the endothiapepsin complex was introduced to the active site of the native structure. Its initial positioning was achieved by employing distance and angle parameters from the active site of endothiapepsin. It is important to note that a complex of HIV-1 proteinase with this inhibitor has not yet been described and analyzed by crystallography. To improve the position of the inhibitor, we employed the force field called cvff[11] with the program DISCOVER.[12] To extend the simulation of the environment's effect, we added a layer of water molecules (to 2.5 Å), which gave 1150 molecules of water. First, the water layer structure was optimized with the field of the stationary protein with two different options for its active-site proton—an inner and outer position. However, we have not altered the position of the "flap" of this enzyme, which is known to change its conformation while interacting with inhibitors. Thus, more water molecules are found close to the active site than in other complexes of HIV-1 proteinase with inhibitors.

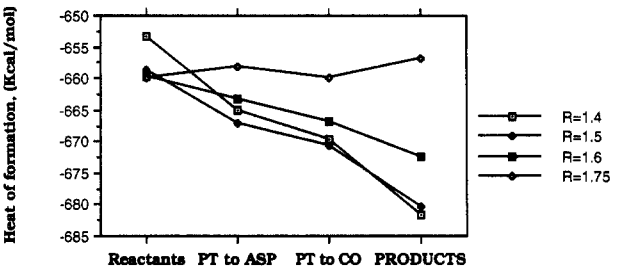


Figure 9. Graphs of the energies four species (Figure 8) at several distances.

The clear-cut preference for one out of the few options for the positioning of protons in the active site–inhibitor complex was continued by a study of the binding preferences for a substrate-derived *gem*-diol (Figure 5) which gave similar results. From this first intermediate on the reaction pathway, we attempted to assign the role of water to each of the two *gem*-diol oxygens and found that only one could assume this role readily. This water molecule attack on the carbonyl of the peptide substrate was attempted at several distances, while following proton transfers toward the intermediate.

Quantum mechanical computations were done on a VAX 9000-210 at the computer center of the Hebrew University of Jerusalem. DISCOVER calculations of the enzyme complex and the environment were run on a Silicon Graphics 4D/280 machine.

RESULTS AND DISCUSSION

Four positions of the proton in the active site are shown, together with the lowest energy positions of *gem*-diol hydroxyls for each, in Figure 6. In these schemes, the differential heats of formation of each species are given. In the upper row, below the schemes, the heats of formation by MNDO/H are given on a relative scale. The conformation denoted "D", with the proton of the active site on an "external" oxygen, is much more stable than any of the others. This external oxygen belongs to Asp-215 of endothiapepsin, which is closer to the two fluorine atoms of the inhibitor, compared to their distance from the other oxygens of the aspartates. We have also demonstrated that the special stabilization of this structure is not due to the fluorine atoms and remains if they are substituted by protons. The inhibitor is bound asymmetrically with respect to an apparent symmetry of the aspartate carboxyl groups (a C_2-axis can be imagined in the plane of the carboxylates, between the two aspartic acids).

In the lower row of Figure 6, the results from force-field calculations for the four options, between a full inhibitor and HIV-1 proteinase (native), are given. The effects of a full enzyme and some of its water environment on the different options (A, B, C, and D) were tested by force-field calculations on the AP of HIV-1, after docking the tripeptide inhibitor into this active site according to its position in *endothiapepsin*. We have studied only the various options for those three protons and assumed that all basic groups (arginines, lysines, and histidines) are protonated while aspartic and glutamic moieties are present as anions. The results for all types of structure optimization, first the protons' positions, then the inhibitor, and finally the water structure, indicate that the same structure that was preferred by the quantum mechanical calculations in endothiapepsin is found to be most stable by this other methodology, that is based on nonbonding (van der Waals and electrostatics, with constant dielectric = 1) and bonding energies (stretch, bend, rotation, and out-of-plane harmonic penalties). Thus, we conclude that form D is the best option

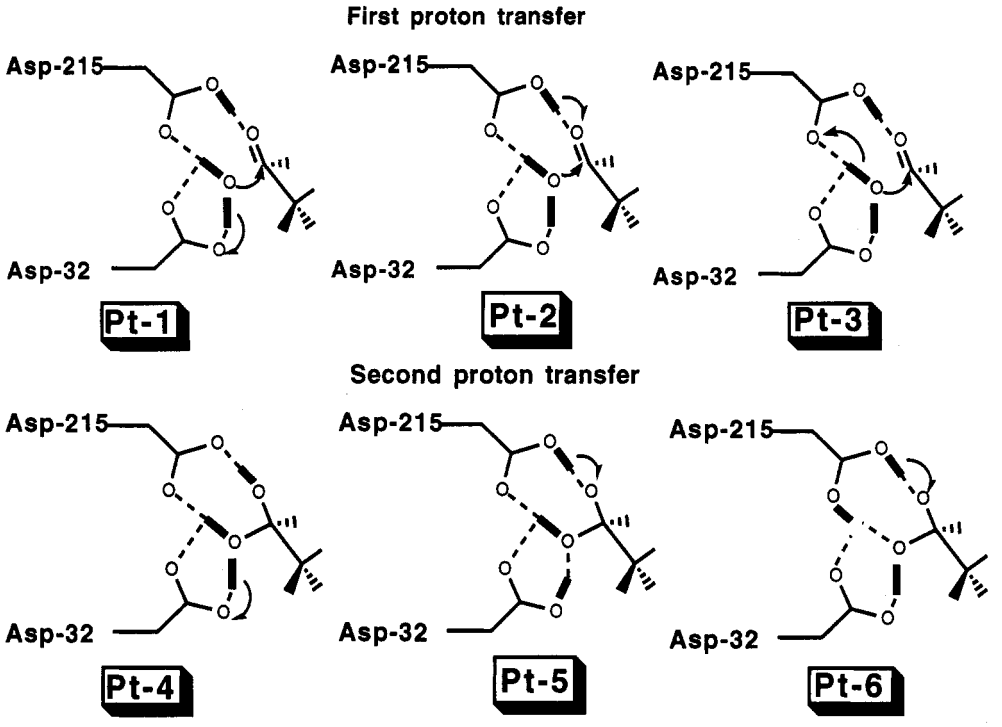


Figure 10. Six proton transfers that were followed by MNDO/H in both the active sites of endothiapepsin and the AP of HIV-1.

for the binding of this inhibitor, and it is unlikely that other computational methods would reverse this clear preference.

The same computer experiments were applied to the substrate's first intermediate, the *gem*-diol shown in Figure 5. It was found to have the same preference as the inhibitor, and thus it is most probable that this intermediate should bind in a similar manner to the inhibitor. Such a hydration intermediate can be produced by an attack of a water molecule that supplies one of two oxygens that comprise the *gem*-diol structure, while the other oxygen is that of the carbonyl. Two main pathways to this diol can be suggested, as depicted in Figure 7. The first, in the upper right part, has the substrate's carbonyl polarized by a proton on the outer oxygen of Asp-215, while a water molecule is polarized by anionic Asp-32. An attack on the carbonyl by water includes, in this case, a proton transfer from water to the carbonyl. In the second case, the carbonyl is polarized by a proton on the outer oxygen of Asp-32, and the attacking water transfers a proton to the outer oxygen of Asp-215.

We have studied by MNDO/H two aspects of the model system: thermodynamics and kinetics. The question posed for the thermodynamic study was as follows: at what distance of the substrate from the water molecule in the center of the active site will the reaction be most favorable? Toward that goal, protons were transferred in the complex of the active site with the substrate's *gem*-diol so that a water molecule and a substrate would interact, in the field of the active site, at several distances of the oxygen of water to the carbonyl carbon of the peptide. For each restricted distance, four species were studied and are presented in Figure 8. From the left side of this figure, with reactants, the reaction could follow to one of the two transient species at the center of the figure, and from each of those it is possible to continue to the *gem*-diol intermediate. Results are given for four distances between the two atoms. We argue that at 1.75 Å, the sequence of reactions (two steps in each case) would not take place because of very little gain of enthalpy (while entropy is certainly negative for the bond formation). At shorter distances, particularly around 1.50 Å, there is a large gain of enthalpy of formation by this reaction. The results are shown in Figure 9.

To facilitate the location of a TS, we studied all the alternative proton transfers by a method developed earlier.[13] The types of proton transfers that were studied are shown in Figure 10. These calculations were followed for both the active site of endothiapepsin and the AP of HIV-1. From the results, it is clear that the proton transfers require very low energies of activation, smaller than 3.0 kcal/mol, and are thus not candidates for the TS. Our present view is that bringing the substrate from an initial binding mode to the active site (which has been only partially characterized by now) to close proximity of the water molecule constitutes the bulk of the activation energy required for transforming the substrate into a *gem*-diol, the first intermediate toward peptide cleavage.

ACKNOWLEDGMENT

This work was supported by Grant 90-00311/1 of the United States–Israel Binational Science Foundation. We are grateful to Dr. T. Blundell for sending us the preprint and supplying the coordinates of the complex of endothiapepsin with inhibitor and of native HIV-1 PR, for suggesting useful ideas, and for discussing the results. We also thank Dr. M. James for sending us the preprint of his manuscript on the structure of the penicillopepsin–difluorostatone and –difluorostatine complexes.

REFERENCES AND NOTES

(1) Veerapandian, B.; Cooper, J. B.; Sali, A.; Blundell, T.; Hoover, D. J.; Rosati, R. L.; Dominy, B. W.; Damon, D. B. Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases. *Protein Sci.* **1992**, *1*, 332–328.

(2) James, M. N. G.; Sielecki, A. R.; Hayakawa, K.; Gelb, M. H. Crystallographic analysis of transition state mimics bound to penicillopepsin:difluorostatine-containing and difluorostatone-containing peptides. *Biochemistry* **1992**, *31*, 3872–3886.

(3) Structure and Function of the Aspartic Proteinases—Genetics, Structures and Mechanisms. In *Advances in Experimental Medicine and Biology*; Dunn, B. M., Ed.; Plenum Press: New York, **1991**, Vol. 306.

(4) Richards, D. D.; Roberts, R.; Dunn, B. M.; Graves, M. C.; Kay, J. Effective blocking of HIV-1 proteinase activity by characteristic inhibitors of aspartic proteinases. *FEBS Lett.* **1989**, *247*, 113–117.

(5) Pearl, L.; Blundell, T. L. The active site of aspartic proteinases. *FEBS Lett.* **1984**, *174*, 96–101.

(6) James, M. N. G.; Sielecki, A. R. Stereochemical analysis of peptide bond hydrolysis catalyzed by the aspartic proteinase penicillopepsin. *Biochemistry* **1985**, *24*, 3701–3713.

(7) Rich, D. H.; Bernatowicz, M. S.; Schmidt, P. G. Direct 13C NMR observation for a tetrahedral intermediate in the binding of a pepstatin analogue to porcine pepsin. *J. Am. Chem. Soc.* **1982**, *104*, 3535–3537.

(8) Goldblum, A.; Glick, M.; Rayan, A. Determining proton positions in an enzyme inhibitor complex is a first step for theoretical mechanistic studies of aspartic proteinases. *Theor. Chim. Acta*, in press.

(9) Goldblum, A. Improvement of the hydrogen bonding correction to MNDO for calculations of biochemical interest. *J. Comput. Chem.* **1987**, *8*, 835–849.

(10) Goldblum, A. Modulation of the affinity of aspartic proteases by the mutated residues in active site models. *FEBS Lett.* **1989**, *261*, 241–244.

(11) Hagler, A. T.; Osguthorpe, D. J.; Dauber-Osguthorpe, P.; Hemple, J. C. Dynamics and conformational energetics of a peptide hormone vasopressin. *Science* **1985**, *227*, 1309–1315.

(12) Biosym Technologies, 10065 Barnes Canyon Road, San Diego, CA 92121.

(13) Goldblum, A. Calculation of proton transfers in hydrogen bonding interactions with semi empirical MNDO/H. *J. Mol. Struct. (THEOCHEM)* **1988**, *179*, 153–163.