# Constructing Protein Models for Ligand−Receptor Binding Thermodynamic Simulations: An Application to a Set of Peptidometic Renin Inhibitors

J. S. Tokarski[‡] and A. J. Hopfinger*

Laboratory of Molecular Modeling and Design (M/C-781), University of Illinois at Chicago, College of Pharmacy, 833 South Wood Street, Chicago, Illinois 60612-7231

Structure-based design is the application of ligand−receptor modeling to predict the activity of a series of molecules that bind to a common receptor for which the molecular geometry is available. Successful structure-based design requires an accurate receptor model which can be economically employed in the design calculations. One goal of the work reported here has been to reduce the size of a model structure of a macromolecular receptor to allow multiple ligand−receptor molecular dynamic (MD) simulations to be computationally economical yet still provide meaningful binding thermodynamic data. A scaled-down 10 Å receptor model of the enzyme renin, when subjected to an alternate atomic mass constraint, maintains the structural integrity of the composite parent crystal structure. A second goal of the work has been to develop schemes to explore and characterize the protonation states of receptors and ligand−receptor systems. Application of the charge state characterization schemes to the hydroxyethylene and statine transition state inhibitors of renin in the training set suggests a monoprotonation state of the two active-site aspartate residues, where the lone proton resides on the outer carboxylate oxygen of Asp226 is most likely. For the reduced amide transition state inhibitors an active site consisting of both aspartates in the totally ionized state, and the ligand carrying a net +1.0 charge, is most stable and consistent with experimental data.

## INTRODUCTION

A current major goal of computational chemistry is the reliable prediction of the binding thermodynamics of a ligand to a receptor of known geometry in a (aqueous) medium. There is an urgency to achieve this goal in order to fully exploit the information being realized from the explosive growth in the availability of receptor geometries and in an increasing number of cases, ligand−receptor complex geometries, from combined recombinant techniques, spectroscopic methods, and homology modeling. The geometry of the receptor provides the fundamental information needed to define necessary but not necessarily sufficient, requirements for ligand binding. This application of computational chemistry is referred to as structure-based design.

The purpose of this paper is to report techniques, methods, and approximations that can be used to make ligand−receptor interaction modeling more robust and reliable. The computer-assisted molecular design (CAMD) tools necessary to extract information from a receptor structure are computationally more economical when applied to smaller systems. One goal of our work has been to reduce the size of a model structure of a macromolecular receptor to allow multiple ligand−receptor molecular dynamic (MD) simulations to be computationally economical yet still yield reliable results. Most receptors can adopt a variable number of protonation states including multiple protonations at the active-site. A second goal of our work has been to develop schemes to characterize the protonation states of receptors and ligand−receptor systems.

The approach to build receptor models presented in this paper is applied to the ligand−receptor system, renin, and a series of inhibitory peptides. Renin, an aspartic acid protease, belongs to a widely distributed family of enzymes that play an important role in fungi, plants, vertebrates, and retroviruses. This class of homologous enzymes includes pepsin, chymosin, cathepsins D and E, related microbial enzymes such as endothiapepsin and penicillopepsin, and the human immunodeficiency virus (HIV) protease. There exists high sequence homology among these aspartic proteases, especially among the residues which comprise the structural core of the active site. This family of aspartic proteases is named as such because they all possess two aspartyl residues at their active site which are involved in the mechanism of peptide cleavage. The binding cleft and active site are at the junction of two structurally similar domains (in HIV protease they are identical). The active-site aspartic acid residues are located centrally in the binding cleft with the carboxyl side chains and surrounding main chain scaffolding related by an approximate interdomain 2-fold axis. A network of hydrogen bonding among the highly conserved residues near the active site keeps these two aspartic acid carboxylates essentially coplanar. The aspartic proteases also share in common a flap region which can open and close upon substrate and inhibitors.

The X-ray crystal structure of recombinant human renin was first determined by Sialecki and co-workers at 2.5 Å resolution.[1] The crystal structure of recombinant human renin in complex with a transition state analog inhibitor was determined by Rahuel et al. to a resolution of 2.4 Å.[2] Dhanaraj and co-workers have also crystallized human renin with a bound peptide inhibitor and defined the structure of the complex at 2.8 Å resolution.[3]

Experimental thermodynamic binding properties, namely, $\Delta H$, $\Delta S$, $\Delta G$, and $K_d$, were reported for a series of 13 renin inhibitory peptides which were made and tested at Upjohn.[4] Several types of inhibitors were synthesized by Epps and
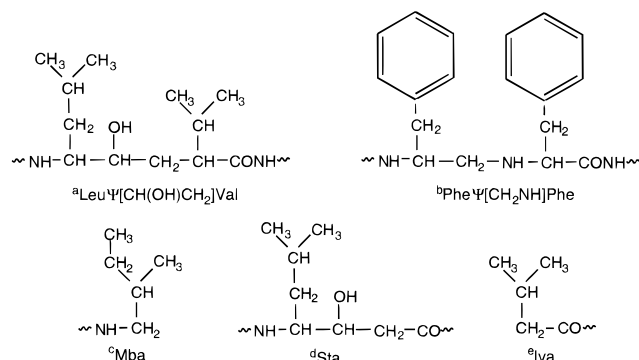
---

* To whom correspondence should be addressed.
‡ Current address: The Chem21 Group, Inc., 1780 Wilson Drive, Lake Forest, IL 60045.

**Table 1.** Chemical Structures of the Renin Inhibitory Peptides Used in the FEFF 3D-QSAR Analysis

| compd | peptide |
|---|---|
| U80631E | Ac-Phe-His-Leu-Ψ[CH(OH)CH$_2$]Val-Ile-NH$_2$[a] |
| U77646E | Ac-Pro-Phe-His-Leu-Ψ[CH(OH)CH$_2$]Val-Ile-NH$_2$ |
| U77647E | Ac-D-Pro-Phe-His-Leu-Ψ[CH(OH)CH$_2$]Val-Ile-NH$_2$ |
| U73777E | Ac-Phe-His-Phe-Ψ[CH$_2$NH]Phe-NH$_2$[b] |
| U71909E | Ac-Pro-Phe-His-Phe-Ψ[CH$_2$NH]Phe-NH$_2$ |
| U77451E | Ac-Pro-Phe-His-Phe-Ψ[CH$_2$NH]Phe-Mba[c] |
| U72407E | Ac-Phe-His-Sta-Ile-NH$_2$[d] |
| U72408E | Ac-Pro-Phe-His-Sta-Ile-NH$_2$ |
| U72409E | Ac-His-Pro-Phe-His-Sta-Ile-NH$_2$ |
| U77455E | Iva-His-Pro-Phe-His-Sta-Ile-Phe-NH$_2$[e] |

**Table 2.** Thermodynamic and Binding Parameters[a] of the Renin Inhibitory Peptides[b] Used in the FEFF 3D-QSAR Analysis

| compd | $K_d$, μm | $-\Delta H°$, kcal/mol | $-\Delta S°$, entropy units | $\Delta G°$, kcal/mol |
|---|---|---|---|---|
| [Leu-Val-OH] | | | | |
| U80631E | 0.37 ± 0.06 | 14.28 ± 0.7 | 75.7 ± 2.7 | 9.2 ± 0.06 |
| U77646E | 0.0054 ± 0.0013 | 28.75 ± 0.6 | 131.1 ± 2.0 | 11.5 ± 0.1 |
| U77647E | 0.0013 ± 0.0002 | 20.33 ± 0.5 | 105.5 ± 1.6 | 12.4 ± 0.2 |
| PheΨ[CH$_2$NH]Phe | | | | |
| U73777E | 0.22 ± 0.01 | 14.20 ± 0.3 | 76.3 ± 0.8 | 9.4 ± 0.1 |
| U71909E | 0.029 ± 0.22 | 13.70 ± 0.5 | 78.4 ± 2.2 | 10.6 ± 0.1 |
| U77451E | 0.0025 ± 0.0015 | 26.70 ± 0.4 | 125.3 ± 1.4 | 12.2 ± 0.1 |
| [Statine] | | | | |
| U72407E | 0.204 ± 0.0012 | 26.10 ± 0.8 | 114.8 ± 2.7 | 9.5 ± 0.1 |
| U72408E | 0.098 ± 0.013 | 14.69 ± 0.9 | 79.6 ± 3.2 | 9.9 ± 0.1 |
| U72409E | 0.023 ± 0.0001 | 22.63 ± 0.5 | 108.0 ± 1.8 | 10.8 ± 0.2 |
| U77455E | 0.0017 ± 0.0001 | 21.36 ± 0.2 | 108.9 ± 0.7 | 12.4 ± 0.04 |

[a] Thermodynamic and binding parameters measured by Epps et al.[8] [b] The renin inhibitory peptides are grouped by class based on their putative transition-state isosteres at the P$_1$–P$_1'$ positions, namely, Leu-Val alcohol compounds, reduced Phe-Phe compounds, and statine compounds.



$^a$LeuΨ[CH(OH)CH$_2$]Val    $^b$PheΨ[CH$_2$NH]Phe    $^c$Mba    $^d$Sta    $^e$Iva

co-workers, by reason of the similarity of the structures to that of angiotensinogen. The structures of the ten inhibitors included in our modeling study are given in Table 1. The thermodynamic binding data of these ten inhibitors are given in Table 2. These inhibitors are transition state analogs because they contain a substructure at the position corresponding to the angiotensinogen cleavage site that mimics the transition state of the natural substrate. The analogs that were modeled in this study include a "hydroxyethylene", "reduced amide", and a statine transition-state isostere. The reduced amide isostere analogs contain a secondary amine at the isosteric replacement (see Table 1). Statine is a rare, natural amino acid found in pepstatin, an inhibitor of pepsin. Each of these transition state isosteres contains a tetrahedral carbon which results in increased rotational flexibility of the inhibitors and perhaps enhanced ability to search for optimal intermolecular interactions in the active site.

An attempt to predict the binding affinities of these analogs based only on the $\Delta H$ values would prove to be unsuccessful. Compound U72407E (see Table 2), has a much more favorable $\Delta H$ of binding than U72408E. Yet, both compounds have about the same $\Delta G$ and $K_d$ measures. The binding entropy, $\Delta S$, of U72407E is much larger than that of U72408E which negates the more favorable $\Delta H$ of binding of U72407E. Thus, this renin-inhibitory peptide data set provides a challenging ligand−receptor system for modeling ligand−receptor interactions.

## METHODS

**1. Thermodynamic Properties of Binding.** The dissociation constants, $K_d$'s, and $\Delta H$, $\Delta S$, and $\Delta G$ values at 37 °C of the renin inhibitors were determined by Epps et al.[4] and are given in Table 2. Ten out of the 13 inhibitors in the Upjohn study were modeled based on the structural diversity and range of activity of the ligands (see Tables 1 and 2).

**2. Building the Molecules.** Rahuel et al. co-crystallized a hydroxyethylene isostere transition state analog, CGP

38'560 (see Figure 1), bound to human recombinant renin at 2.4 Å resolution.[2] The coordinates of the complex have been deposited with the Brookhaven Protein Data Bank[5] under the pdb entry 1RNE. This particular structure of renin was selected as a starting model for the receptor geometry in this analysis for two reasons. Firstly, the geometry of the enzyme with a bound ligand is a more appropriate target for docking the Upjohn inhibitors than attempting to start with the crystal structure of the unbound enzyme. Secondly, the co-crystallized ligand in 1RNE has similar structural features to the Upjohn inhibitors, a property that assists in modeling the ligand−receptor complex.

Most parts of the protein structure are well-defined by the electron density pattern. Two parts of the molecule, which are probably disordered, could not be located unambiguously by the crystallographers, namely the first four N-terminal residues and the loop consisting of residues Asp165 to Ser171 (all residue numbers refer to human renin numbering). The following residue side chains were also not reported in the PDB file, probably due to disorder: Lys50, Arg53, Glu78, Arg139, Lys154, Ser237, Lys249, Lys275, Lys293, and Arg330. All of these residues are on the surface of the protein.

Individual missing amino acid residues were added to the crystal structure of renin where necessary with amino acid fragments from the library of structures in the Chemlab-II molecular modeling program.[6] Any bad contacts within the protein were removed by fixed valence geometry conformational analyses around the torsions of the side chains of the added residues. The missing N-terminal fragment, consisting of Leu-Thr-Leu-Gly, was assembled using residues of the Chemlab-II library. Similar sequences to the N-terminal fragment in the crystal structure of the enzyme were examined for conformational preferences in an effort to establish a suitable conformation for this missing tetrapeptide. The complete N-terminal peptide fragment in a reasonably low energy conformation was added to the renin structure. Again, bad contacts were removed by conformational analyses of the torsions belonging to the fragment.

The missing loop, a heptapeptide sequence consisting of Asp165-Ser166-Glu167-Asn168-Ser169-Gln170-Ser171, was

PEPTIDOMETIC RENIN INHIBITORS

*J. Chem. Inf. Comput. Sci., Vol. 37, No. 4, 1997* **781**

built using the Chemlab-II library. This sequence was modeled in a $\beta$-turn conformation which was created by placing Asp165-Ser166-Glu167 in an extended $\beta$-conformation, Asn168-Ser169 in a type I' $\beta$-turn conformation, and Gln170-Ser171 in an extended $\beta$-conformation. One end of the modeled loop was attached to the crystal structure by forming an amide bond between Arg164 and Asp165. The $\phi$, $\psi$ angles of the loop backbone were scanned in 10° increments between −30° and 30° in order to sample the distance and orientation between the atoms of the unattached residues Ser171 and Leu172. This alignment sampling was done in an effort to join the other end of the missing loop to the rest of the protein crystal structure. Once the main chain of the missing loop was in place, any bad steric contacts between the side chains of the loop and the rest of the structure were removed using a conformational scan of the $\chi$ angles of the appropriate loop side chains.

In native enzymes it is not possible 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. The treatment of protonation states is described below. The appropriate number of hydrogens were "grown" on all atoms of the PDB structure. An option in the molecular modeling package QUANTA[7] was used to assign protons to the crystal structure. All ionizable residues were assigned the charge state which is normally present at the pH of the binding assay experimental conditions (7.4) at which the binding thermodynamics were also measured.[8] Thus, Arg and Lys were assigned a +1.0 charge, and Glu and Asp were each assigned a −1.0 charge. One exception to this assignment was the two active-site aspartic acid residues which were treated separately. The assignment of the protonation states of these residues is considered in a following section. The N-termimus and C-terminus were both modeled as ionized. The total charge of the enzyme, if the two active-site aspartic acid groups are considered ionized, is −8.0. Lone pair electrons were not modeled explicitly. AMBER partial charges[9] were assigned to all atoms of the enzyme structure. Water molecules located in the crystal structure of the renin complex were not explicitly included in the modeling.

In order to check for bad steric contacts of this edited structure, 2000 steps of MD simulation using the program MOLSIM[10] were performed. The side chain of Ile229 was found to be a source of bad steric contacts with other amino acids of the enzyme structure. This unfavorable intramolecular energy was relieved by a conformational scan of the Ile229 $\chi$ angles ± 60° in 10° increments from the values found in the crystal structure. The bad contacts were relieved when $\chi_1$ was rotated −60° and $\chi_2$ was rotated 60°.

The structures of the peptide inhibitors (see Table 2) were also built using the standard amino acid fragments of the Chemlab-II library. The reduced amide isostere, hydroxyethylene isostere, and the N-terminal and C-terminal nonpeptide functional groups were modeled using standard fragments from the Chemlab-II library joined together by standard bond lengths and angles. The isostere replacement, statine, is included in the Chemlab-II amino acid library of structures. The transition state analogs have an additional chirality due to the presence of the isostere replacement fragments. These chiral compounds were modeled as the S-isomer analogs based upon discussions with Epps[11] of the
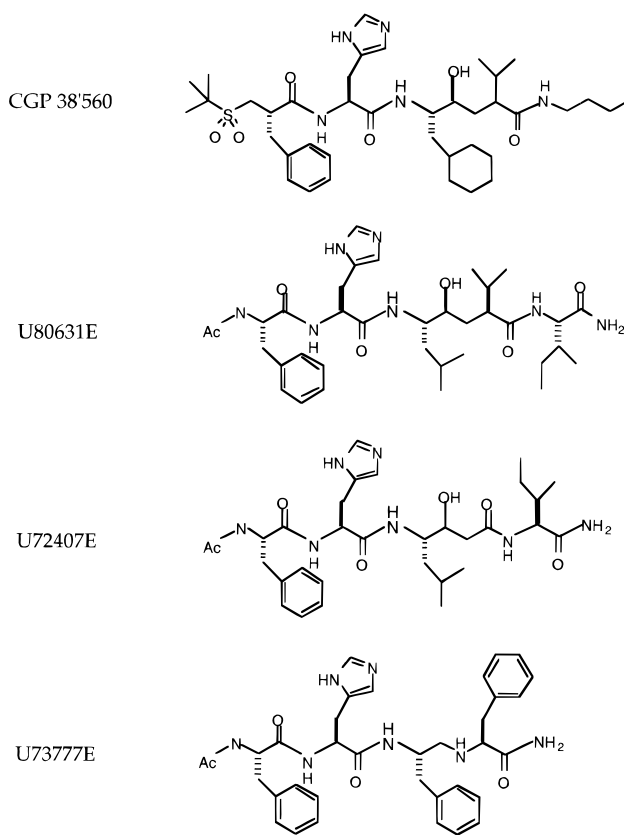


**Figure 1.** The chemical structures of CGP 38′560 and a representative renin inhibitor from each class of transition state isosteres.

Upjohn company. Wherever possible the assignment of AMBER partial atomic charges of the isostere replacements of the analogs were based on similar fragments found among the standard amino acid structures. Semiempirical AM1[12] calculations were performed on the isostere replacements to substantiate the choice of partial charges and/or to establish missing charge assignments. The reduced amide nitrogen was evaluated as both neutral and positively ionized at this point of the study so that both possible sets of partial atomic charges were available. AMBER partial atomic charges on the N-terminal and C-terminal atoms of the inhibitors were adopted from available similar amino acid fragments and scaled accordingly to produce the overall desired charge of neutralilty, or +1.0, where appropriate.

**3. Docking the Ligands.** Ligands of aspartic proteases have generally been found to bind in an extended $\beta$-strand conformation.[13] A starting bound conformation and orientation, i.e., intermolecular alignment, for the modeled Upjohn inhibitors in the active site of renin was established using experimental information. Conformation and alignment information was based upon the bound structure of inhibitor CGP 38′560, which was cocrystallized with recombinant human renin.[2] The structure of CGP 38′560, a hydroxyethylene isostere inhibitor, contains a number of identical amino acid residues to the Upjohn inhibitors. Figure 1 shows the structures of CGP 38′560 and a representative Upjohn inhibitor from each class of transition state isostere studied in this work. Thus, the atomic coordinates, and $\phi$, $\psi$, and $\chi$ torsion angle values, from the crystal structure of this inhibitor were partially used in the initial docking of the Upjohn inhibitors.

Additional information was extracted from the structures of similar ligands cocrystallized with various homologous
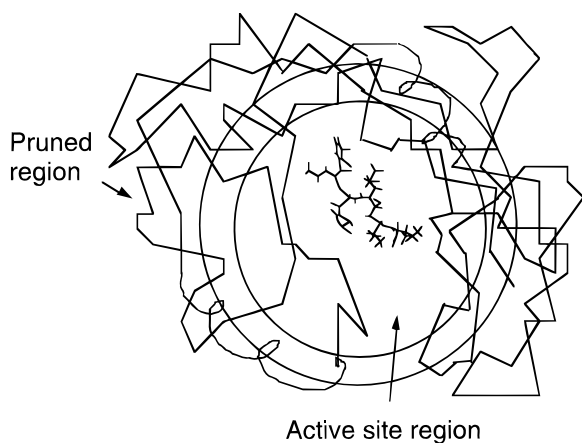
**Figure 2.** A $C_\alpha$-trace representation of the receptor. A ligand is docked in the active site. A scaled-down, or pruned, receptor model includes all amino acid residues within a defined cutoff from any atom of the ligand. Two possible cutoffs are shown.

aspartic proteases. The crystal structure of a hydroxyethylene isostere inhibitor bound to mouse renin[3] and several statine isostere inhibitors cocrystallized with the fungal aspartic protease, endothiapepsin,[13] provided supplementary conformational data which assisted in docking the Upjohn analogs in the active site of the crystal structure of renin. In only a few situations was there a need to relieve bad contacts between ligand and receptor in these initial dockings.

**4. Determination of the Effective Size of a Receptor Model.** The entire model of the renin structure, including protons, contains 5177 atoms. This large number of atoms makes multiple full-scale ligand−receptor MD simulations computationally uneconomical. Thus, modeling approximations were sought to scale down the renin structure to a more manageable size. The analysis was restricted to those amino acid residues of the enzyme near the active-site region. Numerous crystallographic studies of aspartic protease-inhibitor complexes reveal that although binding of the inhibitors is accompanied by significant motion of the flaps of the protease, the active-site structure of the protease in its complexes with inhibitors remains almost unchanged.[14] This observed structural feature of the protease justifies the use of a receptor model centered at the active site.

Three different reduced-size receptor models of renin were modeled to examine how the size of the renin structure could be reduced without losing information regarding the binding process. These models were created by pruning away all residues greater than 12, 10, and 8 Å, respectively, from any atom of the bound ligand. If any one non-hydrogen atom of a residue was within the spherical cutoff, then that entire residue was included in the model. Figure 2 shows a conceptual representation of this receptor model. The pruning operation resulted in a receptor model that was comprised of a number of unconnected peptide fragments. The number of fragments in the receptor model was kept to a minimum by not "cutting" those fragments separated by less than or equal to four residues in the original renin sequence. The largest inhibitor from each class of transition-state isostere was docked in the active site in order to define the largest required receptor model.

Stability of the different size receptor models was evaluated by comparison of a representative structure of the scaled-down models obtained from a MD simulation with that of the crystallographically-determined structure. A MD simula-

tion of 1 ps duration at 300 K was performed on each of the scaled-down models using a nonbonded cutoff corresponding to the size of the model and a time step of 0.5 fs. Both active-site aspartates were assigned the charged state in each of the three pruned receptor models for the size-testing simulations. Compound U73777E was arbitrarily chosen as the ligand for the size-testing simulations and was modeled as positively charged.

In this study all energy minimization and MD calculations used a force field in which the nonbonded, electrostatic, torsional, bond stretching, and bond angle bending force field parameters of the AMBER program[9] were adopted into a MM2 force field[15] potential function representation to compute the parent, nonscaled energy terms. Missing force field parameters (torsional, bond stretching, and bond angle bending) were taken and adjusted from the set proposed by Hopfinger[16] and the MM2 force field. The molecular dielectric was set to a fixed value of 3.5. The "combining" of AMBER and MM2 parameters is accomplished by linear scaling. The most similar set of atoms to those of the missing AMBER parameter is identified for a parameter which has both AMBER and MM2 values. The unknown AMBER parameter value is then scaled against the known MM2 value in the same ratio as the parameter having both AMBER and MM2 values. Since the ultimate goal of our work (see the companion paper[30]) is to develop a refined custom force field for a specific ligand−receptor system, limitations in this linear scaling are considered in the force field refinement. Energy minimization and MD simulations were performed using the MOLSIM[10] program. The hydration shell model proposed by Hopfinger[17] was included in the force field representation to estimate aqueous solvation energetics.

The lowest energy structure for each of the receptor models from the MD simulations was compared by root mean square (RMS) fit to the equivalent portion of the crystal structure of renin. In addition, a 1 ps MD simulation at 300 K employing a nonbonded cutoff of 10 Å was performed on the entire renin crystal structure in complex with positively charged U73777E.

The ligand−receptor interaction energy was examined for each of the scaled-down receptor models using a nonbonded cutoff value of 12 Å. The cutoff of 12 Å was chosen to ensure that all possible interactions up to and including the distance of the largest receptor model size were being considered. Also, because of the low value assigned to the molecular dielectric for use in the MD simulations, potentially ionizable residues beyond the 8 Å of the ligand were modeled as neutral. The assignment of neutral charge for the size-testing simulations ensures that comparisons of interaction energies among the models are valid and is rationalized because long-range electrostatic interactions should appear as neutral due to mobile counterion build-up around ionic sites.

It was found necessary to apply a constraint to some of the atoms in the receptor models to "fix" their positions in space. A number of atoms were each assigned a fictitious mass of 5000 to serve as "momentum reservoirs" and prevent significant departures from the model geometries due to the exclusion of the rest of the protein. The criteria for selecting which atoms to constrain was based on the RMS fit of the structure of the receptor model to the starting crystal structure. Atomic constraints have been found necessary to maintain the structural integrity of a crystal enzyme structure

and to calculate reliable relative free energies of binding, in other MD simulations.[18,19]

**5. Determination of the Protonation State of the Active-Site Aspartic Acid Residues.** Most of the information about the structures of aspartic proteases is based on X-ray crystallographic experiments of the native enzymes and of their complexes with inhibitors. Many of the crystal structures of native aspartic proteases display considerable electron density in the midst of the active-site aspartyl residue pair which has been attributed to the oxygen of a tightly bound water molecule. It is now widely believed that the active-site aspartyl groups assume opposite roles in a general acid−general base catalysis. In this acid−base catalysis, deprotonation of the catalytic water is performed by the unprotonated aspartyl residue, and polarization of the carbonyl oxygen of the natural substrate is achieved by the protonated aspartyl residue. It has been proposed that the aspartyl groups are in opposite states of protonation, such that the catalytically relevant form of the enzyme is "monoprotonated". This conclusion is based on three-dimensional structural data of these enzymes,[20] affinity labeling studies, and the observation of a "bell shaped" pH activity profile.[21,22] However, even when it is assumed that the catalytically competent form of the enzyme is monoprotonated, the position of the lone acidic proton still remains uncertain.

A study of the geometries of enzyme−inhibitor complexes may help in assigning the location of protons at the active site and an improved understanding of the mechanism of action. The direction of approach and the binding of a substrate can be affected by the position of protons in the active site. Conversely, the protonation state of the active site may be altered by ligand binding. Since hydrogen bonding is a dipolar interaction, which is highly dependent on the orientation of interacting atoms, the placement of protons on the enzyme and ligand is critical to the equilibrium alignment of an inhibitor in the active site.

Numerous experimental studies and molecular modeling simulations of aspartic protease−inhibitor complexes have been performed, but the protonation state(s) of the catalytic aspartic acid residues remains unresolved.

In the analysis of ligand binding for five endothiapepsin crystal complexes Blundell et al. list the close multiple interactions between the transition-state mimetics and the catalytic aspartates.[13] The interactions with aspartyl groups represent, in their estimation, the highest contribution to tight binding between the inhibitors and the enzyme. It is suggested that the binding of the transition-state mimetics at the active-site region of the aspartates is the primary site of association and takes precedence over other interactions. This observation emphasizes the importance of modeling complementarity of the ligand−enzyme complex at the active site. Certainly when modeling the ligand/enzyme interaction where, and how, protons are assigned to the active site will influence the calculation of the free energy of binding. Overall, it is worthwhile to review experimental and theoretical research on aspartic proteases to assimilate information that can help specifically in modeling renin and its inhibitors and, more generally, ligand−receptor binding.

First, however, it is useful to point out common nomenclature regarding the active-site aspartates. It is unambiguous when the oxygens of the carboxylates are refered to as "inner" and "outer" based on their interatomic distances. The inner carboxylate oxygens are those of the two aspartyl
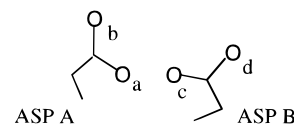


**Figure 3.** The aspartate carboxylate oxygens which are labeled $O_b$ and $O_d$ are referred to as the "outer" oxygens, and the oxygens, $O_a$ and $O_c$, are referred to as the "inner" oxygens.

carboxylates in closer proximity to each other. The assignment of these oxygens in this manner is consistent for all aspartic proteases. Figure 3 depicts the nomenclature for these atoms. The inner carboxylate oxygens are designated "a" and "c", and the outer carboxylate oxygens are designated "b" and "d" in Figure 3.

Suguna et al. cocrystallized a reduced amide isostere transition state analog bound to rhizopuspepsin, a fungal aspartic protease.[23] From an analysis of the crystal structure of the complex, they proposed that a hydrogen bond is formed between the secondary nitrogen atom of the reduced amide isostere and the outer carboxylate oxygen of the active-site aspartate Asp218 (rhizopuspepsin numbering). A corresponding mechanism is proposed for the hydrolysis of the natural substrate in which only one of the catalytic aspartic acid residues in the protease active site is protonated over the course of the catalytic mechanism of rhizopuspepsin.

Veerapandian et al. proposed a general mechanism for proteolytic cleavage of the amide bond of the natural ligand by an aspartic protease based on the analysis of the 2.0 Å resolution X-ray structure of endothiapepsin complexed with a difluorostatone-containing tripeptide renin inhibitor.[20] The ketone of the inhibitor is hydrated to a *gem*-diol in the complex and resembles the transition state of the natural substrate. From an analysis of the positions of the *gem*-diol oxygens of the inhibitor and the aspartyl carboxylate oxygens, several possible models are plausible for the location of the protons at the active site. An accounting of the number of hydrogen bonds to each aspartate oxygen in the various models led the investigators to postulate a mechanism of hydrolysis of substrate (shown in Figure 4) in which Asp32 (endothiapepsin numbering) is the initial proton donor. The Asp215 carboxylate polarizes a water, bound between the catalytic aspartates, for nucleophilic attack. This reaction leads to the formation of the tetrahedral *gem*-diol transition state of the peptide substrate. The reduced amide nitrogen undergoes inversion and a slight rotation of its bond with the carbonyl to accept a proton from one of the aspartic acid residues. This is followed by abstraction of one of the hydroxyl protons of the *gem*-diol which leads to formation of products. This mechanism thus suggests that, in the transition state, one of the catalytic aspartic acid residues exists in the neutral form, whereas another residue is negatively charged in the presence of a substrate. Over the course of the hydrolysis mechanism of the natural ligand, the acid−base role of each aspartic acid residue switches several times. An analysis by Dhanaraj and co-workers[3] of the geometry of the renin active site complexed with a norstatine isostere transition-state analog inhibitor reinforces this view[20] of the protonation state and hydrogen-bonding network of the endothiapepsin-flouroketone hydrate complex as well as the putative tetrahedral intermediate in proteolytic cleavage of the amide bond. In their ligand−receptor binding model, the outer carboxyl oxygen of Asp226 (renin numbering) is hydrogen-bonded to the norstatine ester oxygen and
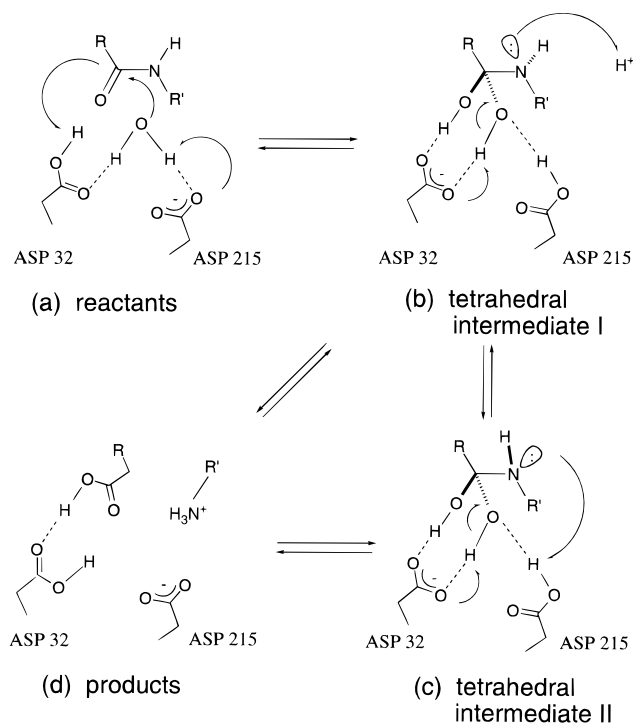
(a) reactants

(b) tetrahedral intermediate I

(d) products

(c) tetrahedral intermediate II

**Figure 4.** A proposed mechanism for cleavage of the amide bond of the natural ligand by an aspartic protease.[20] Aspartic acid residues are numbered according to the endothiapepsin sequence.

is, therefore, protonated. Asp38 is ionized and receives a hydrogen to its inner carboxylate oxygen from the inhibitor hydroxyl group.

Hyland et al. studied pH rate behavior, and solvent kinetic isotope effects, to delineate the details of the chemical mechanism of HIV-1 protease using four oligopeptide substrates and two competitive inhibitors.[21,22] From the pH studies it is concluded that a hydroxyethylene isostere inhibitor preferentially binds to the protease when one of the aspartyl groups is protonated and the other is unprotonated. However, a reduced amide isostere transition state analog, which contains a cationic tertiary amine over the pH range of the study (3–7), binds much more favorably to a form of the protease in which both of these aspartyl groups are unprotonated. The authors propose that this inhibitor forms an ionic bond with the outer carboxylate of an aspartyl residue, the equivalent of Asp125.

Jaskolski et al. also proposed an enzymatic mechanism model based on X-ray crystallographic analysis of a complex between HIV-1 protease and a transition state analog inhibitor which contains a hydroxyethylene isostere replacement at the scissile bond.[24] The OH group of the inhibitor occupies the same position between the carboxylic group of the aspartates as the water molecule found in the active site of all uncomplexed aspartic proteases. These workers conclude that one of the aspartates (the one that does not form the inhibitor C–O–H⋯O bond) must be protonated in order to bind to the inhibitor's hydroxyl group. The outer O atom is more likely to be protonated, in their estimation, since locating the proton on the inner carboxylate oxygen atom would require the other inner oxygen atom to accept three hydrogen bonds (one between the inner oxygens and also from Gly27 and from the inhibitor C–OH). Jaskolski et al. conclude that this proton is also present in the native enzyme, i.e., the proton that is present when an inhibitor is bound is also present in the unbound state. The proposal is made that

when the catalytic water is bound by the two active-site aspartates, the water hydrogens are bound in a symmetrical fashion. A water proton is bound by one of the outer aspartate oxygen atoms, and the other water proton is shared by both inner aspartate oxygen atoms to form a bifurcated hydrogen bond. The other aspartate outer carboxylate oxygen atom would serve as hydrogen bond donor to the water molecule. Movement of the protons in the short outer carboxylate oxygen–water hydrogen bonds would make the active site dynamically symmetric. An extensive dynamic sharing would evenly distribute the acidic proton over all four carboxylate oxygen atoms. Perturbation of this symmetry, for instance, by an approaching substrate/inhibitor molecule, could lock the dynamic equilibrium into one preferred state.

The enzymatic mechanism proposed by Jaskolski et al. is similar to that proposed by others, but there are important differences. These authors suggest that the cleavage process is essentially a one-step reaction during which the nucleophile (water molecule) and electrophile (an acidic proton) attack the scissile bond in a concerted manner. Jaskolski et al. assign this acidic proton to the outer carboxylate oxygen of that aspartate which is proximal to the nitrogen atom of the approaching amide, as opposed to Suguna et al.[23] and Veerapandian et al.[20] who assign this proton to the other outer aspartate.

The assigned protonation states of some models of the active-site aspartates residues have not always been consistent with experimental results. Goldblum et al. stress that determining proton positions in an enzyme–inhibitor complex is critical to the success of theoretical mechanistic studies of aspartic proteases.[25] They performed MNDO/H calculations using a nonsolvated model of the active site of endothiapepsin complexed with a difluorostatone-containing tripeptide renin inhibitor. The strategy used to identify the "real" structures was to characterize the different possible protonation states by computing their energies. All four distinct positions of a proton on an aspartate oxygen were tested since the active site of endothiapepsin is asymmetric. The positions of the hydroxyl protons of the *gem*-diols were also determined. The authors considered different hydrogen-bonding arrangements established by conformational searching and energy minimization. In the active site of the inhibitor complex, stability is much higher for protons on either of the two outer carboxylate oxygens. Specifically, the structure with residue 215 (endothiapepsin numbering) protonated on the outer carboxylate oxygen is found to have the lowest energy of all the alternatives.

Chen and Tropsha addressed the issue of the protonation state of the catalytic aspartates in the presence of an inhibitor as part of their study of a complex of the HIV-1 protease and the *S* and *R* isomers of the hydroxyethylamine transition state mimetic inhibitor U85548E.[26] Different mono- and diprotonated states of the catalytic aspartic acid residues were considered in MD simulations of the inhibitor–protease complex which were solvated with SPC water molecules.[27] The location of a proton(s) substantially influenced the calculated difference in binding constants of the two stereoisomers. Starting with the X-ray structure of the protease-inhibitor complex, they placed a proton successively at all four oxygen atoms and applied molecular mechanics energy minimization to optimize the location of both this proton and the proton of the inhibitor hydroxyl group. They found

PEPTIDOMETIC RENIN INHIBITORS

*J. Chem. Inf. Comput. Sci., Vol. 37, No. 4, 1997* **785**

that the lowest energy for the complex was achieved when the proton was placed at the inner carboxylate oxygen of Asp125 (which is the spatial equivalent to renin Asp226). Similar considerations in treating the diprotonated complex resulted in placing protons on the inner carboxylate oxygen of Asp25 (which corresponds to Asp38, renin) and the outer carboxylate oxygen of Asp125 for one model and on the outer carboxylate oxygen of Asp25 and inner carboxylate oxygen of Asp125 for a second model. Free energy perturbation (FEP) simulation theory was used to evaluate the relative binding affinities of the *S* vs *R* isomers of U85548E and led to the conclusion that only one catalytic aspartic acid residue is protonated. This proposal was based on the fact that the relative binding affinity of the *S* and *R* isomers for this protonation state was similar to that calculated by modeling and suggested by experiment for other HIV-1 protease−inhibitor complexes where the structure of the inhibitor is similar, and the enzyme was modeled in the monoprotonated state.

Harte and Beveridge attempted to predict the protonation state of the active-site aspartyl residues in HIV-1 protease−inhibitor complexes using MD simulations.[28] They examined two types of inhibitors, a hydroxyethylamine transition-state analog (U85548E) and a reduced amide type transition-state analog (MVT-101). In an earlier simulation of the enzyme/U85548E complex they found that the position of the inhibitor in the binding pocket deviated considerably from that found in the crystal structure. They had used a monoprotonated active site which they denote the $(-1, 0)$ state. The results from this earlier study prompted the authors to explore the protonation state of the aspartyl groups as a possible source of the discrepancy in binding geometry. They performed MD simulations using the SPC model for water[27] on the aforementioned two complexes examining all possible protonation states of the two catalytic aspartates which they refered to as the dianionic form $(-1, -1)$, the monoanionic forms $(-1, 0)$ and $(0, -1)$ in which one of the catalytic aspartic acid groups is anionic and the other protonated, and the diprotonated, or neutral, form $(0, 0)$. No mention is made as to whether the proton is placed on the inner or outer carboxylate oxygen of the neutral aspartyl residue. The MD simulation geometries were compared to the corresponding crystal structures, and it was found that only one of the protonation states of the enzyme complexed with the different inhibitors is in good agreement with the crystallographic data. The basis for the geometric comparisons is the residual RMS deviation of the average MD structure for the inhibitor geometry superposed on the X-ray inhibitor structure. The RMS deviations are significantly different for the structures of each of the possible protonation states. The simulation results support the diprotonated state $(0, 0)$ for U85548E, the same hydroxyethylamine transition-state analog modeled by Chen and Tropsha,[26] and the dianionic state $(-1, -1)$ for MVT-101, the reduced amide type transition-state analog. Harte and Beveridge rationalize the different enzyme protonation states for the two inhibitors by proposing that each ligand presents a different environment to the aspartyl side chains. The hydroxyethylene peptide isostere is neutral, while the secondary amine of the reduced amide isostere has a positive charge.

Ferguson and Kollman[18] also explored the protonation state of the aspartyl groups in their modeling of HIV-1 protease. Three possible protonation models were considered in the

determination of the relative binding free energies of two isomeric inhibitors by FEP theory. One model left both aspartyl residues anionic, while the remaining models had one, or the other, aspartate protonated. These investigators conclude that the most likely protonation model is the single protonation state at Asp125, because this model yields a calculated free energy difference in the range of those observed experimentally. Protonation on the inner or outer carboxylate oxygen, as defined in the starting crystal structure, is not reported. The authors believe it would be difficult to unequivocally rule out the dianionic state from the FEP theory calculations alone. However, they state that the dianionic state is less likely, since two negative charges would be in close proximity to each other and would preclude the formation of a hydrogen bond network with the inhibitor.

Yamazami et al.[29] determined the p$K_a$ of all Asp and Glu residues in a complex of HIV protease with a non-peptide cyclic urea-based inhibitor by use of chemical shift titrations, H/D isotope shift measurements, and the X-ray crystal structure. The data provides strong evidence that both active-site aspartyl residues are protonated over the pD range 2−7 and form a network of strong hydrogen bonds with the diol hydroxy groups of the inhibitor.

A review of the experimental and theoretical work on aspartic proteases in the native state or in complexes clearly shows that there is not a consensus choice for the protonation states of the active-site aspartates. Neutron diffraction experiments might resolve the uncertainty of the location of the proton(s) on the carbonyl oxygen atoms of the aspartic acid side chains; however, this has not been done. If the correct protonation state is not assigned for these important recognition-site residues, then the computed thermodynamic binding energies may not be reasonable. Thus, as part of the preparatory analysis of using our scheme to compute ligand−receptor binding thermodynamics, the free energy force field (FEFF) 3D-QSAR method,[30] we have also explored the protonation state problem for the Upjohn set of renin inhibitors.

Different mono- and diprotonation states of the active-site aspartic acid residues were examined for the enzyme alone and in complex with a representative of the different types of transition state analogs from the Upjohn data set. This analysis included the state in which one of the two aspartates was negatively charged, and the other was neutral. The scenarios where both aspartates were ionized and both were neutral were also explored.

Representations of the protonation states for the hydroxyethylene and statine isostere type inhibitors are given in Figure 5. The diprotonated states consisting of a protonated inner carboxylate oxygen of one aspartate coupled with a protonated outer carboxylate oxygen of the other aspartate were also examined. Differentiation between the oxygens for the placement of a proton is necessary, because the renin active site is not "symmetrical" in the presence of an inhibitor. This asymmetry is maintained by an identifiable intramolecular hydrogen bonding network involving the aspartates and neighboring residues. Furthermore, the carboxylate oxygens differ in their orientation to both the scissile bond of the natural ligand and to the isostere replacements found in the transition-state analog inhibitors. In crystal structures of complexes of the aspartic proteases the oxygens are distinguishable by their distances to the atoms of the isostere replacement. It is possible that under normal binding

conditions near free rotation about the Asp $\chi_2$ torsion angles, and also vibrational motions, may render these oxygens indistinguishable with respect to the position and "ownership" of the acidic proton(s). However, free rotation about the side-chains bonds of the Asp residues would disrupt the favorable intramolecular hydrogen bond network that contributes to the structural stability of the catalytic center. Therefore, it is likely that the aspartate oxygens are distinct with respect to their orientations to the inhibitor.
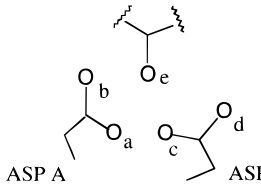
The protonation state of the reduced amide transition-state isostere inhibitors was also investigated because it is uncertain whether the protonation state of the secondary nitrogen is neutral or charged. The pH of the binding assay was 7.4 for the thermodynamic binding parameters reported by the Upjohn workers.[8] The measurement of the p$K_a$ of the secondary amine in question has not been reported. Therefore, the secondary amine nitrogen could be modeled as neutral or ionized. In the neutral state the single proton could be placed in one of two spatial directions since the nitrogen possesses a lone pair of electrons, i.e., in one of two available sp³ orbitals. These two orbitals are differentiable with respect to their orientation to the active-site aspartate oxygens. Suguna et al. proposed the possibility of an intramolecular hydrogen bond between the secondary nitrogen and the $P_2$ carbonyl oxygen of a reduced amide isostere inhibitor that was cocrystallized with the aspartyl protease of *Rhizopus chinesis*.[23] This hydrogen bond would locate the position of at least one of the possible proton(s) belonging to the nitrogen. We sampled both possible placements of the lone proton in conjunction with the various protonation models of the active-site aspartates. The protonation state models considered for the reduced amide isostere analogs are given in Figure 6.

In order to postulate the likely protonation state at the active site for the inhibitor−renin complex, the X-ray structures of similar transition-state analogs bound to various aspartic proteases were examined. Atomic distances and the orientation and geometries of the inhibitor atoms with respect to the aspartate oxygens were computed in these crystal structures in order to assess the most likely position(s) of the proton(s). The interatomic distances between the aspartic acid residue carboxylate oxygens of one residue and those of the other residue were also calculated. Tables 3 and 4 show the results of this analysis for the hydroxyethylene and statine isostere analogs and the reduced amide isostere analogs, respectively. In Tables 3 and 4 the aspartate group referred to as "Asp A" and "Asp B" correspond to Asp38 and Asp226 in renin, respectively. Short interatomic distances between a potential hydrogen bond donor and acceptor pair of 3.0 Å, or less, may indicate a hydrogen bond, and, consequently, the location of a proton.

Starting from the crystal structure of the renin−inhibitor complex, proton(s) were placed at the appropriate heteroatom(s) of the active site to create an initial model of the desired protonation state. Several methods were used to evaluate the robustness of each possible protonation state. In one method, 100 steps of molecular mechanics steepest descent minimization, followed by 300 steps of conjugate gradient minimization, were performed using the program MOLSIM to optimize the location of the aspartate proton(s) and the interacting proton(s) of the transition-state analogs.

In a second approach, starting from the crystal structure of each receptor model and ligand complex, 20 ps of MD
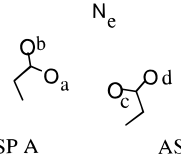
**Table 3.** Interatomic Distances of Aspartic Protease Complexes Cocrystallized with Either a Hydroxyethylene or Statine Isostere Transition State Inhibitor



| distance (Å) | | | | | | enzyme | PDB code[a] | res[b] (Å) |
|---|---|---|---|---|---|---|---|---|
| ae | be | ce | de | ac | bd | | | |
| 2.68 | 3.51 | 2.98 | 2.45 | 2.80 | 5.68 | human renin | 1RNE | 2.4 |
| 2.69 | 3.33 | 3.06 | 2.66 | 3.00 | 5.63 | mouse renin | 1SMR | 1.9 |
| 2.67 | 3.21 | 3.09 | 2.68 | 2.96 | 5.49 | endothiapepsin | 2ER7 | 1.6 |
| 2.70 | 3.20 | 3.08 | 2.67 | 2.88 | 5.54 | endothiapepsin | 5ER2 | 1.8 |
| 2.66 | 3.21 | 2.87 | 2.74 | 2.98 | 5.74 | endothiapepsin | 4ER1 | 2.0 |
| 2.92 | 3.81 | 2.89 | 2.70 | 2.81 | 5.92 | porcine pepsin | 1PSA | 2.9 |
| 2.66 | 3.15 | 2.95 | 2.75 | 3.11 | 5.64 | endothiapepsin | 2ER9 | 2.2 |
| 2.68 | 3.21 | 2.95 | 2.69 | 2.94 | 5.61 | endothiapepsin | 3ER5 | 1.8 |
| 2.75 | 3.14 | 2.84 | 2.65 | 2.90 | 5.52 | endothiapepsin | 2ER0 | 3.0 |
| 2.71 | 3.31 | 2.97 | 2.67 | 2.93 | 5.64 | av | | |

[a] Protein Data Bank entry code. [b] Crystallographic resolution of crystal structure.

**Table 4.** Interatomic Distances of Aspartic Protease Complexes Cocrystallized with a Reduced Amide Isostere Transition State Inhibitor



| distance (Å) | | | | | | enzyme | PDB code[a] | res[b] (Å) |
|---|---|---|---|---|---|---|---|---|
| ae | be | ce | de | ac | bd | | | |
| 4.51 | 4.54 | 4.13 | 2.94 | 2.95 | 5.64 | endothiapepsin | 2ER6 | 2.0 |
| 4.18 | 4.52 | 3.78 | 2.79 | 3.09 | 5.97 | endothiapepsin | 4ER4 | 2.1 |
| 4.53 | 4.44 | 3.78 | 2.82 | 2.96 | 5.91 | rhizopuspepsin | 3APR | 1.8 |
| 4.37 | 4.32 | 3.64 | 2.78 | 3.01 | 5.72 | endothiapepsin | 1ER8 | 2.0 |
| 4.40 | 4.46 | 3.83 | 2.83 | 3.00 | 5.83 | av | | |

[a] Protein Data Bank entry code. [b] Crystallographic resolution of crystal structure.

simulation were performed at 200 K with a time step of 0.5 fs. A nonbonded interaction cutoff of 10 Å was employed. MD simulations were done in order to study the motions of the active-site atoms near the X-ray crystal geometry and accomplish energy-space sampling under the influence of different protonation states. The resultant total energies of the complex and the RMS deviations of the active-site atoms from those of the crystal structure were examined. Key ligand−enzyme and enzyme−enzyme atomic distances including the active-site hydrogen bonding network were also monitored. In addition, torsion angle rotations for the aspartates in the active site as well as the torsion angles of the transition-state isostere were also monitored over the course of the MD simulation. The overall hydrogen bonding patterns at the active sites of the resultant models were assessed for feasibility with respect to the potential hydrogen bond donors and acceptors of the initial crystal structure.

The X-ray geometry serves as an experimental reference from which one can hypothesize a reasonable placement of proton(s). Significant RMS deviations from the X-ray

PEPTIDOMETIC RENIN INHIBITORS

*J. Chem. Inf. Comput. Sci., Vol. 37, No. 4, 1997* **787**

geometry over the course of a MD simulation may indicate a particular protonation state is not reasonable. We consider here as "wrong" any protonation state in which the structure changed significantly (RMS > 1.5 Å) during the MD simulations. It must be remembered, however, that the experimental binding data is not for the crystalline state but rather for the complex in solution at 37 °C. The crystalline state may, in fact, be only one of a number of energetically favorable complex states. Therefore, the X-ray geometry was used as a reference with the limitations mentioned above in mind. The pH of the binding assay, which is different from the pH of crystallization, must also be recognized as influencing the protonation state. Moreover, the local pH environment of the active site, and the binding geometry of the transition-state analog in a complex, will depend on the type of isostere fragment and likely influence the protonation state of the aspartates differently.

## RESULTS

The three reduced-size receptor models (8, 10, and 12 Å radii) behave similarly with respect to RMS fit of all non-hydrogen atoms to the crystal structure over the course of the trial MD simulations. The RMS fit of the lowest energy 8, 10, and 12 Å receptor models found in the MD simulations is 1.27, 1.10, and 1.03, respectively, compared to the equivalent part of the crystal structure. The RMS fit of the whole enzyme structure from the MD simulations with respect to the crystal geometry is 1.39. A comparison of the ligand−receptor interaction energy of each of the receptor models using a nonbonded cutoff value of 12 Å reveals no significant difference in intermolecular energy among the scaled-down receptor models. The interaction energies did not differ by more than 1% among the three models. The intermediate-sized receptor model of 10 Å was, therefore, adopted in all production MD simulations as a compromise between reliability and computational efficiency. Stability of the receptor structure was optimal when the alternate heavy mass constraint was assigned to all main chain atoms of the entire receptor model, regardless of their distance from the active site. This constraint was used for all simulations. In addition, preliminary MD evaluation studies of the complex indicated that trajectories longer than 20 ps did not improve the convergence of the total potential energy and in some cases increased the divergence of the model from the crystal structure. The choice of 10−20 ps MD simulations and simulation temperatures of no greater than 200 K was found to be the best compromise in generating trajectory geometries that remained close to the X-ray structure of renin for reasonable amounts of CPU time yet gave stable low-energy states relative to the parent X-ray structure.

An analysis of the MD simulations performed on each possible protonation state model for the different types of transition-state analogs indicates that some models show more stability than others. A measure of stability of a particular protonation model is determined by evaluating how closely the simulation structure replicates the available experimental geometries (in terms of RMS fit) as well as the model's relative total potential energy compared to the other possible protonation states of the same net charge.

Energy minimization does not resolve the choice of protonation state. For example, the energetically most favorable protonation state model for different ligands

**Table 5.** Energetics of the Complex Structures of U77646E and U73777E Possessing Different Protonation States

a. U77646E

| model[a] | $E_T$[b] | $E_{inter}$[c] | $E_L$[d] | $E_R$[e] | RMS[f] | RMS[g] |
|---|---|---|---|---|---|---|
| 1 | 943.0 | −69.8 | 127.1 | 885.7 | 0.89 | 0.53 |
| 2 | 860.0 | −104.3 | 116.1 | 848.5 | 0.98 | 0.42 |
| 3 | 883.9 | −145.4 | 121.6 | 907.7 | 0.96 | 0.32 |
| 4 | 926.8 | −108.3 | 123.4 | 911.7 | 0.89 | 0.62 |
| 5 | 1031.9 | −93.3 | 128.3 | 996.9 | 0.85 | 0.64 |
| 6 | 935.9 | −132.9 | 129.5 | 939.3 | 0.82 | 0.34 |
| 7 | 862.6 | −142.1 | 111.7 | 893.0 | 1.41 | 0.45 |

b. U73777E

| model[h] | $E_T$[b] | $E_{inter}$[c] | $E_L$[d] | $E_R$[e] | RMS[f] | RMS[g] |
|---|---|---|---|---|---|---|
| 1 | 841.4 | −185.6 | 76.8 | 950.2 | 1.45 | 0.46 |
| 2 | 925.7 | −100.6 | 75.3 | 951.0 | 1.49 | 0.56 |
| 3 | 830.4 | −135.9 | 79.8 | 886.5 | 1.08 | 0.72 |
| 4 | 991.5 | −204.0 | 88.4 | 1107.1 | 1.09 | 0.42 |
| 5 | 828.7 | −149.8 | 78.7 | 899.7 | 1.29 | 0.75 |
| 6 | 846.0 | −135.4 | 71.1 | 910.3 | 1.43 | 0.53 |
| 7 | 969.4 | −95.3 | 84.1 | 980.6 | 0.85 | 0.70 |
| 8 | 901.8 | −131.2 | 80.5 | 952.5 | 0.68 | 0.42 |
| 9 | 872.4 | −199.1 | 82.8 | 988.7 | 0.41 | 0.25 |

[a] Protonation state model as shown in Figure 5. [b] Total energy, including solvation energy, of complex. [c] Ligand−receptor intermolecular interaction energy. [d] Intramolecular energy, including solvation energy, of bound ligand. [e] Intramolecular energy, including solvation energy, of bound receptor. [f] RMS deviation of receptor backbone compared to crystal structure. [g] RMS deviation of two active-site aspartates compared to crystal structure. [h] Protonation state model as shown in Figure 6.

containing the identical type of transition-state isostere replacement was not the same. This also held true for a comparison of the preferred protonation state model for the hydroxyethylene isostere and the statine isostere transition-state inhibitors. Since both the hydroxyethylene and statine transition-state inhibitors have in common a hydroxyl group at the isosteric replacement, they present a similar chemical environment to the enzyme and, hence, an identical protonation state of the active-site aspartates would be anticipated.

The MD simulations provide a better vehicle to compare protonation states than energy minimization. The total energy, including solvation energy, along with a breakdown of the total energy in terms of intermolecular ligand−receptor interaction energy, intramolecular bound ligand energy, and intramolecular bound receptor energy is given in Table 5 for the renin−inhibitor data set. Inhibitors U77646E and U73777E were arbitrarily chosen as representatives of the hydroxyethylene and reduced amide transition-state analogs for the calculations reported in Table 5 (parts a and b, respectively). Data are not presented for the statine analogs, because only certain protonation state models were examined for these inhibitors. It is assumed that since both the hydroxyethylene and statine type of inhibitors contain a hydroxyl group at the transition state isostere that the protonation state at the ligand−receptor active site would be identical for these analogs.

The intermolecular ligand−receptor interaction energy listed in Table 5 is equal to the sum of the van der Waals, electrostatic, and hydrogen bonding energies. The differences of the MD receptor structure from the crystal structure, in terms of RMS deviations of the backbone atoms of the receptor and of the two active-site aspartates, are also given in Table 5. The lowest energy complex state sampled during

**Table 6.** MD Average Interatomic Distances of U77646E and U73777E Receptor Models

a. U77646E[a]

| model[b] | ae | be | ce | de | ac | bd | a[c] | b[d] | c[e] |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2.96 | 3.03 | 4.13 | 3.70 | 3.93 | 5.98 | 3.05 | 2.80 | 2.91 |
| 2 | 3.07 | 3.25 | 3.80 | 3.01 | 2.68 | 5.06 | 3.10 | 2.82 | 3.10 |
| 3 | 3.13 | 3.19 | 3.55 | 2.80 | 3.07 | 5.09 | 2.90 | 2.84 | 2.94 |
| 4 | 2.92 | 3.27 | 3.77 | 2.78 | 2.85 | 4.59 | 3.81 | 2.77 | 3.10 |
| 5 | 3.45 | 2.78 | 3.69 | 2.79 | 3.17 | 3.90 | 3.01 | 3.55 | 3.03 |
| 6 | 2.99 | 2.81 | 3.66 | 2.89 | 2.67 | 5.04 | 3.05 | 2.91 | 2.95 |
| 7 | 2.87 | 3.02 | 3.56 | 2.92 | 2.96 | 4.99 | 3.57 | 2.85 | 3.05 |

b. U73777E[f]

| model[f,g] | ae | be | ce | de | ac | bd | a[c] | b[d] | c[e] |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 4.02 | 4.42 | 3.19 | 2.72 | 3.57 | 6.08 | 2.99 | 2.84 | 3.42 |
| 2 | 4.22 | 4.31 | 4.34 | 2.76 | 2.69 | 4.91 | 3.02 | 2.77 | 3.07 |
| 3 | 4.46 | 4.75 | 4.27 | 3.57 | 3.34 | 4.76 | 3.08 | 2.86 | 3.13 |
| 4 | 4.53 | 4.48 | 3.64 | 2.75 | 2.79 | 4.74 | 2.93 | 2.94 | 3.12 |
| 5 | 4.45 | 4.80 | 3.62 | 3.24 | 4.21 | 5.66 | 3.01 | 2.87 | 3.67 |
| 6 | 4.29 | 4.62 | 4.26 | 2.85 | 2.73 | 4.96 | 3.06 | 2.86 | 3.07 |
| 7 | 4.00 | 4.41 | 3.48 | 2.93 | 3.56 | 6.34 | 2.91 | 2.83 | 3.30 |
| 8 | 5.29 | 4.88 | 5.03 | 3.57 | 2.77 | 4.81 | 3.22 | 2.90 | 3.07 |
| 9 | 5.44 | 4.84 | 5.15 | 3.72 | 3.07 | 4.80 | 2.90 | 2.86 | 2.97 |

[a] See Table 3 for definition of interatomic distances. [b] Protonation state model as shown in Figure 5. [c] Distance between Asp38 inner carboxylate oxygen and Gly40 amide N. [d] Distance between Asp38 outer carboxylate oxygen and Ser41 hydroxyl oxygen. [e] Distance between Asp226 inner carboxylate oxygen and Gly228 amide N. [f] See Table 4 for definition of interatomic distances. [g] Protonation state model as shown in Figure 6.

the MD simulations was used to represent the receptor structure.

A common finding in the X-ray crystal structures of all aspartic proteases is the coplanarity of the aspartic acid carboxylate groups which is maintained by an intramolecular hydrogen bonding network. The torsion angles of the active-site aspartates were used to monitor whether coplanarity between the carboxylates is maintained over the course of MD simulations. Overall "permanent" rotation of a protonated aspartate $\chi_2$ torsion angle of 90−180° from the crystal geometry is taken to indicate instability for a protonation state. For example, when a proton is placed on the inner carboxylate oxygen of Asp38 of a complex with U77646E (model 4 in Figure 5), the intramolecular hydrogen bond between this oxygen and the amide portion of Gly40 is lost as the MD simulation progresses. This hydrogen bond breaking appears to be the result of a rotation of Asp38 $\chi_2$ with a subsequent loss in coplanarity of the two aspartate side chains. In contrast, the protonation state consisting of a neutral outer carboxylate oxygen of Asp226 with the same inhibitor was found to maintain the enzyme's intramolecular hydrogen bonding network in the active site.

Key ligand−receptor interatomic distances were also monitored in the MD simulations. Average values of the atomic distances referred to in Tables 3 and 4 were calculated from the MD simulations of inhibitors U77646E and U73777E and are shown in Table 6(parts a and b, respectively). Also listed in Table 6 are the average distances between the active-site aspartate carboxylate oxygens and their respective intramolecular hydrogen bonding partners identified in the crystal structure. The enzyme intramolecular hydrogen bonding partners and their respective distances, observed in the crystal complex 1RNE, are the Asp38 inner carboxylate oxygen and Gly40 amide N (distance = 2.76

Å), the Asp38 outer carboxylate oxygen and Ser41 hydroxyl oxygen (distance = 2.48 Å), and the Asp226 inner carboxylate oxygen and Gly228 amide N (distance = 2.74 Å).

In some instances large translational movements of the bound ligand prevented any reasonable opportunity for intermolecular contacts with the receptor as typically seen in the crystal complexes of aspartic proteases with similar ligands. This strange alignment behavior occurs for the protonation state model consisting of a monoprotonated aspartate with the proton residing at the outer carboxylate oxygen of Asp38 in complex with a neutral reduced amide type analog (model 9 in Figure 6). It is not unusual in a simulation for a bound inhibitor to "explore" the active site and, therefore, undergo excursions from its starting position. This behavior is reasonable since the initial bound alignments for the inhibitors in this study are approximated by the binding modes of similar ligands in the crystal structures of various aspartic protease complexes. However, irreversible drifting of the ligand away from key binding groups of the receptor is viewed as a fatal instability of the model.

The structural and thermodynamic results accumulated over the course of the test MD simulations, coupled to an analysis of experimental findings, led to the selection of the most plausible protonation states. The energy calculations of the protonation state models for the hydroxyethylene and statine isostere type analogs reveals that a few models are more stable than others. The monoprotonated states consisting of a proton on the inner and outer carboxylate oxygen of Asp226 (models 2 and 3 in Figure 5, respectively) and the state comprised of neutral aspartates (protons on the inner Asp38 and outer Asp226 carboxylate oxygens, model 7 in Figure 5) have more favorable total energies than the other models (see Table 5a). The protonation state consisting of a monoprotonated active-site model where the lone proton is situated at the outer carboxylate oxygen of Asp226 (model 3 in Figure 5) was chosen to be representative of the bound receptor for a number of reasons. The intermolecular interaction energy is the most favorable for this model. The diprotonated model (model 7 in Figure 5) and model 2 (see Figure 5) did not reproduce the close intramolecular receptor contacts as well as the monoprotonated Asp226 outer carboxylate oxygen protonation state model as shown by a comparison of the data in Table 3 and Table 6a, whereas model 3 maintains the intramolecular hydrogen bonding network of the active-site aspartates over the course of the MD simulations. Furthermore, an analysis of the available renin-ligand X-ray crystal structures indicates that each of the active-site aspartate oxygens has an intramolecular hydrogen bonding partner, except for the Asp226 outer carboxylate oxygen. This observation reinforces the decision to model this oxygen as neutral. Coincidentally, this protonation state model has also been proposed by experimentalists[3,21,22] and from results of theoretical calculations.[25]

An examination of various crystal aspartic protease complexes consisting of ligands with the reduced amide isostere reveals that the secondary nitrogen of the inhibitor and outer carboxylate oxygen of Asp226 are very likely hydrogen bonding partners (see Table 4). Models 3, 5, 8, and 9 (Figure 6) produce average distances between these atoms (the range is 3.24−3.72 Å) that are greater than typical hydrogen bonding distances. The protonation state model consisting of a positively-charged inhibitor and negatively-charged active-site aspartates (model 1 in Figure 6) yields
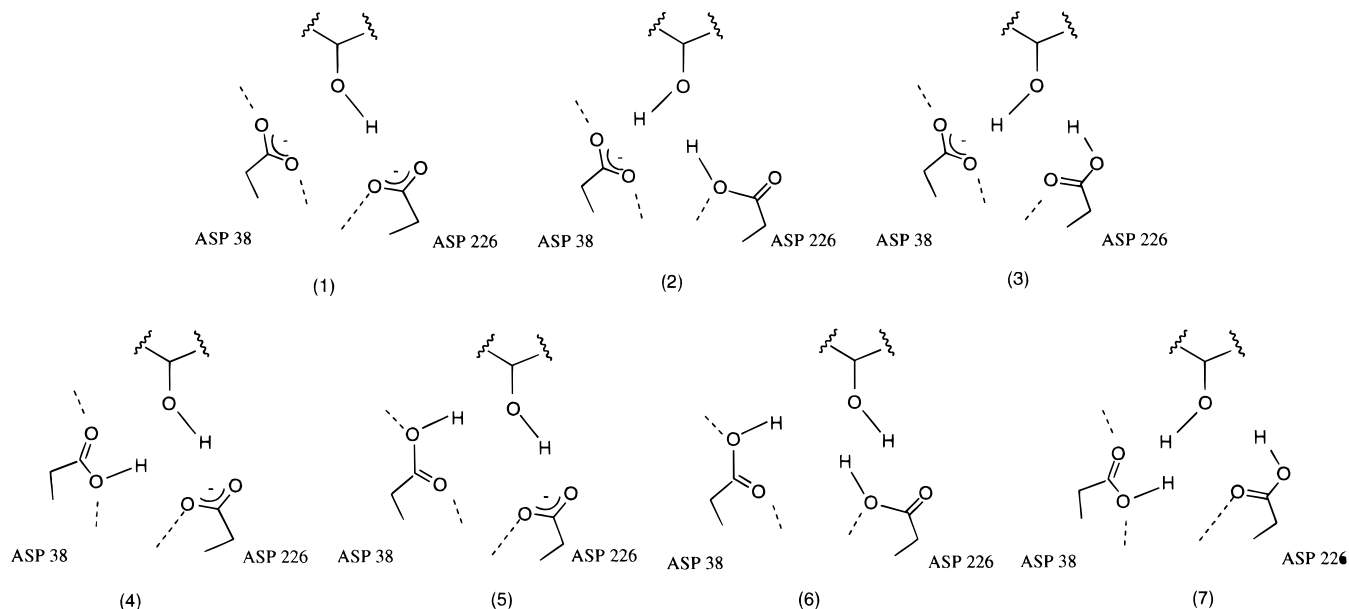
PEPTIDOMETIC RENIN INHIBITORS

*J. Chem. Inf. Comput. Sci., Vol. 37, No. 4, 1997* **789**

**Figure 5.** Representation of the possible protonation states for the active-site aspartic acid residues in the presence of a hydroxyl group of a hydroxyethylene or statine transition state isostere analog. Dashed lines to the aspartate carboxylates represent hydrogen bonding to other enzyme active-site residues.
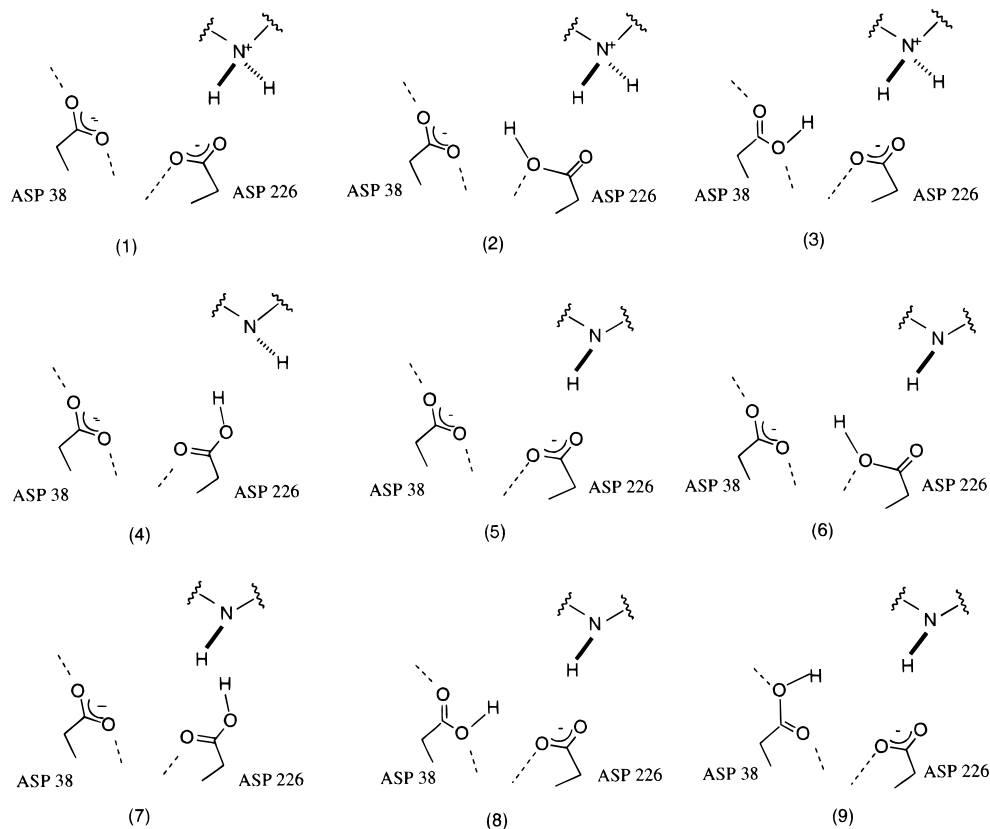
**Figure 6.** Representation of possible protonation states for the active-site aspartic acid residues in the presence of a reduced amide transition state isostere analog. The secondary nitrogen of the reduced amide isostere is shown. Dashed lines to the aspartate carboxylates represent hydrogen bonding to other enzyme residues of the active site.

one of the lowest energy complexes for this class of inhibitors. This model also has one of the most favorable interaction energies. Only models 4 and 9 have more favorable interaction energies. However, the data in Table 6b show that both of these models display significant deviations in one or more of the key intermolecular distances listed in Table 4. In contrast, a feasible hydrogen bonding network at the active site is maintained for model 1 over the course of the MD simulations that nearly replicates the

average crystal interatomic distances between the heteroatoms listed in Table 4. Therefore, this protonation state model was chosen for the reduced amide isostere analogs. This protonation state is in agreement with that proposed by experimentalists[21,22] and from results of a molecular modeling study[28] for this class of inhibitors.

Overall, the renin receptor model is assumed to possess different protonation states for the active-site aspartates depending on the bound ligand. However, the overall total

charge of each of the complexes is identical which is important when making comparisons of calculated intermolecular interactions of different ligand−receptor systems.

## DISCUSSION

Several critical observations regarding the "mechanics" of modeling ligand−receptor systems could be made over the course of this study. Firstly, the force field and the representation of the system (e.g., limited solvation of the system and the use of a scaled-down enzyme model) is a significant approximation with respect to simulation momentum. The geometry of the receptor has a tendency to drift from the X-ray structure under these conditions unless constrained. Solvation energy contributions, calculated using a hydration shell solvation model, were not included in the course of the MD simulations but evaluated for representative low energy structures only. This protocol is adopted because the solvation model may incorrectly balance, over the course of a MD simulation, the optimization of solvation of two interacting hydrophilic groups with the opposing competition of formation of hydrogen bonds between these groups, which may require some desolvation. Optimal behavior of the system occurs when all main chain atoms of the peptide receptor are assigned an alternate heavy mass which results in decreased movement for these particular atoms over the course of a MD simulation. Other choices of atomic constraints, such as constraining only those main chain atoms greater than 8 Å away from any atom of the inhibitor, were tested but proved to be less suitable. It is very possible that the criteria for constraining atom movement will be ligand−receptor system dependent. Thus, the constraints, namely free-space MD simulations and alternate heavy atomic mass assignment, were designed to minimize the "drift" from the experimental structure, while allowing the efficient sampling of the corresponding phase space. The use of fictitious masses is virtually the same as using Cartesian constraints, particularly when the masses are chosen to be very large. Fictitious masses provides a convenient way to "tune in" MD simulation motions in the pruned protein model that are similar to those of the complete parent protein by selectively varying the mass values. However, no constraint set is perfect, and, consequently, our results are only relevant if the system retains a geometry "close" to that observed in the experimentally determined crystal state during the MD simulation.

The results of this renin−inhibitor study also suggests that the local chemical environment of the active site should be taken into consideration in ligand−receptor modeling. The chemical nature of the ligand may influence the ionization state of the active site. Therefore, the receptor and ligand may exist in different protonation states in the bound and unbound forms. In this study the protonation state of the unbound receptor was also modeled. The two active-site aspartates in the complex were left unchanged when modeling the unbound receptor. Different protonation states of the unbound receptor were examined using the crystal structure from the complex as the initial geometry. Several protonation models appear more stable than others in trial MD simulations including the completely ionized state as well as the monoprotonated state where the lone proton resides at either of the two inner carboxylates oxygens. However, no one particular protonation state stood out as

the most favorable. It is likely that a more appropriate approach to modeling the unbound state of renin is to include the explicit water molecule proposed to be bound between the two active-site aspartates of all unbound aspartic proteases. Placing proton(s) on the active-site aspartates in the absence of this bound water molecule may not accurately reproduce the environment of the native enzyme. Ultimately, the known X-ray geometry of the unbound enzyme may be the more appropriate starting structure to use in modeling this state of the receptor.

As has been proposed from experimental findings,[24] the position of protons at the active site may not be "permanent" but rather "dynamical". This dynamic behavior may explain why some protonation states produce similar energetics as they represent one of the possible dynamic states which are in equilibrium with one other.

In the next paper[30] the renin receptor models constructed in this study are used in developing free energy force field (FEFF) 3D-QSAR models to predict ligand−receptor binding thermodynamics.

## REFERENCES AND NOTES

(1) Sialecki, A. R.; Hayakawa, K.; Fujinaga, M.; Murphy, M.; Fraser, M.; Muir, A. Carilli, C.; Lewicki, J. A.; Baxter, J. D.; James, M. N. G. Structure of Recombinant Human Renin, a Target for Cardiovascular-Active Drugs, at 2.5 Å Resolution. *Science* **1989**, *243*, 1346−1351.
(2) Rahuel, J.; Priestle, J. P.; Grütter, M. G. The Crystal Structures of Recombinant Glycosylated Human Renin Alone and in Complex with a Transition State Analog Inhibitor. *J. Struct. Biol.* **1991**, *107*, 227−236.
(3) Dhanaraj, V.; Dealwis, C. G.; Frazao, C.; Badasso, M.; Sibanda, B. L.; Tickle, I. J.; Cooper, J. B.; Driessen, H. P. C.; Newman, M.; Aguilar, C.; Wood, S. P.; Blundell, T. L.; Hobart, P. M.; Geoghegan, K. F.; Ammirati, M. J.; Danley, D. E.; O'Connor, B. A.; Hoover, D. J. X-ray Analyses of Peptide-Inhibitor Complexes Define the Structural Basis of Specificity for Human and Mouse Renins. *Nature* **1992**, *357*, 466−472.
(4) Epps, D. E.; Cheney, J.; Schostarez, H.; Sawyer, T. K.; Prairie, M.; Krueger, W. C.; Mandel, F. Thermodynamics of the Interaction of Inhibitors with the Binding Site of Recombinant Human Renin. *J. Med. Chem.* **1990**, *33*, 2080−2086.
(5) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protetin Databank, a Computer Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535−542.
(6) Pearlstein R. A. CHEMLAB-II Users Guide, Version 11.1; Molecular Simulations Inc.: 16 New England Executive Park, Burlington, MA 01803, 1991.
(7) QUANTA Version 3.3; Molecular Simulations Inc.: 16 New England Executive Park, Burlington, MA 01803, 1993.
(8) Epps, D. E.; Schostarez, H.; Argoudelis, C. V.; Poorman, R. A.; Hinzmann, J.; Sawyer, T. K.; Mandel, F. An Experimental Method for the Determination of Enzyme-Competitive Inhibitor Dissociation Constants form Displacement Curves: Application to Human Renin Using Flourescence Energy Transfer to a Synthetic Dansylated Inhibitor Peptide. *Anal. Biochem.* **1989**, *181*, 172−181.
(9) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230−252.

PEPTIDOMETIC RENIN INHIBITORS

*J. Chem. Inf. Comput. Sci., Vol. 37, No. 4, 1997* **791**

(10) Doherty, D. C. MOLSIM User Guide; The Chem21 Group; 1780 Wilson Dr., Lake Forest, IL 60045, 1994.

(11) Epps, Personal communication, 1994.

(12) Stewart J. J. P.; Seiler, F. K. QCPE #455 (Ver 4.0), 1987.

(13) Blundell, T. L.; Cooper, J.; Foundling, S. I.; Jones, D. M.; Atrash, B.; Szelke, M. On the Rational Design of Renin Inhibitors: X-ray Studies of Aspartic Proteinases Complexed with Transition-State Analogues. *Biochemistry* **1987**, *26*, 5585−5590.

(14) Swain. A. L.; Gustchina, A.; Wlodawer, A. A Comparison of Three Inhibitors of Human Immunodeficiency Virus Protease. *Adv. Exp. Med. Biol*. **1992**, *306*, 433−441.

(15) Allinger, N. L. Conformational Analysis. 130. MM2. A Hydrocarbon Force Field Utilizing $V_1$ and $V_2$ Torsional Terms. *J. Am. Chem. Soc*. **1977**, *99*, 8127−8134.

(16) Hopfinger, A. J. In *Conformational Properties of Macromolecules*; Academic Press: New York, 1973; p 38.

(17) Hopfinger, A. J. In *Conformational Properties of Macromolecules;* Academic Press: New York, 1973; p 71.

(18) Ferguson, D. M.; Radmer, R. J.; Kollman, P. A. A Determination of the Relative Binding Free Energies of Peptide Inhibitors to the HIV-1 Protease. *J. Med. Chem*. **1991**, *34*, 2654−2659.

(19) Miyamoto, S.; Kollman, P. Absolute and Relative Binding Free Energy Calculations of the Interaction of Biotin and its Analogs with Streptavidin Using Molecular Dynamics/Free Energy Pertubation Approaches. *Proteins: Structure, Function, and Genetics* **1993**, *16*, 226−245.

(20) Veerapandian, B.; Cooper, J. B.; Sali, A.; Blundell, L. T.; Rosati, R. L.; Dominy, B. W.; Damon, D. B.; Hoover, D. J. Direct Observation by X-ray Analysis of the Tetrahedral "Intermediate" of Aspartic Proteinases. *Protein Science* **1992**, *1*, 322−328.

(21) Hyland, L. J.; Tomaszek, T. A. Jr.; Roberts, G. D.; Carr, S. A.; Magaard, V. W.; Bryan, H. L.; Fakhoury, S. A.; Moore, M. L.; Minnich, M. D.; Culp, J. S.; DesJarlais, R. L.; Meek, T. D. Human Immunodeficiency Virus-1 Protease. 1. Initial Velocity Studies and Kinetic Characteriztion of Reaction Intermediates by $^{18}$O Isotope Exchange. *Biochemistry* **1991**, *30*, 8441−8453.

(22) Hyland, L. J.; Tomaszek, T. A, Meek, T. D. Human Immunodeficiency Virus-1 Protease. 2. Use of pH Rate Studies and Solvent Kinetic Isotope Effects to Elucidate Details of Chemical Mechanism. *Biochemistry* **1991**, *30*, 8454−8463.

(23) Suguna, K.; Padlan, E. A.; Smith, C. W.; Carlson, W. D. Binding of a Reduced Peptide Inhibitor to the Aspartic Proteinase from *Rhizopus Chinensis*: Implications for a Mechanism of Action. *Proc. Natl. Acad. Sci. U.S.A*. **1987**, *84*, 7009−7013.

(24) Jaskolski, M.; Tomaselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R.L.; Schneider, J.; Kent, S. B. H.; Wlodawer A. Structure at 2.5 Å Resolution of Chemically Synthesized Human Immunodeficiency Virus Type 1 Protease Complexed with a Hydroxyethylene-Based Inhibitor. *Biochemistry* **1991**, 30, 1600−1609.

(25) Goldblum, A.; Rayan, A.; Fliess, A.; Glick, M. Extending Crystallographic Information with Semiempirical Quantum Mechanics and Molecular Mechanics: A Case of Aspartic Proteinases. *J. Chem. Inf. Comput. Sci*. **1993**, *33*, 270−274.

(26) Chen, X.; Tropsha, A. Relative Binding Free Energies of Peptide Inhibitors of HIV-1 Protease: The Influence of the Active Site Protonation State. *J. Am. Chem. Soc*. **1995**, *38*, 42−48.

(27) Berendsen, H. J. C.; Postman, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys*. **1984**, *81*, 3684−3690.

(28) Harte, W. E.; Beveridge, D. L. Molecular Dynamics of HIV-1 Protease. *J. Am. Chem. Soc*. **1993**, *115*, 3883−3886.

(29) Yamazaki, T.; Nicholson, L. K.; Torchia, D. A.; Wingfield, P.; Stahl, S. J.; Kaufman, J. D.; Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Ru, Y.; Jadhav, P. K.; Chang, C.; Weber, P. C. NMR and X-ray Evidence that the HIV Protease Catalytic Aspartyl Groups are Protonated in the Complex Formed by the Protease and a Non-Peptide Cyclic Urea-BAsed Inhibitor. *J. Am. Chem. Soc*. **1994**, *116*, 10791−10792.

(30) Tokarski, J. T.; Hopfinger, A. J. Prediction of Ligand-Receptor Binding Thermodynamics by Free Energy Force Field (FEFF) 3D-QSAR Analysis: An Application to a Set of Renin Inhibitors. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 792−811.