HIV-1 Protease Cleavage Mechanism Elucidated with Molecular Dynamics Simulation

David C. Chatfield* and Bernard R. Brooks

Contribution from the Laboratroy of Structural Biology, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20892

Received November 2, 1994[⊗]

Abstract: The cleavage mechanism of HIV-1 protease is investigated with molecular dynamics simulation of substrate-, inhibitor-, and *gem*-diol intermediate-bound protease. Initial structures are based on X-ray crystallographic coordinates for the protease bound to the inhibitor JG-365.[1,2] The conformation space explored by atoms near the active site on the 100 ps time scale at 300 K is analyzed for structures likely to initiate reaction. Conformations suitable for reaction initiation are generated for both general acid–general base and direct nucleophilic attack mechanisms. The simulations suggest that (1) both types of mechanism are plausible; (2) the catalytic Asp of monomer B is protonated when reaction begins; (3) if the mechanism is general acid-general base, the catalytic Asp of monomer A is protonated when the second reaction step is initiated; (4) the carbonyl oxygen is more likely than the scissile nitrogen to be protonated in the early stages of reaction; (5) water 301[1] stabilizes productive conformations of reactants and intermediates, but it does not participate directly in reaction; and (6) a lytic water, if present, has very little mobility.

Introduction

In the course of reproduction of the human immune deficiency virus type 1 (HIV-1), the viral protease (HIVPR) cleaves a polyprotein encoded by the *gag* and *pol* genes at eight different sites.[3,4] This function is essential to the maturation of the virus. Consequently, the protease is a promising target for AIDS therapies and the subject of intense experimental and theoretical scrutiny.[5–7] This study reports molecular dynamics (MD) simulations of a number of ligands (a model substrate, Ac-Ser-Gln-Asn-Tyr-Pro-Ile-Val-OMe; a model *gem*-diol intermediate; and the inhibitor JG-365[1,2]) bound at the active site of HIVPR to elucidate the catalytic mechanism.

HIVPR belongs to the class of aspartic proteases, which are characterized by a highly conserved pair of catalytic triads Asp-Thr(Ser)-Gly and facilitate the hydrolysis of peptide bonds.[8,9] The catalytic triads are located at the periphery of a substrate-binding pocket and opposite a flap-like region believed to open and close to promote entry and binding of a substrate. HIVPR is a homodimer, and each monomer contributes one triad to the active site. The viral protease is related to human aspartic proteases such as pepsin, renin, and cathepsin D[5,6,9] which, though monomeric and somewhat larger, are also characterized by two domains and a flap region.

Aspartic proteases have been known and studied for more than a century.[9–12] Numerous catalytic mechanisms have been proposed, and they fall into two general categories: general acid–general base catalysis and catalysis via a covalently-bound intermediate. Proteolysis was long thought most likely to proceed via direct nucleophilic attack by an aspartate,[12] resulting in a covalent intermediate subsequently hydrolyzed. More recently, evidence has weighed strongly in favor of general acid–general base catalysis with water acting as the nucleophile, leading to an intermediate that is not covalently bonded to the protease.[13–16] Many related mechanisms of this type have been proposed for aspartic proteases in general[10,15,17–21] and for HIVPR in particular.[1,22,23] Recently, Meek and co-workers[13,14,24] have suggested a general acid–general base mechanism for HIVPR in which the scissile nitrogen is protonated at an intermediate step, on the basis of a variety of experimental studies.

Three representative mechanisms are shown in Figure 1: general acid–general base catalysis with a neutral (a) or zwitterion (b) intermediate and direct nucleophilic attack with a covalent intermediate (c). In mechanism (a), the first step consists of nucleophilic attack at the scissile carbon by a water

⊗ Abstract published in *Advance ACS Abstracts,* April 15, 1995.

(1) Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805–8809.
(2) Rich, D. H.; Green, J.; Toth, M. V.; Marshall, G. R.; Kent, S. B. H. *J. Med. Chem.* **1990**, *33*, 1285–1288.
(3) Darke, P. L.; Nutt, R. F.; Brady, S. F.; Garsky, V. M.; Ciccarone, T. M.; Leu, C.-T.; Lumma, P. K.; Freidinger, R. M.; Veber, D. F.; Sigal, I. S. *Biochem. and Biophys. Res. Comm.* **1988**, *156* (1), 297–303.
(4) Veronese, F. D.; DeVico, A. L.; Copeland, T. D.; Oroszlan, S.; Gallo, R. C.; Sarngadharan, M. G. *Science* **1985**, *229*, 1402–1405.
(5) Wlodawer, A.; Erickson, J. W. *Annu. Rev. Biochem.* **1993**, *62*, 543–585.
(6) Debouck, C. *Res. Hum. Retroviruses.* **1992**, *8*, 153–164.
(7) Huff, J. R. *J. Med. Chem.* **1991**, *34*, 2305–2314.
(8) Fitzgerald, P. M. D.; Springer, J. P. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 299–320.
(9) Davies, D. R. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 189–215.
(10) Polgár, L. *Mechanisms of Protease Action*; CRC Press: Boca Raton, FL, 1989.
(11) Fruton, J. S. In *Hydrolytic Enzymes*; Neuberger, A.; Brocklehurst, K., Eds.; Elsevier: New York, 1987; pp 1–37.
(12) Fruton, J. S. In *Advances in Enzymology and Related Areas of Molecular Biology*; Meister, A., Ed.; John Wiley and Sons, Inc.: New York, 1976; Vol. 44, pp 1–36.
(13) 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. *Biochemistry* **1991**, *30*, 8441–8453.
(14) Hyland, L. J.; Tomaszek, Jr., T. A.; Meek, T. D. *Biochemistry* **1991**, *30*, 8454–8463.
(15) Antonov, V. K.; Ginodman, L. M.; Rumsh, L. D.; Barshevskaya, Y. V. K. T. N.; Yavashev, L. P.; Gurova, A. G.; Volkova, L. I. *FEBS Lett.* **1981**, *117*, 195–200.
(16) Antonov, V. K.; Ginodman, L. M.; Kapitannikov, Y. V.; Barshevskaya, T. N.; Gurova, A. G.; Rumsh, L. D. *FEBS Lett.* **1978**, *88* (1), 87–90.
(17) Suguna, K.; Padlan, E. A.; Smith, C. W.; Carlson, W. D.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7009–7013.
(18) Polgár, L. *FEBS Lett.* **1987**, *219* (1), 1–4.
(19) Bott, R.; Subramanian, E.; Davies, D. R. *Biochemistry* **1982**, *21*, 6956–6962.
(20) Rich, D. H. *J. Med. Chem.* **1985**, *28*, 263–273.
(21) James, M. N. G.; Saliecki, A. R. *Biochemistry* **1985**, *24*, 3701–3713.

0002-7863/95/1517-5561$09.00/0

molecule, with the unprotonated Asp acting as a general base to abstract a proton from the water while the protonated Asp, acting as a general acid, donates a proton to the carbonyl oxygen (increasing the partial positive charge on the carbonyl carbon, making it more attractive to the nucleophile). In the second step, the *gem*-diol intermediate is cleaved. The catalytic Asps reverse the roles of general acid and general base, the former protonating the scissile nitrogen and the latter abstracting a proton from the diol. Catalysis might also proceed via an oxyanion intermediate, the carbonyl oxygen hydrogen bonding with the protonated Asp but never abstracting a proton. Mechanism (b) also begins with nucleophilic attack by a water molecule, but the scissile nitrogen rather than the carbonyl oxygen is protonated by the general acid. The intermediate in this case is a zwitterion, which undergoes internal rearrangement to form products. In mechanism (c), the nucleophile is an aspartate rather than a water molecule. Nucleophilic attack at the carbonyl carbon produces an anionic intermediate that abstracts a proton from the protonated Asp, cleaving the peptide bond. The resulting acyl–enzyme intermediate is then hydrolyzed, with the unprotonated Asp acting as general base.

This article reports MD simulations that explore the conformations of the active site accessible just prior to the first step of reaction or, in the case of mechanism (a), just prior to the second step. The conformations sampled are analyzed to determine which ones would be likely to initiate reaction. These we call "productive conformations". A large number of simulations were performed in order to explore the many different initial conditions compatible with the mechanisms outlined above.

The simulations incorporate either a hydrolyzable model substrate, a model *gem*-diol intermediate, or the peptido-mimetic inhibitor JG-365 at the active site. The model substrate and *gem*-diol intermediate are closely related to JG-365, for which a high-resolution crystal structure is available.[1] The HIVPR/JG-365 system was studied to verify that MD simulation maintains the active-site interactions reasonably well. The HIVPR/substrate and HIVPR/*gem*-diol intermediate systems were probed under a variety of conditions to determine which conditions favor a productive conformation and which reaction mechanisms are consistent with the productive conformations observed. The criteria for a productive conformation are that interatomic distances be appropriately small for the hydrogen bonding, proton transfer, and nucleophilic attack processes of one of the proposed mechanisms. These criteria are defined more precisely in the Methods section. The variables in the simulation conditions are which of the catalytic Asps is protonated, whether a lytic water is included (simulations of general acid–general base mechanisms incorporate a lytic water, while those probing direct nucleophilic attack by an aspartate do not), and the type of atomic charge parameters used for the catalytic aspartate (see below). Simulations in a like spirit have been reported, for example, for *Streptomyces* R61 DD-peptidase,[25] acyl-chymotrypsins,[26,27] and carboxypeptidase A.[28]

Consideration of the protonation state of the catalytic Asps was critical to these studies. Aspartic proteases in their active forms have one of the two catalytic Asps protonated.[10,14,17,19] All of the mechanisms shown in Figure 1 require that one Asp be protonated and one be unprotonated in the first reaction step. Because of their proximity and ability to share a hydrogen, it is likely that proton transfer between the two Asp side chains in facile.[22] However, productive conformations may be favored by protonation of a particular Asp, and the protonated Asp has not been identified by experiment. For these reasons, we considered the four monoprotonated states in the simulations of HIVPR/substrate and HIVPR/*gem*-diol intermediate complexes. We distinguished between the side-chain oxygens of each Asp because the C_β–C_γ bond did not generally rotate by 180° to exchange the side-chain oxygens during the simulations and because these classical MD simulations require that the hydrogen remain bonded to a particular oxygen even when in reality it may be transferred between the side-chain oxygens.

There is even greater uncertainty about the catalytic Asp protonation states in HIVPR/inhibitor complexes. Simulations of HIVPR complexed with U-85548e and MVT-101 have suggested that the former is diprotonated and the latter is unprotonated,[29] although this has not been confirmed by experiment. In our simulations of HIVPR/JG-365, we considered all possible monoprotonated states, the unprotonated state, and one diprotonated state.

On the basis of the simulations, the following questions are addressed: Which mechanisms are consistent with simulation? Which of the catalytic Asps is protonated when reaction begins? Is the substrate's carbonyl oxygen or scissile nitrogen, if either, more likely to be protonated by the acidic Asp in the early stages of reaction? What is the role of water 301 (a crystallographic water located near the active site in many HIVPR/inhibitor complexes[1]) in catalysis? Is a *gem*-diol a likely intermediate, and, if so, which protonation state of the catalytic Asps is most likely to initiate cleavage of the intermediate?

Theoretical techniques have been used previously to study several aspects of HIVPR structure and function. The structure and large-scale motions of both the monomeric[30] and dimeric[31–34] forms of HIVPR in solution have been studied with MD simulation. Inhibitor binding to HIVPR has been studied with free energy perturbation techniques.[35–37] Modeling of substrate[22,38] and intermediate[13,39] complexes, based on protease/inhibitor crystal structures, has been used to support proposals concerning the reaction mechanism, but these studies have been primarily restricted to energy-minimization techniques and to a static view of the system. The protonation state of HIVPR complexed with various inhibitors has been studied with MD[29] and minimization[36] techniques. (Previous studies of the pro-

(22) Jaskólski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. *Biochemistry* **1991**, *30*, 1600–1609.
(23) Ido, E.; Han, H. P.; Kezdy, F. J.; Tang, J. *J. Biol. Chem.* **1991**, *266*, 24359–24366.
(24) Rodriguez, E. J.; Angeles, T. S.; Meek, T. D. *Biochemistry* **1993**, *32*, 12380–12385.
(25) Boyd, D. B.; Snoddy, J. D.; Lin, H.-S. *J. Comput. Chem.* **1991**, *12*, 635–644.
(26) Bemis, G. W.; Carlson-Golab, G.; Katzenellenbogen, J. A. *J. Am. Chem. Soc.* **1992**, *114*, 570.
(27) Nakagawa, S.; Yu, H.-A.; Karplus, M.; Umeyama, H. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 172.
(28) Banci, L.; Bertini, I.; Penna, G. L. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 186–197.
(29) Harte, W. E., Jr.; Beveridge, D. L. *J. Am. Chem. Soc.* **1993**, *115*, 3883–3886.
(30) Venable, R. M.; Brooks, B. R.; Carson, F. W. *Proteins: Struct., Funct., Genet.* **1993**, *15*, 374–384.
(31) York, D. M.; Darden, T. A.; Pedersen, L. G.; Anderson, M. W. *Biochemistry* **1993**, *32*, 1443–1453.
(32) Harte, W. E., Jr.; Swaminathan, S.; Mansuri, M. M.; Martin, J. C.; Rosenberg, I. E.; Beveridge, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8864–8868.
(33) Swaminathan, S.; Harte, W. E., Jr.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 2717–2721.
(34) Harte, W. E.; Swaminathan, S.; Beveridge, D. L. *Proteins: Struct., Funct., Genet.* **1992**, *13*, 175–194.
(35) Ferguson, D. M.; Radmer, R. J.; Kollman, P. J. *J. Med. Chem.* **1991**, *34*, 2654–2659.
(36) Tropsha, A.; Hermans, J. *Protein Eng.* **1992**, *5*, 29–33.
(37) Reddy, M. R.; Varney, M. D.; Kalish, V.; Viswanadhan, V. N.; Appelt, K. *J. Med. Chem.* **1994**, *37*, 1145.
(38) Harrison, R. W.; Weber, I. T. *Protein Eng.* **1994**, *7*, 1353–1363.
(39) Goldblum, A.; Glick, M.; Rayan, A. *Theor. Chim. Acta* **1993**, *85*, 231–247.

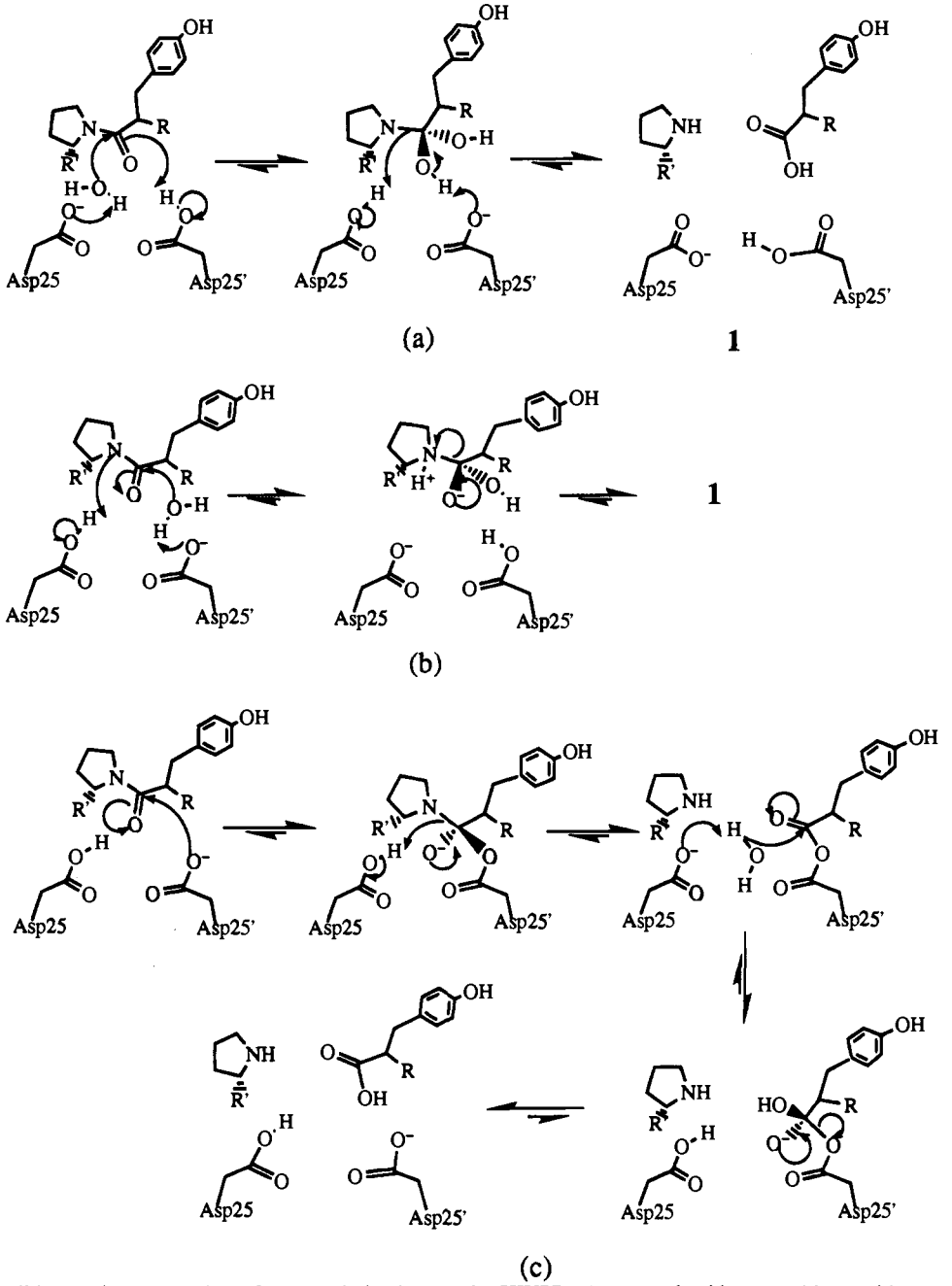

Figure 1. Three possible reaction mechanisms for proteolytic cleavage by HIVPR: (a) general acid–general base with neutral intermediate, (b) general acid–general base with zwitterion intermediate, and (c) direct nucleophilic attack with covalent intermediate. The Asp labels 25 and 25′ are arbitrary.

tonation state of JG-365 complexed with HIVPR were restricted to energy minimization and only considered monoprotonation of the catalytic Asps.[36]) The work reported here is the first to study the reaction mechanism in a systematic way on the basis of multiple MD simulations.

In the discussion to follow, atom types are consistent with Brookhaven notation.[40,41] Protease atoms are designated by the amino acid abbreviation, residue number, and atom type. The residues of protease monomers A and B are numbered 1–99 and 101–199, respectively. Ligand residues are designated with an L preceding the residue number and waters with a W. The labeling of selected active-site atoms is shown in Figure 2. Atoms L 4 C, L 5 N, and L 4 O are also referred to as the scissile carbon, scissile nitrogen, and carbonyl oxygen, respectively. The crystallographic water labeled water 301 in the JG-365 crystal structure is W 1 in our labeling scheme and will be referred to either way in this article; the lytic water, if present, is W 2. In the discussion of ligand backbone hydrogen bonding, ligand residues are labeled P_n (N-terminal side) and P'_n (C-terminal side) according to their sequential distance from the scissile bond to facilitate comparison with other studies.

In the HIVPR/JG-365 crystal structure, the carboxylate groups of the catalytic Asps are coplanar, and the outer and inner side-chain oxygens are labeled OD1 and OD2, respectively. The same labeling scheme is used here whenever the Asps are approximately coplanar. When they are not, OD1 and OD2 indicate which atoms were outer and inner in the initial (*i.e.*, crystal) structure.

(40) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 525–542.

(41) Abola, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. In *Crystallographic Databases-Information Content, Software Systems, Scientific Applications*; Allen, F. H., Bergerhoff, G., Sievers, R., Eds.; Data Commision of the Int'l Union of Crystallography: Bonn/Cambridge/Chester, 1987; pp 107–132.

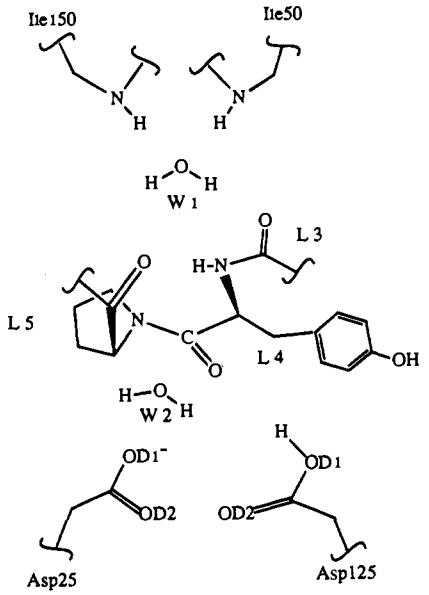

Figure 2. Labeling scheme for atoms near active site.

Methods

Definition of Ligands. The seven-residue model substrate chosen for study, Ac-Ser-Gln-Asn-Tyr-Pro-Ile-Val-OMe, incorporates the Tyr-Pro cleavage site corresponding to the p17-p24 junction of the *gag* fusion protein of HIV-1. This is one of three similar sites cleaved by the protease, the other two substituting Phe for Tyr.[3] Seven residues is the minimum length for a readily hydrolyzable substrate of HIVPR, and it has been shown that the model substrate above undergoes facile cleavage in the presence of HIVPR.[3]

The model substrate differs from JG-365 in only two residues. Initial substrate coordinates were generated from the JG-365 crystal structure by replacing Leu with Gln and the phenylalanine analog Phs with Tyr (Figure 3). The Leu side-chain δ-carbons were deleted, and the remainder of the Gln side chain was built from standard PARM20[42] parameters. To form Tyr, the extra CH_2 in the Phs backbone was deleted, and the phenolic hydroxy group was built from standard parameters. The Phs hydroxy group was converted to the lytic water, if present, or to the Tyr carbonyl oxygen otherwise. In the former case, the carbonyl oxygen initially replaced the deleted carbon. Unrealistic conformations near the scissile carbon were corrected quickly upon simulation. The lytic water, when present, generally moved quickly to a location between the catalytic Asps and the scissile carbon. A water molecule in such a position has been observed in crystal structures of unliganded HIVPR[43] and of other aspartic proteases.[44] In simulations of JG-365, the proline nitrogen was protonated because it is not part of a peptide bond and is in the environment of a negatively charged Asp.

It was necessary to define parameters for three nonstandard residues. Atomic charges for the Phs residue of JG-365 were identical to those for Phe except for the backbone $CHOHCH_2$ unit. The $CHOHCH_2$ charges were chosen by analogy to similar structures in standard residues.[45] Charges and bond and angle parameters for the *gem*-diol functionality of the intermediate were chosen by comparison with those for hydroxy-group-containing amino acids (Ser and Thr) in the PARM20 parameter set.[46] The scissile nitrogen was pyramidal but not protonated. For the protonated catalytic Asp, charges developed by Bellido and Rullmann[47] (obtained by a dipole-conserving population analysis of Hartree–Fock SCF LCAO-MO wave functions) were modified to be consistent with PARM20 charges. The modifications were made so as to reproduce the differences between PARM20 charges and charges obtained by Bellido and Rullmann for an unprotonated Asp. PARM20 charges for the unprotonated Asp and the charges used for the protonated Asp are given in Table 1.

Table 1. Catalytic Aspartic Acid Side-Chain Atomic Charges (au)

atom	Asp[a]	AspH[b]
CB	−0.36	−0.40
HB1	0.10	0.165
HB2	0.10	0.165
CG	0.36	0.54
OD1	−0.60	−0.41
OD2	−0.60	−0.51
HD2		0.45

[a] Unprotonated Asp charges from PARM20.[42] [b] Protonated Asp charges of Bellido and Rullmann,[47] modified to be consistent with other PARM20 charges as described in text. The modified charges for the unprotonated Asp were identical to these, except for the omission of HD2.

Simulations. The MD simulations were carried out with CHARMM,[48] using the all-atom parameter set PARM20.[42] Waters were represented with a modified[49] TIP3P model.[50] A constant dielectric ($\epsilon = 1$) was used, and electrostatic forces were treated with the force switch method[51] and a switching range of 8–12 Å. Van der Waals forces were calculated with the shift[51] method and a cutoff of 12 Å. Nonbond lists were kept to 14 Å and updated heuristically. Hydrogens were placed using the HBUILD[52] routine of CHARMM, and the dynamics was propagated with a Verlet algorithm using a time step of 1 fs.

The simulation protocols are summarized in Table 2. The protease/ligand complex was first hydrated with 500 water molecules (95 crystallographic plus 405 solvent) following a protocol published previously.[53] On the 100 ps time scale, such a minimal solvation layer hydrates surface charged groups and has been shown for myoglobin to maintain a conformation and fluctuations comparable to those obtained in simulations with 3830 waters.[49] Initial coordinates for the solvent waters were assigned by placing the protease system in an equilibrated box of water molecules[54] and deleting all waters except those with oxygens (1) at least 2.5 Å from the nearest protease, ligand, or

(42) Parameter file for CHARMM Version 20, Polygen Corp. (now Molecular Simulation Inc.), Waltham, MA, released Oct 20, 1987.

(43) Lapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Merson, J. R.; Whittle, P. J.; Danley, D. E.; Geoghegan, K. F.; Hawrylik, S. J.; Lee, S. E.; Scheld, K. G.; Hobart, P. M. *Nature (London)* **1989**, *342*, 299–302.

(44) Miller, M.; Jaskólski, M.; Ral, J. K. M.; Leis, J.; Wlodawer, A. *Nature (London)* **1989**, *337*, 576–579.

(45) Atomic charges for the $CHOHCH_2$ functionality of the Phs backbone were 0.1, 0.1, −0.5, 0.3, −0.2, 0.1, and 0.1 au, respectively.

(46) The *gem*-diol intermediate required the addition of three atomic charge, two angle, and two improper dihedral parameters: the hydroxy oxygen and hydrogen and the adjoining carbon had charges of −0.65, 0.4, and 0.5 au; angles OT–CT–OT and NT–CT–OT had force constants of 160.0 kcal/mol-rad^2 and θ-eq of 109.47°; improper dihedrals NT–CT–CT–CT and CT–X–X–OT had force constants of 32.0 and 110.0 kcal/mol-rad^2, respectively, and ω-eq of 35.26°.

(47) Bellido, M. N.; Rullmann, J. A. C. *J. Comput. Chem.* **1989**, *10*, 479–487.

(48) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1982**, *4* (2), 187–217.

(49) Steinbach, P. J.; Brooks, B. R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9135–9139.

(50) Jorgensen, W. L.; Chandrasekhar, J.; Medura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(51) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667–683.

(52) Brünger, A.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 148–156.

(53) Loncharich, R. J.; Brooks, B. R. *J. Mol. Biol.* **1990**, *215*, 439–455.

(54) Reiher, W. E. Ph.D. Thesis, Harvard University, Cambridge, MA, 1985.

Table 2. Simulation Protocols

no.	operation	mobile atoms	temp (K)	heat/cool rate (K/ps)	time or no. steps
1	initialize[a]				
2	langevin MD	A[b]	300		10 ps
3	minimize	all			20 steps[c]
4	heat	B[d]	150 → 300	40	37.5 ps
5	heat	B	300 → 600	40	75 ps
6	cool	B	600 → 300	2	150 ps
7	MD[e]	B	300		100 ps

[a] Mutate JG-365 for substrate simulations, protonate appropriate catalytic Asp, initially place solvent molecules. [b] A: solvent waters only (X-ray waters fixed). [c] Steepest descents method. [d] B: ligand, groups within 4 Å of ligand, and groups within 12 Å of scissile carbon. [e] For *gem*-diol intermediate simulations, mutate substrate and equilibrate for 20 ps prior to step 7.

crystallographic water heavy atom and (2) within 16.6 Å of the scissile carbon or within 4.12 Å of the protease or ligand. With the positions of the protease, ligand, and crystallographic waters fixed, the solvent waters were equilibrated at 300 K for 10 ps to allow them to assume favorable positions on the protein surface. The entire structure was then energy-minimized briefly to relieve bad contacts.

The inhibited structure was prepared with steps 1−4 (Table 2), after which a 100 ps MD trajectory at 300 K was collected for analysis (step 7). The initial locations of the solvent waters for all of the substrate systems were computed as described above without a lytic water present. The lytic water, if required, was generated prior to step 3 (minimization). Simulated annealing to relax the substrate conformation (steps 5 and 6) was followed by collection of the 100 ps, 300 K trajectory for analysis. Harmonic restraints were placed on water 301 and the lytic water during simulated annealing, tethering them to nearby atoms to prevent them from moving far from their binding sites. These restraints were removed prior to 300 K dynamics. The *gem*-diol intermediates were prepared by mutating substrate structures after annealing. Equilibration for 20 ps preceded the trajectory used for analysis.

Reactant and intermediate conformations are likely to be very similar to the transition state, which the HIVPR/JG-365 crystal structure is believed to represent,[5] except for local changes in the immediate vicinity of the active site. Since molecular dynamics simulation does not preserve crystal structures exactly, and to limit the computer time needed, the locations of atoms in groups more than 4 Å from the ligand and 12 Å from the scissile carbon were fixed following equilibration of the minimal solvation layer, preserving the crystal structure. (Related studies of other proteins have also limited the size of the flexible portion of the protein/ligand system.[27,28]) Thus local conformational changes in the vicinity of the ligand or the active site were allowed, but global conformational changes of the protease were not. The minimal solvation layer, rather than a large box of water in conjunction with periodic boundary conditions, was used because images and extra solvent would have had little if any interaction with the mobile atoms. (This is particularly true of the catalytic site, whose distance from all fixed atoms is greater than the nonbond cutoff distance, 12 Å.)

We emphasize that simulations with only local flexibility would be unlikely to sample productive conformations if global conformational changes were necessary to achieve them. Several of the simulations reported here maintained productive conformations for their entire duration, suggesting that global conformational changes are not important for cleavage once the substrate is bound at the active site.

Analysis. Productive conformations for the HIVPR/substrate and HIVPR/*gem*-diol intermediate simulations were characterized as follows. For direct nucleophilic attack by an aspartate, a productive conformation was defined as a close approach (generally less than 3.5 Å) of an aspartate to the scissile carbon concurrent with formation of a hydrogen bond between the protonated Asp and either the carbonyl oxygen or the scissile nitrogen. For general acid−general base catalysis, a productive conformation consisted of a close approach of the lytic water to the scissile carbon, while hydrogen bonds were formed between the aspartate and the lytic water and between the protonated Asp and either the carbonyl oxygen or the scissile nitrogen. For the *gem*-diol intermediate, formation of a hydrogen bond between the protonated Asp and the scissile nitrogen constituted a productive conformation.

Table 3. Ligand Backbone Hydrogen Bonding

ligand	bb	enzyme	crystal	JG-365[a]	direct[a]	ga-gb[a]
P_4 (Ser)	O	Gly48 N	3.1	3.1	3.3	3.1
P_3 (Gln[b])	N	Asp29 OD2	3.1	3.2	3.1	3.3
..	O	Asp29 N	3.1	4.4[c]	3.0	5.3[c]
P_2 (Asn)	N	Gly48 O	2.9	3.1	3.1	3.0
P'_2 (Ile)	N	Gly127 O	3.0	3.2	3.8	4.6
..	O	Asp129 N	3.5	3.7	5.8	4.0
P'_3 (Val)	N	Gly148 O	2.6	3.2	3.0	4.4
..	O	Gly148 N	2.9	3.0	7.4[d]	7.6[d]

[a] Average interatomic distance (Å) for modified charge simulations: JG-365 (inhibitor, Asp 125 OD1 protonated), direct (substrate, Asp 125 OD2 protonated, no lytic water), and ga-gb (substrate, Asp 125 OD2 protonated, lytic water present). [b] Leu for JG-365. [c] P_3 O interacts with P_1 N. [d] P'_3 O is involved in hydrogen bond with Arg 8 NH2.

For HIVPR/substrate simulations that sampled productive conformations, the structure most likely to approximate the transition state was selected from the MD trajectory. For simulations with no lytic water, this structure was defined as the frame with the minimum quantity

$$(max(0,(min(r_1,r_2) - 2.4)))^2 + 3(max(0,(r_3 - 2.5)))^2 \quad (1)$$

where r_1 and r_2 are the distances (in Å) between atom L 4 C (the carbonyl carbon) and atoms OD1 and OD2 of the unprotonated Asp, and r_3 is the distance between atom L 4 O (the carbonyl oxygen) and atom O_H, the hydrogen-bearing oxygen of the protonated Asp. For simulations with a lytic water, the quantity minimized by frame selection was

$$(max(0,(r_4 - 2.4)))^2 + 3(max(0,(r_5 - 2.5)))^2 + 3(max(0,min(r_6,r_7) - 2.5))^2 \quad (2)$$

where r_4, r_5, r_6, and r_7 are the distances between atoms W 2 O and L 4 C, O_H and L 4 O, and between W 2 O and atoms OD1 and OD2 of the unprotonated Asp, respectively. The value 2.5 Å represents a lower limit for a hydrogen bond heavy atom−heavy atom distance,[54] and 2.4 Å is probably close to the distance from the nucleophile to the scissile carbon in the transition state.[55] The factor of 3 adjusts the weighting of the three criteria (the hydrogen bond distances are probably more important), but it is not critical. We call trajectory frames chosen according to these criteria "best" frames.

Asp Coplanarity and Aspartate Charges. A characteristic feature of crystal structures of HIVPR and other aspartic proteases is the coplanarity of the catalytic Asp carboxylate groups. In the initial simulations of both inhibitor and substrate, however, one of the Asp side chains rotated and bent toward the other, creating a scissoring conformation in which coplanarity was destroyed and the Asps interacted primarily with each

(55) Bernandi, F.; Olivucci, M.; Poggi, G. *Chem. Phys. Lett.* **1988**, *144*, 141−144.

Table 4. JG-365 Simulations

AspH O_H[a]	Asp[a] charges[b]	ϕ[c]	a[d]	b[d]	c[d]	d[d]	e[d]	no. water 301 H-bonds[e]
crystal[f]		4	2.8	3.0	3.3	3.5	2.8	4
125 OD1	s	74	3.9	2.9	4.6	5.5	3.0	2
25 OD1	m	44	2.9	2.9	3.1	3.7	2.8	0
25 OD2	m	9	2.8	3.0	3.0	3.4	2.7	2
125 OD1	m	4	2.9	3.1	2.9	3.3	3.3	(1)
125 OD2	m	6	2.8	3.1	3.0	3.3	2.8	0
25 OD2, 125 OD1		4	2.9	3.1	2.9	3.2	3.0	1
	m	14	4.6	4.6	2.8	4.5	3.0	3

[a] Asp and AspH refer to unprotonated and protonated catalytic aspartic acid residues, respectively, and O_H refers to a protonated side-chain oxygen. [b] The letters s and m indicate standard (PARM20[42]) and modified charges. [c] Angle in degrees between the planes defined by the side-chain atoms O−C−O of each catalytic Asp (simulation average structure). [d] Columns a, b, c, and d give the average distance (Å) between the inhibitor's hydroxy oxygen and each of the catalytic Asp side-chain oxygens (25 OD1, 25 OD2, 125 OD1, and 125 OD2, respectively), and column e gives the average distance between the inner Asp oxygens (25 OD2 and 125 OD2). [e] We have defined a hydrogen bond as an average heavy atom−heavy atom distance of less than 3.5 Å; numbers in parentheses include distances of up to 4.0 Å and may be associated with hydrogen bonds formed transiently. [f] Data pertain to crystal structure.

other (note the values of the angle between the two Asp O−C−O planes in Tables 4−6). The protonated Asp invariably formed a bifurcated hydrogen bond with both side-chain oxygens of the aspartate, a conformation not seen in the crystal structures. The short hydrogen bonds between Asp25(125) OD2 and Gly27(127) N, which are present in the crystal structure and may help maintain the coplanarity of the Asp carboxylates, were broken. No conformation likely to initiate reaction was realized by these simulations, probably because the details of interactions in the active site were unrealistic.

The cause of these difficulties was thought to be the atomic charge parameters for the catalytic Asp side chains, particularly the aspartate. The environment of the catalytic Asps is unique: they are buried (their accessible surface area, calculated by the Lee and Richards algorithm,[56] is 0.0 Å^2), poorly solvated, and close to each other. Since in such an environment Asps are probably less polarized than elsewhere, we tested the sensitivity of the active-site conformation to the choice of Asp side-chain charges. The degree of charge polarization in the aspartate side chain was reduced from its value in the initial simulations by using the same charge parameters as for the protonated Asp but removing the hydrogen with its partial positive charge of 0.45 au (see Table 1). The inner and outer side-chain oxygens had slightly different charges, which is consistent with their having different environments. The partial negative charge on the aspartate is unphysical, but this model does provide a decreased polarization and a charge difference between the side-chain oxygen charges.

Simulation with the modified charges generally maintained Asp coplanarity as well as other hydrogen bonding patterns in the catalytic triads (*e.g.*, the hydrogen bonds between Asp25(125) OD2 and Gly27(127) N). The crystal structure of JG-365 was better maintained, and substrate simulations led to potentially productive conformations.

That a minor adjustment in the charges of the unprotonated Asp has such a large effect on the dynamics is surprising. To better understand this behavior, we calculated atomic charges with the method of Truhlar, Cramer, and co-workers,[57] which scales Mulliken charges from an AM1 or MNDO calculation according to empirically derived scaling factors and which gives results very similar to those obtained with the ChelpG[58] method. For isolated protonated and unprotonated Asp residues, charges obtained with this method agree well with the charges listed in Table 1. (Carboxylate oxygen charges calculated for CH_3CH_2-CO_2H differ from the protonated Asp charges in Table 1 by only 0.00 and 0.03 au.) When environments similar to those of the HIVPR catalytic Asps were included in the calculations, however, the carboxylate oxygen charges varied. For these calculations, only residues containing a heavy atom or polar hydrogen within 4 Å of the catalytic Asp carboxylate oxygens were retained: Asp25, Thr26, Gly27, and Ala28 of both monomers; Pro and Tyr(Phs) of the ligand; and the lytic water, if present. Severed covalent bonds were terminated with a hydrogen atom. Productive conformations corresponding to energy-minimized "best" frames (see eqs 1 and 2) with and without a lytic water and to the crystal structure, briefly minimized, were analyzed. In all of these calculations, Asp25 was unprotonated, and Asp125 was protonated.

Two noteworthy trends in the Asp25 and Asp125 charges calculated in these environments were observed. First, the Asp-25 OD1 and OD2 charges differed by 0.13 au on average and by as much as 0.16 au, primarily as a consequence of a hydrogen bond involving Asp25 OD2 and Gly27 N. This supports the modified charges, for which the difference is 0.10 au. Second, the charge on the hydrogen-bearing Asp125 oxygen varied by as much as 0.05 au, depending on its position, and the Asp25 OD1 and OD2 charges differed from those of an isolated Asp by as much as 0.13 au. Thus the carboxylate oxygen charges vary significantly according to the chemical environment.

Table 5. Substrate Simulations

AspH O_H[a]	lytic water	Asp[a] charges	ϕ[a]	nucleophile[b] to scissile C	AspH[a] O to carbonyl O	AspH O to scissile N	Asp O to lytic water O[c]	no. water 301 H-bonds[a]
25 OD2	no	s	87	3.2[d]	5.2[d]	4.3[d]		2
25 OD2	yes	s	85	3.9	5.9	4.5	3.0[d]	1
125 OD2	no	s	58	3.9	3.3[e]	3.6[e]		1(2)
125 OD2	yes	s	56	3.1	4.0	4.5	2.8	1(2)
25 OD1	no	m	63	5.1	6.1	4.2		2
25 OD1	yes	m	29	3.2	3.9	4.0	2.8	1(2)
25 OD2	no	m	17	4.7	7.2	5.5		2
25 OD2	yes	m	47	3.3	4.9	5.4	2.8	1(2)
125 OD1	no	m	23	3.3	2.9	5.1		4
125 OD1	yes	m	29	3.2	3.0	4.8	2.8	1
125 OD2	no	m	14	3.1	3.0	5.0		2(4)
125 OD2	yes	m	37	3.2	3.7	5.3	2.8	1(2)
best frame[f]	yes	m	22	2.5	2.7	4.9	2.6	0(1)
minimum[f]	yes	m	13	2.8	2.8	4.9	2.7	0(1)

[a] As in Table 4. [b] Lytic water when present; otherwise nearest catalytic aspartate oxygen. [c] Shorter of the distances from the lytic water oxygen to the two aspartate side-chain oxygens. [d] Distance in Å. [e] Although this heavy atom−heavy atom distance is small, analysis of the hydrogen position indicates that no hydrogen bond is formed. [f] The last two rows correspond to the "best" frame from the simulation immediately preceding them and the local minimum corresponding to it (Figure 6).

Table 6. *Gem*-Diol Simulations

AspH O_H[a]	chirality of scissile N	ϕ[a]	AspH[a] O to scissile N	no. water 301 H-bonds[a]
25 OD1	S	39	4.4[b]	4
25 OD2	S	36	3.5	4
125 OD1	S	23	4.9	2
125 OD2	S	30	4.9	3(4)
25 OD2	R	24	5.5	2

[a] As in Table 4. [b] Distance in Å.

However, simulations with the Asp25 and Asp125 side-chain charges obtained in this manner failed to maintain coplanarity of the Asp side chains and sampled no productive conformations. With Asp125 OD1 protonated and no lytic water, the carbonyl oxygen swung away from the Asps and toward water 301, apparently to avoid the high negative charge of the region of the Asp side chains. With a lytic water present (Asp125 OD1 protonated), the Asps sampled both coplanar and noncoplanar conformations, but Asp125 OD1 quickly swung away from the carbonyl oxygen and toward Asp25, and the carbonyl oxygen swung away from the Asps, so that no productive conformations were sampled.

The inability of simulation with our best assessments of the catalytic Asp side-chain charges (based either on calculations for isolated molecules, as in the case of the PARM20 charges, or on calculations that include environmental effects, such as those just discussed) to maintain observed conformations or sample productive conformations is surprising. A dynamical model that includes charge polarizability might be essential, since the conformation of the active site and the precise charge distribution appear to be sensitive to each other. However, such a method is not yet practicable for large simulations.

Our solution was empirical: to choose a charge distribution according to the criterion that the observed conformation, namely one with coplanar Asp carboxylates, be maintained. One may view the unphysically small total charge of the unprotonated Asp side chain as compensating for the inability of the charges to adjust during dynamics, an adjustment that may be necessary to maintain the Asp carboxylates in a coplanar orientation. The observation that a modest perturbation in the atomic charge parameters effects a major change in the Asp conformations and hydrogen-bonding patterns suggests that metastability of the protease active-site conformation may be important for catalysis. In the work described in remainder of this article, modified charges were used unless otherwise specified.

Results

The reliability of the simulations was assessed on the basis of the root mean square deviation (rmsd) of the protease and ligand backbone atoms from their crystal structure positions, ligand backbone and active-site hydrogen bonding, and comparison of the simulated active-site conformation with crystal structures.

The average rmsd of the atomic coordinates from the crystal coordinates was calculated for three of the modified charge simulations: for JG-365 (Asp125 OD1 protonated) and for the substrate both with and without lytic water (Asp125 OD2 protonated). The rmsd was 0.95, 1.23, and 1.36 Å, respectively, for the mobile protease backbone atoms, and 1.39, 1.63, and 1.98 Å for the ligand backbone (including only the nitrogen of Tyr or Phs because of the backbone mutation). The larger values for the substrate probably reflect accommodation to the mutations. In any case, these rmsd values indicate that the protease conformation was well maintained and the ligand did not undergo large backbone deformation during simulation.

Hydrogen bonds between the ligand backbone and the protease for the same three simulations are compared with those of the JG-365 crystal structure in Table 3. (Ligand backbone atoms omitted from the table do not form hydrogen bonds with the protease or are proximal to the backbone mutation.) The simulations preserved the crystal structure hydrogen bonding pattern well on the N-terminal side of the scissile bond. The only shift in hydrogen bonding is for P_3 O in two of the simulations. The C-terminal side is more variable. The JG-365 simulation conserves the crystal pattern, but several hydrogen bonds were broken or replaced in the substrate simulations, to a greater extent when a lytic water was present. Apparently, inserting the lytic water into the active site caused some shifting of the ligand backbone during simulated annealing.

Five important enzyme–enzyme hydrogen bonds were also monitored for the simulations discussed above: the "fireman's grip"[5,22] (Thr26 N-Thr126 OG1 and Thr126 N-Thr26 OG1) and those involving the catalytic Asp side chains (Gly27 N-Asp25 OD2 and Gly127 N-Asp125 OD2), all of which stabilize the position of the catalytic Asps, and the one between the tips of the flaps (Gly51 N-Ile150 O). In the JG-365 simulation all were maintained, although Gly27 N-Asp25 OD2 was weakened (average heavy atom–heavy atom distance of 3.7 Å). Without a lytic water, simulation with substrate maintained all of these hydrogen bonds well. With a lytic water included, Thr126 N-Thr26 OG1 was broken (4.9 Å), and Gly51 N-Ile150 O and Gly27 N-Asp25 OD2 were weakened (3.6 and 3.4 Å).

The rmsd and hydrogen bonding analyses confirm that simulation maintained the protease conformation and that the mutations introduced to convert the inhibitor to the substrate did not cause large changes in the orientation of the ligand. Some adjustment did take place, presumably to accommodate the mutated ligand and the inclusion of a lytic water.

Results of the JG-365 simulations are summarized in Table 4. Figure 4 shows the active site of the crystal structure and of time-averaged structures for simulations with standard charges

(56) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379–406.
(57) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Design* **1995**, *9*, 87–110.
(58) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894–905.

(59) The differences caused by protonating the outer oxygen (OD1) rather than the inner oxygen (OD2) of Asp125 at the beginning of a simulation are easily understood. For simulations without a lytic water (direct nucleophilic attack), the scissile carbon is brought closer to Asp25 OD1, the nucleophile, when the carbonyl oxygen hydrogen bonds with the inner side-chain oxygen of Asp125 than when it hydrogen bonds with the outer oxygen. Therefore, the simulation with Asp125 OD2 protonated sampled conformations with a short nucleophile-to-scissile carbon distance more frequently. With a lytic water present (general acid–general base mechanism), protonation of the outer oxygen of Asp125 was favored to such an extent that even when the simulation was begun with the inner oxygen protonated, the side chain of Asp125 flipped so that the protonated oxygen became outer. (This was the only occurrence of such a flip during the 300 K simulations.) The simulations still differed in one respect, though: the charges on the side-chain oxygens of Asp25. When the inner oxygen of Asp125 was protonated to begin the simulation, the inner oxygen of Asp25 was given the more negative charge (−0.51 vs −0.41 au) because of the stabilizing influence of the nearby acidic proton of Asp125. When the outer oxygen of Asp125 was protonated to begin the simulation, the outer oxygen of Asp25 was given the more negative charge because of the destabilizing influence of positioning two large negative charges (the inner Asp oxygens) near each other. When the outer oxygen of Asp25 had the larger negative charge, it formed a stronger hydrogen bond with the lytic water, which in turn shifted the hydrogen-bonding preference of Asp125 from the lytic water to the carbonyl oxygen. On the other hand, a smaller negative charge on the outer oxygen of Asp25 weakened its hydrogen bond with the lytic water, and so Asp125 tended to hydrogen bond with the lytic water to a greater extent than with the carbonyl oxygen. These subtle differences emphasize the sensitivity of simulations to the precise distribution of charge in the active site.

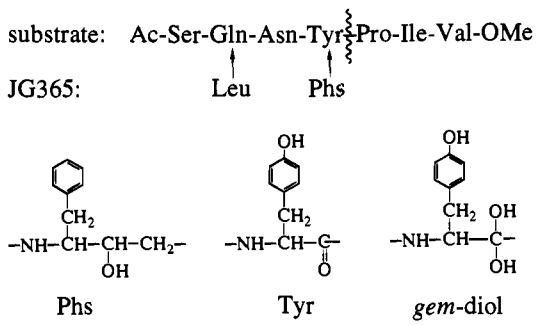

Figure 3. Model substrate, inhibitor, and *gem*-diol intermediate. Substrate was formed by mutating residues Leu and Phs of inhibitor JG-365. Diol was formed by mutating residue Tyr of substrate.

(Asp125 OD1 protonated) and with modified charges (Asp125 OD2 protonated). Selected average interatomic distances are indicated. Note that the carboxylate groups of the catalytic Asps adopt a scissoring conformation in the standard charge simulation but are coplanar in both the crystal structure and the modified charge simulation. The active-site distances are fairly constant as functions of time, indicating that the average distances are representative of the dynamical structure. (For example, the standard deviation for the distances between the inhibitor's hydroxy oxygen and the four Asp side-chain oxygens for the modified charge simulation with Asp125 OD2 protonated, shown in Figure 4c and Table 4, are 0.15, 0.20, 0.21, and 0.27 Å.)

Table 5 summarizes the substrate simulations. The active site for selected simulations is depicted in Figure 5. In the standard charge simulations, the carboxylate groups of the catalytic Asps assumed a nonproductive scissoring conformation, as they did in the JG-365 simulations. Figure 5c depicts the average conformation for one such simulation (Asp125 OD2 protonated and no lytic water). With modified charges, productive conformations are sampled when Asp125 is protonated, but not when Asp25 is protonated. Average structures that are productive are shown for two simulations: with Asp125 OD2 protonated and no lytic water (Figure 5a,d) and with Asp 125 OD1 protonated and a lytic water present (Figure 5b,e). For these simulations, the "best" frame (*i.e.*, the one most closely approximating a transition state) from the MD trajectory was chosen according to eqs 1 and 2. This "best" structure was then energy-minimized. The energy-minimized structure for the simulation with a lytic water (whose average structure is shown in Figure 5b) is shown in Figure 6.

Gem-diol simulations were run for each of the Asp monoprotonated states and for both chiralities of the scissile nitrogen (which is pyramidal). The results are summarized in Table 6. The average structure found to be most favorable for initiating a subsequent reaction step, generated when Asp25 OD2 is protonated and the scissile nitrogen has an S chirality, is shown in Figure 7.

Discussion

JG-365. The JG-365 calculations verify that simulation can maintain the crystal structure, and they address the question of catalytic Asp protonation state. The crystal structure protonation state of JG-365-bound protease is not known, although it has

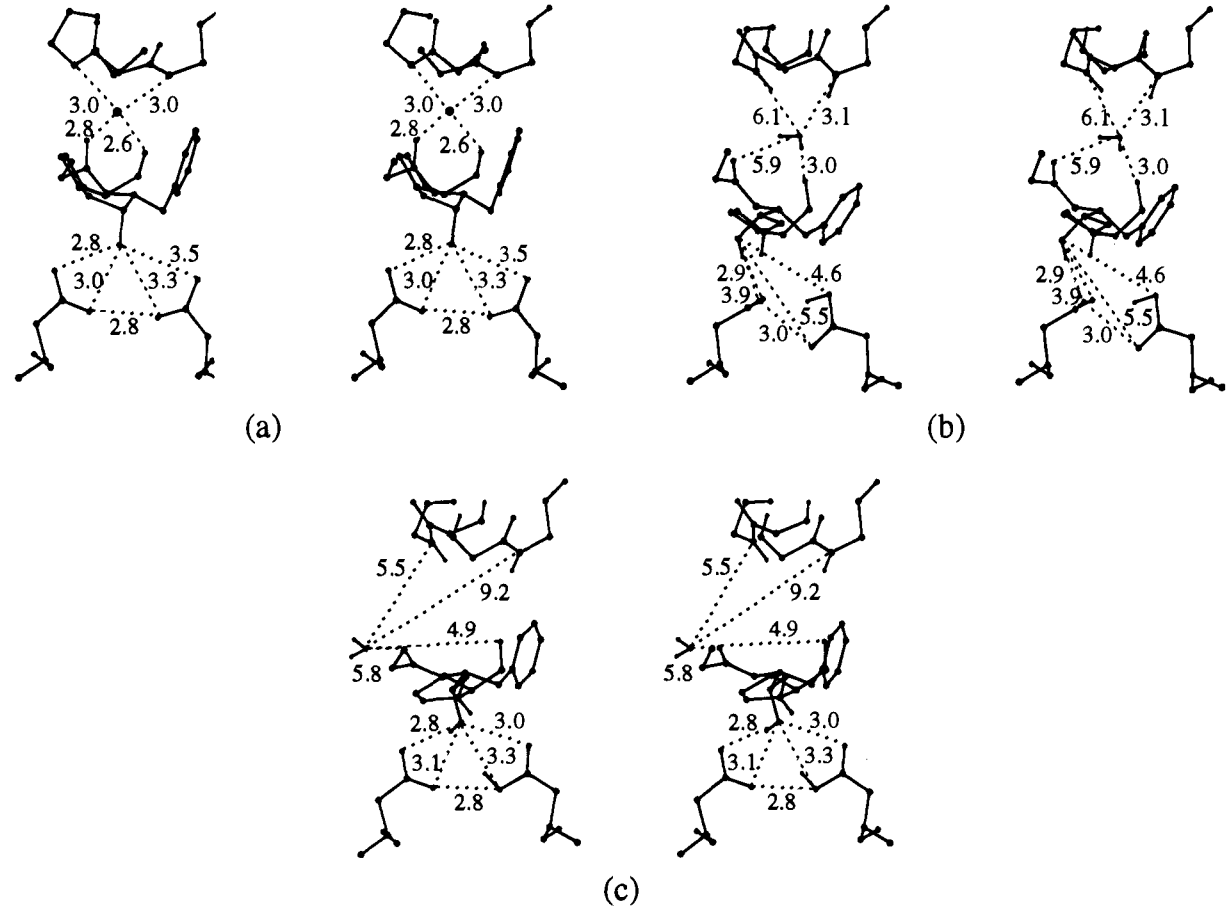

Figure 4. Active site of HIVPR/JG-365, corresponding to labeling scheme in Figure 2. The stereo images are for wall-eyed viewing. (a) Crystal structure. Numbers and dotted lines indicate interatomic distances in Å. (b) Average structure for standard charge simulation with Asp125 OD1 protonated. Numbers indicate distances averaged over 100 ps of simulation (rather than corresponding to the average structure). (c) Average structure for modified charge simulation with Asp125 OD2 protonated.

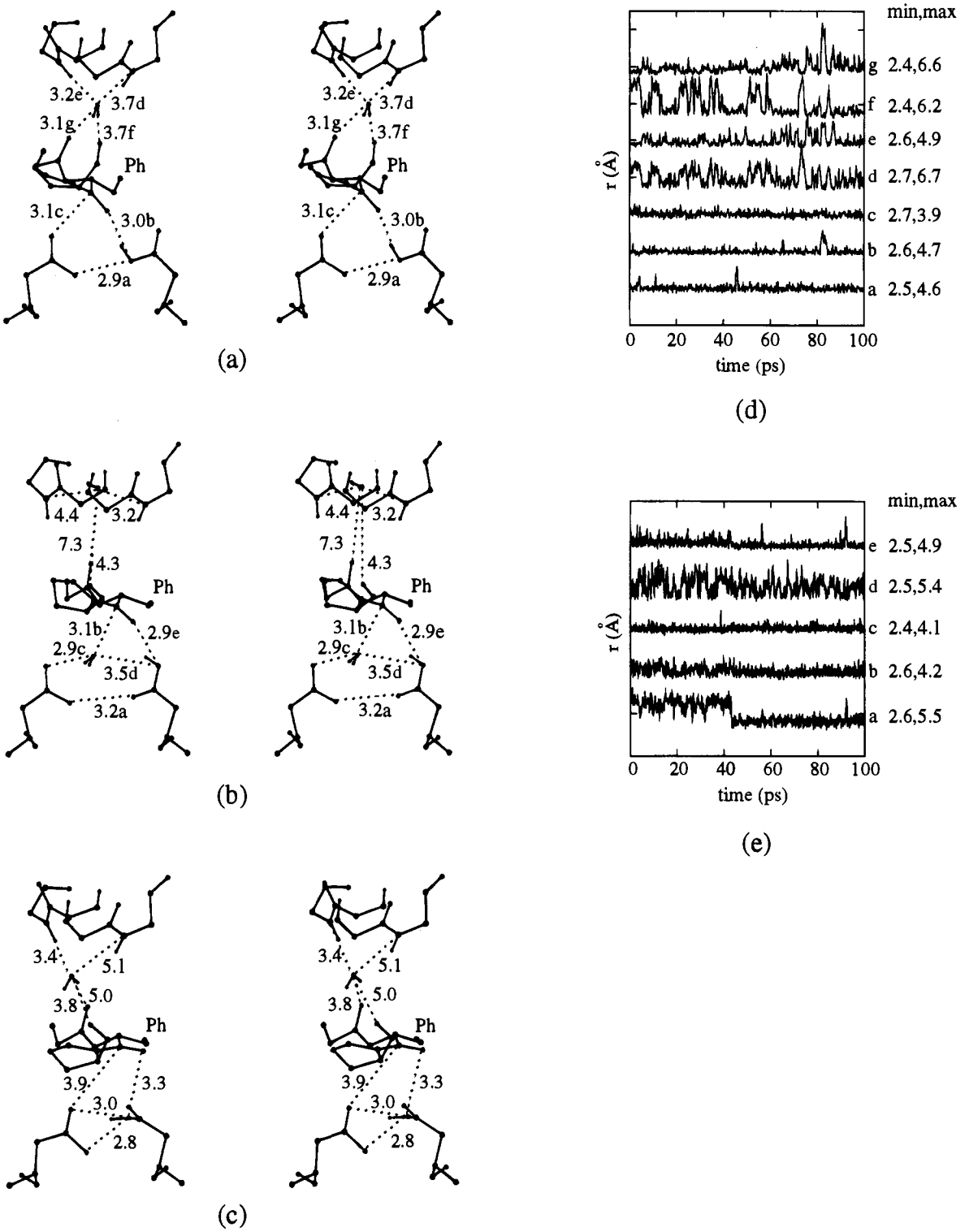

Figure 5. Active site of HIVPR/substrate: average structure for simulations with (a) Asp125 OD2 protonated, modified aspartate charges, and no lytic water; (b) modified charges, Asp125 OD1 protonated, and lytic water present; and (c) Asp125 OD2 protonated, standard charges, and no lytic water. (d) and (e) show selected interatomic distances as functions of time for the simulations corresponding to (a) and (b), respectively. Each distance is labeled by its letter in (a) or (b). The *y*-axis ticks represent increments of 3 Å. The lines are offset, each being positioned such that its average value (shown in (a) or (b)) coincides with a tick. The minimum and maximum values for each distance are shown at the right. The tyrosine side chain is abbreviated for clarity, Ph indicating the position of the phenyl ring. The average structure and distances in (b) were computed for the interval from 40 to 100 ps only, which represents a different structure than the interval from 0 to 40 ps.

been suggested that it is monoprotonated like the active form of the protease.[36] We simulated the HIVPR/JG-365 complex with six different Asp protonation states, using modified charges (Table 3). The Asps maintained a conformation similar to the crystal structure when Asp25 OD2, Asp125 OD1, Asp125 OD2, or both Asp25 OD2 and Asp125 OD1 were protonated (the Asp side chains were coplanar and positioned symmetrically about the inhibitor's hydroxy oxygen). In the other simulations, the Asps did not remain coplanar, or the hydroxy group swung away from the vicinity of the Asps to form a hydrogen bond with water 301 (when both Asps were unprotonated). It has been suggested that there is a preference for protonation of Asp125;[36]

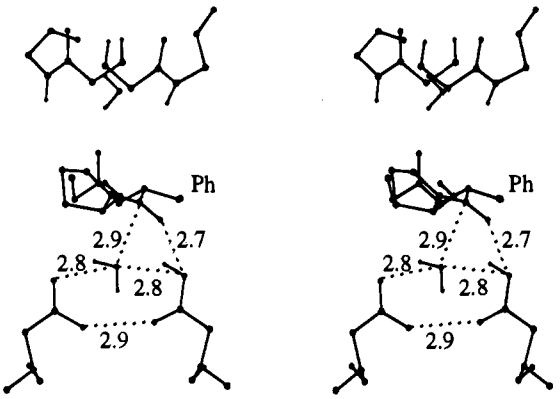

Figure 6. (a) Active site of energy-minimized "best" frame from simulation with modified aspartate charges, Asp125 OD2 protonated, and a lytic water present. The frame was chosen so as to minimize the expression given by eq 1.

it is also possible that the proton is not localized on one oxygen but jumps rapidly from a donor to its hydrogen-bonding acceptor. The simulations are consistent with either of these possibilities.

The simulations do not maintain the four hydrogen bonds with water 301 that are found in the crystal structure. This is surprising because most of the substrate and *gem*-diol simulations that maintain productive conformations also maintain the four water 301 hydrogen bonds (see below). The disruption of the water 301 binding site appears to be caused by a repositioning of the ligand proline, whose nitrogen is protonated and forms a hydrogen bond with Gly127 O. This hydrogen bond is not present in the crystal structure. To test whether insufficient hydration of the proline nitrogen causes the proline to shift its position during simulation, two waters were inserted into pockets 2.6 and 4.5 Å from the nitrogen, and the simulation was rerun (Asp125 OD1 protonated). The proline shifted less and the number of water 301 hydrogen bonds increased to two, but the water 301 binding site was still not as ideal as in the crystal structure. Nevertheless, aside from some difficulties with water 301, the JG-365 simulations were able to maintain the active-site conformation reasonably well.

Substrate. Substrate simulations were run for all possible Asp monoprotonated states, with and without lytic water. The protonation states do not generally interconvert on the 100 ps time scale at 300 K. The simulations were analyzed for productive conformations. Values for criteria pertinent to the assessment of productivity (see Methods section) are reported in Table 5.

Table 5 shows that none of the simulations supports formation of a hydrogen bond between the scissile nitrogen and the protonated Asp. Further analysis of the trajectories indicates that even transient formation of such hydrogen bonds does not occur (data not shown). This finding is in agreement with experimental evidence that the scissile nitrogen is not protonated in the first reaction step.[24] In the four cases in which Asp25 was protonated, the carbonyl oxygen remained too far from Asp25 for hydrogen bonding, and Asp125 was too far from the scissile carbon to be a likely nucleophile even when there was no lytic water.

With Asp125 protonated, energetically favorable conformations consistent with either the direct nucleophilic attack or general acid–general base mechanisms were found. The average structures shown in Figure 5a,b represent productive conformations stable on the 100 ps time scale: the former for direct attack by an aspartate and the latter for a general acid–general base mechanism. These classical MD simulations suggest that reactants may explore conformations likely to initiate either type of mechanism, which is perhaps surprising since interpretation of experimental evidence has favored general acid–general base catalysis.[13,14,24] Calculations of barrier heights will be required to better determine which type of mechanism is favored by theory, but the results here suggest that direct nucleophilic attack as well as general acid–general base catalysis is a possibility.

Each of the simulations represented in Figure 5a,b is characterized by a local energy minimum close to the productive average structure. The energy-minimized "best" frame (Figure 6) from the simulation corresponding to Figure 5b is one example. The high degree of symmetry and the bifurcated hydrogen bond formed by one of the lytic water's hydrogens with the inner Asp oxygens is very much like the productive conformation suggested in ref 22. This energy minimum is relatively insensitive to the precise choice of catalytic aspartate atomic charge parameters. When the structure was re-energy-minimized using the standard charges, the conformation changes very little.

A quite stringent definition of a productive conformation was also used to characterize the simulations: heavy atom–heavy atom distances of 3.0 Å or less for the hydrogen bonds and nucleophilic attack that pertain to the mechanism under investigation. According to this definition, the simulations of direct nucleophilic attack maintained productive conformations 22.3% (Asp 125 OD2 protonated) and 7.3% (Asp 125 OD1 protonated) of the time, and simulations of a general acid–general base mechanism did so 23.6% (Asp 125 OD1 protonated) and 3.7% (Asp 125 OD2 protonated) of the time. With Asp 25 protonated, no productive conformations were sampled. Since productive conformations were observed only when Asp125 was protonated, the substrate simulations suggest that reaction is initiated by formation of a hydrogen bond between the carbonyl oxygen and Asp125. This is consistent with experimental results.[13,14,24]

The differences between simulations with alternate side-chain oxygens of Asp125 protonated are easily rationalized in terms of the electrostatic environment of the catalytic Asps.[59] These differences underscore the sensitivity of the simulations to the precise distribution of charge in the active site.

Two other aspects of the substrate simulations warrant remark: the mobility and hydrogen bonding of the lytic water and of water 301.

The lytic water is held very tightly in place by hydrogen bonds with the Asps and the substrate. During dynamics, hydrogen bonds involving the lytic water undergo even less variation in length than those involving water 301 in most simulations (Figures 5d,e and 7b). The lytic water oxygen remained approximately midway between the outer Asp oxygens in all simulations in which it was included. This position seems to be quite favorable: in one standard charge simulation without an explicit lytic water (Asp25 OD2 protonated), another water migrated to this position during simulated annealing. It seems likely that a lytic water, if generally present, would be seen in X-ray crystallographic structures for inhibitors that do not displace it (the Phs hydroxy group of JG-365 is thought to mimic an attacking water molecule in a transition-state structure[1]). As noted previously, a water molecule in such a position has been observed in crystal structures of the unliganded protease[43] and of other aspartic proteases.[44]

The simulations suggest that hydrogen bonding between the lytic water and the substrate may be important in initiating nucleophilic attack. The lytic water invariably formed a strong hydrogen bond with atom L 4 N. It would be natural for this

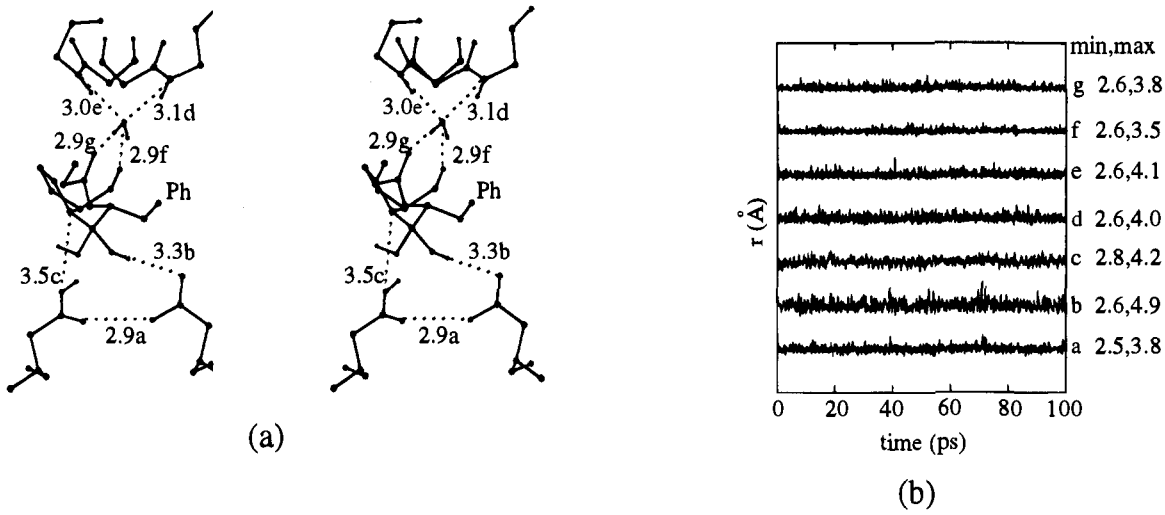

Figure 7. (a) Active site average structure for *gem*-diol simulation with Asp25 OD2 protonated and a scissile nitrogen chirality of S. (b) Selected interatomic distances as functions of time.

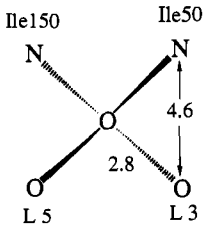

Figure 8. Interatomic distances for ideal water 301 hydrogen bonding. When water oxygen-heavy atom distances are 2.8 Å, the heavy atoms are all 4.6 Å apart.

interaction to facilitate the approach of the lytic water to the scissile carbon by orienting one of the water oxygen's lone pairs toward the scissile carbon and pulling the water away from the Asps and toward the substrate. For example, in the energy-minimized "best" frame of Figure 6, the distance from atom W 2 O to both atoms L 4 N and L 4 C is 2.8 Å, and the L 4 N-W 2 O-L 4 C angle is 48°.

Water 301 is thought to be important for catalysis because it is located near the active site in many crystal structures of inhibited protease, where it forms hydrogen bonds with both protease and ligand. In the crystal structure of JG-365, it hydrogen bonds with the substrate atoms L 3 O and L 5 O and with the flap atoms Ile50 N and Ile150 N.[22] Water 301 probably does not participate in the reaction directly; in none of the simulations did it approach to within 4 Å of the scissile carbon for more than about 0.1 ps. Its importance is believed to be in stabilizing the transition state and in straining the substrate's peptide linkage away from coplanarity, thereby weakening the peptide bond and promoting cleavage.[22]

The mobility of water 301 varied widely among the simulations, reflecting the suitability of the site for binding a water. To quantify this, we observed that for water to form four strong hydrogen bonds (*i.e.*, a heavy atom−heavy atom distance of about 2.8 Å), the four heavy atoms must be located near the vertices of a regular tetrahedron 4.6 Å on a side (Figure 8). The average length of the six sides of the tetrahedron formed by atoms L 3 O, L 5 O, Ile50 N, and Ile150 N was calculated for the crystal structure and the 24 simulations of Tables 4−6, and it is compared with the number of water 301 hydrogen bonds in Table 7. The average value is 4.6 Å for the crystal structure, 5.2 Å or less for the six simulations that support formation of four hydrogen bonds, and 5.3 Å or more for all others. Thus this distance is a useful indicator of the suitability

Table 7. Water 301 Hydrogen Bonding

no. H-bonds[a]	no. simulations	av dist[b]	50 N-L 5 O[c]
2(4),4	6	4.6−5.2	4.8−6.2
2,3	6	5.3−6.0	6.3−8.2
0−1(2)	13	5.7−6.4	6.8−8.6

[a] As in Table 4. [b] Average of the six average interatomic distances (Å) between the four heavy atoms involved in hydrogen bonding with water 301 in the HIVPR/JG-365 crystal structure (see Figure 8). [c] Average distance between atoms Ile 50 N and L 5 O.

of the site for binding water 301, which varies significantly among the simulations.

The most critical of the four water 301-binding heavy atoms is L 5 O. It is the most mobile, and in every simulation but one, the longest of the six heavy atom−heavy atom distances is between L 5 O and Ile50 N. This is because the position of L 5 O is related to the conformation of the ligand Pro, which varies among the simulations, and the carbonyl oxygen, which due to the rigidity of the peptide bond may affect the Pro conformation.

A greater number of hydrogen bonds with water 301 was associated with conformations likely to initiate reaction (Table 5). The two simulations without a lytic water that sampled productive conformations (Asp125 protonated) were the only ones to support formation of four hydrogen bonds with water 301. This suggests that the positions of the carbonyl oxygen and atom L 5 O are correlated, since a stable carbonyl oxygen-protonated Asp hydrogen bond is critical for productivity, and the position of atom L 5 O is critical for water 301 binding. When atom L 5 O is pulled "upward" to improve the conformation of the water 301 binding site for hydrogen bonding, the carbonyl oxygen may be moved in lever-like fashion "downward" to interact with the Asp. This cooperativity suggests that water 301 may stabilize productive reactant conformations.[22]

It is not clear whether such cooperative movements characterize the lytic water system. Even simulations that sample productive conformations (Asp125 protonated) do not bind water 301 well, probably due to shifting of the substrate backbone during simulated annealing. Visual inspection of average structures reveals that atom L 5 O is far out of position for hydrogen bonding, a consequence of adjustment in the Pro and Ile conformations (*cf.* Table 3).

This brings up the question of whether repositioning of the substrate backbone during simulated annealing with a lytic water present was an artifact of simulation. In order to assess this, a

lytic water was inserted into a structure created by simulated annealing *without* a lytic water (Asp125 OD2 protonated). The water oxygen was placed midway between the outer Asp oxygens, its location in the other simulations with lytic water. Within 20 ps the water left its binding site, drifting far from the Asps and the scissile carbon. This demonstrates the necessity of performing simulated annealing with a lytic water included and with its position restrained to allow the binding site to adjust to its presence.

***Gem*-Diol Intermediate.** For general acid–general base catalysis, a *gem*-diol, zwitterion, or oxyanion are all possible reaction intermediates. Studies of protease-catalyzed ^{18}O exchange into the substrate strongly suggest that the steps preceding peptide bond cleavage are reversible,[13] implying that the intermediate species is at least moderately stable. The simulations in Table 6 explore the conformation space accessible to a *gem*-diol intermediate for each of the four likely Asp protonation states. In four of the simulations, the chirality of the pyramidal scissile nitrogen (S) was chosen so that its three bonds pointed away from the Asps, and the fourth coordination site, where the lone pair would be located, was exposed to the Asps. This would favor proton transfer from the protonated Asp oxygen to the scissile nitrogen, commencing the next step in reaction. The distance between these two atoms was smallest (3.5 Å) when Asp25 OD2 was protonated. This is consistent with Asp25 acting as general base in the first reaction step, receiving a proton which it is then ready to give up in a later step. The average conformation and selected interatomic distances as functions of time for this simulation are shown in Figure 7. When Asp125 was protonated, a diol hydroxy group was positioned between the scissile nitrogen and Asp125, which would impede proton transfer.

Arguments based on antiperiplanarity[21,60,61] suggest that if attack of a water at the scissile carbon occurs from the general vicinity of the Asps, the scissile nitrogen's lone pair will tend to be oriented away from the direction of attack, *i.e.*, away from the Asps. If such a conformation were maintained, the scissile nitrogen would be less accessible to protonation by a neutral Asp in the ensuing step.

In light of this argument, a *gem*-diol simulation was performed with Asp25 OD2 protonated and a chirality of R for the scissile nitrogen so that the lone pair would be pointing away from the Asps. Not surprisingly, the nitrogen and protonated Asp oxygen (25 OD2) remained farther apart, 5.5 Å on average. The more intriguing result was that water 301 maintained only two hydrogen bonds rather than the four it maintained when the nitrogen had S chirality (same Asp protonation state), apparently due to unfavorable positioning of atom L 5 O. This suggests that water 301 may provide the stabilization necessary to overcome the barrier to inversion at the nitrogen (probably similar to the 8 kcal/mol barrier of pyrrolidine,[62] about the energy of two hydrogen bonds) and thereby favor the productive chirality of the scissile nitrogen.

(60) Deslongchamps, P. *Tetrahedron* **1975**, *31*, 2463–2490.
(61) Bizzorero, S. A.; Dutler, H. *Bioorg. Chem.* **1981**, *10*, 46–62.

Conclusion

We have performed MD simulations of HIVPR complexed with a model substrate, a model *gem*-diol intermediate, and the inhibitor JG-365 under a variety of conditions. Distinguishing the simulations were the protonation state of the catalytic Asps, the presence or absence of a lytic water to initiate nucleophilic attack (substrate simulations), and the chirality of the scissile nitrogen (*gem*-diol simulations). The substrate simulations suggest that, in the first reaction step, Asp125 is the general acid, and the carbonyl oxygen is more accessible to protonation than the scissile nitrogen. The *gem*-diol simulations suggest that, if the mechanism is general acid–general base, it is Asp25 that initiates peptide bond cleavage by transferring a proton to the scissile nitrogen in a later step. Simulation also suggests that water 301 may stabilize productive reactant and intermediate conformations as well as the transition state but does not participate in the reaction directly. Meanwhile, the lytic water, if present, assumes a position well suited to nucleophilic attack and is held tightly in place. The latter observation suggests that a lytic water would be observed in crystal structures of appropriately designed inhibitors.

The conformations identified by these simulations as likely to be productive may be useful starting points for theoretical investigations of reaction barriers. Finally, while simulation alone cannot demonstrate the feasibility of a mechanism definitively, the productive conformations observed here may be useful in the design of new, more potent inhibitors.

Acknowledgment. The authors thank Frederick W. Carson, Kirsten P. Eurenius, and Martin J. Field for many useful discussions. Support from the NIH Intramural AIDS Targeted Antiviral Program and time provided by the Computational Bioscience and Engineering Laboratory on their Intel iPSC/860 computer and by the NCI's Biomedical Supercomputing Center are gratefully acknowledged. D.C.C. thanks the National Research Council for a postdoctoral associateship.

JA943572Q

(62) Lambert, J. B. In *Topics in Stereochemistry*; Allinger, N. L., Eliel, E. L., Eds.; Wiley-Interscience: New York/London/Sydney/Toronto, 1971; Vol. 6, p 39.