# Molecular Dynamics Simulation of HIV-1 Protease in a Crystalline Environment and in Solution†

D. M. York,*,‡,§ T. A. Darden,‡ L. G. Pedersen,‡,§ and M. W. Anderson‡

*Laboratory of Molecular Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, and Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290*

*Received September 11, 1992; Revised Manuscript Received November 10, 1992*

ABSTRACT: Simulations of the unbound form of the human immunodeficiency virus type 1 protease have been carried out to 200 ps in a crystalline environment and in solution. Solution simulations were performed with and without charge-balancing counterions. The results are compared with the 2.8-Å crystallographic structure of Wlodawer et al. [(1989) *Science 245*, 616], and a proposed model for the solution structure which involves local refolding of the flap regions is presented. The simulations suggest the crystal packing environment of the protease dimer stabilizes the flaps in an extended conformation. Solvation of the dimer leads to local refolding of the flaps which contract toward the active site, forming increased overlap and stronger *inter*subunit hydrogn bonding at the tips. The degree to which the flaps overlap in solution is observed to depend on the charge state of the system.

Human immunodeficiency virus type 1 protease (HIV-1 PR)[1] is a 99 amino acid virally encoded protease necessary for the maturation of the HIV-1 virus, a causative agent of AIDS and related disorders. Certain viral gene products initially translated by the host cell's machinery must undergo posttranslational processing to yield mature catalytic and structural proteins. One enzyme involved in this processing is HIV-1 PR which cleaves the virally encoded *gag* and *gag–pol* fusion polyproteins into functional products (Henderson et al., 1988; Debouck et al., 1987; Graves et al., 1990). Inhibition of HIV-1 PR function results in the production of immature noninfectious virus particles in vitro (Kohl et al., 1988). Consequently, the maturation process mediated by HIV-1 PR has been identified as an essential step of the HIV-1 retroviral life cycle, which makes this enzyme an attractive therapeutic target (Debouck, 1992).

HIV-1 PR has been classified as an aspartyl protease on the basis of sequence homology to cellular proteases (Toh et al., 1985), catalytic pH studies (Hyland et al., 1991), and its sensitivity to inhibition by aspartyl protease inhibitors such as pepstatin (Seelmeir et al., 1988; Richards et al., 1989). The enzymatic form of the bacterially expressed protein behaves as a dimer (Darke et al., 1989; Meek et al., 1989). Crystallographic data of a related protease encoded by the Rous sarcoma virus (Miller et al., 1989a; Jaskolski et al., 1990) enabled initial structural models to be proposed for the HIV-1 protease (Weber et al., 1989). The subsequent progress in unraveling the structure and mechanism of inhibition of HIV-1 PR has been greatly facilitated by the elucidation of X-ray crystal structures of the enzyme both unbound (Navia et al., 1989; Wlodawer et al., 1989; Lapatto et al., 1989) and bound to synthetic inhibitors (Miller et al., 1989b; Fitzgerald et al., 1990; Swain et al., 1990; Erickson et al., 1990; Jaskolski et al., 1991; Bone et al., 1991). These structures have provided the starting point of several theoretical investigations designed to answer specific questions about the structure, dynamics, and energetics of the protein at the molecular level (Harte et al., 1990, 1992; Swaminathan et al., 1991; Ferguson et al., 1991; Reddy et al., 1991; Rao et al., 1992).

The topology of the HIV-1 PR monomer is similar to that of a single domain of pepsin-like aspartic proteases and can be described by analogous nomenclature conventions (Blundell et al., 1985). Substrates and inhibitors bind in the active site cleft (Figure 1), a cavity formed between subunits of the dimer. Loops at the base of the cleft toward the interior of the protein contain the active site triads (residues 25–27 in each monomer), conserved Asp-Thr-Gly sequences characteristic of aspartyl proteases. Two flexible flaps (residues 42–58), one from each monomer, envelop the triads and presumably regulate substrate entry into the active site. Each flap consists of an antiparallel $\beta$ sheet with an intervening glycine-rich loop at the tip. In crystallographic structures the flaps are observed to interact with inhibitors and help bind them to the active site.

Substantial differences between crystal structures in the unbound versus the bound form are observed in the region of the flaps. Analysis of crystallographic data of the unbound protease suggests the flaps have substantial thermal motion relative to the rest of the molecule (Wlodawer et al., 1989). Analogous flap domains in other aspartic proteases are also observed to be flexible (Sali et al., 1992). Comparison of the unbound and bound form of the protease indicates the conformation of the flaps must undergo significant rearrangement upon inhibitor binding (Gustchina & Weber, 1990). To study the structure and motion of this region as well as other domains, and predict conformational changes that occur upon solvation, the molecular dynamics (MD)

† This work was supported in part by Grant HL27995 from the National Heart, Lung, and Blood Institute.

* Author to whom correspondence should be addressed.

‡ National Institute of Environmental Health Sciences.

§ University of North Carolina.

[1] Abbreviations: Aba, L-$\alpha$-amino-$n$-butyric acid; AIDS, acquired immunodeficiency syndrome; HIV-1, human immunodeficiency virus type 1; MD, molecular dynamics; MDC, molecular dynamics crystal simulation; MDS, molecular dynamics solution simulation (neutralized with $Cl^-$ counterions); $MDS^+$, molecular dynamics solution simulation (+4 charge, no counterions); ⟨MDC⟩, crystal simulation average structure; ⟨MDS⟩, neutral solution simulation average structure; ⟨$MDS^+$⟩, charged solution simulation average structure; PR, protease; rms, root mean square.

0006-2960/93/0432-1443$04.00/0 © 1993 American Chemical Society

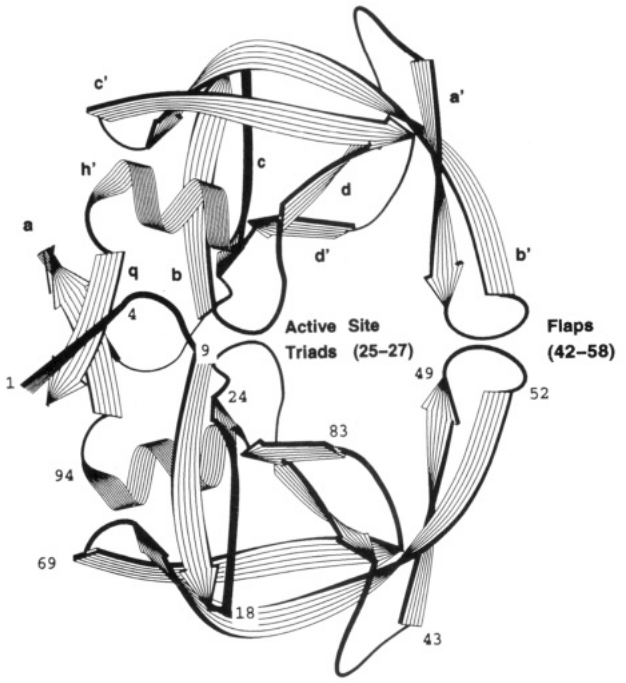


FIGURE 1: Ribbon drawing of the HIV-1 PR dimer. The upper subunit shows the nomenclature employed for the secondary structural domains (Wlodawer et al., 1989) [$\beta$ strands a (1–4), b (9–15), c (18–24), d (30–35), a′ (43–49), b′ (52–66), c′ (69–78), d′ 83–85), and q (95–99); helix h′ (86–94)]. The lower subunit shows the residue numbering. The active site triads and the flap regions are also shown. The program PAP available through the Quantum Chemistry Program Exchange (QCPE 594) was used to make the drawing.

methodology has been employed.

Molecular dynamics is a technique that permits the limited simulation of macromolecules (McCammon & Harvey, 1987). Comparisons of crystallographic and 2D NMR data (Bax, 1989; Baldwin et all, 1991) clearly show that solvation can significantly alter protein structure. It is not clear what structural changes might generally be expected upon solvating a protein crystal. However, regions which are involved in intermolecular crystal packing contacts might be expected to rearrange in solution. If changes do not involve major reorganization of the protein, it is possible that early refolding events might be observed in the subnanosecond time domain. Under these circumstances the molecular dynamics technique is useful as a predictive tool for examining structural and dynamical properties involved in early protein refolding.

## METHODS

All molecular mechanics and dynamics calculations were performed using a modified version of the AMBER3.0 (Revision A) software package (Weiner et al., 1984; Revision A by George Seibel, UCSF). The all-atom force field of Weiner et al. (1986) was employed for all standard residues. Solvent was treated explicitly using the TIP3P water model (Jorgensen et al., 1983). Chloride ion parameters were obtained from Lybrand et al. (1986). Electrostatic and van der Waals interactions were treated using a "twin-range" (9/18-Å) residue-based cutoff, described below, updated every 20 steps. A 1-fs time step was used in the integration. Molecular dynamics simulations were performed at constant temperature by coupling the systems to a thermal bath with thermal relaxation time $\tau_T$ = 0.4 ps. Simulations in the crystalline environment and in solution were carried out to 200 ps.

Several modifications to the AMBER code were made [for a more detailed description of the modifications, see Foley et al. (1992)]. The evaluation of van der Waals and electrostatic interactions was extended to include a long-range force correction by employing a twin-range cutoff. At the time of the nonbonded list update, a correction term which includes interactions between 9.0 and 18.0 Å was added to each of the force components. This term was assumed to be constant until the next nonbond update when it is recomputed. This correction proved to be essential for the stability of protein structures lacking disulfide bonds over a long simulation. Several test simulations of the HIV-1 PR were initially performed using a 9.0-Å cutoff. These systems developed instabilities early in the simulations and were not equilibrated after 100 ps. After inclusion of the twin-range method, the same systems equilibrated within 75 ps and remained stable for the duration of the simulations (200 ps). Recent studies have shown that the cutoff size for electrostatic interactions strongly influences the stability of polypeptides in solution (Schreiber & Steinhauser, 1992). Consequences of various implementations of the long-range cutoff have also been examined by Loncharich and Brooks (1989) and by Smith and Pettitt (1991).

Additional modifications to AMBER were made to allow gentle heating and ensure nondivergent solute and solvent temperatures. Further code modifications in the computation of the nonbonded list and solvent–any atom interactions were included to increase throughput [see Foley et al. (1992)]. All calculations were performed on a Cray Y-MP supercomputer (National Cancer Institute or North Carolina Supercomputing Center) or in parallel on a Silicon Graphics Iris 4D/380-VGX workstation.

(*i*) *Crystal Simulation.* The synthetic [Aba$^{67,95}$]HIV-1 protease (Wlodawer et al., 1989) modeled in this study crystallizes in the tetragonal space group $P4_12_12$, with unit cell parameters $a = b = 50.24$ Å and $c = 106.56$ Å. The unit cell contains eight monomers, the asymmetric unit consisting of a single protein monomer. The unrefined unit cell was constructed by applying the $P4_12_12$ symmetry operations to the crystallographic structure. The net charge of each monomer was assumed to be +2, consistent with the normal protonation states of the component amino acids at neutral pH. The active enzyme dimer, however, requires one of the catalytic aspartate residues to be protonated. It is possible that the active site aspartates share a proton near neutral pH since maximum protease activity occurs in the pH range 4.5–6.0 (Hyland et al., 1991). We have chosen to treat both aspartates as being fully charged for several reasons: to preserve the crystallographic symmetry within dimers, to examine the possibility of structural waters bridging the aspartates as has been observed in crystallographic structures, and to allow consistent comparison with earlier simulations of HIV-1 PR in the unbound form which also treated the aspartates as fully charged (Harte et al., 1990). In order to neutralize the net positive charge of the unit cell, 16 chloride counterions were added around positively charged surface residues and minimized in the field of the unsolvated crystal. Water molecules were then packed around the protein–ion complexes until the experimental crystal density [1.17 g/cm$^3$, assumed to be that of the isomorphous crystal structure reported by Navia et al. (1989)] was attained (5703 water molecules). The water molecules were relaxed with 200 steps of steepest descents minimization with the protein/ion positions fixed, equilibrated with 20 ps of MD, and reminimized with another 200 steps of steepest descents minimization. Positional constraints on the solute and ions were then removed, and the entire system was relaxed with 200 steps of steepest descents

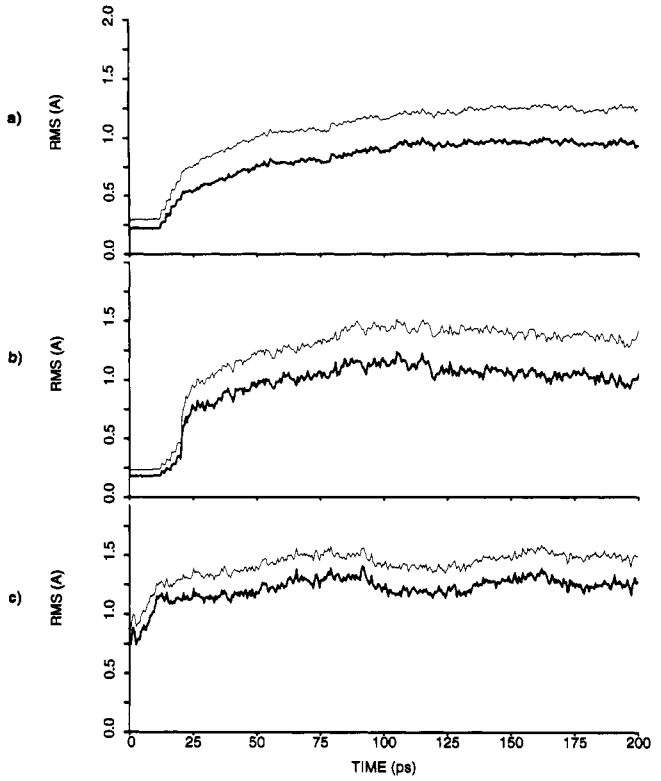


FIGURE 2: Time evolution of the rms $\alpha$-carbon (solid line) and heavy atom (broken line) deviation from the crystallographic structure of Wlodawer et al. (1989) for (a) MDC, (b) MDS, and (c) MDS$^+$. Heavy atoms include all nonproton atoms. rms values are for the *inter*monomer average (average over all monomers in the system) at each point in time.

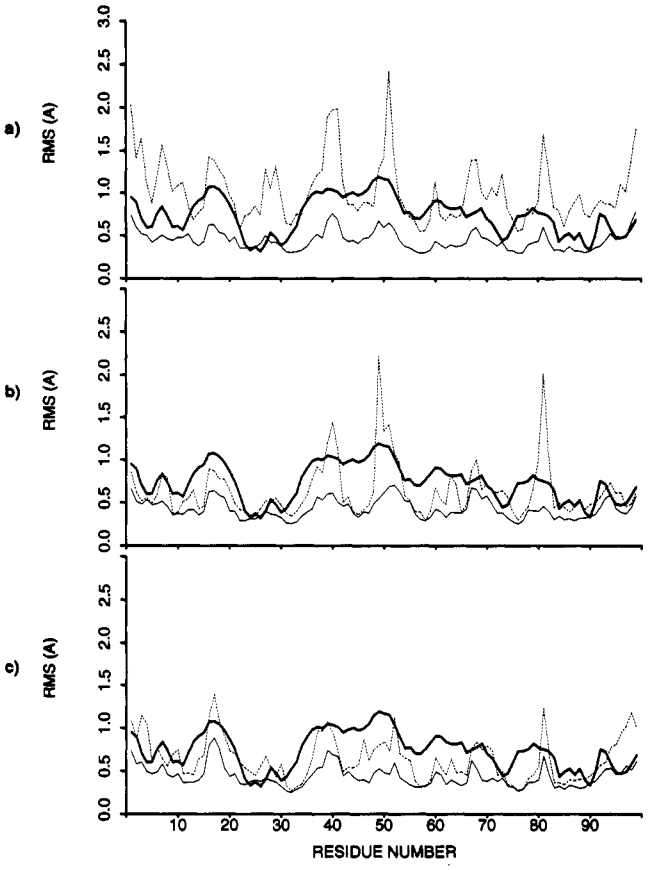


FIGURE 3: Comparisons of the estimated rms fluctuations obtained from the crystallographic isotropic temperature factors (solid line) and the total *inter*monomer rms (broken line) and average *intra*monomer rms (thin solid line). rms fluctuations in the simulations were computed over the time interval from 150 to 200 ps for (a) MDC, (b) MDS, and (c) MDS$^+$.

minimization to give the starting configuration for molecular dynamics. Initial velocities were obtained from a Maxwellian distribution at 1 K, and stepwise heating to the crystallographic temperature (293 K) was performed over 10 ps. During the heating step the positions of the solute and ions were constrained with harmonic force constants (2.0 kcal/Å$^2$) to the starting geometry to ensure that ions did not drift. These constraints were then relieved by exponential decay over the interval from 10 to 20 ps. Unconstrained dynamics of the crystal unit cell (MDC) was then performed at constant volume and temperature to 200 ps.

(*ii*) *Solution Simulations.* Two solution simulations were performed on the HIV-1 PR dimer. The first system consisted of the dimer and water and had a net +4 charge. The second system contained $Cl^-$ ions to neutralize the charge. Chloride ion positions in the neutral simulation (MDS) were obtained in the same manner as in the crystal simulation. Solvation was accomplished by immersing the protein/protein–ion coordinates in a large water box such that a layer at least 13 Å thick surrounded the solute. Water molecules were placed so as to have no water oxygen closer than 2.8 Å or water hydrogen closer than 2.0 Å to a protein atom or ion. This procedure resulted in 8002 water molecules for the charged system (MDS$^+$) and 8013 water molecules for the neutral system (MDS). Equilibration of water and energy refinement were conducted in the same manner as for the crystal simulation to arrive at the starting configuration for MD. Solution simulations were kept at constant pressure by coupling the systems with a pressure bath at 1.0 bar with pressure relaxation time $\tau_p$ = 0.6 ps. Stepwise heating to 300 K over 10 ps was performed using constrained dynamics for the simulation containing $Cl^-$ ions (MDS) as described earlier and unconstrained dynamics for the simulation without counterions (MDS$^+$). Unconstrained dynamics was then performed at constant pressure and temperature to 200 ps.

## RESULTS AND DISCUSSION

Simulations of the crystallographic structure reported by Wlodawer and co-workers (1989) were performed for 200 ps in a crystalline environment and in solution. We have examined the structure and dynamics of the protein and predicted the effects of solvation and counterions on the early refolding.

(*i*) *Structural Equilibration of the Protease.* Figure 2 shows the time evolution of the root mean square (rms) deviation of the MD structures with respect to the crystallographic structure for $\alpha$ carbons (bold line) and for heavy atoms (thin line). Since each system contains multiple asymmetric units (eight monomers in MDC, two monomers in MDS and MDS$^+$), the problem of treating each monomer explicitly was simplified by constructing an instantaneous monomer average for which the rms was obtained. The rms values in the simulations containing counterions (Figure 2a,b) show a stepwise increase between 10 and 20 ps due to the positional constraints employed during the heating phase (see Methods). The crystal simulation required an equilibration of ~120 ps with the $\alpha$-carbon rms reaching a stable asymptotic value of ~0.9 Å (Figure 2a). The solution simulations appear well equilibrated after ~75 ps of unconstrained dynamics with stable $\alpha$-carbon rms values of ~1.0 Å for MDS (Figure 2b) and ~1.2 Å for MDS$^+$ (Figure 2c).

(*ii*) *Atomic Fluctuations.* The rms fluctuation in atomic positions can be estimated from the experimental isotropic

```
          1        10        20        30        40        50        60        70        80        90        99
          |        |         |         |         |         |         |         |         |         |         |
          PQITLWQRPLVTIRIGGQLKEALLDTGADDTVLEEMNLPGKWKPKMIGGIGGFIKVRQYDQIPVEIAbaGHKAIGTVLVGPTPVNIIGRNLLTQIGAbaTLNF

X-ray     ---      ------TT-------      ----      ------  TT  ------    ----T  T----  ----      -- HHHHHHHT    ---
<MDC>     --       ------TT-------      --        --------T---------------T  T---------          -- HHHHHHHT    ---
<MDS>              ------TT-------      --        -----       -------------T  T---------          -- HHHHHHHT
<MDS+>             ------TT-------      ---       ------      -------------T  T---------          -- HHHHHHHT    ---
```

FIGURE 4: Secondary structure map as determined by the Kabsch and Sander program DSSP (Kabsch & Sander, 1983) showing $\beta$ chains (–), helices (H), and hydrogen-bonded turns (T) for the crystallographic structure by Wlodawer et al. (1989) (X-ray) and for the average structures ⟨MDC⟩, ⟨MDS⟩, and ⟨MDS⁺⟩.

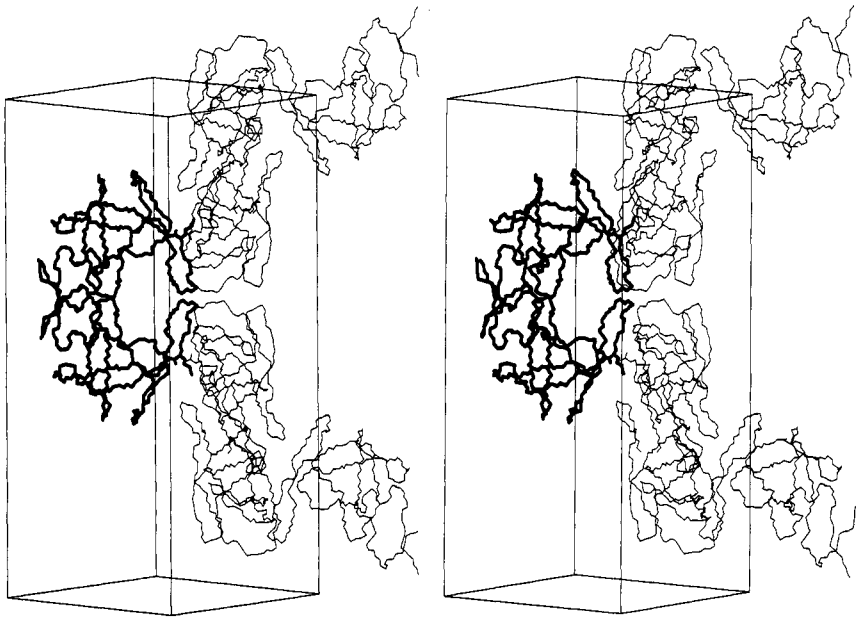

FIGURE 5: Stereoview of the HIV-1 PR crystallographic unit cell ($P4_12_12$) reported by Wlodawer et al. (1989). Unit cell dimensions are $a = b = 50.24$ Å and $c = 106.56$ Å. Backbone atoms of each monomer are shown, as well as the crystal packing contact between $Lys^{55}$ and $Gln^{92''}$ of the dimer shown in bold.

temperature factors ($B$ values) using the relation (Koehler et al., 1987; Karplus & Petsko, 1990):

$$\langle \Delta r_i^2 \rangle^{1/2} = \left[\frac{3}{8}\pi^2 B_i\right]^{1/2} \quad (1)$$

where $\langle \Delta r_i^2 \rangle^{1/2}$ is the root mean square fluctuation for atom $i$ and $B_i$ is the corresponding experimental temperature factor. Figure 3 compares the estimated rms fluctuations of $\alpha$ carbons in the crystallographic structure (bold line) with the rms fluctuations computed from the MD simulations (thin lines). To differentiate between atomic fluctuations within monomers that involve only a time dependence from atomic fluctuations which include variations between monomers, both the average *intra*monomer rms (thin solid lines) and total *inter*monomer rms (thin broken lines) are shown. The rms values calculated from the crystal simulation tend to bracket the values derived from the crystallographic data. Regions of high fluctuation in the simulations occur in loops and turns between regions of stable secondary structure and agree qualitatively with the crystallographic values.

In general, the fluctuations calculated from the MD show good correlation with the crystallographic values, except that the magnitudes differ. Fluctuations involving only a time dependence (*intra*monomer fluctuations) are generally smaller than the fluctuations derived from the experimental $B$ values. Similar results have been reported for the HIV-1 protease in solution (Harte et al., 1992). The reason for these observations relates to the isotropic harmonic approximation used to obtain eq 1, which must break down for highly anisotropic and anharmonic motions (Ichiye & Karplus, 1987). This presumably leads to overestimation of the thermal factors in the crystallographic refinement (Harte et al., 1992). On the other hand, when one considers positional fluctuations which include variations between asymmetric units (broken lines in Figure 3), the values are generally larger than the values derived from the experimental $B$ values. This type of variation is pronounced in regions not involved in regular secondary structure that adopt different local conformations between monomers. Similar observations have been reported in other crystal simulations containing multiple asymmetric units (Koehler et al., 1987; Nilsson et al., 1990).

Table I: Interdimer Contacts in the HIV-1 PR Crystallographic Structure of Wlodawer et al. (1989)[a]

| dimer 1 | | dimer 2 | | | |
|---|---|---|---|---|---|
| atom | $B$ value (Å²) | atom | $B$ value (Å²) | dist (Å) | interaction |
| $Trp^{6}$ CH2 | 20.7 | $Lys^{55''}$ O | 15.0 | 3.10 | vdW[b] |
| $Trp^{6}$ CH2 | 23.4 | $Trp^{42''}$ CH2 | 18.8 | 3.13 | vdW |
| $Trp^{6}$ CE3 | 18.9 | $Arg^{57''}$ NH2 | 3.2 | 3.46 | vdW |
| $Arg^{14}$ NH2 | 39.9 | $Arg^{14''}$ NH2 | 39.9 | 3.56 | vdW |
| $Gly^{49}$ O | 38.3 | $Gln^{61''}$ NE2 | 33.3 | 3.31 | vdW |
| $Ile^{54}$ CD1 | 18.3 | $Ile^{72''}$ O | 6.0 | 3.43 | vdW |
| $Ile^{54}$ CD1 | 18.3 | $Gln^{92''}$ OE1 | 16.0 | 3.57 | vdW |
| $Lys^{55}$ N | 16.2 | $Gln^{92''}$ OE1 | 16.0 | 3.03 | H-bond |
| $Lys^{55}$ O | 15.0 | $Gln^{92''}$ NE2 | 7.0 | 3.10 | H-bond |
| $Pro^{79}$ CG | 19.0 | $Gln^{92''}$ NE2 | 7.0 | 3.24 | vdW |

[a] Only the closest heavy atom distances <4.0 Å are shown, except for crystal contacts which involve H-bonding. Isotropic temperature factors ($B$ values) for each atom are shown. [b] vdW = van der Waals.

(*iii*) *Secondary Structure.* Secondary structural analysis of the simulation average structures and the crystallographic structure (Figure 4) was performed using the Kabsch and

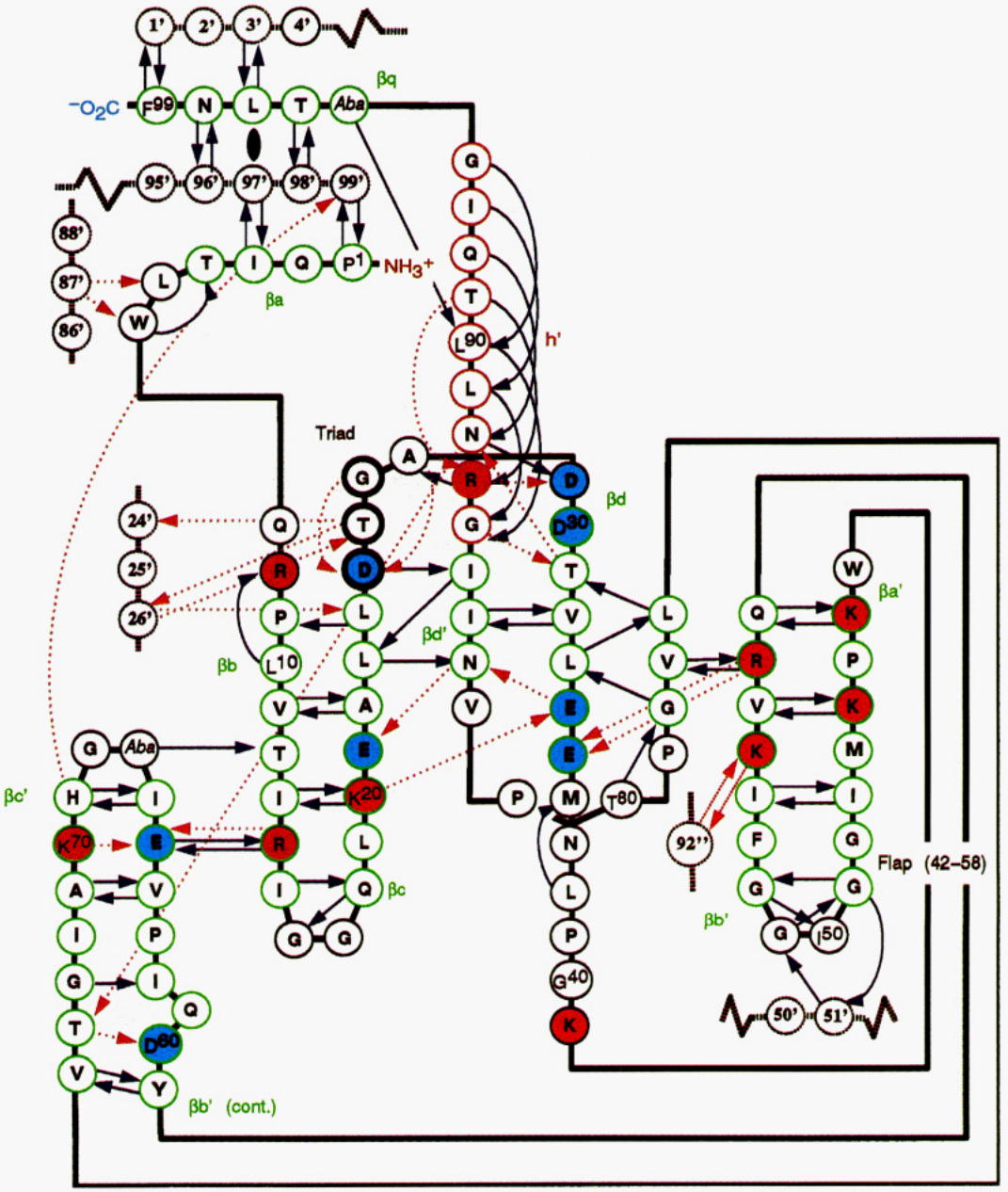


FIGURE 6: Schematic diagram of the primary sequence of one monomer of the HIV-1 protease showing secondary structure and hydrogen bond contacts observed in the crystal simulation [based on Figure 5 of Wlodawer et al. (1989)]. Solid lines (dark blue) indicate hydrogen bonds between backbone atoms, and broken lines (magenta) indicate hydrogen bonds involving side chains. Positively and negatively charged residues are colored red and blue, respectively. $\beta$ chains are outlined in green, and helical regions are outlined in magenta. Residues outlined in broken lines and labeled with primed numbers indicate neighboring monomers (single primes are used for the dimer-related monomer and double primes for neighboring dimers).

Sander program DSSP (Kabsch & Sander, 1983). Average structures from the MD, abbreviated ⟨MDC⟩, ⟨MDS⟩, and ⟨MDS$^+$⟩, were constructed by transforming each asymmetric unit to a local principal axis system and averaging the positions of corresponding atoms over the time interval 150–200 ps. The secondary structural assignment of the ⟨MDC⟩ structure shows close agreement to that of the crystallographic structure. Regions of the b (10–15), c (18–24), and d′ (84–85) $\beta$ chains and the h′ helix (87–93) are conserved in all the simulations. The solution structures differ in that partial melting of the $\beta$-sheet framework at the amino and carboxyl termini occurs. This is probably due to solvent end effects which have been observed in other simulations of HIV-1 PR in solution (Harte et al., 1992). Conversely, the average structures from the MD predict an extended $\beta$-sheet structure at residues 60–62 and 73–74.

(*iv*) *Solvation of the HIV-1 PR Crystal.* The HIV-1 PR unit cell (Figure 5) contains eight protein monomers (four dimers) and approximately 60% solvent. This solvent concentration is at the high end of the distribution normally observed for protein crystals. Nonetheless, several protein–protein contacts are observed between dimers (Table I). The only *inter*dimer contact of the hydrogen bond type occurs between Lys$^{55}$ of the flap region (residues 42–58) and Gln$^{92''}$ (double primes indicate residues of different dimers) of the h′ helix (residues 86–94). The remaining crystal packing contacts are of the van der Waals type.

In solvating the HIV-1 PR dimer, crystallographic *inter*dimer contacts are disrupted and replaced by solvent or *intra*dimer interactions. Figure 6 illustrates the hydrogen bond contacts observed during the last 50 ps of the crystal simulation. Tables II and III list the *inter*monomer and

Table II: Intermonomer Hydrogen Bonds in the HIV-1 PR Crystallographic Structure and in the Simulations[a]

| | | simulation (150–200 ps) | | | | |
|---|---|---|---|---|---|---|
| monomer 1 | monomer 2 | X-ray | MDC | MDS | MDS+ | interaction |
| $Pro^{1}$ | $Phe^{99'}$ | P | P+ | P+ | P+ | B ↔ B |
| $Gln^{2}$ | $Asn^{98'}$ | P | A | A | A | S → S |
| $Ile^{3}$ | $Leu^{97'}$ | P | P | P+ | P+ | B ↔ B |
| $Leu^{5}$ | $Arg^{87'}$ | P | P+ | P | P+ | B ← S |
| $Trp^{6}$ | $Arg^{87'}$ | A | P+ | P– | P– | B ← S |
| $Leu^{24}$ | $Thr^{26'}$ | P | P– | P+ | P | B ← S |
| $Thr^{26}$ | $Thr^{26'}$ | P | P+ | P+ | P+ | B → S, S ← B |
| $Gly^{49}$ | $Gly^{51'}$ | A | P– | P | P+ | B ← B |
| $Gly^{51}$ | $Gly^{51'}$ | A | P– | P+ | A | B ↔ B |
| $Gly^{51}$ | $Gly^{52'}$ | A | A | A | P | B ← B |
| $Lys^{55}$ [b] | $Gln^{92''}$ | P | P+ | A | A | B ← S |
| $Lys^{55}$ [b] | $Gln^{92''}$ | A | P+ | A | A | S → B |
| $His^{69}$ | $Phe^{99'}$ | P | P– | A | A | S → B |
| $Thr^{96}$ | $Asn^{98'}$ | P | P | P+ | P+ | B ← B |

[a] H-Bond interactions are designated A (absent: <10% observed) or P (present: P–, 10–40%; P, 40–70%; P+, 70–100%). The type of interaction is indicated as involving either backbone (B) or side-chain (S) atoms (arrows indicate the direction of the H-bond: donor → acceptor). [b] Dimer–dimer interaction.

Table III: Intramonomer Hydrogen Bonds in the HIV-1 PR Crystallographic Structure and in the Simulations[a]

| | | simulation (150–200 ps) | | | | |
|---|---|---|---|---|---|---|
| residue 1 | residue 2 | X-ray | MDC | MDS | MDS+ | interaction |
| $Gln^{2}$ | $Asn^{98}$ | A | A | A | P | S ← S |
| $Pro^{9}$ | $Leu^{24}$ | P | P | P+ | P | B ← B |
| $Val^{11}$ | $Ala^{22}$ | P | P+ | P+ | P+ | B ↔ B |
| $Thr^{12}$ | $Aba^{67}$ | A | P– | P– | P | B ← B |
| $Ile^{13}$ | $Lys^{20}$ | P | P+ | P+ | P+ | B ↔ B |
| $Arg^{14}$ | $Glu^{65}$ | P | P+ | P+ | P+ | B ↔ B, S → S |
| $Lys^{20}$ | $Glu^{34}$ | A | P | P+ | P+ | S → S |
| $Glu^{21}$ | $Asn^{83}$ | P | P– | A | A | B ← S |
| $Leu^{23}$ | $Asn^{83}$ | P | P+ | P+ | P+ | B → B |
| $Leu^{23}$ | $Ile^{85}$ | P | P | P+ | P+ | B ← B |
| $Asp^{25}$ | $Ile^{85}$ | P | P– | P– | P | B → B |
| $Ala^{28}$ | $Arg^{87}$ | P | P+ | P+ | P+ | B ← B |
| $Asp^{29}$ | $Arg^{87}$ | P | P– | P+ | P– | S ← S |
| $Asp^{29}$ | $Asn^{88}$ | P | P | P+ | P+ | B ← B |
| $Thr^{31}$ | $Leu^{76}$ | P | P+ | P+ | P+ | B ← B |
| $Thr^{31}$ | $Gly^{86}$ | P | P | P+ | P+ | S ← B |
| $Thr^{31}$ | $Asn^{88}$ | P | P+ | P+ | P+ | S → S |
| $Val^{32}$ | $Ile^{84}$ | P | P+ | P+ | P+ | B ↔ B |
| $Leu^{33}$ | $Leu^{76}$ | P | P | P+ | P+ | B → B |
| $Leu^{33}$ | $Gly^{78}$ | P | P | P+ | P+ | B ← B |
| $Glu^{34}$ | $Asn^{83}$ | P | P+ | P+ | P+ | B → S |
| $Glu^{35}$ | $Arg^{57}$ | P | P+ | P | P | S ← S |
| $Met^{36}$ | $Arg^{57}$ | A | A | P | P | B ← S |
| $Lys^{43}$ | $Gln^{58}$ | P | P+ | P+ | P+ | B ← B |
| $Lys^{45}$ | $Val^{56}$ | P | P+ | P+ | P+ | B ↔ B |
| $Lys^{47}$ | $Ile^{54}$ | P | P+ | P+ | P+ | B ↔ B |
| $Arg^{57}$ | $Val^{77}$ | P | P+ | P+ | P+ | B ↔ B |
| $Tyr^{59}$ | $Val^{75}$ | P | P+ | P+ | P+ | B ↔ B |
| $Asp^{60}$ | $Thr^{74}$ | P | P– | A | A | B ← S |
| $Ile^{62}$ | $Gly^{73}$ | P | P+ | P+ | P+ | B ← B |
| $Val^{64}$ | $Ala^{71}$ | P | P+ | P | P+ | B ↔ B |
| $Glu^{65}$ | $Lys^{70}$ | A | P | P | P | S ← S |
| $Ile^{72}$ | $Gln^{92}$ | A | A | P+ | P– | B ← S |
| $Thr^{74}$ | $Asn^{88}$ | A | P | P+ | P+ | B ← S |
| $Leu^{90}$ | $Aba^{95}$ | A | P+ | P | P+ | B ← B |

[a] H-Bond interactions are designated A (absent: <10% observed) or P (present: P–, 10–40%; P, 40–70%; P+, 70–100%). The type of interaction is indicated as involving either backbone (B) or side-chain (S) atoms (arrows indicate the direction of the H-bond: donor → acceptor). H-Bonds between residues more than four residues separated are listed.

*intra*monomer hydrogen bonds, respectively, observed in the crystallographic structure and in the last 50 ps of the simulations. Hydrogen bond interactions in the simulation were defined as follows: a hydrogen bond was initiated when the distance between hydrogen-bonding atom types (atom 1 = N or O; atom 2 = H attached covalently to N or O) became less than 2.2 Å and broken when the distance became greater than 2.4 Å.

Table IV: Intersubunit Hydrogen Bonding of the Flap Tips (Residues 48–52)[a]

| simulation | H-bond pair | % time | ⟨r⟩ (Å) | ⟨$E_{nb}$⟩ (kcal/mol) |
|---|---|---|---|---|
| MDC | $Gly^{49}$ ← $Gly^{51'}$ | 35.9 | 2.24 | –4.77 |
| | $Gly^{51}$ ← $Gly^{51'}$ | 31.1 | 2.16 | –4.99 |
| MDS | $Gly^{48}$ ← $Gly^{51'}$ | 13.3 | 2.65 | –3.91 |
| | $Gly^{49}$ ← $Gly^{51'}$ | 95.0 | 2.06 | –5.02 |
| | $Gly^{51}$ ← $Gly^{51'}$ | 100.0 | 1.99 | –5.16 |
| MDS+ | $Gly^{49}$ ← $Gly^{51'}$ | 100.0 | 1.91 | –5.37 |
| | $Gly^{49}$ → $Gly^{51'}$ | 100.0 | 1.93 | –5.33 |
| | $Gly^{51}$ ← $Gly^{52'}$ | 88.5 | 2.09 | –4.95 |

[a] H-bond interactions are denoted by the percentage time they were observed over the last 50 ps of the simulations (% time) and the corresponding average pairwise distance (⟨r⟩) and nonbonded energy (⟨$E_{nb}$⟩).

The primary *inter*monomer hydrogen bond contacts observed in the crystallographic structure are conserved in the simulations. These include interactions at the amino- and carboxyl-terminal β strands (residues 1–4, 95–99) which form a dove-tailed β-pleated sheet and in the region of the active site triads (residues 25–27) which interlock in the so-called "fireman's grip" characteristic of aspartyl proteases (Weber, 1990). These interactions are maintained in the simulations. The only dimer–dimer hydrogen bond interactions in the crystal occur between $Lys^{55}$ and $Gln^{92''}$. In the crystallographic structure these residues form two backbone–**side-chain** hydrogen bonds ($Lys^{55}$ HN → **$Gln^{92''}$ OE1**, $Lys^{55}$ O ← **$Gln^{92''}$ HNE2**; arrows indicate H-bond donors/acceptors and boldface type indicates residues in which side-chain atoms are involved in the H-bond). In the crystal simulation an additional hydrogen bond forms between **$Lys^{55}$ HNZ1** and $Gln^{92''}$ O. This dimer–dimer contact stabilizes the flaps in an extended conformation in the crystal away from the active site triads. On solvation of the HIV-1 PR dimer, this contact is lost, and local refolding of the protein occurs, resulting in rearrangement of the flaps.

Major *intra*monomer backbone–backbone hydrogen bonds in the crystallographic structure are maintained in the simulations. The molecular core of the protein contains four β strands arranged into two "ψ"-shaped structures characteristic of aspartyl proteases (Wlodawer et al., 1989; Blundell et al., 1985) and is stabilized by a network of backbone hydrogen bonds (Figure 6). In addition, hydrogen bond interactions involving side chains (shown by dotted lines) connect structural domains distant in primary sequence and help to stabilize the protein tertiary structure in the crystal. The **$Glu^{35}$-$Arg^{57}$**, **$Asp^{29}$-$Arg^{87}$**, and **$Arg^{14}$-$Glu^{65}$** contacts form salt bridge interactions. The ionic interaction between $Asp^{29}$ and $Arg^{87}$ is believed to be important for catalytic activity (Lapatto et al., 1989; Jaskolski et al, 1990; Weber, 1990) by stabilizing the $Asp^{29}$ carboxylate in an orientation which would allow it to mediate substrate binding. Mutational analysis shows that substitution at either of these positions inactivates the protease (Loeb et al., 1989). $Thr^{31}$ is a buried polar residue which forms a backbone–backbone H-bond with $Leu^{76}$ and also interacts with its side chain to form H-bonds with $Gly^{86}$ and $Asn^{88}$. These important structural contacts were all conserved in the crystal simulation.

Several contacts appear in the crystal simulation that were not apparent in the crystallographic structure. These include interactions at the dimer interface between the flap tips ($Gly^{49}$ ← $Gly^{51'}$, $Gly^{51}$ → $Gly^{51'}$) and formation of *intra*monomer

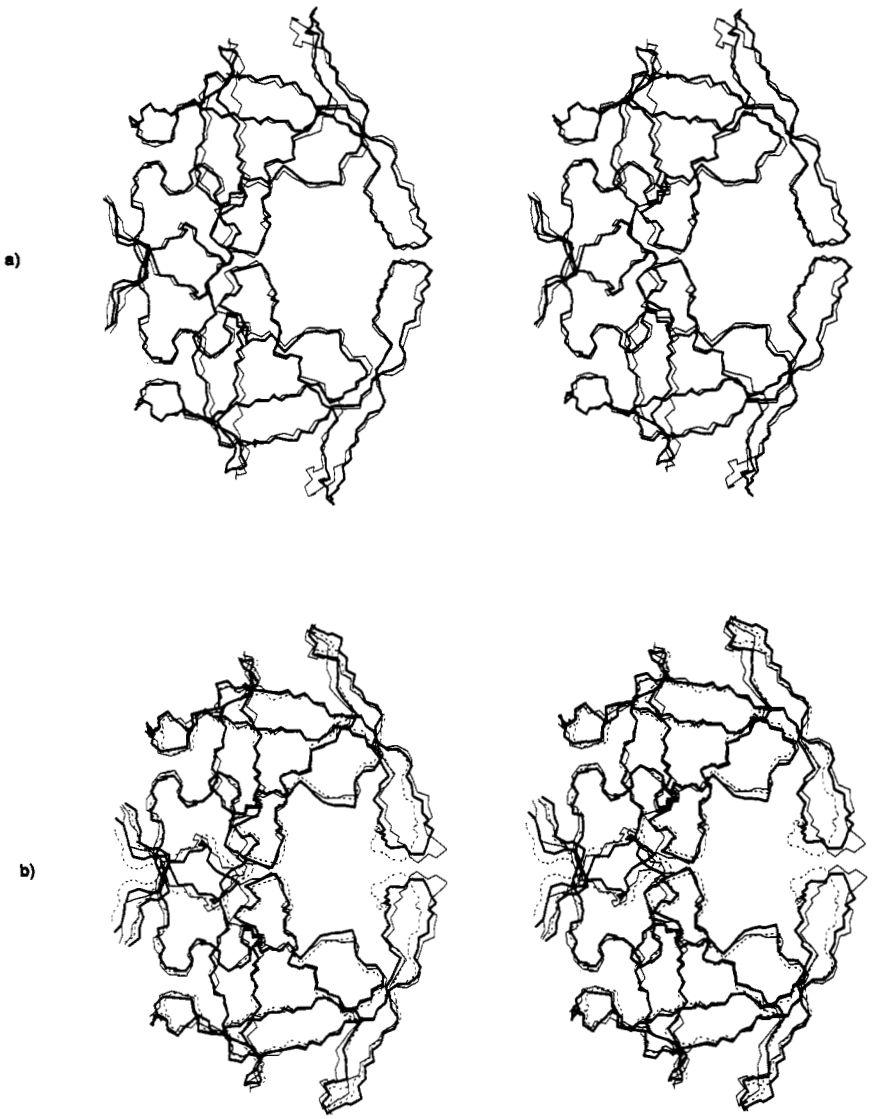


FIGURE 7: Backbone trace of the simulation average structures superimposed on the crystallographic structure of Wlodawer et al. (1989) (fine line): (a) ⟨MDC⟩ (bold); (b) ⟨MDS⟩ (bold) and ⟨MDS⁺⟩ (dotted). Simulation averages were obtained from 150 to 200 ps.

salt bridges between **Lys**$^{20}$ → **Glu**$^{34}$ and **Glu**$^{65}$ ← **Lys**$^{70}$. Alternately, the H-bond contact between side chains of **Gly**$^{2}$ → **Asn**$^{98'}$ was present in the crystallographic structure but is negligible in the simulations.

(*v*) *Changes at the Dimer Interface upon Solvation of the Crystal.* The largest difference observed in the *intra*dimer H-bonds between the crystalline and solution structures occurs in the region of the flap tips (Tables II and IV). In the crystallographic structure, the flaps are observed to be highly mobile as indicated by relatively large rms positional fluctuations (Figure 3). Gly$^{51}$ and Gly$^{51'}$ of the symmetry-related monomer are in close proximity (N–N distance 2.70 Å). It has been suggested that the flaps are stabilized by H-bonds between backbone atoms of these residues in the crystal (Wlodawer et al., 1989). The crystal simulation indicates that a weak interaction between these residues exists (Table IV). Similar interactions between Gly$^{49}$ ← Gly$^{51'}$ are also observed. In solution, reordering of the flaps occurs due to the loss of the dimer–dimer contacts between Lys$^{55}$ and Gln$^{92''}$ which stabilizes the flaps in an extended conformation in the crystal. As a result, the flaps relax into a more tightly coupled structure with increased intersubunit hydrogen bonding. In MDS significant coupling occurs between Gly$^{49}$ ← Gly$^{51'}$ and Gly$^{51}$ → Gly$^{51'}$. In MDS$^{+}$ strong coupling occurs between Gly$^{49}$ ↔ Gly$^{51'}$, which form two stable intersubunit hydrogen bonds, and between Gly$^{51}$ ← Gly$^{52'}$.

(*vi*) *Intramonomer Contacts That Change upon Solvation of the Crystal. Intra*monomer contacts which melt on solvation include backbone–side-chain H-bonds between Glu$^{21}$ ← **Asn**$^{83}$ and Asp$^{60}$ ← **Thr**$^{74}$. The side chain of Asn$^{83}$ can act as both a hydrogen bond donor and acceptor. In the crystallographic structure this side chain acts as an acceptor with the backbone HN of Glu$^{34}$ and as a donor with the backbone carbonyl of Glu$^{21}$. The former interaction was present in all the simulations whereas the latter was absent in the solution simulations. The interaction between Asp$^{60}$ ← **Thr**$^{74}$ connects the b′ and c′ $\beta$ chains located on the surface of the protein. These residues are in close proximity to crystal packing contacts in the crystallographic structure (Gly$^{49}$-Gln$^{61''}$ and Ile$^{54}$-Ile$^{72''}$) with dimer–dimer (heavy-atom) distances less than 3.5 Å (Table I). *Intra*monomer hydrogen bonds that develop in solution include backbone–side-chain interactions between Met$^{36}$ ← **Arg**$^{57}$ and Ile$^{72}$ ← **Gln**$^{92}$. These interactions arise from rearrangement due to loss of the dimer–dimer crystal packing contacts between Lys$^{55}$ and Gln$^{92''}$ (Tables I and II).

(*vii*) *Effect of Solvation on Backbone Structure.* Early refolding of the protease backbone can be examined by comparing the simulation average structures to the crystal-

Table V: rms Deviation of the 150–200-ps Simulation Average Structures from the Crystallographic Structure

| atom set | no. of atoms | rms (Å) ⟨MDC⟩ | ⟨MDS⟩ | ⟨MDS⁺⟩ |
|---|---|---|---|---|
| α carbons | 99 | 0.94 | 0.97 | 1.24 |
| backbone N–Cα–C | 297 | 0.91 | 0.94 | 1.30 |
| heavy atoms (non-hydrogens) | 762 | 1.22 | 1.30 | 1.45 |
| α carbons, neglect flap (42–58) | 82 | 0.93 | 0.85 | 0.85 |
| α carbons, neglect loops (15–19, 36–42) | 87 | 0.74 | 0.93 | 1.24 |
| α carbons, neglect contact regions (4–8, 40–44, 47–59, 70–74, 77–81, 90–94) | 61 | 0.99 | 0.74 | 0.91 |

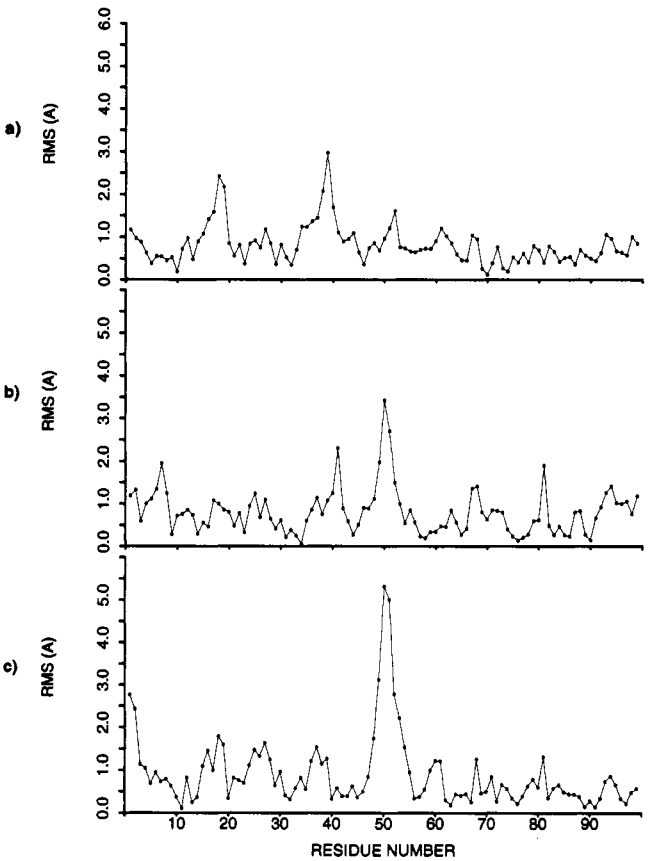


FIGURE 8: rms α-carbon deviation of the simulation average structures from the crystallographic structure (Wlodawer et al., 1989): (a) ⟨MDC⟩; (b) ⟨MDS⟩; (c) ⟨MDS⁺⟩.

lographic structure (Figure 7). Table V compares the corresponding rms deviation for several atom sets. The ⟨MDC⟩ structure has the lowest overall α-carbon rms (0.94 Å), followed by ⟨MDS⟩ (0.97 Å) and ⟨MDS⁺⟩ (1.24 Å). Figure 8 shows the α-carbon rms plotted as a function of residue number for each comparison. ⟨MDC⟩ shows its largest deviation in segments involving loops or turns (residues 15–19, 36–42), and these are observed in both the MD and the crystallographic structure to have large fluctuation (Figure 3). The largest deviations in the solution structures (⟨MDS⟩, ⟨MDS⁺⟩) occur in the flap regions (42–58), particularly at $Gly^{50}$ and $Gly^{51}$ at the tips. This change in backbone structure accompanies the rearrangement resulting from loss of the dimer–dimer contact between $Lys^{55}$ and $Gln^{92''}$. In general, regions of large backbone deviation correlate with corresponding regions of large fluctuation (Figure 3).

*(viii) Effect of Counterions on the Structure of the Flaps in Solution.* The flap regions contain significant positive charge located on residues 43, 45, 55, and 57 (Figure 6). Not surprisingly, the behavior of the flaps in solution was sensitive to the presence of charge-balancing counterions. The ⟨MDS⁺⟩ structure shows increased α-carbon rms (maximum = 5.3 Å) in the flap region (Figure 8) relative to the ⟨MDS⟩ structure (maximum = 3.4 Å). The flaps rearrange from the crystal geometry by contracting inward toward the active site and forming greater intersubunit overlap at the flap tips (Table VI). This motion was accompanied by adjustment of the χ torsion angle of $Lys^{55}$, resulting in movement of its side chain farther away from other positively charged residues at the base of the flap ($Lys^{41}$, $Lys^{43}$). In the neutral simulation (MDS), $Cl^-$ ions partially neutralized the flaps and alleviated some of the interresidue repulsions. Consequently, the flaps remained more extended (crystal-like) with less overlap of the tips. Slightly weaker intersubunit hydrogen bonding was observed with the flaps in the extended position (Table IV). In order for a peptide substrate to enter the active site, presumably this type of intersubunit hydrogen bonding must break and the flaps separate.

Table VI: Overlap and Contraction of the Flap Tips (Residues 48–52) in the Simulations[a]

| simulation | overlap (Å$^{-1}$)[b] | contraction (Å)[c] |
|---|---|---|
| MDC | 0.127 | 24.74 |
| MDS | 0.148 | 20.46 |
| MDS⁺ | 0.156 | 19.27 |

[a] Time-average structures for each dimer were obtained from the MD between 150 and 200 ps and used to compute the average overlap and contraction in each simulation. The backbone atoms (N, HN, Cα, C, and O) of the flap tips were used in calculating overlap and contraction. [b] Overlap = $1/S$; $S = (1/n^2)\sum_{i=1}\sum_{j=1} r_{ij}$, where $r_{ij}$ is the distance between atom $i$ of the flap tip of monomer 1 and atom $j$ of the flap tip of monomer 2 in the dimer. The indices $i$ and $j$ sum over the $n$ ($n$ = 25) backbone (N, C, Cα, HN, O) atoms of each flap tip. [c] Contraction = $(1/2n)(\sum_{i=1}R_i + \sum_{j=1}R_j)$, where $R_i = |\mathbf{r}_i - \mathbf{r}_0|$ ($\mathbf{r}_i$ is the position vector of atom $i$ and $\mathbf{r}_0$ is the position vector of the geometric center of the dimer). The indices $i$ and $j$ sum over the $n$ ($n$ = 25) atoms of the flap tips in monomer 1 and monomer 2, respectively.

The flexibility of the flaps may play an important role in substrate binding. Examination of the crystal structures of HIV-1 PR in the unbound and inhibitor-bound states shows that a large rearrangement (~7 Å) of the flaps occurs upon inhibitor binding. It has been estimated that the flaps must adjust by approximately 15 Å to allow a polypeptide substrate into the active site (Gustchina & Weber, 1990). It is possible that conditions which enhance the uncoupling of the flaps in solution might more readily allow a substrate into the active site cleft of the enzyme. It has been observed experimentally that the Michaelis–Menton constants for HIV-1 PR cleavage of peptide substrates are sensitive to salt concentration (Wondrak et al., 1991). With an increase in certain salt concentrations, the protease activity increases while $K_m$ [$K_m = k_2/(k_1 + k_{cat})$] decreases and $k_{cat}$ is unaffected. These data imply that $k_1/k_2$ increases with salt and, taken together with the MDS simulation result (weaker intersubunit H-bonding and less overlap of the flaps in the presence of counterions), provides a consistent view of a requirement for substrate binding.

*(ix) Structural Waters in the Crystal and in Solution.* It is an open question as to the degree in which water stabilizes protein structure. Previous studies of the HIV-1 protease in solution and in vacuo suggest water plays a significant role in stabilizing the structure (Harte et al., 1992). Furthermore, defined water structure might be expected to differ in crystalline and solution environment. Positions of crystallographic waters were not available in the structure reported by Wlodawer et al. (1989); however, several of the X-ray

Table VII: Intramonomer Water Bridges in the HIV-1 PR Inhibitor-Bound Crystallographic Structures and in the Simulations[a]

| | Inhibitor-bound Crystal Structures[b] | | | | | Simulations | | |
|---|---|---|---|---|---|---|---|---|
| Res. no. | MV 12 | U8 12 | JG 12 | L7 12 | AC 12 | MDC 12345678 | MDS 12 | MDS+ 12 |
| **A) Crystallographic waters:** | | | | | | | | |
| 4-6 | | | | XX | XX | *...**X* | .* | XX |
| 4-7 | | | | XX | XX | **XXX*X* | XX | XX |
| 6-7 | | | | *X | XX | **. **XX | .* | XX |
| 12-21 | | | X. | | .* | X*.**.*. | .* | *X |
| 12-67 | | | | | *X | ..... | X* | .. |
| 16-63 | | | *X | .X | XX | XXXXX.*X | XX | X. |
| 18-37 | | | | | .X | .....*. | .. | .. |
| 20-34 | | | | XX | XX | .*X*.**X | .* | X. |
| 20-35 | | | | .X | .X | .*XX..X* | X. | .X |
| 20-83 | | | .X | .* | XX | *...*..* | .* | .. |
| 21-83 | .X | X. | | | XX | X*.X.X*X | X* | *X |
| 26-87 | X. | X. | XX | XX | XX | *...*..* | . | ** |
| 27-29 | X. | XX | .X | XX | XX | •.....*. | *. | .. |
| 30-31 | .* | | X. | | .* | XX..X.XX | .X | .X |
| 30-45 | | | | XX | .X | ...*X*.* | .. | .. |
| 30-74 | .* | X. | X. | | ** | .XX**XXX | .* | X. |
| 30-88 | | XX | X. | | .* | .X.***XX | .. | .. |
| 31-74 | XX | | XX | XX | .X | .X..*.*X | X. | .. |
| 31-88 | *. | | X* | .* | .* | .X..*..X | *. | .. |
| 34-80 | | | .X | | .* | •.....X | *. | .. |
| 34-83 | | | | | .X | ..**..** | .* | *. |
| 35-36 | | | | | .X | XXXX.X.X | .. | .. |
| 35-37 | | | | | .X | ..... | .. | .. |
| 35-57 | | | | X* | X. | X.*XXX.X | .X | .. |
| 35-77 | | | | X. | X. | ..... | .. | .. |
| 35-78 | | | X. | *. | .X | •.....*X | .. | .. |
| 36-37 | | | | | .X | ....*... | .. | .. |
| 37-38 | | | | | .X | X*****.. | ** | ** |
| 38-40 | | | | X. | .X | ** | • | ** |
| 38-59 | X. | .X | | | .X | .*.**X.* | XX | X* |
| 39-60 | | | | | .X | ..... | .. | .. |
| 41-42 | | | | | X. | ...*X.XX | X* | X. |
| 41-43 | XX | | | | | .....• | .* | X. |
| 41-60 | X* | ** | XX | XX | XX | *XX.XX*X | X. | XX |
| 43-58 | | X. | | | | *..X.**. | .. | .. |
| 43-60 | X* | X* | | | .X | ..... | .. | .. |
| 57-77 | X. | | | X. | XX | ..... | X. | .X |
| 58-60 | | X. | | XX | X. | .*...**. | .• | *. |
| 61-72 | | | | | .X | .....X.* | .. | .* |
| 61-73 | | | | | X. | ..... | .. | .. |
| 61-74 | X* | | | XX | XX | X..... | X. | .. |
| 65-67 | .X | X* | X. | .* | | XX....X. | .* | .. |
| 65-68 | .X | X. | X. | .* | X. | **XXXX.X | XX | *X |
| 67-68 | .X | X. | | .* | | ..... | .* | .. |
| 70-92 | | | | | .X | ..... | .. | .. |
| 72-92 | | | | X. | X. | X.XXX*X. | *X | ** |
| 73-74 | | | | | X. | ..... | .. | .. |
| 74-88 | ** | X. | XX | XX | XX[c] | XXXX*XXX | XX | XX |
| 88-92 | | | X. | .X | .X | .****..X | •. | .. |
| 94-96 | .* | | | XX | .X | *******X | ** | X* |
| **B) Non-crystallographic waters:** | | | | | | | | |
| 6-7 | | | | | | **..**XX | .* | XX |
| 14-17 | | | | | | *.X**.XX | X. | ** |
| 15-18 | | | | | | XX...X*. | .. | .* |
| 15-20 | | | | | | XXX*.... | .. | X. |
| 17-65 | | | | | | *.X..... | .X | .X |
| 25-28 | | | | | | ...X.X.X | .. | .. |
| 26-27 | | | | | | *..X*..* | XX | ** |
| 27-87 | | | | | | *XX.X.X* | .. | *X |
| 29-30 | | | | | | XXXXXXX* | XX | XX |
| 29-87 | | | | | | ***.**** | .. | XX |
| 30-58 | | | | | | ..*XXX*X | XX | X* |
| 34-35 | | | | | | XXX*XX. | .X | X. |
| 38-57 | | | | | | ....X*X. | .X | XX |
| 45-56 | | | | | | ....*..* | .. | XX |
| 57-59 | | | | | | XXXXX*.X | .X | X. |
| 60-61 | | | | | | **.XXXXX | XX | ** |
| 60-74 | | | | | | .X*XXX** | X* | XX |
| 64-71 | | | | | | XX..X**X | .. | *. |
| 65-70 | | | | | | X.XX*..* | .. | .* |
| 66-69 | | | | | | .•..XXX. | .. | .. |
| 91-92 | | | | | | .X.XX**. | ** | *. |
| 98-99 | | | | | | .X****** | XX | X. |

[a] Bridges are designated X, *, or ., indicating strong, weak, and negligible bridging interactions, respectively. Strong bridge (X): in crystallographic structures, protein–water–protein H-bond (heavy atom) distances <3.3 Å; in simulations, protein–water–protein H-bonds maintained more than 40% of the time over the last 50 ps. Weak bridge (*): in crystallographic structures, protein–water–protein H-bond (heavy atom) distances >3.3 and <3.5 Å; in simulations, protein–water–protein H-bonds maintained >10% and <40% of the time over the last 50 ps. Subunits in the dimers are labeled 12; subunits in the crystal cell are labeled 12345678. [b] MV, 2.3-Å resolution (Miller et al., 1989); U8, 2.5-Å resolution (U-85548E, Jaskolski et al., 1990); JG, 2.4-Å resolution (JG-365, Swain et al., 1990); L7, 2.1-Å resolution (L-700, Bone et al., 1991); AC, 2.0-Å resolution (acetyl pepstatin, Fitzgerald et al., 1990). [c] Chloride ions.

structures of the inhibitor-bound protease do contain crystallographic waters (Miller et al., 1989b; Jaskolski et al., 1991; Swain et al., 1990; Bone et al., 1991; Fitzgerald et al. 1990). Tables VII and VIII list *intra*monomer and *inter*monomer residue pairs, respectively, in which structural waters are involved in bridging for the crystallographic and simulation structures.

Significant water bridging is observed in the crystallographic structures within monomers between residues 4–6/7, 16–63, 20–34, 21–83, 26–87, 27–29, 30/31–74, 41–60, 61–74, 65–

Table VIII: Intermonomer Water Bridges in the HIV-1 PR Inhibitor-Bound Crystallographic Structures and in the Simulations[a]

| | Inhibitor-bound Crystal Structures[b] | | | | | Simulations | | |
|---|---|---|---|---|---|---|---|---|
| Res. no. | MV 12 | U8 12 | JG 12 | L7 12 | AC 12 | MDC 12345678 | MDS 12 | MDS+ 12 |
| **A) Crystallographic waters:** | | | | | | | | |
| 5-95' | XX | | XX | XX | XX | .X.X...* | X. | .. |
| 6-87' | X. | | | | XX | *....*.. | ** | XX |
| 8-27' | | | | XX | X. | ..*X.*.. | *. | .. |
| 8-29' | | | | XX | X. | ...*.... | .. | X* |
| 50-50' | .* | | X. | X. | X. | ........ | .. | .. |
| 50-79' | | | | | X. | ........ | .. | .. |
| 51-51' | | | | | | ..*..... | .. | .. |
| 51-53' | .* | X. | X. | X. | | ........ | .. | .. |
| 51-79' | .X | | .X | X. | X. | ........ | .. | .. |
| 93-99' | | | .X | XX | .X | .X.X.... | ** | .. |
| 94-98' | .* | | | XX | .* | .****..X | *. | *. |
| 96-98' | .X | | | XX | .* | .**X.*.X | *. | ** |
| **B) Non-crystallographic waters:** | | | | | | | | |
| 3-97' | | | | | | .*XX... | .. | .. |
| 5-91' | | | | | | .*X....X | X* | .. |
| 5-87' | | | | | | ........ | X. | *X |
| 6-35"[c] | | | | | | *...X... | .. | .. |
| 8-26' | | | | | | .X..... | .. | .. |
| 24-26' | | | | | | .X.X*... | .. | X. |
| 25-25' | | | | | | .XXXXX.X | XX | XX |
| 25-26' | | | | | | ...X.X.* | .. | .. |
| 25-27' | | | | | | .....X.. | .. | .. |
| 26-27' | | | | | | .....X.. | .. | .. |
| 48-50' | | | | | | ........ | .X | .. |
| 50-60"[c] | | | | | | ..X..... | .. | .. |
| 50-61"[c] | | | | | | X***X..* | .. | .. |
| 51-52' | | | | | | ..X..*.. | .. | .. |
| 53-72"[c] | | | | | | XXXXXXX. | .. | .. |
| 53-92"[c] | | | | | | *.XXX.X. | .. | .. |
| 55-91"[c] | | | | | | *.X..X.. | .. | .. |

[a] Bridges are designated X, *, or ., indicating strong, weak, and negligible bridging interactions, respectively. Subunits in the dimers are labeled 12; subunits in the crystal cell are labeled 12345678. [b] MV, 2.3-Å resolution (MVT-101, Miller et al., 1989); U8, 2.5-Å resolution (U-85548E, Jaskolski et al., 1990); JG, 2.4-Å resolution (JG-365, Swain et al., 1990); L7, 2.1-Å resolution (L-700, Bone et al., 1991); AC, 2.0-Å resolution (acetyl pepstatin, Fitzgerald et al., 1990). [c] Dimer–dimer water bridges.

67/68, and 74–88 (charged residues are underlined) and at the dimer interface between residues 5–95′, 50–50′, 51–53′, 51–79′, and 93–99′. Similar water bridges are also observed in the crystal simulation. Exceptions occur for bridges that involve residues of the active site or the flaps. Solvation of the active site is expected to be different in the crystallographic structures since this region is occupied by an inhibitor. Similarly, water bridges between subunits might be expected to be different in the flaps of the dimer, particularly at the flap tips (48–52), since this region is observed to undergo significant rearrangement in the inhibitor-bound state relative to the unbound state. In addition to the water bridges observed in the crystallographic structures, *intra*monomer water bridges in the crystal simulation also occur between residues 14–63/65, 20–35, 27–87, 30–58, 30–88, 35–57, 34/36/38–59, 60–74, 64–71, 65–70, and 69–92. It is clear that ionic residues (underlined) play a large role in forming stable water bridges in the simulations. These waters mainly bridge regions of secondary structure in the crystal simulation (usually grooves formed between $\beta$ chains). Structural waters at the dimer interface in the crystal simulation occur at the amino and carboxyl termini (3–97′, 5–95′, 93–99′, and 94/96–98′). Water structure in this region is significantly reduced upon solvation. In the crystal simulation, dimer–dimer water bridges were observed between residues 50–61″, 53–72″, 53–92″, and 55–91″.

An interesting feature observed in the simulations concerns the solvation of the active site. Structural waters in all the simulations are observed to bridge the catalytic aspartate residues ($Asp^{25}$-$Asp^{25'}$, Table VIII). Density for a similar water was observed in the crystallographic structure of the unbound HIV-1 protease (Wlodawer et al., 1989) and the protease from Rous Sarcoma virus (Miller et al., 1989a; Jaskolski et al., 1990).

## CONCLUSION

Simulations of the HIV-1 protease have been performed in a crystalline environment and in solution in order to examine the effects of solvation on the protein structure. Of particular interest to us was the examination of the structure of the protease "flaps", which have been observed crystallographically to undergo significant rearrangement upon inhibitor binding (Gustchina & Weber, 1990). The flaps presumably regulate entry of substrates and inhibitors into the active site; hence the understanding of their behavior in solution is of great importance. The results for the crystalline simulation (MDC) are in close agreement with crystallographic data, lending credence to the methodology and providing a basis for which the solution simulations (MDS, MDS+) could be compared. In the crystalline state, the orientations of the flaps were found to be stabilized in an extended conformation away from the active site triads by a crystal packing contact between $Lys^{55}$ and $Gln^{92''}$ of neighboring dimers in the unit cell. The flaps in the crystal interact within the dimer through weak intersubunit hydrogen bonds formed at the tips. In solution, the crystal packing contacts are lost, and the flaps relax into a more contracted conformation with increased overlap and intersubunit hydrogen bonding. The flap rearrangement, however, was observed to be sensitive to the presence of charge-balancing counterions in solution. In the presence of the counterions, the flaps, which are highly charged, exhibited less contraction and overlap and weaker intersubunit hydrogen bonding. These observations are consistent with kinetic data which indicate that high salt stabilizes the association of the enzyme and substrate (Wondrak et al, 1991). Taken together, these results suggest that charged species in solution may play a role in the regulation of substrate/inhibitor binding through interaction with the flaps.

## ACKNOWLEDGMENT

We thank Dr. A. Wlodawer for providing us with several crystallographic structures and for useful comments. We acknowledge grants of supercomputer time from the National Cancer Institute and the North Carolina Supercomputing Center. D.M.Y. thanks the University of North Carolina Medical School for support through the M.D./Ph.D. program. L.G.P. thanks the NIEHS, RTP, NC, for support for this project and acknowledges NIH Grant HL27995. We also thank Howard Smith at NIEHS for essential computer graphics assistance.

## REFERENCES

Baldwin, E. T., Weber, I. T., St. Charles, R., Xuan, J.-C., Appela, E., Yamda, M., Matsushima, K., Edwards, B. F. P., Clore, G. M., Groneborn, A. M., & Wlodawer, A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 502–506.

Bax, A. (1989) *Annu. Rev. Biochem. 58*, 223–256.
Blundell, T. L., Jenkins, J., Pearle, L., Sewell, T., & Pedersen, V. (1985) in *Aspartic Proteinases and Their Inhibitors* (Kostka, V., Ed.) pp 151–161, Walter de Gruyter, Berlin.
Bone, R., Vacca, J. P., Anderson, P. S., & Holloway, M. K. (1991) *J. Am. Chem. Soc. 113*, 9382–9384.
Darke, P. L., Leu, C. T., Davis, L. J., Heimbach, J. C., Diehl, R. E., Hill, W. S., Dixon, R. A., & Sigal, I. S. (1989) *J. Biol. Chem. 264*, 2307–2312.
Debouck, C. (1992) *AIDS Res. Hum. Retroviruses 8*, 153–164.
Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W., & Rosenberg, M. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8903–8906.
Erickson, J., Neidhart, D. J., VanDrie, J., Kempf, D. J., Wang, X. C., Norbeck, D. W., Plattner, J. J., Rittenhouse, J. W., Turon, M., Wildeburg, N., Kohlbrenner, W. E., Simmer, R., Helfrich, R., Paul, D. A., & Knigge, M. (1990) *Science 249*, 527–533.
Ferguson, D. M., Randall, J. R., & Kollman, P. A. (1991) *J. Med. Chem. 34*, 2654–2659.
Fitzgerald, P. M., McKeever, B. M., VanMiddlesworth, J. F., Springer, J. P., Heimbach, J. C., Leu, C. T., Herber, W. K., Dixon, R. A., & Darke, P. L. (1990) *J. Biol. Chem. 265*, 14209–14219.
Foley, C. K., Pedersen, L. G., Charifson, P. S., Darden, T. A. *Wittinghofer*, A., Pai, E. F., & Anderson, M. W. (1992) *Biochemistry 31*, 4951–4959.
Graves, M. C., Meidel, M. C., Pan, Y. C., Manneberg, M., Lahm, H. W., & Gruninger-Leitch, F. (1990) *Biochem. Biophys. Res. Commun. 168*, 30–36.
Gustchina, A., & Weber, I. T. (1990) *FEBS Lett. 269*, 269–272.
Harte, W. E., Jr., Swaminathan, S., Mansuri, M. M., Martin, J. C., Rosenberg, I. E., & Beveridge, D. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8864–8868.
Harte, W. E., Jr., Swaminathan, S., & Beveridge, D. L. (1992) *Proteins 13*, 175–194.
Henderson, L. E., Copeland, T. D., Sowder, R. C., Schultz, A. M., & Oroszlan, S. (1988) in *Human Retroviruses, Cancer, and AIDS: Approaches to Prevention and Therapy*, pp 135–147, Alan R. Liss, Inc., New York.
Hyland, L. J., Tomaszek, T. A., & Meek, T. D. (1991) *Biochemistry 30*, 8454–8463.
Ichiye, T., & Karplus, M. (1987) *Proteins 2*, 236–259.
Jaskolski, M., Miller, M., Mohana Rao, J. K., Leis, J., & Wlodawer, A. (1990) *Biochemistry 29*, 5889–5898.
Jaskolski, M., Tomasselli, A. G., Sawyer, T. K., Staples, D. G., Heinrikson, R. L., Schneider, J., Kent, S. B., & Wlodawer, A. (1991) *Biochemistry 30*, 1600–1609.
Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983) *J. Chem. Phys. 79*, 926–935.
Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.
Karplus, M., & Petsko, G. (1990) *Nature 347*, 631–639.
Koehler, J. E., Saenger, W., & van Gunsteren, W. F. (1987) *Eur. Biophys. J. 15*, 197–210.
Kohl, N. E., Emini, E. A., Scheif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M., & Sigal, I. S. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4686–4690.
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. (1989) *Nature 342*, 299–302.
Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E., & Hutchison, C. A., III (1989) *Nature 340*, 397–400.
Loncharich, R. J., & Brooks, B. R. (1989) *Proteins 6*, 32–45.
Lybrand, T., McCammon, J., & Wipff, G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 833–835.
McCammon, J. A., & Harvey, S. (1987) in *Dynamics of Proteins and Nucleic Acids*, pp 176–180, Cambridge University Press, Cambridge.
Meek, T. D., Dayton, B. D., Metcalf, B. W., Dreyer, G. B., Strickler, J. E., Gorniak, J. G., Rosenberg, M., Moore, M., Magaard, V. W., & Debouck, C. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1841–1845.
Miller, M., Jaskolski, M., Rao, J. K. M., Leis, J., & Wlodawer, A. (1989a) *Nature 337*, 576–579.
Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshal, G. R., Clawson, L., Selk, L., Kent, S. B., & Wlodawer, A. (1989b) *Science 246*, 1149–1152.
Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C.-T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., & Springer, J. P. (1989) *Nature 37*, 615–620.
Nilsson, O., Tapia, O., & van Gunsteren, W. F. (1990) *Biochem. Biophys. Res. Commun. 171*, 581–588.
Rao, B. G., Tilton, R. F., & Singh, U. C. (1992) *J. Am. Chem. Soc. 114*, 4447–4452.
Reddy, R. R., Viswanadhan, V. N., & Weinstein, J. N. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10287–10291.
Richards, A. D., Roberts, R., Dunn, B. M., Graves, M. C., & Kay, J. (1989) *FEBS Lett. 247*, 113–117.
Sali, A., Veerapandian, B., Cooper, J. B., Moss, D. S., Hofmann, T., & Blundell, T. L. (1992) *Proteins 12*, 158–170.
Schreiber, H., & Steinhauser, O. (1992) *Biochemistry 31*, 5856–5860.
Seelmeir, S., Schmidt, H., Turk, V., & von der Helm, K. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6612–6616.
Smith, P. E., & Pettitt, B. M. (1991) *J. Chem. Phys. 95*, 8430–8441.
Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Schneider, J., Kent, S. B., & Wlodawer, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8805–8809.
Swaminathan, D., Harte, W. E., Jr., & Beveridge, D. L. (1991) *J. Am. Chem. Soc. 113*, 2717–2721.
Toh, H., Kikuno, R., Hayashida, H., Miyata, T., Kugimiya, W., Inouye, S., Yuki, S., & Saigo, K. (1985) *EMBO J. 4*, 1267–1272.
Weber, I. T. (1990) *J. Biol. Chem. 265*, 10492–10496.
Weber, I. T., Miller, M., Jaskolski, M., Leis, J., Skalka, A. M., & Wlodawer, A. (1989) *Science 243*, 928–931.
Weiner, S. J., & Kollman, P. A. (1986) *J. Comput. Chem. 7*, 230–252.
Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Chio, C., Alagona, G., Profeta, S., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765–784.
Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., & Kent, S. B. H. (1989) *Science 245*, 616–621.
Wondrak, E., Louis, J., & Oroszlan, S. (1991) *FEBS Lett. 280*, 344–346.