PROTEINS: Structure, Function, and Genetics 13:175–194 (1992)

Molecular Dynamics of HIV-1 Protease

William E. Harte, Jr.,[1,2] S. Swaminathan,[2] and David L. Beveridge[1]
[1]Chemistry Department, Hall-Atwater Laboratories, Wesleyan University, Middletown 06457 and [2]Bristol-Myers Squibb Pharmaceutical Research Institute, Wallingford, Connecticut 06492

ABSTRACT **Molecular dynamics simulations have been carried out based on the GROMOS force field on the aspartyl protease (PR) of the human immunodeficiency virus HIV-1. The principal simulation treats the HIV-1 PR dimer and 6990 water molecules in a hexagonal prism cell under periodic boundary conditions and was carried out for a trajectory of 100 psec. Corresponding in vacuo simulations, i.e., treating the isolated protein without solvent, were carried out to study the influence of solvent on the simulation. The results indicate that including waters explicitly in the simulation results in a model considerably closer to the crystal structure than when solvent is neglected. Detailed conformational and helicoidal analysis was performed on the solvated form to determine the exact nature of the dynamical model and the exact points of agreement and disagreement with the crystal structure. The calculated dynamical model was further elucidated by means of studies of the time evolution of the cross-correlation coefficients for atomic displacements of the atoms comprising the protein backbone. The cross-correlation analysis revealed significant aspects of structure originating uniquely in the dynamical motions of the molecule. In particular, an unanticipated through-space, domain-domain correlation was found between the mobile flap region covering the active site and a remote regions of the structure, which collectively act somewhat like a molecular cantilever. The significance of these results is discussed with respect to the inactivation of the protease by site-specific mutagenesis, and in the design of inhibitors.**

Key words: AIDS, HIV-1 protease, molecular dynamics, atomic fluctuations, domain communication

INTRODUCTION

HIV-1 protease (HIV-1 PR) is a 198-residue protein dimer that is vitally important to polyprotein processing in the lifecycle of the AIDS virus.[1] This enzyme is currently one of the few macromolecular components of the AIDS virus for which detailed structural information is available[2–4] and a target of considerable pharmaceutical interest in the quest for AIDS therapies.[5] A current structural model for HIV-1 PR developed from crystallography,[3] depicted in Figure 1, has firmly established the essential secondary and tertiary structure of the enzyme. The observed isotropic temperature factors indicate certain dynamical elements in the structure, particularly the flexibility in the flap regions opposite the active site Asp-Thr-Gly triads. The recent structure of an HIV-1 PR inhibitor complex[6] provides evidence of flap motion and other substantial changes in the enzyme. The presence of mobile domains in the protease is just one indication of the need, in developing a detailed understanding of HIV-1 PR action, for consideration of the structural dynamics.

We describe herein a theoretical model for HIV-1 protease dimer in dilute aqueous solution developed from molecular dynamics (MD) simulation and report the properties of secondary and tertiary structural elements in the enzyme emerging from a 100-psec dynamical trajectory. Analysis of the results reveals new knowledge about the structure and functional energetics that originates uniquely in the dynamical motions, and in particular, evidence for contact and through-space correlations between domains of well-defined secondary structure. Also discussed is the extent to which the dynamical structure of HIV-1 PR plays a significant role in the functional energetics of the enzyme.

A preliminary account of this work has recently been reported.[7,8] We provide herein a full and complete description of the calculation and present new details on (1) the conformational and helicoidal dynamics, (2) the nature and time evolution of domain structure and interactions as measured by dynamical cross correlation coefficients, and (3) comparison of the results with a corresponding MD simulation neglecting the solvent. On the basis of (3), the role of solvent in the protein structure and domain correlations is discussed in detail and provides new perspectives on the role of water in protein stability and the extent to which water is necessary in an MD simulation to obtain an accurate dynamical model.

Received June 24, 1991; revision accepted September 24, 1991.
Address reprint requests to Dr. David L. Beveridge, Department of Chemistry, Hall-Atwater Laboratories, Wesleyan University, Middletown, CT 06459.

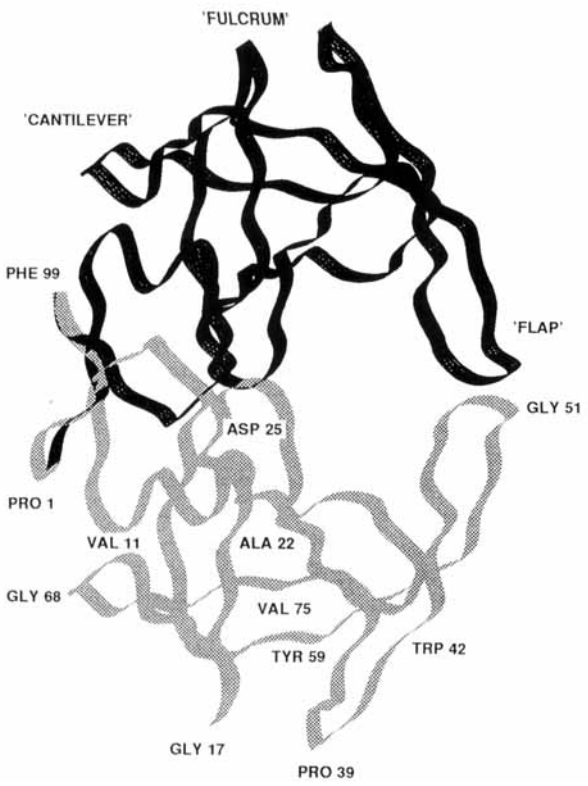

Fig. 1. Ribbon tracing of the x-ray crystal structure of HIV-1 PR dimer as reported by Wlodawer et al.[3] Names of regions and amino acid sequence numbers important for the discussion of dynamical structure in the protein are indicated.

BACKGROUND

The replication pathway of the AIDS virus (HIV-1) involves a number of steps susceptible to possible therapeutic intervention.[5,9] At this point, one of the few components of the virus for which detailed information is available at the level of actual molecular structure is the virus encoded HIV-1 PR.[2–4] (The structure of the ribonuclease H domain of the viral reverse transcriptase[10] and the structure of a portion of the cell surface receptor protein CD4[11,12] have also recently been reported.) HIV-1 PR is responsible for the proteolytic processing of viral *gag* and *pol* precursor polyproteins at or near the point at which the virus buds from the host cell.[13] A role for HIV PR has also been suggested in the hydrolysis of nucleocapsid proteins before integration of the viral RNA into the host cell nucleus.[14] The HIV-1 PR is essential for viral replication, and if the PR is inactive or absent, noninfectious virus particles with aberrant structures are produced.[15,16] As such, inhibition of HIV-1 PR is an attractive target in the search of an AIDS therapy.

The HIV-1 PR belongs to the general class of aspartyl protease enzymes.[17,18] This classification is based on the presence of the characteristic active site triad Asp-Thr-Gly, inhibition by pepstatin A, and sequence homology with other enzymes of this class. The PR monomer in HIV-1 consists of 99 amino acids, but the catalytic unit is inferred, by comparison with the bilobal structures of analogous cellular aspartyl proteases, to be a dimer.[19] The active site involves a triad contributed by each monomeric unit and is capable of accommodating 7 residues of a substrate peptide sequence. The active site cleft is bounded by a so-called flap region of the structure, which has been recognized in other aspartyl proteases and appears to be dynamically flexible.[20] The exact role of the flap region in substrate and inhibitor binding contacts has yet to be unequivocally established, but suggestions include guarding the entrance to the active site cleft[13] and exclusion of water from the binding site.[20] It seems likely that the flexibility of the flap has something to do with accommodation of amino acid residues of various sizes as the polyprotein chain is processed. A thermodynamic role for the flap is also possible.

A series of models have now been proposed to explore the nature of HIV-1 PR at the molecular level. The first was developed by Pearl and Taylor,[19] based on a comparison of the HIV-1 primary sequence with the microbial enzymes, rhizopuspepsin, and endothiapepsin. The active site triad, well conserved in aspartyl proteases,[21] and a second conserved sequence at the start of a short α-helix in the vicinity of the carboxy-terminus of HIV-1 PR were the main points of alignment. Some deletions of large surface loops and some residues at the C-terminus were required to model the smaller HIV-1 PR from the larger microbial enzymes. When the structure of the retroviral PR from Rous sarcoma virus (RSV PR) was obtained,[22] Weber et al.[23] developed a refined model of HIV-1 PR by close comparison. The overall topology was observed to be similar to that proposed by Pearl and Taylor,[19] considerably more molecular detail was set forth. The nature of the flap region was less certain in the Weber model as this turns out to be disordered in the RSV PR structure, presumably due to high mobility.

Recently models for the structure of HIV-1 PR have been obtained directly from X-ray crystallographic structure determination. Navia et al.[2] reported a crystal structure at 3.0 Å resolution for the NY5 isolate of HIV-1 cloned from *E. coli*. The protein crystallized in essentially the form proposed for the active dimer. Large regions of the structure including the active site were homologous with microbial aspartyl proteases, but regular secondary structure was neither observed in the dimer interface nor in an expected α-helix near the carboxy terminus, in contrast with the structure reported in these regions for the highly homologous RSV PR. The flap region was observed to be ordered crystallographically, but subject to intermolecular contacts in the crystal.

A determination of the structure of chemically synthesized [Aba 67,95] HIV-1 PR, SF2 isolate, was subsequently reported by Wlodawer and coworkers.[3]

Data on the crystalline form was obtained to 2.7 Å resolution and indicated the N-terminal region to be ordered with the dimer interface to be a dovetailed β-pleated sheet involving both C-terminal and N-terminal regions. Wlodawer et al.[3] also observed the α-helix region missing in the Navia structure.[2] Other elements of the Wlodawer structure were in close agreement with the earlier result. Again, the flap region was observed to be ordered, in contrast to RSV PR. Comparison of the Wlodawer structure with the proximal structure of Pearl and Taylor[19] resulted in an RMS deviation of 2.7Å, and general but not necessarily specific agreement. The RMS correspondence with the model structure reported by Weber et al.[23] was 1.3 Å for 82 of the 99 C^{α} atoms, excluding the flap region and frameshifting residues 1–5 by two positions. Navia[24] subsequently indicated that the discrepancies between the two HIV-1 PR crystal structures on the order in the N-terminal region and the presence of the homologous α-helix in the structure have been resolved. A third structure of this enzyme has been reported[4] and is in substantial agreement with the Wlodawer structure.

The models based on homology mapping and determined from X-ray crystallography ultimately resulted in the proposal of a specific molecular structures for HIV-1 PR, and the crystallographic temperature factors indicated dynamical range of motion of the various atoms in the crystal. Moreover, there is an indication of interesting sequence-dependent correlated dynamical motion in the structure[7,8] that could have considerable implications in HIV-1 PR dimer formation, enzyme mechanism, and inhibitor interactions. Thus in addition to the previous crystallographic models proposed for HIV-1 PR, it is of considerable interest to obtain an accurate dynamical model for this important system.

Swaminathan et al.[8] recently described the domain structure in HIV-1 PR based on the cross-correlation of atomic fluctuations obtained from an MD simulation. This approach was used earlier to study residue displacements in cytochrome c by Ransom-Wright and McCammon[25] and the active site dynamics of ribonuclease by Brunger et al.[26] The normal mode structure of pancreatic trypsin inhibitor was analysed on this basis by Nishikawa and Go[27] and concurrently by Ichiye and Karplus.[28] The determinant of the cross-correlation matrix has been related by Karplus and Kushick to the configurational entropy.[29] Our previous report[8] is extended herein to a consideration of the time evolution of the domain correlations in HIV1-PR, and the sensitivity of these results to the inclusion of solvent.

CALCULATIONS

Molecular dynamics calculations in this study were performed with the Monte Carlo (MC) and MD simulation program WESDYN[30] using the RT37C4 force field of GROMOS86[31] and the SPC model for water.[32] Switching functions were used to make the long-range, nonbonded interactions go smoothly to zero between 7.5 and 8.5 Å and applied on a group by group basis to avoid artificially splitting dipoles. The point of departure for this study was the X-ray crystal structure of HIV-1 PR solved by Wlodawer et al.[3] The protein was solvated with 6,990 molecules of water in a hexagonal prism of height 78.928 Å with an inscribed circular radius of 34.562 Å. The system was treated in the simulation under periodic boundary conditions. The volume of the system was chosen to produce a solvent density of 1 gm/cc. The simulation protocol included an initial solvent relaxation of 21 M configurations of Metropolis Monte Carlo simulation,[33] followed by 50 steps of conjugate gradient minimization of the total system. The MD simulation[34] involved a heating to 300 K over 1.5 psec, a Gaussian equilibration step of 2.5 psec, and a trajectory involving 96 psec of free MD with a temperature window of 5K. The calculation was carried out on the CRAY Y-MP/832 at the Pittsburgh Supercomputer Center and required ca. 100 hours of machine time.

The convergence characteristics of the simulation are collected in Figure 2. The block averaged total

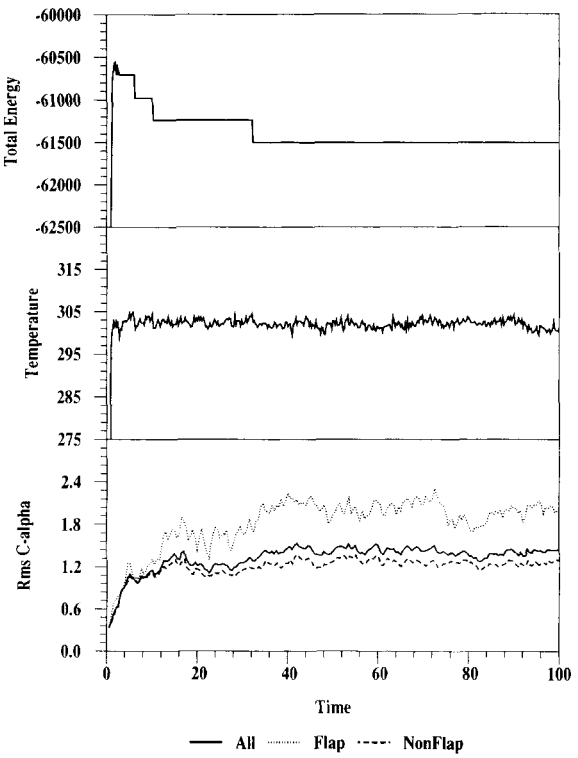

Fig. 2. Calculated total energy, temperature, and RMS C^{α} deviation of dynamical structures from the 3HVP crystal geometry vs. time for the MD simulations on HIV-1 PR (solvated model). For the rms deviations: ——, all; ····, flap; – – –, nonflap.

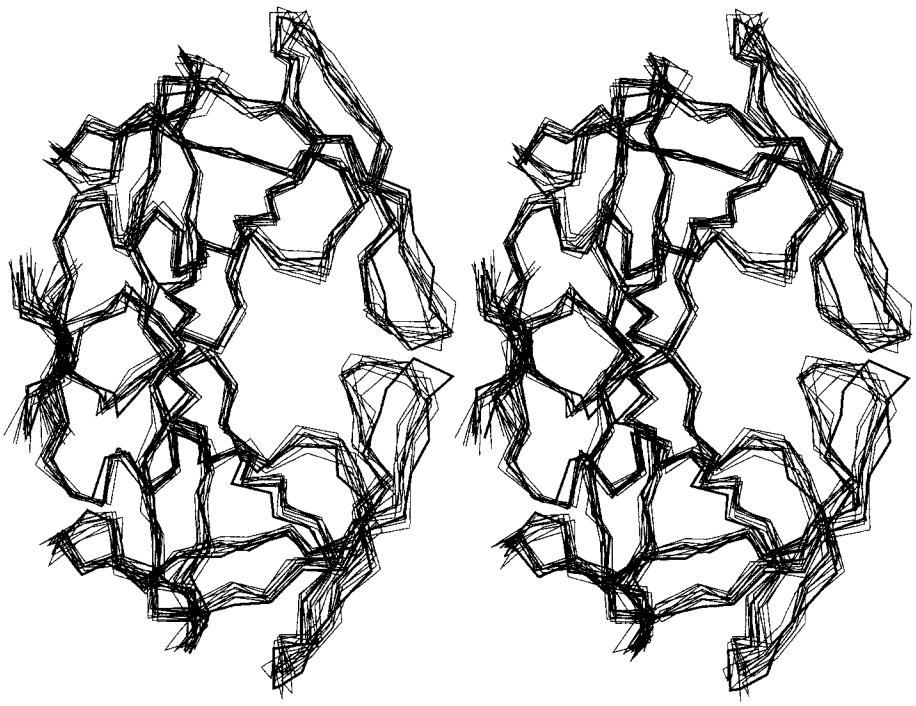

Fig. 3. Dynamical range of motion subsumed by the HIV-1 PR dimer in the MD simulation (solvated model), showing the superposition of α-carbon traces for 9 snapshots equally spaced along the MD trajectory. The bolder solid line is the α-carbon trace of the 3HVP crystal geometry.

energy is plotted as a function of time in Figure 2a. The total energy increases rapidly during the heating phase as kinetic energy is supplied to the system and subsequently shows evidence of a series of velocity rescalings as the system makes structural transitions en route to locating a thermally bounded state. As a consequence of the rescalings, the temperature is seen to be constant (Fig. 2b). The simulation is observed to be stabilized after some 33 psec, and no further rescaling was necessary after that point in the trajectory. The root mean square (RMS) deviation of the MD structures with respect to the crystal structure is shown as a function of time in Figure 2c. After the initial period of equilibration, the RMS is also is seen to be stable and oscillatory around a mean. Collectively, this information indicates the simulation is well stabilized, an essential condition for further analysis.

We have also carried out two additional MD simulations on HIV-1 PR in the absence of solvent (in vacuo models) in order to assess the role of water in the simulation and the effect of solvent on protein stability. These simulations utilized the RT37C4 and the RT37D4 force fields of GROMOS, respectively. The former was set up to correspond directly to the simulation including water as described above, and the latter uses a variation of the GROMOS force field designed especially for in vacuo simulations. These simulations were also carried out for 100 psec each. The convergence characteristics for each of these calculations were examined and indicate them to be of comparable stability to the solvated case. However, the results based on the in vacuo models are significantly different than those obtained with solvent included. A detailed description and analysis of the results from each of the simulations are presented in the following section.

RESULTS

Description and Comparison With the X-Ray Structure

The MD structures from the solvated model are compared with the crystal structure (initial configuration in the simulation) in Figure 2c. The RMS deviation is quite stable over the last 66 psec of the trajectory, and at the termination of the run is 1.3 Å. This indicates that the dynamical structure of HIV-1 PR obtained by the simulation including solvent remained generally in the realm of the crystal geometry during the course of the simulation and serves to characterize the inherent stability and accuracy of the model. The main discrepancy is found in the flap region.

The overall motion and dynamical structure of the protein is presented in Figure 3, a superposition of 9 snapshots obtained at equally spaced intervals from the MD trajectory. The dynamic range of the MD structures brackets the crystal geometry (bold line),

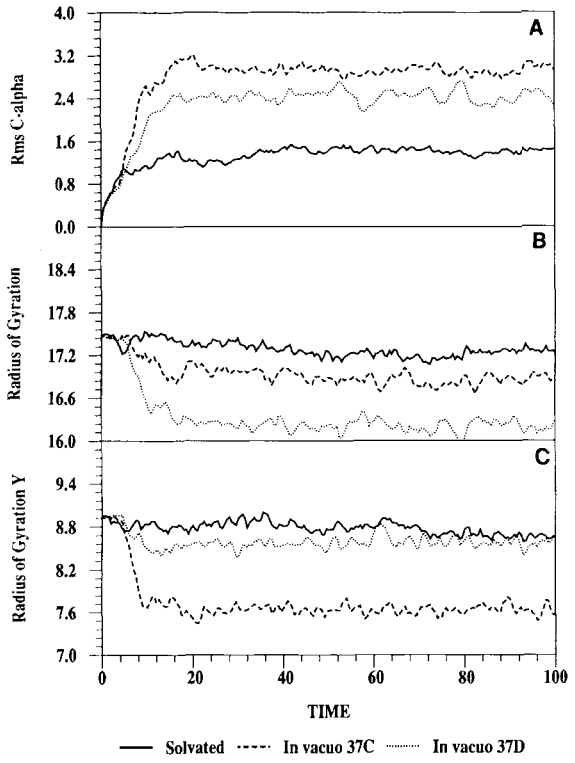

Fig. 4. Calculated **(A)** RMS C^{α} deviation of dynamical structures from the 3HVP crystal geometry, **(B)** radius of gyration, **(C)** y component of the radius of gyration vs. time for the MD simulations on HIV-1 PR (in vacuo model).

except in the flap region where concerted changes are observed. Aspartyl proteases have mobile flaps in general, and in the case of the highly homologous RSV protease, the flaps were too disordered to solve positionally. The unique solution of flap location in the HIV-1 PR crystal structure seems to be due to the influence of intermolecular contacts in the crystal.[3] The RMS C^{α} deviation of the MD structures from the crystal values in the flap region is broken out in Figure 2c. The deviation in this region is considerably larger than that of the other atoms in the structure. In the simulation, the flap adjusts to a geometry predicted hereby to be favored in aqueous solution.

The results of the corresponding in vacuo simulations are shown in Figures 4–6. The RMS C^{α} deviations of the in vacuo and solvated MD structures with respect to the crystal geometry are compared in Figure 4A. The RT37C4 in vacuo results stabilize at 3.0 Å RMS, nearly three times that of the solvated case. The corresponding results from the RT37D4 force field lie at 2.3 Å RMS from the crystal structure. Neither in vacuo simulation ends up as close to the crystal structure as the solvated MD run. The superposition of MD snapshots from the RT37C4 in vacuo simulation, Figure 5, shows considerably more deviation from the crystal structure than the solvated case, characterized particularly by a net contraction of the structure in the absence of solvent forces. The RT37D4 MD structures, shown in Figure 6, also show evidence of considerable deformation vis a vis the crystal form.

The differences in the various simulations can be seen in more detail from the time evolution of the radius of gyration R_G of the structures, Figure 4B,C. The magnitude of R_G, Figure 4B, is smaller than in the solvated case, consistent with the contraction observed in the structures of Figure 5 compared with Figure 3. Figure 4C shows the component exhibiting the greatest deviation. The structure is especially contracted in the principal y direction for RT37C4, and in the x direction for RT37D4. The radius of gyration vectors for the last MD structure in the three cases are:

$R_G = 14.1\,i + 8.8\,j + 6.1\,k$; $|R_G| = 17.7$ (solvated model, RT37C4 force field),

$R_G = 14.1\,i + 7.8\,j + 6.2\,k$; $|R_G| = 17.3$ (in vacuo model, RT37C4 force field),

$R_G = 12.9\,i + 8.7\,j + 5.9\,k$; $|R_G| = 16.6$ (in vacuo model, RT37D4 force field),

as compared with the crystal structure values of

$R_G = 14.2\,i + 9.0\,j + 6.1\,k$; $|R_G| = 17.9$ (Wlodawer et al.[3] crystal structure).

All these values are computed in the respective center-of-mass/principle axes coordinate frames. Collectively, the radius of gyration results confirm that the solvated model gives a much better account of the crystal geometry than either of the respective in vacuo models. Thus further analysis of results in this section is presented only for the solvated case.

Figure 7a shows a topological representation of the HIV-1 PR crystal structure and Figure 7b shows the dynamical secondary structure from the solvated MD simulation. The backbone hydrogen bonds are represented as solid lines. In the MD topology, Figure 7b, the thickness of the lines is proportional to the calculated average bond strength. Following the previously suggested descriptors of secondary structure in HIV-1 PR,[3] the interface residues (1–5, 95–99) form a dovetailed β-pleated sheet, slightly tighter than that obtained by an energetic analysis of the crystal geometry in this region. Residues 9 through 24 make up the b and c strands of an extended β-sheet. Note also the β-linkage between the active site (25–29) and the N-terminus of the helix (86–94). The β-sheet in the flap region (42–58) remains regular, with lower end of the flap (49–52) forming weak but persistent intermolecular backbone hydrogen bonds with the analogous region of the other monomer. In the crystal structure, flap contacts are made between and not within dimers.

Comparisons of the calculated and observed (crystallographic) temperature factors are shown in Fig-

Fig. 5. Dynamical range of motion subsumed by the HIV-1 PR dimer in the MD simulation (in vacuo model, RT37C4 force field), showing the superposition of α-carbon traces for 9 snapshots equally spaced along the MD trajectory. The bolder solid line is the α-carbon trace of the 3HVP crystal geometry.

ure 8 and indicate regions of high and low thermal mobility. The flaps are flexible, but other regions of regular secondary structure show less mobility. The active site triad motions are particularly narrow in both calculated and observed structures. The qualitative trends in comparing the calculated and observed temperature factors obtained from crystallography are in good accord, but the latter are generally higher than the simulation values. The isotropic symmetry constraints in the solution of the crystal structure leads to larger reported thermal factors if considerable asymmetric motions are present.

Conformational and Helicoidal Analysis

The analysis of all HIV1-PR conformational and helicoidal parameters from the simulation as a function of time was carried out using the procedure "Curves, Dials and Windows" (CDW).[35] The conformational analysis involves consideration of all ψ, ϕ, and ω parameters for each amino acid residue and side chain torsions χ where relevant. In the CDW procedure, the time evolution of each conformational parameter is displayed on a Klyne-Prelog[36] conformational wheel (dial) in a composite form using computer graphics. The helicoidal analysis is based on the determination of a unique, curved axis for the protein using the Curves procedure of Sklenar et al.,[37] and a full set of helicoidal parameters, which locate the mean plane of each peptide unit uniquely with respect to this axis. The time evolutions of the helicoidal parameters are displayed as "windows" on the appropriate intervals, also in composite plots produced by computer graphics. Full details of CDW analysis on proteins are described in a recent article by Swaminathan et al.[38] The complete results of CDW analysis of the MD trajectory for HIV1-PR described herein are too lengthy to include directly in this study and are provided as a supplement available on request from the authors. Readers interested in the original details are urged to consult this material. For others, we provide herein a concise summary of the results.

For a description of the conformational analysis, we divide the space into the three conventional sectors of 120° each, with the standard designations *gauche*$^+$ (g^+), *gauche*$^-$ (g^-), and *trans* (*t*). For the backbone ψ and ϕ angles, 90% of the residues remain in the same sector of conformation space as the corresponding crystal structure values, and in most cases show stable oscillatory behavior with a mean quite close to the value observed crystallographically. The exceptions are Asp 29, Ile's 50 and 93, and Gly's 40, 49, 52, and 94. The C-terminal and N-ter-

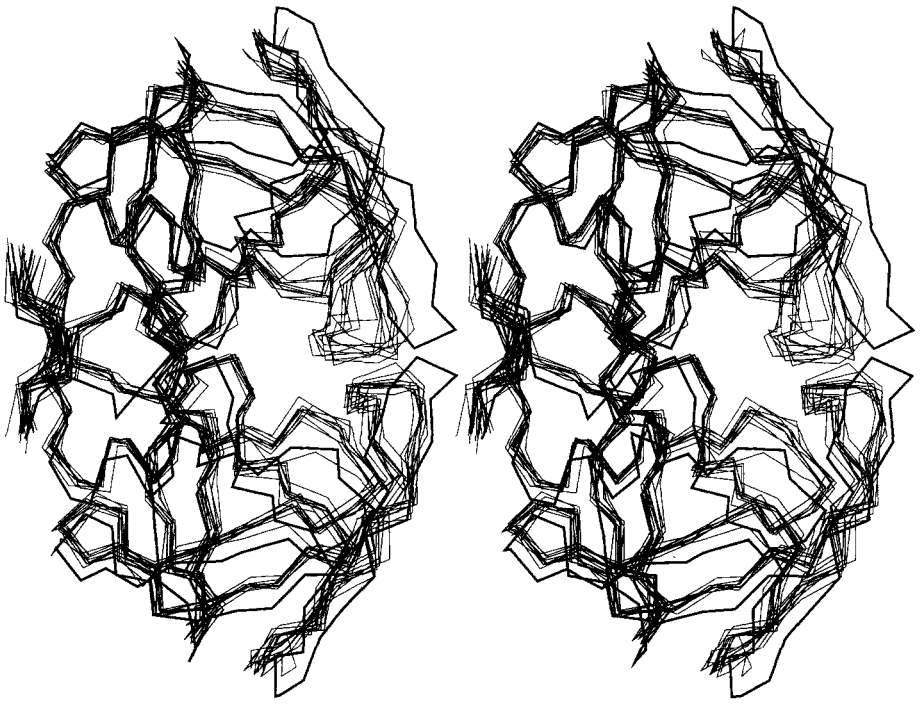

Fig. 6. Dynamical range of motion subsumed by the HIV-1 PR dimer over the interval of 40–100 psec in the MD simulation (in vacuo model, RT37D4 force field), showing the superposition of α-carbon traces for 9 snapshots equally spaced along the MD trajectory. The bolder solid line is the α-carbon trace of the 3HVP crystal geometry.

minal residues also show deviations from the crystal structure, which may be due to end effects in the simulation.

Asp 29 makes a g^- to g^+ transition early on in the simulation and remains in a stable oscillation at g^+ for the duration of the trajectory. The two Ile's, 50 and 93, show g^+ to g^- in ϕ and *cis* to *t* in ψ, respectively. Gly 40 goes g^- to *t* in ϕ, Gly 49 makes a concerted change from g^- to *t* in ϕ, coupled with a g^+ to g^- transition in ψ. Gly 52 moves from g^- to t in ψ while ϕ moves from the upper bound of *t* to the lower bound. Gly 94 goes g^+ to g^- in ϕ and ψ moves from midrange g^+ to the borderline with *t*. All of these changes occur very early on in the simulation, during heating or in the very early stages of equilibration. The significance of this is considered further in the discussion section.

The parameter ω, the torsion angle around the bond connecting the side chain to the backbone of a given amino acid residue, is stable and tightly oscillatory about *t*, in complete accord with the values for ω found in the crystal structure.

Significant conformational activity is seen in about half the residues exhibiting side chain torsion angles, once again discounting possible end effects. For example, Gln flips from g^- to g^+ in χ_2, coupled with a change of g^+ to g^- in χ_3. In Arg 8, χ_4 goes from g^+ to g^-, and then later to *t*. In some cases, such as Leu 19 and Leu 24, there is an early transition away from the crystal value of g^+ to g^-, and later a return to g^+. Met 36 goes from *t* to g^- about halfway along the trajectory. Sidechain transitions are seen in a total of 15 cases per monomer, half of which occurred in heating or equilibration and remained stable thereafter, and half showed further conformational activity at subsequent points in the trajectory. The residues Gln 18, Leu 19, Lys 20, Leu 33, Lys 43, and 45, Met 46, Lys 55, Gln 61, Glu 65, Lys 70, Leu 76, and Asn 98 showed side chain conformational activity that continued throughout the simulation. The overall pattern is thus one of considerable conformational activity in the sidechains, even on the 100 psec timescale of this simulation, and in spite of the presence of solvent water.

The helicoidal axis of the protein crystal structure and snapshots representative of the dynamical simulation are shown in Figure 9, and a panel of snapshots from the MD simulation including solvent, presented in terms of the peptide planes obtained from Curves, is shown in Figure 10. The helicoidal parameters of the analysis can be considered in three classes: (1) the axis-peptide parameters x-displacement (XDP), y-displacement (YDP), inclination (INC), and tip (TIP); (2) the interpeptide parameters shift (SHF), slide (SLD), rise (RIS), tilt (TLT), roll (ROL), and twist (TWS); and (3) the axis-junction parameters x-displacement (AXD), axis y-displacement (AYD), axis inclination (AIN), and axis

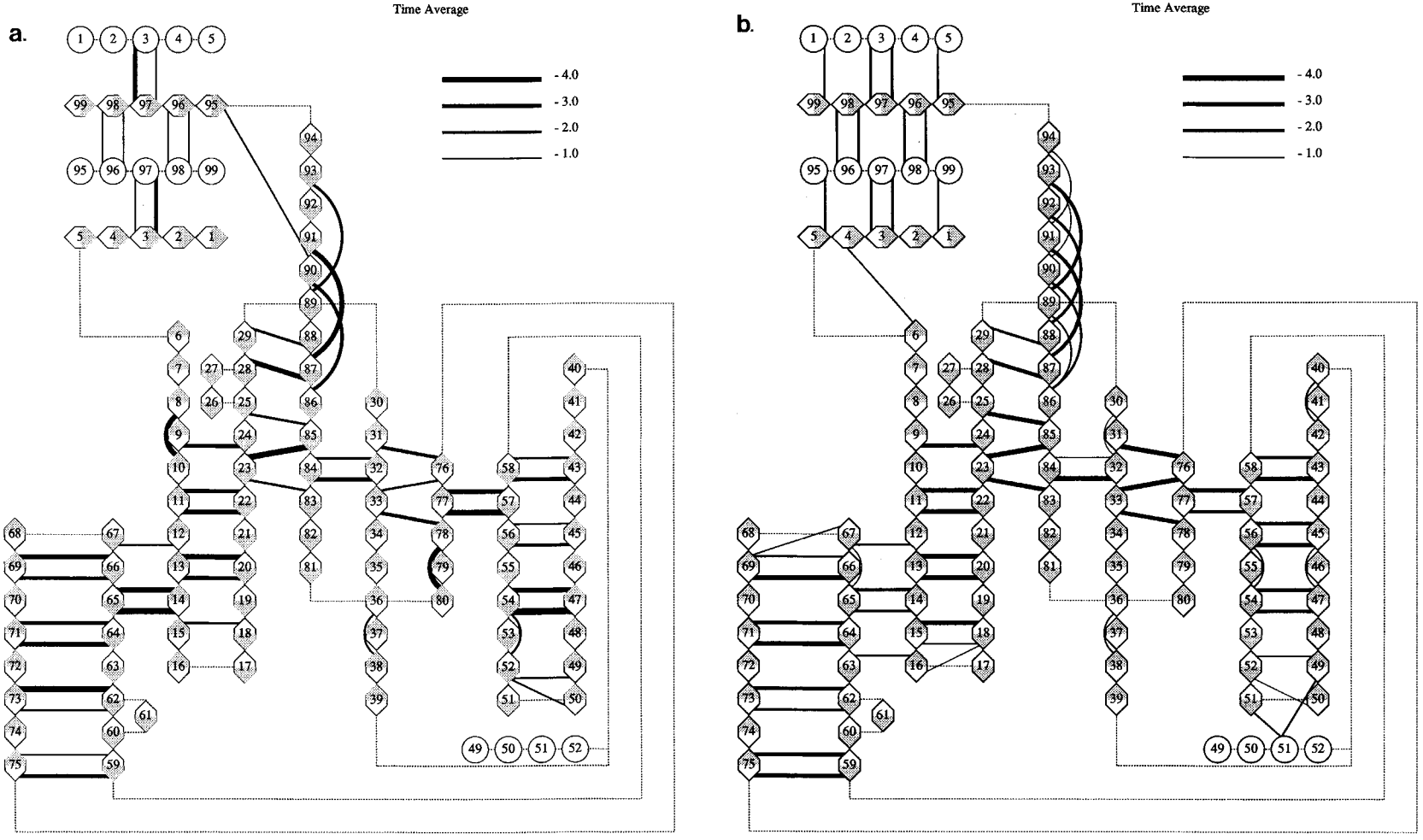

Fig. 7. Topology of the secondary structure and the calculated hydrogen bonding patterns for the backbone atoms of HIV-1 PR in **(a)** the 3HVP crystal structure and **(b)** MD simulation (solvated model). Solid lines indicate the hydrogen bonds and their relative strengths; broken line indicates the backbone peptide bonds, with the residues in hexagons; residue numbers in circles are for the other chain in the dimer.

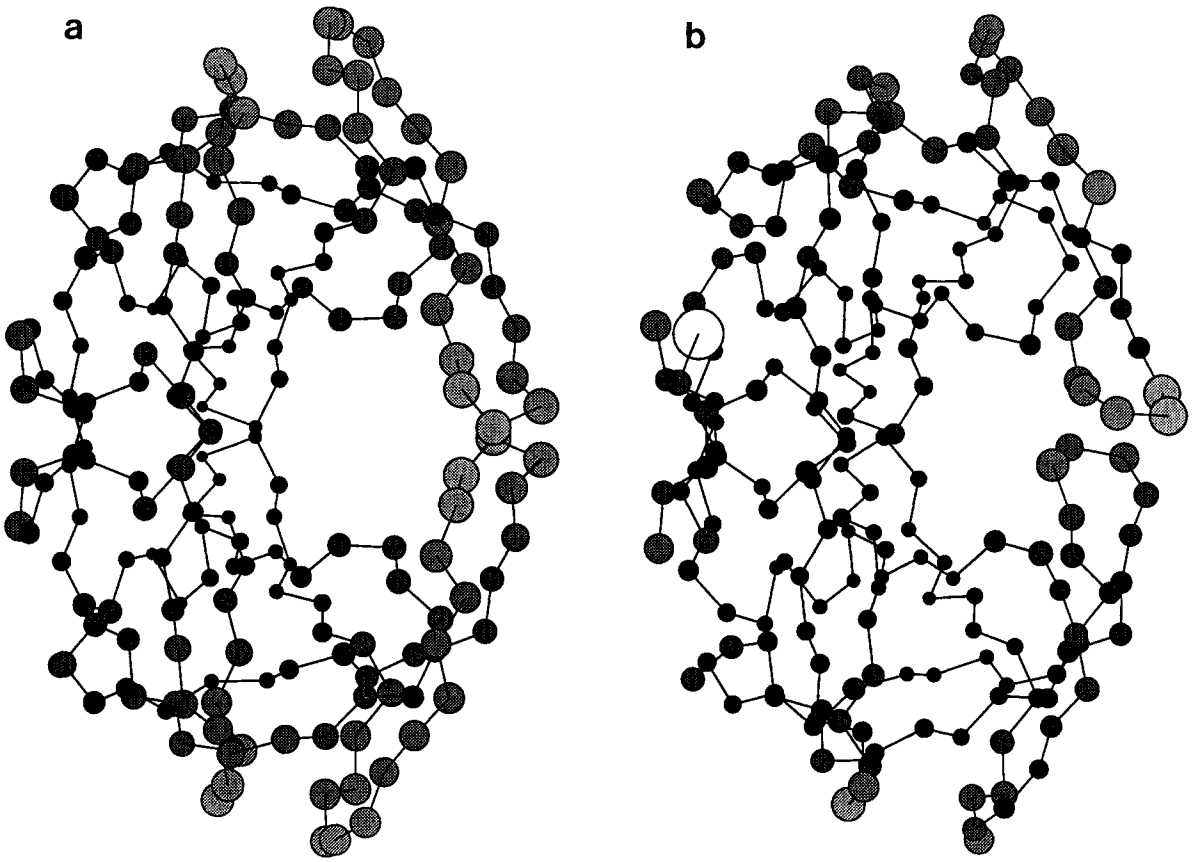

Fig. 8. Comparison of the observed and calculated temperature factors for HIV-1 PR; **(a)** crystallographically observed and **(b)** MD calculated (solvated model). The magnitude of the temperature coefficients is indicated by the size and shading of the circles; low temperature factors have small circles and dark shading.

tip (ATP). The helicoidal analysis shows the axis-peptide parameters to be remarkably stable over time in the simulation and no significant deviation is observed from the corresponding values in the crystal structure. The characteristic patterns in peptide INC and TIP values for α helices and β sheets in BPTI are generally observed in the corresponding regions of HIV1-PR as well.

The peptide-junction parameters are likewise consonant in nearly all cases with the crystal structure values. The only notable exception is a major flip in TWS at the Gly 49/Ile 50 junction. However, high amplitude fluctuations in TLT, ROL, and TWS occur at the turn regions of the structure as seen for BPTI as well. The axis-junction parameters also execute high amplitude fluctuations at turns, but in ordered regions of secondary structure are found to be generally quite close to values observed crystallographically.

Correlated Motion and Molecular Communication

The constituent atoms of a protein motif or microdomain may be expected to execute concerted, well-correlated motions. A molecular dynamics trajectory for a protein contains a description of the dynamical motions of all atoms included in the model. Concerted atomic motions can be identified by analyzing the cross-correlation coefficients for atomic displacements,[39] defined between any two atoms by the expression

$$C_{ij} = \langle \Delta r_i \cdot \Delta r_j \rangle / (\langle \Delta r_i \rangle^2 \langle \Delta r_j \rangle^2)^{1/2}$$

where Δr is the displacement from the mean position of the ith atom. A contour plot of the matrix $[C_{ij}]$ constitutes a "dynamical cross-correlation map" (DCCM), in which atomic motions with strong correlations will have correspondingly large off-diagonal cross peaks. A DCCM depends also on the time scale over which data on the correlations was gathered and different types of interactions can be expected to emerge on different time scales. Details of DCCMs applied to the investigation of domain structure in proteins have recently been given by Swaminathan et al.[8]

The DCCM analysis of domain structure in the form we use herein is illustrated schematically in Figure 11, indicating how the hypothetical case of a

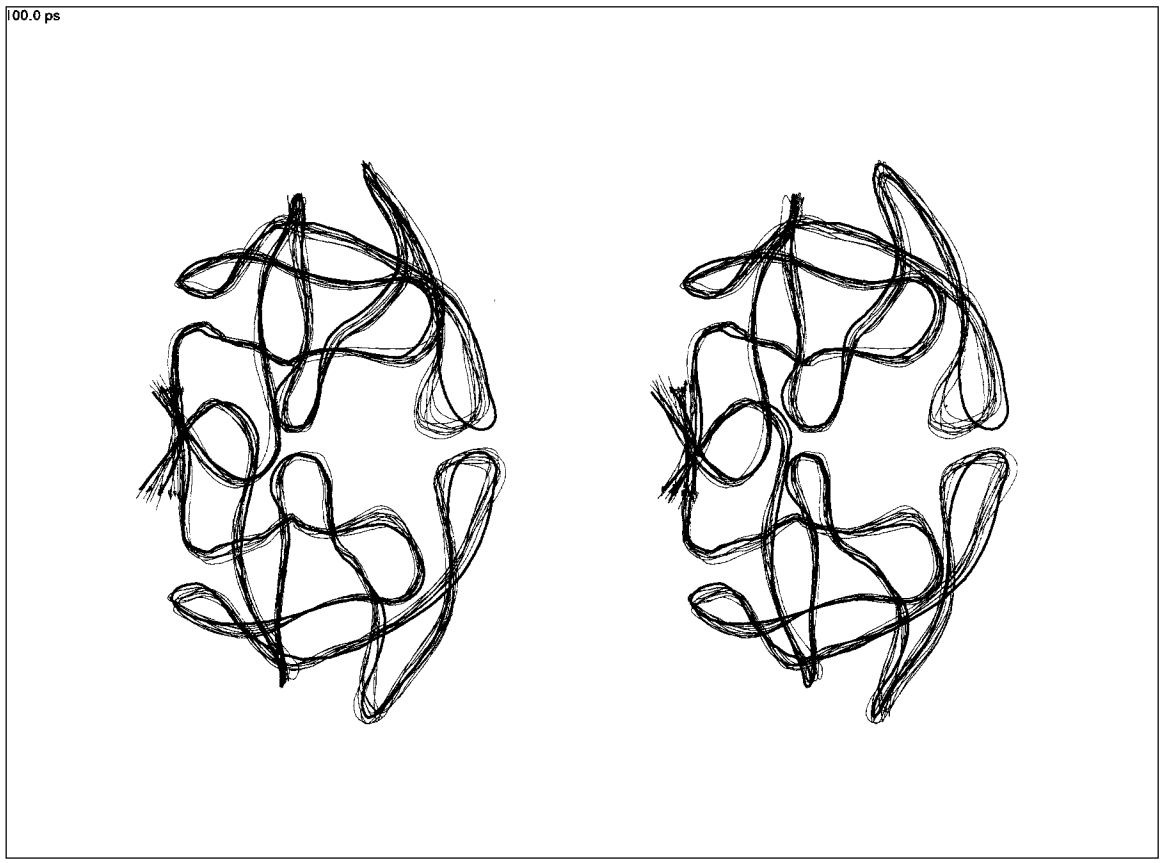

Fig. 9. Dynamical superposition of the helicoidal axes of HIV-1 PR obtained from MD simulation as compared with the crystal structure (darker line).

protein with two domains, A and B, would be handled. We have chosen to present the C_{ij} matrix in transposed form, to emphasize a certain analogy with the presentation of 2D NOESY spectra in FT NMR. Positive correlations are collected in the upper triangle of the map, and negative correlations in the lower triangle. Domains of contiguous residues should give rise to regions of positive correlations along the diagonal of the DCCM (A and B in Fig. 11) indicating strong correlations among backbone atoms situated sequentially along the backbone. In the simplest cases, such as an α-helix, these regions will be essentially triangular in shape. Regions of antiparallel β-sheet will give rise to plumes of positive correlations emanating from the diagonal. In fact each distinct type of protein domain will have a signature appearance on a DCCM. Intradomain correlations between residues not contiguous in the primary sequence appear as off-diagonal matrix elements, also expected to be positive since the residues move in concert.

The remaining off-diagonal cross-peaks can be identified with domain-domain correlations in proteins and may be positive or negative, AB^+ and AB^-, respectively, in Figure 11. This information reflects aspects of structure originating uniquely in the dynamical motions of the protein and may not be apparent at all from a visual inspection of the average or the crystal structure.

The C_{ij} are computed as averages over successive backbone N, C, and C atoms to give one entry per pair of amino acid residues. There is a time scale implicit in the as well since $<\Delta r_i>$ is the mean position of the ith atom for a specific block of time. DCCMs for HIV-1 PR were calculated as block averages over intervals of time from 5 to 40 psec from the MD trajectory. The 5, 25, and 40-psec block averages are represented in the DCCMs shown in Figures 12–14. Only correlations above a threshold value of 0.25 are included. The intensity of the shading is proportional to the magnitude of the cross-correlation coefficient. The positive correlations are given in the upper triangle, and the negative correlations in the lower triangle.

In the 5-psec DCCM, Figure 12, three plumes of correlations emanating from the diagonal of the

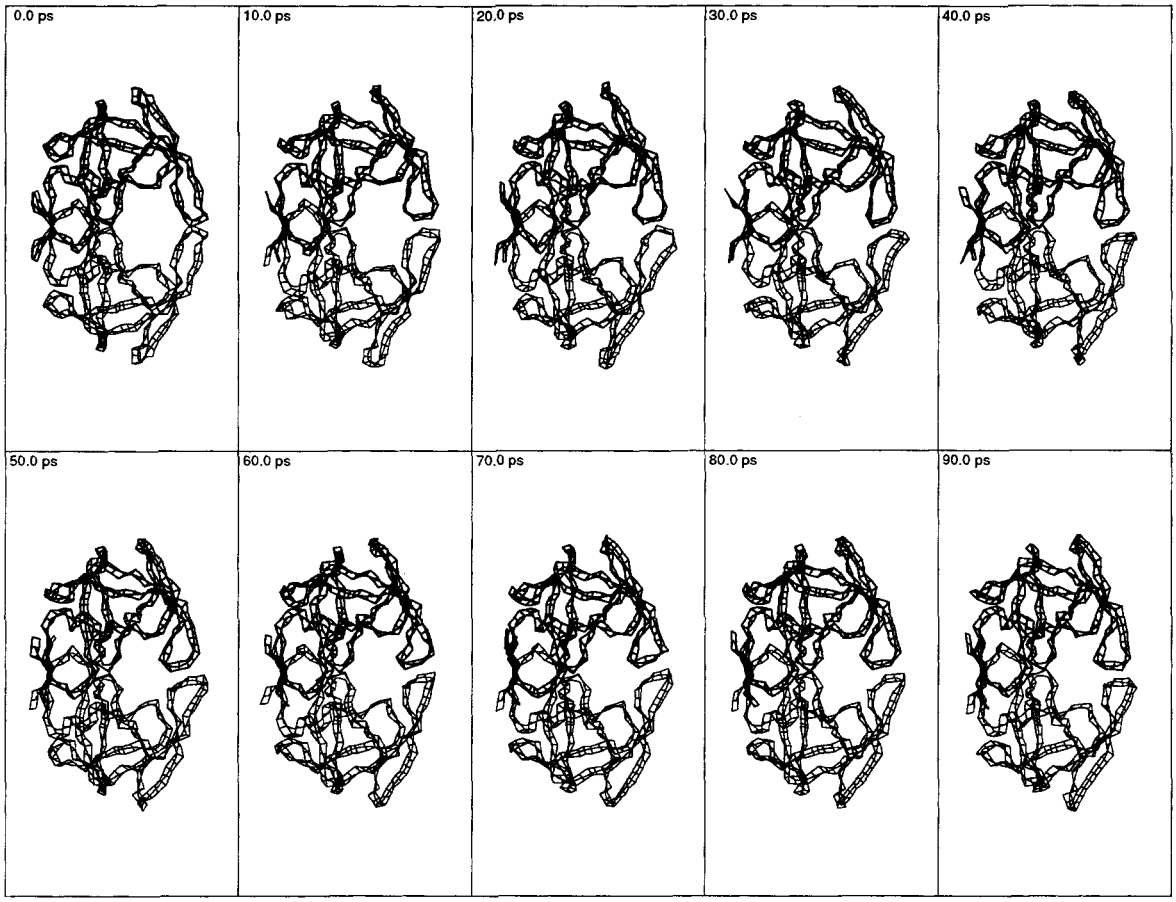

Fig. 10. Snapshots from the Md simulation on HIV-1 PR (solvated model) presented in terms of the peptide planes generated in Curves analysis.

DCCMs into the upper triangle are immediately evident. These are labelled (a), (b) and (c) on the map and correspond to three distinct contiguous sequences of antiparallel β-sheet domains. (Henceforth, lower case letters in parentheses refer to correlations indicated on the various DCCM.) Moving up the diagonal, the first region (a) is comprised of residues 9–24, with correlated motion extending over 5 residues within the domain. (The width of a plume represents correlation of proximal residues in the β-sheet rather than that of the whole domain.) The second region (residues 42–58) is the flap (b), with motions correlated over up to nine residues in the strongly pleated sheet section. Residues 59–75 form the longest and widest domain of β-sheet (c), with correlations extending over 10 residues. The terminal and α-helix regions are characterized by triangular regions of positive correlation situated along the diagonal of the DCCM and are barely discernable in the 5-psec map.

The main patch of off-diagonal peaks (d,e) in the DCCM of Figure 12 corresponds to an intradomain correlation of between residues noncontiguous in the primary amino acid sequence. The major crosspeaks found in the area of the DCCM between active site residues 22–25 and residues 81–85 (d), and between residues 28–42 and 76–86 (e) arise from the interaction of noncontiguous residues that fold to form parallel β-sheets. A small peak corresponding to the correlation of noncontiguous residues 9–24 and 59–75 (f) is only weakly represented on the 5 psec map. This will later emerge as an interesting through-space correlation in the dynamical structure.

The 25 psec map (Fig. 13), formed from averaging over three successive blocks in the simulation, shows additional definition in the plume domains and their contiguous interactions, and an increase in the above-mentioned through-space correlation (f). Also, a correlation between the dovetailed N-terminal and C-terminal regions of the protein has emerged (g). Dynamical correlations amongst residues 43–46 and 56–58 on the flap and 76–78 in the ψ region, which form a β-type hydrogen bond network, now appear (h); the topological connection can be seen in Figure 7. A weak communication between

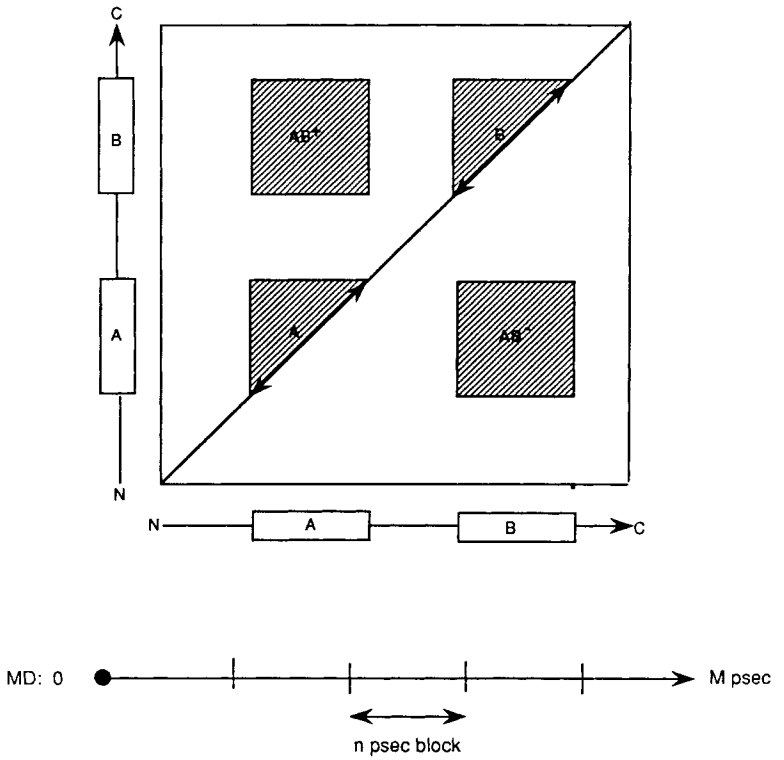

Fig. 11. Schematic diagram of a dynamical cross-correlation map for a hypothetical protein with two distinct domains, A and B.

the domain (c) and the α-helix is indicated (j), and a correlation is developing between the active site triads and the α-helix. The DCCM are seen to be quite sensitive to this type of interactions. Similarly, a correlation between residues 20–24 and 76–86 (h) as well as 28–34 (i) is indicated and is of a similar nature.

The 40 psec map, Figure 14, is based on two block averages and is the most elaborate DCCM we can construct on the basis of a 100 psec trajectory. The active site along with the flanking residues (22–29), now have distinct correlated motion indicated at (k) with residues of the h-helix (86–90) with a width spanning up to five residues. Along with the flap (42–58) the two other plume domains, residues 19–21 and 59–75, form an interesting ensemble. The positive correlations seen between residues 19–21 and 59–75 (k) have now become a major cross-peak. The onsets of the flap region (42–55) and (55–58) correlate with certain areas of β-sheet (59–65, 73,85) of the third plume domain (l,m). Negative domain-domain correlations are indicated between the flap and both other plume domains (n,o).

The observed interdomain correlations suggest that flap motion occurs with compensatory changes residues in residues 59–75 in the manner of a "cantilever." Pursuing this analogy, the β-sheet region comprised of residues 9–21, correlated to both the flap and the cantilever, mediates as a "fulcrum." This correlated motion and domain communication is recognizable in the superposition of MD structures shown in Figure 3, where the flap is observed to close down as the cantilever moves up. Negative compensatory correlations exist between the middle of the flaps and the center of the cantilever and reinforce this interpretation. The active site triad shows a distinct positive correlation to the h-helix and to adjacent residues in the Ψ-region (p).

The calculated 5, 25, and 40 psec DCCMs between residues of the two different monomeric units of the HIV-1 PR dimer are shown in Figures 15–17. The 5 psec map, Fig. 15, shows little correlations other than where the N-terminal and C-terminal ends monomer dovetail together to form the β-pleated dimer interface (a,b) and a weak flap-flap together to form the β-pleated dimer interface (a,b) and a weak flap-flap correlation (c). At 25 psec (Fig. 16), additional proximal intermonomeric correlations have formed between the two active site triads (d) and between the two flap regions, which are indicated by this simulation to be hydrogen bonded in

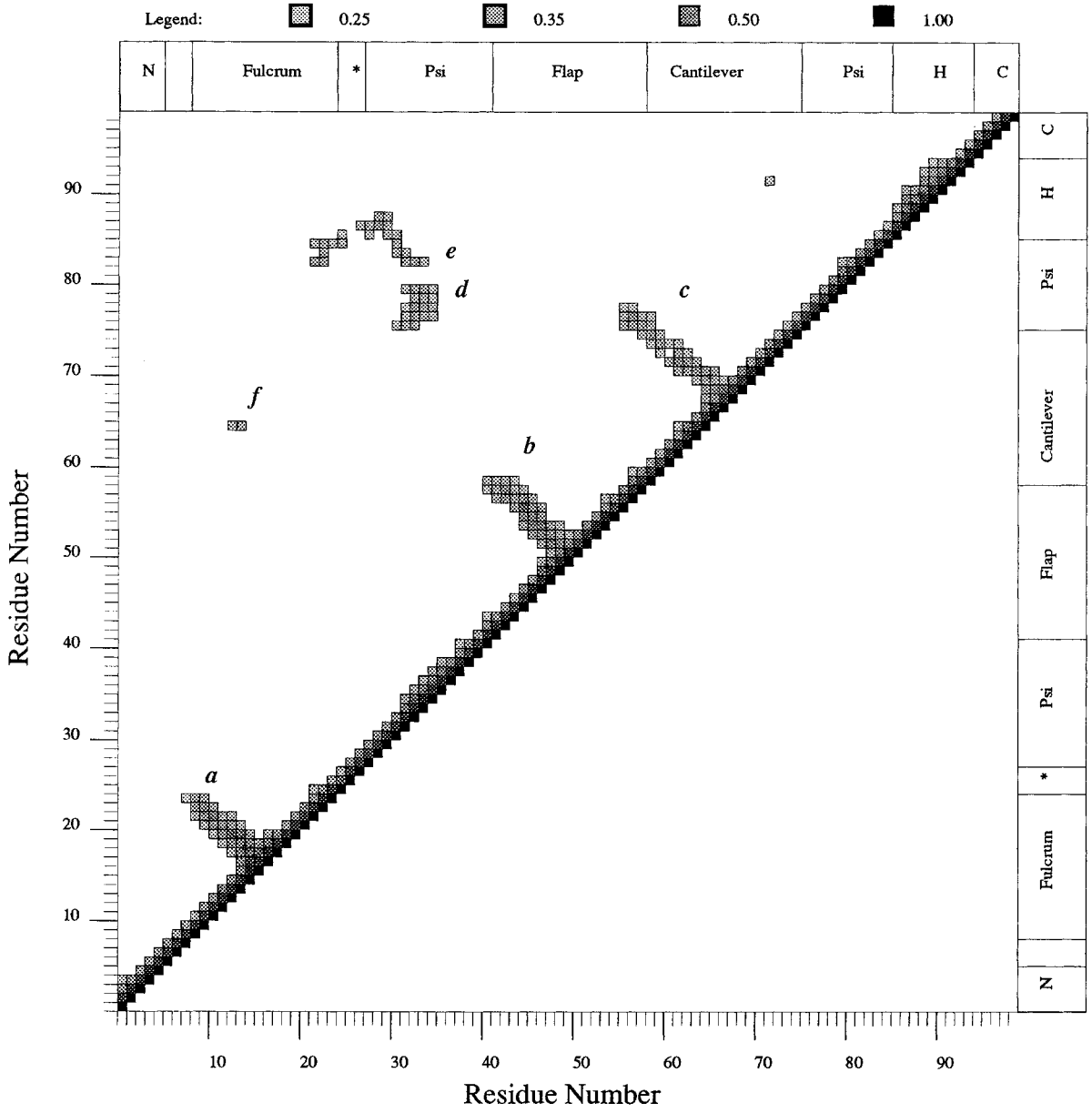

Fig. 12. Calculated DCCM for a monomeric unit of the HIV-1 protease dimer, averaged over sixteen 5-psec intervals in the MD simulation (solvated model). Only correlation > 0.25 is shown, and the intensity of shading is proportional to the size of the cross-correlation coefficient for each element, as indicated by the key at the top. Positive correlations are in the upper-left triangle, and negative correlations in the lower-right triangle.

aqueous solution. A correlation between the cantilever domains of the two monomers, separated by some 30 Å, is seen to be forming (e). The 40 psec map for the correlations between monomeric units is shown in Figure 17. The contiguous flap and interface correlations become clearer, and the domain communication between cantilevers and from the cantilever of one monomer to the fulcrum of the other is indicated (f).

The 40 psec DCCMs for the in vacuo results were also examined and compared with the results of the solvated case. Many similarities with the maps for the solvated model are apparent, since many of the correlations are due to contact interactions present in both cases. However, inspection reveals that the maps for the in vacuo models show more extensive internal correlations, a consequence of the spurious contraction of the dynamical structure. Since the in vacuo results are in poorer agreement with the crystal structure, we expect that the increased level of correlations in this model are not of biophysical interest.

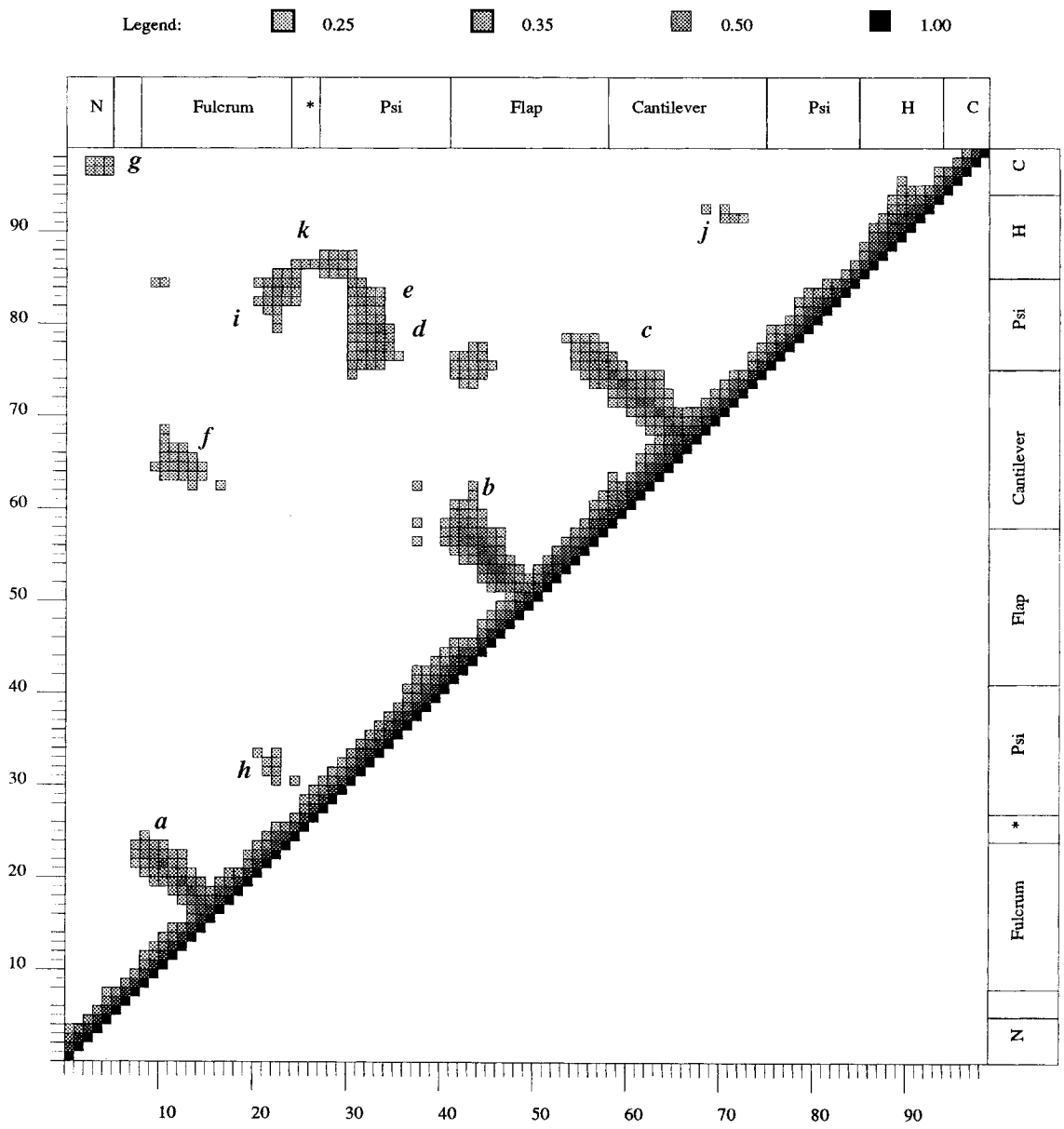

Fig. 13. Calculated DCCM for a monomeric unit of the HIV-1 protease dimer, averaged over three 25-psec intervals in the MD simulation (solvated model). See Figure 12 legend for details.

DISCUSSION

Overall, these MD results on HIV-PR indicate that a stable dynamical model has been obtained. The solvated form is in close agreement (1.3 Å rms) with the crystal structure, and thus is considered to be a reasonably accurate theoretical account of the structure of the protein. The MD results on the protein without solvent show considerably larger deviation from the crystal structure for this model. This leads us to conclude that the presence of water is essential for obtaining an accurate dynamical model for a protein using MD based on the GROMOS force field. This conclusion is consistent with results reported in several other recent MD simulations in both crystalline and solution environments.[40–43]

The conformational and helicoidal analysis of the solvated form indicates close agreement with the essential geometrical elements point by point in the structure, consistent with the low RMS value. The discrepancies that do occur are seen for the most part at the outset of the simulation. This could be a simple adjustment of the structure to accommodate local preferences dictated by the force field, but the possibility that these changes are a consequence of

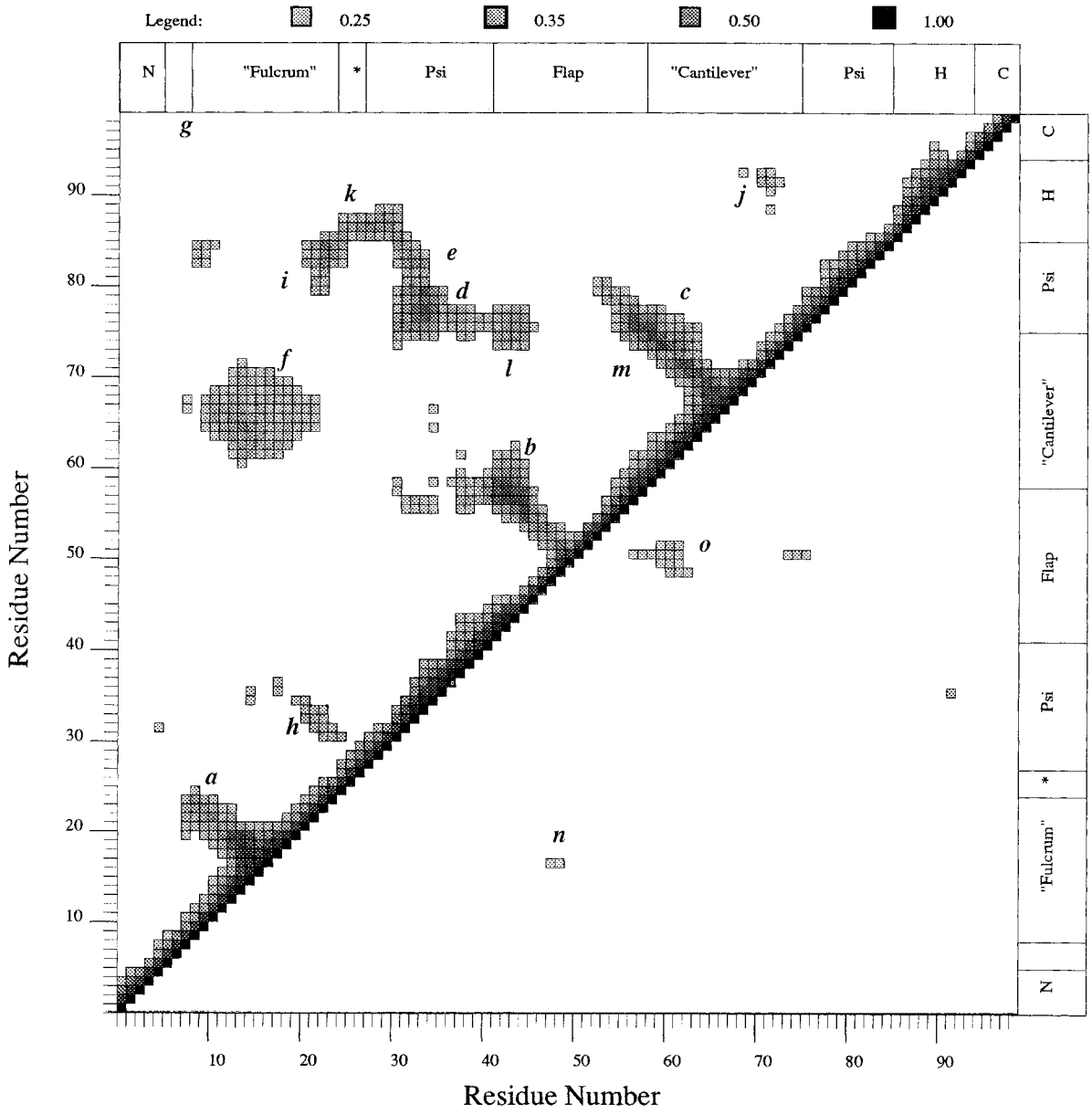

Fig. 14. Calculated DCCM for a monomeric unit of the HIV-1 protease dimer, averaged over two 40-psec intervals in the MD simulation (solvated model). See Figure 12 legend for details.

the heating protocol cannot be ruled out at this time. The sensitivity of MD results to heating protocol, essentially a problem with popping the structure into another local minimum in configurational hyperspace, has not been considered previously and we intend to examine this more carefully in future simulations on this system.

The presence of correlated motions in the protein, both within and between monomeric units, is probably the most important idea emerging from this study. The correlations between contiguous units, albeit expected, nevertheless turn out to be highly sensitive to the hydrogen bonded interconnections between the various domains and are indicated clearly on the DCCM. The domain communication between the flap region and remote areas of the protein and other correlations linked into the active site raise the possibility that the enzymatic activity is in fact a holistic property of the entire structural entity and not just local to the active site.

Experimental confirmation of domain communication is by no means simple to obtain, but one possibility is that site specific changes in regions of an enzyme remote from the active site but correlated

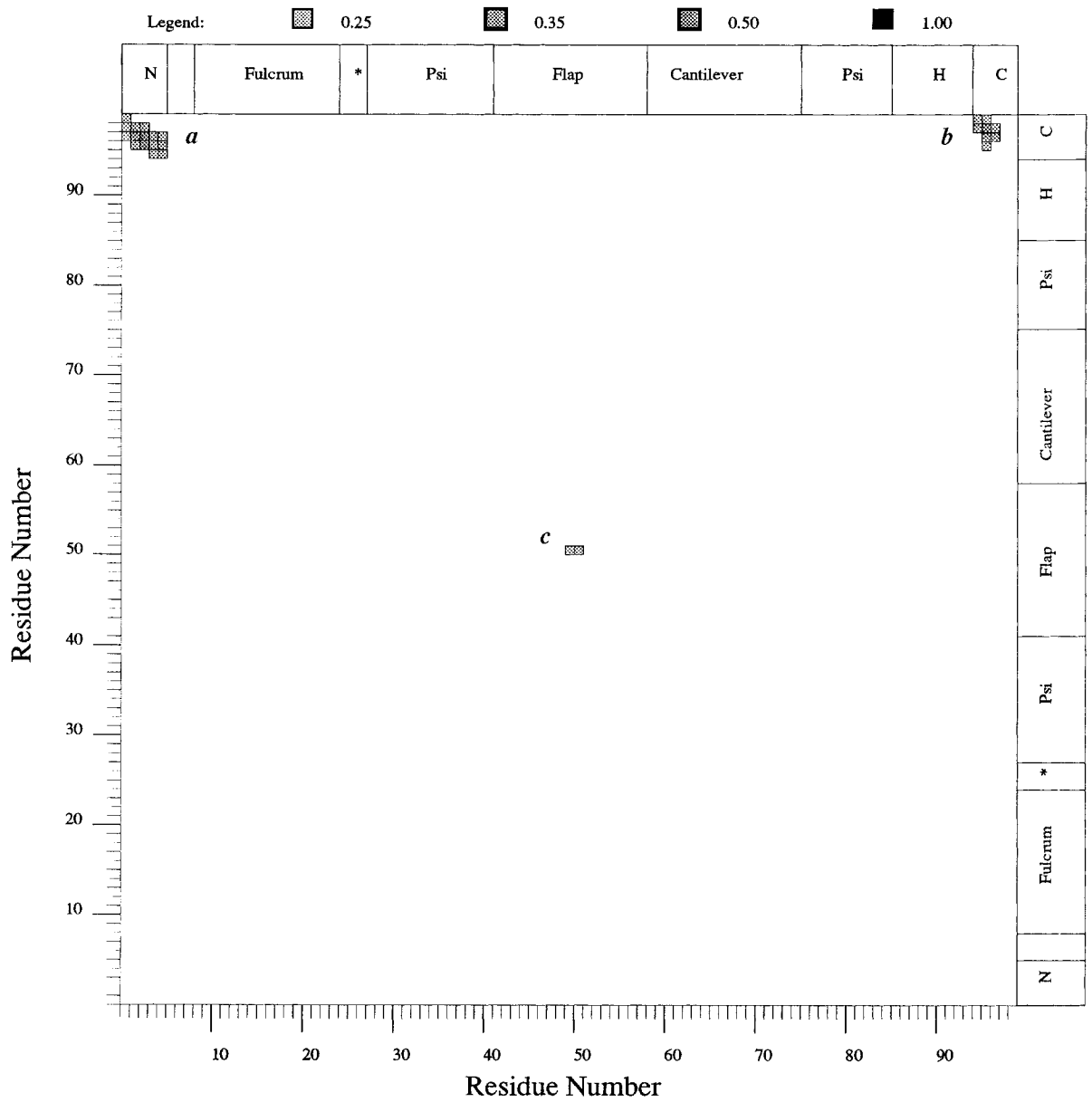

Fig. 15. Calculated DCCM between monomeric units of the HIV-1 PR dimer, averaged over sixteen 5-psec intervals in the MD simulation (solvated model). Only correlation > 0.25 is shown, and the intensity of shading is proportional to the size of the cross-correlation coefficient for each element, as indicated by the key at the top. Positive correlations are in the upper-left triangle, and negative correlations in the lower-right triangle.

thereto via dynamical correlations could alter enzyme activity. As at least leading evidence, it should be noted that site-specific mutagenesis of certain residues in the cantilever region of HIV-1 PR deactivates the protease to some extent and leads to non-infectious AIDS virus.[44] Our results suggest a possible mechanism for this otherwise unexplained phenomena. Another possibility is, of course, a general destabilization of the enzyme. In any case, further investigations on this point are currently in progress.

We still regard the idea of domain communication in proteins to be somewhat speculative and in need of further investigation before the existence and implications of these effects are consider established. These kinds of ideas have been discussed for some time in various forms,[45] but we obtain here direct evidence of such an effect, at least in our model system. We also find, in preliminary results on MD simulations the HIV-PR complex with MVT101 inhibitor[6] that in spite of the considerable reorganization of the flaps the through-space internal correlations

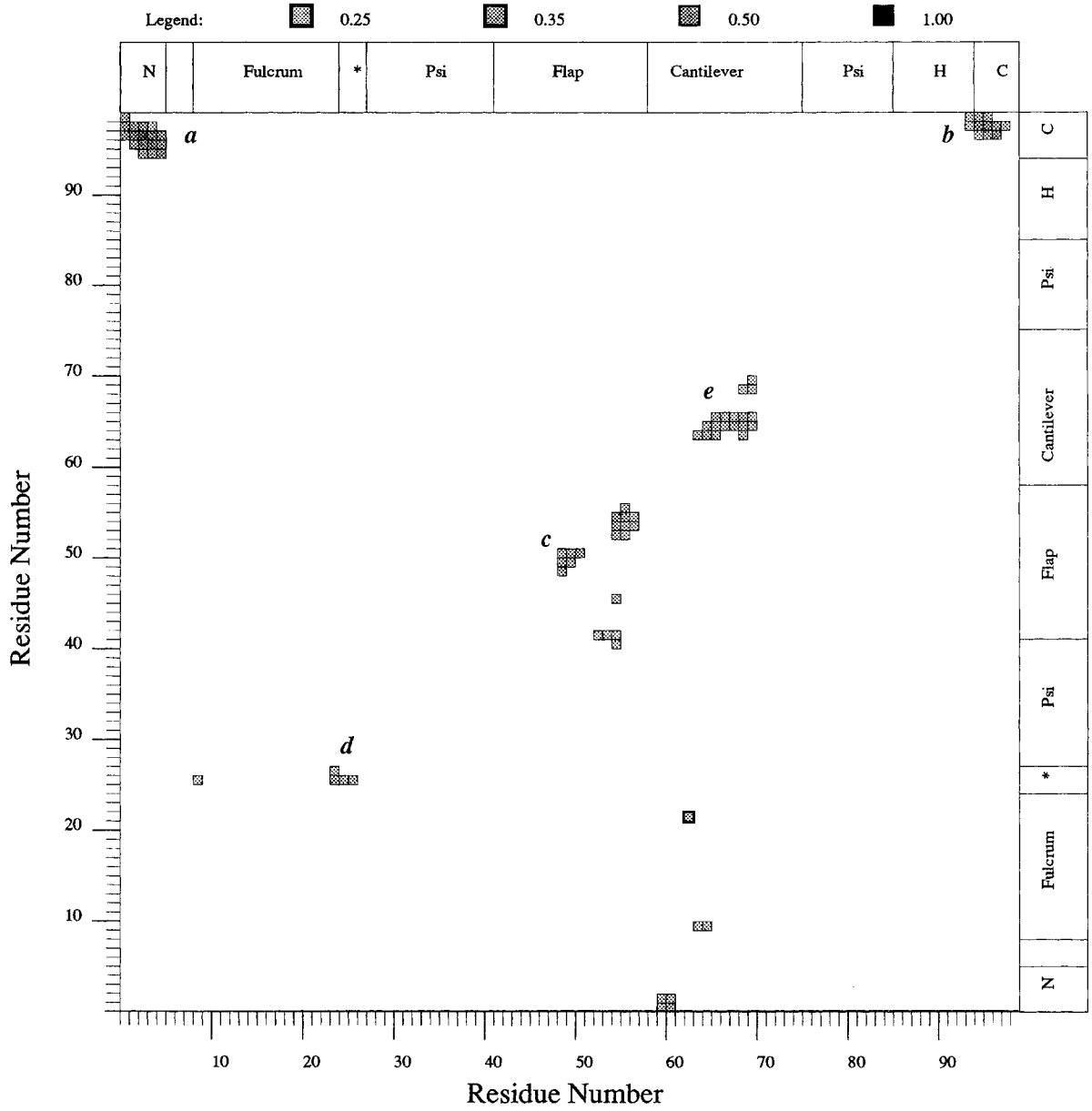

Fig. 16. Calculated DCCM between monomeric units of the HIV-1 protease dimer, averaged over three 25-psec intervals in the MD simulation (solvated model). See Figure 15 legend for details.

as evidenced by region (k) on Figure 14 are maintained.

SUMMARY AND CONCLUSIONS

A 100 psec trajectory for HIV-1 protease in water has been obtained from MD simulation and found to give a quite reasonable overall account of the observed crystal structure. Corresponding simulations neglecting the solvent water were found to be in poorer accord with experiment, indicating that explicit consideration of water is necessary to obtain an accurate dynamical model of HIV-1 PR. Detailed conformational and helicoidal analysis was carried out to determine the exact nature of the dynamical model. The main point of discrepancy with the crystal structure was found to be in the flap region, which is influenced in the crystal by intermolecular contacts not available to the protein in our simulation. The flap of one monomer then forms hydrogen bonds with the flap of the other monomer, which can be considered a prediction of aqueous solution behavior of the protein dimer.

The cross-correlation analysis calculated over various intervals of time indicate that the basic domain

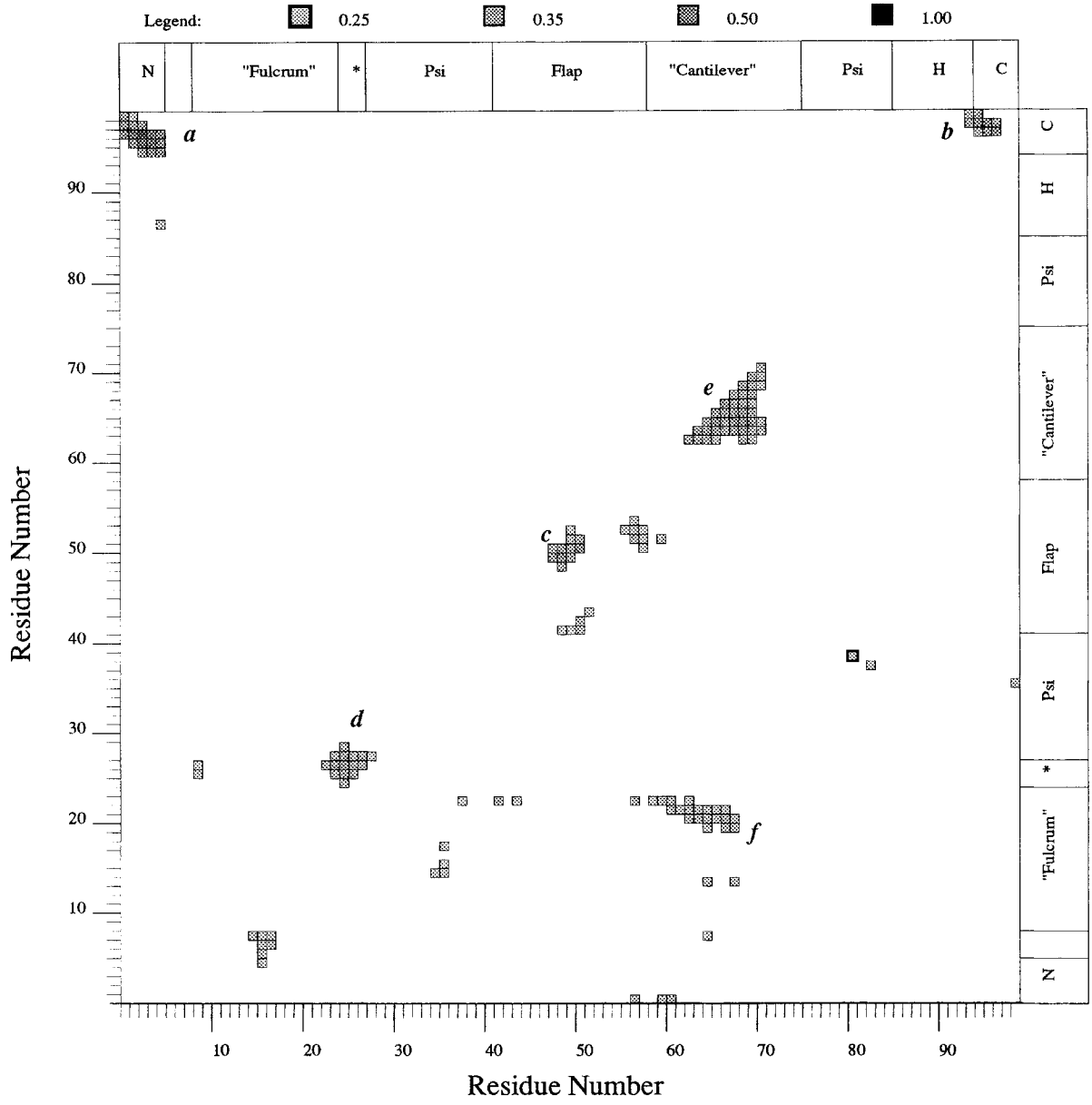

Fig. 17. Calculated DCCM between monomeric units of the HIV-1 protease dimer, averaged over two 40-psec intervals in the MD simulation (solvated model). See Figure 15 legend for details.

structure of the protein is well established on a 5 psec time scale, and the domain-domain correlations do not emerge until the 25–40 psec timescale. Of course, other significant domain correlations may be present on a timescale beyond the covered by the present trajectory. The correlations for through-space interactions obtained in the absence of solvent are more extensive than seen for the solvated case and are thought to be a consequence of the unphysical contraction of structure in the in vacuo case.

The cross-correlation analysis revealed significant aspects of structure originating uniquely in the dynamical motions of the molecule. In particular, an unanticipated through-space, domain-domain correlation was found between the mobile flap region covering the active site and a remote region of the structure, which act collectively somewhat like a molecular cantilever. The active site triad characteristic of aspartyl proteases is correlated with a region of helix near the C-terminus. These correlations involving active site regions of the protease indicate that dynamical structure could play an integral role in the mechanism of catalysis and the functional enegetics of HIV-1 PR processes. The

next step in this project will be a similar study of HIV-1 PR inhibitor complexes and a model of the activated complex for bound substrate, mapping out the dynamical structural changes and differential correlations expected on complexation and catalysis.

ACKNOWLEDGMENTS

This research was supported by Grant GM 37909 from the National Institutes of Health and a Cooperative High Technology Research and Development from the State of Connecticut to Wesleyan University and Bristol-Myers Squibb. Computer facilities were provided by Wesleyan University and the Pittsburgh Supercomputer Center. The atomic coordinates of the crystal geometry of HIV-1 PR (3HVP) were generously provided in advance of release to the Brookhaven Data Bank by Dr. Alex Wlodawer.

REFERENCES

1. Kohl, N., Emini, E.A., Schlief, W.A., Davis, L.J., Heimbach, J., Dixon, R.A.F., Scolnik, E.M., Sigal, I.S. Active human immunodeficiency virus protease is required for viral infectivity. Proc. Natl. Acad. Sci. (USA) 85:4686–4690, 1988.
2. Navia, M.A., Fitzgerald, P.M.D., McKeever, B.M., Leu, C.-., Heimbach, J.C., Herber, W.K., Segal, I.S., Darke, P.L., Springer, J.P. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature 337:615–620, 1989.
3. 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. Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease. Science 245:616–621, 1989.
4. Lipatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J.R., Whittle, P.J., Danley, D.E., Geogharan,K.F., Hawrylik, S.J., Lee, S.E., Scheld, K.G., Hobart, P.M. X-ray analysis of HIV-1 protease at 2.7 A resolution confirms structural homology among retroviral enzymes. Nature (London) 342:299–301, 1989.
5. Yarchoan, R., Mitsuya, H., Broder, S. AIDS therapies. Sci. Am. 259:110–119, 1988.
6. Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H., Wlodawer, A. Structure of a complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. Science 246:1149–1151, 1989.
7. Harte Jr., W.E., Swaminathan, S., Mansuri, M.M., Martin, J.C., Rosenberg, I.E., Beveridge, D.L. Domain communication in the dynamical structure of HIV-1 protease. Proc. Natl. Acad. Sci. (USA) 87:8864–8868, 1990.
8. Swaminathan, S., Harte Jr., W.E., Beveridge, D.L. Identification of domain structure in proteins via molecular dynamics simulation: Application to HIV-1 protease. J. Am. Chem. Soc. 113:2717–2721, 1991.
9. Baum, R.M., Dagani, R. AIDS vaccine, drug research advancing on several fronts. Chem. & Eng. News 68(50):7–15, 1990.
10. Davies II, J.F., Hostomska, Z., Hostomsky, Z., Jordan, S.R., Matthews, D.A. Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. Science 252:88–95, 1991.
11. Wang, J.H., Yan, Y.W., Garrett, T.P.J., Liu, J.H., Rodgers, D.W., Garlick, R.L., Tarr, G.E., Husain, Y., Reinherz, E.L., Harrison, S.C. Atomic structure of a fragment of human CD4 containing 2-immunoglobulin-like domains. Nature 348:411–418, 1990.
12. Ryu, S.E., Kwong, P.D., Truneh, A., Porter, T.G., Arthos, J., Rosenberg, M., Dai, X.P., Xuong, N.H., Axel, R., Sweet, R.W., Hendrickson, W. Crystal structure of an HIV binding recombinant fragment of human CD4. Nature 348: 419–425, 1990.
13. Skalka, A.M. Retroviral proteases: First glimpses of the anatomy of a processing machine. Cell 56:911–913, 1989.
14. Oroszalan, S. The biology and biochemistry of retroviral proteases. 2d International Conference on Drug Research in Immunologic and Infectious Diseases: AIDS. 1989.
15. Katoh, I., Yoshinaka, Y., Rein, A., Shibuya, M., Odaka, T., Oroszalan, S. Marine leukemia virus maturation: Protease region required for conversion from "immature" to "mature" core form and for virus infectivity. Virology 145: 280–292, 1985.
16. Crawford, S., Goff, S. A deletion mutation in the 5′ part of the pol gene of moloney murine leukemia virus blocks proteolytic processing of the gag and pol polyproteins. J. Virology 53:899–907, 1985.
17. Toh, H., Kikuno, R., Hayashida, H., Miyata, T., Kugimiya, W., Inouye, S., Yuki, S., Saigo, K. Close structural resemblance between putative polymerase of a drosophila transposable genetic element 17.6 and pol gene product of MOloney murine leukaemia virus. EMBO J. 4:1267–1272, 1985.
18. Power, M.D. Nucleotide sequence of SRV-1, a type D simian aquired immune deficiency syndrome retrovirus. Science 231:1567–1572, 1986.
19. Pearl, L.H., Taylor, W.R. A structural model for the retroviral proteases. 329:351–354, 1987.
20. Blundell, T., Pearl, L. A second front against AIDS. Nature 337:596–597, 1989.
21. Teng, M., Usman, N., Frederick, C.A., Wang, A.H. H 33258. Nucl. Acids Res. 16:2671–2690, 1988.
22. Miller, M., Jaskolski, M., Rao, J.K.M., Leis, J., Wlodawer, A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. Nature 337:576–579, 1989.
23. Weber, I.T., Miller, M., Jaskolski, M., Leis, J., Skalka, A.M., Wlodawer, A. Molecular modeling of the HIV-1 protease and its substrate binding site. Science 243:928–931, 1988.
24. Navia, M. Three dimensional structure of the aspartyl protease of HIV-1, 2d International Conference on Drug Research in Immunologic and Infectious Diseases: AIDS. 1989.
25. McCammon, J.A. Protein dynamics. Rept. Prog. Phys. 47: 1–46, 1984.
26. Brunger, A.T., Brooks, C.L., Karplus, M. Active site dynamics of ribonuclease. Proc. Natl. Acad. Sci. (USA) 82: 8458–8462, 1985.
27. Nishikawa, T., Go, N. Normal modes of vibration in bovine pancreatic trypsin inhibitor and its mechanical property. Proteins 2:308–329, 1987.
28. Ichiye, T., Karplus, M. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins 11:205–217, 1991.
29. Karplus, M., Kushick, J.N. Method for estimating the configurational entropy of macromolecules. Macromolecules 14:325–332, 1981.
30. Swaminathan, S. WESDYN. 1990.
31. van Gunsteren, W.F., Berendsen, H.J.C. GROMOS86: Groningen Molecular Simulation System. 1986.
32. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Hermans, J. "Interaction models for water in relation to protein hydration." Intermolecular Forces, Pullman ed. Dordrecht: D. Reidel, 1981.
33. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. Equation of state calculations by fast computing machines. J. Chem. Phys. 21(6):1087–1092, 1953.
34. van Gunsteren, W.F., Berendsen, H.J.C. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. Angew. Chem. Int. Ed. Engl. 29:992–1023, 1990.
35. Ravishanker, G., Swaminathan, S., Beveridge, D.L., Lavery, R., Sklenar, H. Conformational and helicoidal analysis of 30 psec of molecular dynamics on the d(CGCGAATTCGCG) double helix. J. Biomol. Struct. Dyn. 6:669–699, 1989.
36. Klyne, W., Prelog, V. Description of steric relationships across single bonds. Experimentia 16:521–523, 1960.

37. Sklenar, H., Etchebest, C., Lavery, R. Describing protein structure: A general algorithm yielding complete helicoidal parameters and a unique overall axis. Proteins 6:46–60, 1989.
38. Swaminathan, S., Ravishanker, G., Beveridge, D.L., Lavery, R., Etchebest, C., Sklenar, H. Conformational and helicoidal analysis of the molecular dynamics of proteins. Proteins 8:179–193, 1990.
39. McCammon, A.J., Harvey, S.C. "Dynamics of Proteins and Nucleic Acids." Cambridge: Cambridge University Press, 1986.
40. van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hol, W.G.J., Postma, J.P.M. Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray data. Proc. Natl. Acad. Sci. (USA) 80:4315–4319, 1983.
41. Berendsen, H.J.C., van Gunsteren, W.F., Swinderman, H.R.J., Geurtsen, R.G. Simulations of proteins in water. Ann. N.Y. Acad. Sci. 482:269–286, 1986.
42. Levitt, M., Sharon, R. Accurate simulation of protein dynamics in solution. Proc. Natl. Acad. Sci. (USA) 85:7557–7561, 1988.
43. Avbelj, F., Moult, J., Kitson, D.H., James, M.N.G., Hagler, A.T. Molecular dynamics study of the structure and dynamics of a protein molecule in a crystalline ionic environment, streptomyces griseus protease A. Biochemistry 29:8658–8676, 1990.
44. Loeb, D.D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S.E., Hutchinson III, C.A. Complete mutagenesis of HIV-1 protease. Nature 340:397–400, 1989.
45. Welch, G.R. The fluctuating enzyme: Nonequilibrium properties in the physical sciences and biology. V: 496, 1986.