JCTC Journal of Chemical Theory and Computation

Study of the Conformational Dynamics of the Catalytic Loop of WT and G140A/G149A HIV-1 Integrase Core Domain Using Reversible Digitally Filtered Molecular Dynamics

Sarah L. Williams and Jonathan W. Essex*

School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K.

Received May 14, 2008

Abstract: The HIV-1 IN enzyme is one of three crucial virally encoded enzymes (HIV-1 IN, HIV-1 PR, and HIV-1 RT) involved in the life-cycle of the HIV-1 virus, making it an attractive target in the development of drugs against the AIDS virus. The structure and mechanism of the HIV-1 IN enzyme is the least understood of the three enzymes due to the lack of three-dimensional structural information. X-ray cystallographic studies have not yet been able to resolve the full-length structure, and studies have been mainly focused on the catalytic domain. This central domain possesses an important catalytic loop observed to overhang the active site, and experimental studies have shown that its dynamics affects the catalytic activity of mutant HIV-1 IN enzymes. In this study, the enhanced sampling technique, Reversible Digitally Filtered Molecular Dynamics (RDFMD), has been applied to the catalytic domain of the WT and G140A/G149A HIV-1 IN enzymes and has highlighted significant differences between the behavior of the catalytic loop which may explain the decrease of activity observed in experimental studies for this mutant.

1. Introduction

The HIV-1 integrase (HIV-1 IN) enzyme is one of three key enzymes involved in the life-cycle of the HIV-1 virus. The full length HIV-1 integrase enzyme comprises 288 residues which, based on partial proteolysis, functional, and structural studies, can be divided into three domains, the N- and C-terminal domains and the catalytic domain.[1−3] The N-terminal domain (residues 1−50) consists of three α-helices and a zinc binding site where conserved histidine and cysteine residues coordinate a zinc ion, stabilizing the interaction between the helices. The binding of zinc is thought to promote the multimerisation of the enzyme which enhances the enzyme's activity.[4−6] The third, C-terminal domain (residues 212−288) is a nonspecific binding site of DNA[7−11] and additionally contributes to the multimerization, which is essential to the integration process.[12] In isolation, both the C- and N-terminal domains are dimeric in solution, but the C-terminal is monomeric when attached to the catalytic domain.[13]

The central catalytic domain (residues 50−212) is composed of a mix of α-helix and β-sheet structures (Figure 1). This domain is the most conserved of the three and contains the active motif, comprising residues Asp64, Asp116, and Glu152 (DDE). The domain possesses a catalytic loop (residues 140−149) which overhangs the active site and is postulated to play an important role in the catalytic activity of the enzyme through its role in positioning the substrates in the active site for processing.

The HIV-1 IN enzyme catalyzes two stages of the HIV-1 virus life-cycle which occur after the viral-RNA genome is reverse transcribed to produce the double-stranded DNA,[14,15] a process carried out by the HIV-1 reverse transcriptase (HIV-1 RT) enzyme. The first, 3′-processing stage involves the removal of 2 deoxynucleotides from each of the 3′-ends of the viral DNA, and the second, strand transfer stage then covalently ligates these processed 3′-ends to the host chromosomal DNA via transesterification reactions. The final

* Corresponding author e-mail: J.W.Essex@soton.ac.uk.

10.1021/ct800162v CCC: $40.75
Published on Web 01/23/2009

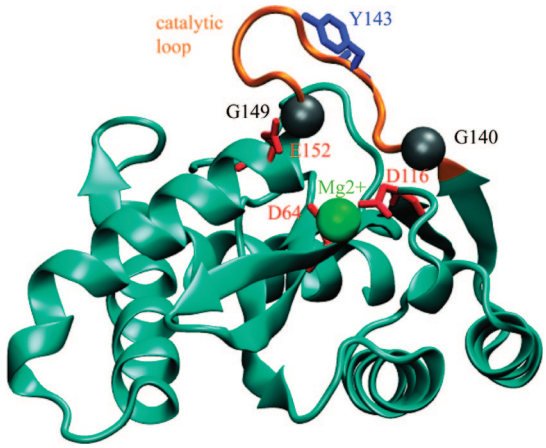

Figure 1. Catalytic core domain of the HIV-1 integrase enzyme (PDB ID: 1BL3) with the following highlighted residues: active site residues, D64, D116, and E152 (red), catalytically important tyrosine residue, Y143 (blue), and α-carbon atoms of hinge residues, G140 and G149 (silver), Mg^{2+} ion (green). The catalytic loop residues (140−149) are shown in orange.

product is the provirus, which comprises the intact double-stranded host DNA containing the inserted viral genetic information.

All three of the domains of the full-length HIV-1 IN are required to carry out the catalytic function of the enzyme.[2] However, in isolation, the catalytic domain has the ability to initiate the process of disintegration *in vitro*.[16,17] Disintegration is the reverse process of strand transfer, whereby the catalytic domain is able to release partially integrated viral DNA from the host DNA.

Owing to the same active site being used for the two distinct types of processes which utilize different substrates, it is likely that the active site undergoes a conformational change for each stage. Evidence for this hypothesis is given by the ability of diketo acid inhibitors to selectively inhibit the strand transfer while not affecting the 3′-processing process.[18]

The arrangement of the three domains in the full-length integrase enzyme still remains unknown due to the low solubility and the tendency of the enzyme to aggregate. Therefore, to gain structural information, the three domains have been determined and studied in isolation by crystallographic[19−22] and NMR methods.[23−26] The catalytic core was the first domain to be determined, and there are currently approximately 15 structures of this domain available in the Protein Data Bank (PDB),[27] each containing either the F185K or F185H solubility enhancing mutation. A few, more recent studies have published a two-domain structure comprising the catalytic core and C-terminal domains.[28−30] As with the determination of the core domain in isolation, not all residues have been resolved, with missing or poorly defined loop residues. It is believed that HIV-1 IN requires at least one divalent metal ion (Mg^{2+} or Mn^{2+}), placed between the carboxylate groups of catalytic residues Asp64 and Asp116 in the active site (E152 is not involved),[31−35] and an activated water to behave as a nucleophile to be fully functional. The divalent ions are required for both the reactions and for the formation of the HIV-1 IN complex with the viral DNA.[12,32] However, the first crystallographic studies of the catalytic domain were carried out in the presence of cacodylate (dimethylananic acid),[19] which was critical in the enhancement of the solubility of the enzyme. It is now thought that the presence of this chemical provided a distorted representation of the active conformation where the active site aspartate residues are not close enough to be able to bind a metal ion. Experimental and MD studies have also found the absence of the metal ion to significantly increase the flexibility of the loop, presumably because of the absence of the interactions between the ion and catalytic residues. Disruption of the secondary structure has been noted in some MD studies.[22,36−38]

The location of a second metal ion in the presence of the DNA substrate has been suggested based on the similarities of the function with DNA polymerase, with the second ion being bound by D116 and E152. It is proposed that this ion is required for the stabilization of the active conformation in the presence of DNA.[39] However, in the absence of any structural data of integrase with the DNA substrate, there is currently no evidence to prove the existence of the second metal ion.

Studies of the activity of HIV-1 IN have been focused on the catalytic domain since residues 50−190 are sufficient to promote disintegration in vitro,[16,17] which has been shown to occur irrespective of whether or not cacodylate has been used in the crystallization process.[40] However, the structure of this domain has not yet been fully determined unambiguously by experimental techniques, especially the highly mobile loop region which shows a high degree of disorder, which is vital for the activity of the domain. As a consequence, only a few crystal structures exist with the entire loop resolved.[35,22,41] Bujacz et al.[41] determined a crystallographic structure said to contain the active loop in an extended conformation, with E152 shown to point away from the other two catalytic carboxylates. According to the assumed roles of these residues in the binding of metal ions, based on comparison with the related avian sarcoma virus (ASV) integrase, the authors surmise that the conformation of the loop they observe is not that seen during catalysis. Additionally, this structure does not contain the catalytically important divalent metal ion. In the structure resolved by Maignan et al.,[35] the location of a Mg^{2+} ion has been resolved, and the loop is positioned over the active site in a "closed" and more compact conformation, likely to correspond to an active conformation of the loop. The loop structure overhangs the active site, and although the active conformation adopted by the surface loop during the integration reaction is unknown, correlation has been found between the flexibility of the loop and enzymatic activity.[42] Mutagenesis experiments carried out by Greenwald et al.[42] replaced the loop hinge residues G140 and G149 (individually and as the double mutant) with alanine, resulting in reduced flexibility of the loop, demonstrated by the lower B values. The greatest reduction is seen in the double mutant, and they attribute the diminished catalytic HIV-1 IN activity observed in their experimental studies as a consequence of this decreased mobility. The authors suggest two possible mecha-

nisms causing the changes in enzyme activity, one being the alteration of the equilibrium between the different conformations required at different steps in the catalytic cycle compared with the WT. Alternatively, they suggest that the structure adopted by the constrained loop mutants may represent a nonfunctional conformation which is more stable. However, the crystallographic determination of these structures was carried out in the presence of cacodylate which is known to affect the conformational dynamics of the loop.[36]

1.1. Theoretical Studies of the Conformational Dynamics of the WT and G140A/G149A Mutant HIV-1 IN Core Domain. Lee et al.[43] performed a number of MD and locally enhanced sampling (LES)[44,45] simulations of the HIV-1 IN core domain of the WT and mutant containing loop hinge mutants (G140A, G149A, and G140A/G149A), using the AMBER2003 forcefield[46] and explicit solvent. The WT crystal structure (PDB code: 1QS4)[47] of the HIV-1 IN core domain used in this study contains the Mg^{2+} ion coordinated by the D64 and D116 catalytic residues. The mutant HIV-1 IN structures were created through the mutation of the appropriate residues using the SwissPDB software.[48] This structure has two unresolved loop residues (I141 and P142), which have been modeled in, again using the SwissPDB software[48] from chain B of the 1BIS[22] (PDB code) crystal structure, which contains these missing residues. Initially, a number of short MD simulations were performed on the WT and mutant structures, followed by the extension (to 40 ns) of the most interesting trajectory. Their results report the WT loop to move from its initial open state to a closed state at ~8 ns, in which it remains stable for almost 30 ns. After applying LES to this closed state, the reopening of the loop is seen within 4 ns.

In the same study, MD and LES simulations involving three mutants (G140A, G149A, and the double mutant, G140A/G149A) produced results in agreement with experimental studies of Greenwald et al.,[42] demonstrating the hinge movement of the loop to be less prominent in the structures containing the single mutants, and completely eliminated in the structure possessing the double mutant. They suggest that the opening/closing ability of the loop is vital for catalytic activity, and the mutants studied hinder this loop mobility, thus affecting the activity of the enzyme.

1.2. Role of Tyr143 in Catalysis. The Tyr143 (Y143) residue is situated at the top of the loop of the core domain (Figure 1). Experimental studies have shown the presence of a conserved tyrosine residue in the catalytic domain near the active site in several retroviral integrase structures.[1] Its role in the mechanism of catalysis is based on the proposed structural arrangement of the active site of *E. coli* polymerase, where the residue is suggested to stabilize the activated water molecule.[49] Studies involving the mutation of this residue have resulted in alteration of the preference of the nucleophile during the 3′-processing reaction, from water (3′-processing) to alcohol (strand transfer)[50,51] thus demonstrating its importance. The side chain of the Y143 residue pointing toward the active site is generally assumed to be the active orientation, and simulations carried out by Lee et al.[43] and De Luca et al.[52] suggest that loop flexibility is required to position Y143 in close proximity to the substrate DNA when the loop is closed, thus suggesting a correlation between the function of Y143 and loop dynamics.

In summary, this enzyme is relatively less well understood compared with the HIV-1 protease (HIV-1 PR) and HIV-1 reverse transcriptase (HIV-1 RT) enzymes of the HIV life cycle. Owing to difficulties in obtaining structural information, the catalytic core domain is the main focus of the majority of studies. The catalytic mechanism of this domain is not fully understood, but theoretical and experimental studies have demonstrated the importance of the side chain orientation of Y143 residue in the loop and the dynamics of the loop itself. However, the active form of the domain is still not accurately known, and there is some ambiguity concerning the highly mobile loop conformation and the number of metal ions present. The G140A/G149A mutant HIV-1 IN has been shown by an experimental[42] and theoretical[43] study to diminish the catalytic activity of the apoenzyme through the rigidifying of the loop. This reduction in loop mobility prevents the loop assuming active conformations and prevents the Y143 residue approaching the active site.

A frequent problem in the process of inhibitor design is the emergence of mutant variants affecting the binding ability of the inhibitor. Several studies of mutant forms of the catalytic domain have attributed a variation in enzyme activity and inhibition compared with the WT enzyme to a change in the dynamics of the important catalytic loop structure.[42,43,53−55] In this study, the conformational dynamics of the catalytic domain has been investigated for the WT and G140A/G149A HIV-1 IN enzymes to gain an increased understanding of the operation of this mutation which is reported to significantly decrease the catalytic activity of the enzyme.

2. Methodology

The starting structure, chain C of 1BL3,[35] an apo-form of the wild type catalytic domain of HIV-1 IN, was taken from the Protein Data Bank.[27] This structure was chosen as it was the only crystal structure available at the time possessing all the residues of the loop (residues 140−149) and included the catalytically important Mg^{2+} ion. Three end residues, 210−212, are missing from this structure, but this was not considered significant as they are located away from the active loop, the focus of this study. The WHATIF[56] program was used to add polar hydrogens and to check the structure. The AMBER utility XLEAP[57] was used to add other hydrogen atoms and to solvate the system, with a minimum distance of 12 Å from the protein, in a box of 8693 TIPS3P[58] water molecules. One chloride counterion was added to neutralize the overall charge of the system.

This structure contains two solubility enhancing mutants, F185K and W131E, which were mutated back to their native forms using the SCAP[59,60] software, as it has been suggested that they may cause a deformation of the native structure. A previous experimental study has suggested that the F185K mutant resulted in the mutant protein being more active than the WT[61] and simulations carried out by Lee et al. suggest

the mutation to increase the flexibility of the catalytic loop through disruption of this region.[43]

All simulations, unless otherwise stated, have been carried out using the NAMD[62] molecular dynamics package and the CHARMM27 forcefield.[63]

Minimization was carried out in stages, starting with the protein only (5000 steps), followed by solvent (30,000 steps), ions (1000 steps), solvent and ions (20,000 steps), and finally the entire system (40,000 steps), giving a total of 96,000 steps. Two minimization algorithms were used, initially the steepest descent algorithm, followed by the conjugate gradient method. The minimized system was heated to 300 K in the NVT ensemble. The procedure employed a Langevin thermostat with a 10 ps^{-1} damping parameter. The heating was carried out gradually in stages at 50 K intervals, each interval being 20,000 steps long. Following this, a further 50,000 steps in the NVT ensemble were carried out using a 5 ps^{-1} thermostat damping parameter to control the temperature of the system.

Equilibration simulations, using a Nosé-Hoover barostat in the NPT ensemble were then carried out for 50,000 steps, with a target pressure of 1 atm. A decay parameter of 100 fs and a piston period of 200 fs were used. A further 50,000 steps were run, with a decay parameter of 300 fs and a piston period of 500 fs. The final equilibrated system had box dimensions of 66.57, 68.17, and 61.88 Å.

The apo mutant system, possessing the G140A/G149A double mutant, has also been studied. Since there are no complete crystal structures of this double mutant which possess the Mg^{2+} ion, the equilibrated WT structure was taken, and the appropriate residues mutated using SCAP.[59,60] Careful minimization and equilibration of this mutated structure was carried out before use in simulations. Initially, the mutated residues were minimized for 1000 steps while restraining the rest of the system, followed by the minimization of the rest of the protein for 5000 steps, the solvent and counterion for 1000 steps, and, last, the entire system for 10,000 steps. Heating and equilibration of the system were then carried out as for the WT structure.

The final equilibrated system had box dimensions of 66.63, 68.27, and 61.91 Å.

For each system, one MD production simulation, 20 ns in length, has been carried out. All production MD simulations were run in the NVT ensemble using a 2 fs time step, a Langevin thermostat with a 5 ps^{-1} damping parameter at a temperature of 300 K. Periodic boundary conditions were used, along with a particle mesh Ewald treatment of electrostatic interactions, using a interpolation order of 6, and switching function applied to the Lennard-Jones interactions between 9 Å and the 10.5 Å cutoff. PME gridsizes of 69 × 72 × 64 Å were used, similar values to those of the boxsizes. SHAKE[64] was applied to all bonds containing hydrogen, using a tolerance of 10^{-8} Å.

2.1. RDFMD Simulation Details. Reversible Digitally Filtered Molecular Dynamics (RDFMD) enhances conformational change through amplification of the low frequency motions of specific structural regions of a protein. Prior to this study, the method has been successfully applied to a number of protein systems, including *E. coli* dihydrofolate reductase[65] and apo WT HIV-1 PR.[66] The details of the method have been described previously in the literature,[65,67,68,66] and a protocol of parameters for use on regions of interest in proteins has been developed previously.[65] The parameters are heavily interrelated and some are system dependent, and, therefore, a suitable set of parameters has been optimized through trial and error for the study of the WT and mutant HIV-1 IN systems simulated here. These include the use of a digital filter designed to amplify frequencies between 0−100 cm^{-1} using 201 coefficients, an amplification factor of 2, a temperature cap of 700 K, and a delay between filter applications of either 50 or 100 steps. The filter sequences were separated by 4 ps of molecular dynamics simulation in the NVT ensemble. This is sufficient time for the system temperature to return to 300 K, and it is during this period of time that conformations for analysis are generated. The final results are taken from piecing together the individual 4 ps MD runs. Each RDFMD simulation produces 100 4 ps MD sections, totalling 400 ps. Since the dynamics of the catalytic loop is thought to play a fundamental role in the activity of the enzyme, all of the atoms of residues in this region (140−149) have been selected as the target region of the filter in the RDFMD simulations. Simulations were run in the NVT ensemble using the Langevin thermostat with a 5 ps^{-1} damping parameter.

A total of 12 RDFMD simulations have been carried out using six different starting structures with different velocities. The last stage of the equilibration process was extended for a further 60,000 steps, taking the velocities and starting coordinates after every 10,000 steps, to generate each of the six different starting structures.

3. Results

3.1. Flexibility of the Core Domain. Analysis of the root mean squared fluctuations (RMSF) of the α-carbon atoms of the residues of the catalytic domain over the length of the MD and RDFMD simulations shows two main regions of flexibility in both the WT and mutant HIV-1 integrase enzyme. Both of these regions are loop structures (residues 140−149 and 186−194), with residues 140−149 comprising the catalytic loop. The residues toward the ends of the catalytic domain, which would be connected to the N- and C-terminal domains in the full length HIV-1 IN, are also shown to possess flexibility. The overall relative higher flexibility of these regions is consistent with the profile of B-factors calculated from MD simulations by Lee et al.[43]

Figure 2 compares the RMSF of the core domain of WT and G140A/G149A HIV-1 IN over the length of the MD and RDFMD simulations. In both the MD and RDFMD simulations, the catalytic loop is shown to be slightly more flexible for the mutant HIV-1 IN compared with the WT, although this difference is perhaps negligible in the case of the RDFMD simulations. These results are in contrast to those reported in the theoretical studies of Lee et al.[43] and an X-ray crystallography study by Greenwald et al.[42] which concluded that the double mutant possesses significantly reduced flexibility.

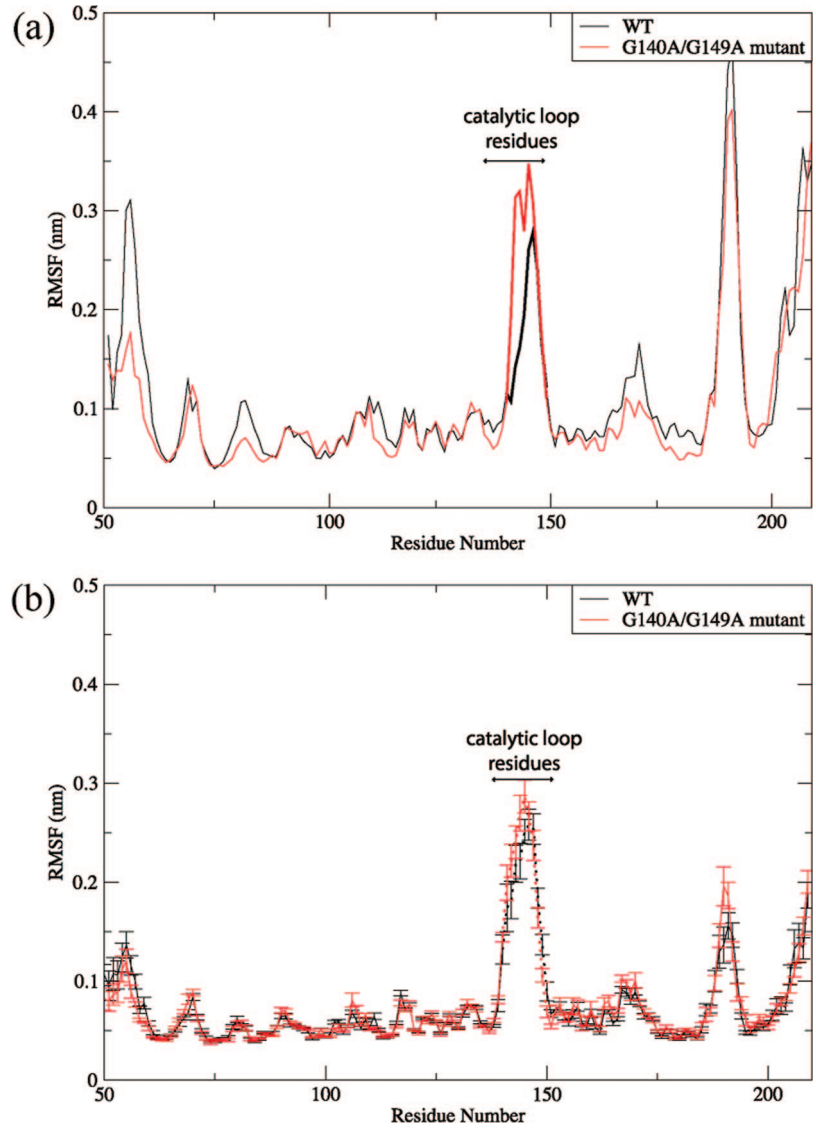

Figure 2. RMSF of the α-carbon atoms of all residues of the core domain for the WT and G140A/G149A mutant: (a) MD simulations and (b) RDFMD simulations. The standard error for each residue for each of the 12 RDFMD simulations has been calculated.

3.2. Conformational Dynamics of WT and G140A/G149A HIV-1 IN Catalytic Loop. Principal component analysis (PCA) was used to identify the major motions of the catalytic loop and highlight any differences in the conformational dynamics which may occur as a result of the mutation from glycine to alanine. In addition, this method has proved useful in the evaluation of the conformational sampling as a result of the RDFMD methodology used. All PCA analysis has been carried out using the utilities provided in tools of the gromacs simulation package.[69,70]

The trajectories of the α-carbon atoms of the catalytic loop (residues 140−149) of each of the HIV-1 IN systems have been used as the data set in the calculation of the covariance matrix for PCA, with fitting carried out using the nonloop residues. The noncatalytic loop residues have been disregarded for this analysis to avoid the dynamics of the region of interest being obscured by the dynamics of the rest of the protein. Additionally, to compare the eigenvectors of the different MD and RDFMD trajectories, the same reference structure has been used; the α-carbon atoms of the first equilibrated WT HIV-1 IN system.

For each of the RDFMD and MD simulations, the number of eigenvectors chosen for study was based on the proportion of the total motion captured and by visual inspection of the motions they represent. As a result, the first two eigenvectors were selected (represent >70% of total motion in MD and RDFMD simulations), and further eigenvectors have been disregarded as they represent higher frequency motions and were harder to define.

3.2.1. Cross-Correlation Analysis. Comparison of the inner-products between two eigenvectors has been used to indicate the level of correlation between them, and correlation coefficents provide quantitative information to describe this correlation. A correlation coefficient value of 1 demonstrates the two eigenvectors being compared to be identical, and a value of 0 means the eigenvectors are orthogonal.

Intially, the twelve RDFMD trajectories of the WT and mutant HIV-1 IN systems were concatenated to form two

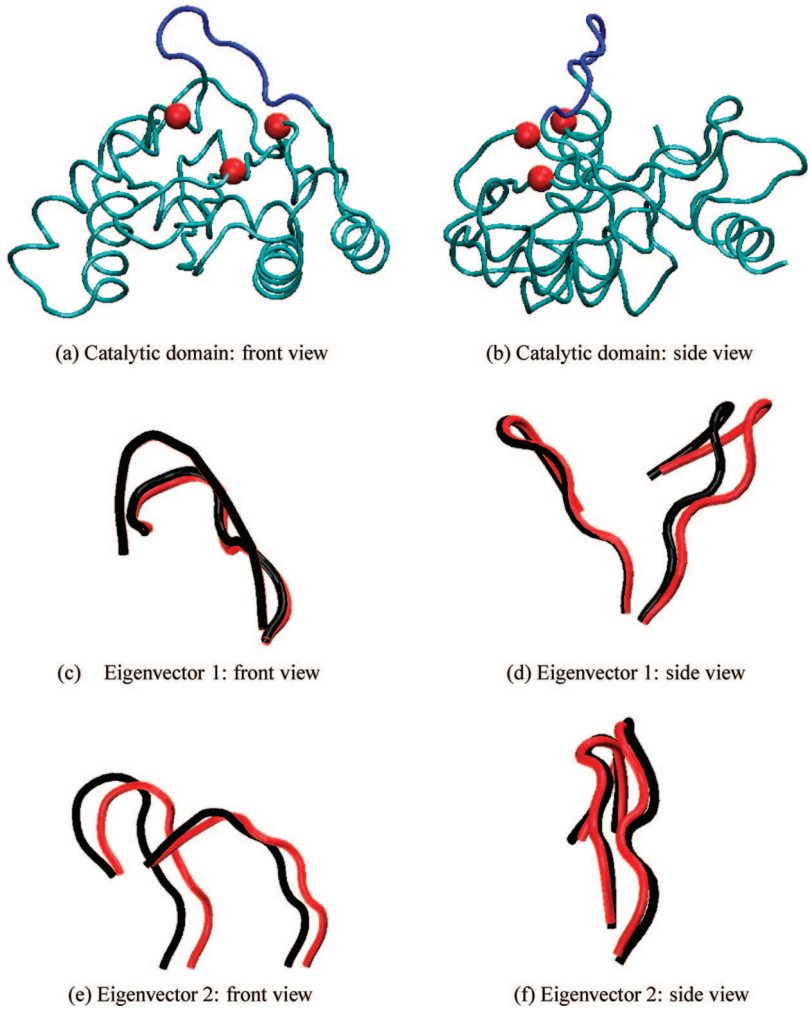

Figure 3. Extreme conformations sampled by the concatenated WT and G140A/G149A HIV-1 IN RDFMD trajectories projected on the first two eigenvectors (black: WT, red: G140A/G149A HIV-1 IN). Eigenvectors generated from the concatenation of the WT and G140A/G149A HIV-1 IN RDFMD trajectories.

single trajectories (WT and mutant), and PCA was performed on the catalytic loop as described earlier. Comparison of the first few eigenvectors between the WT and mutant are shown to be highly comparable (correlation between inner-products of eigenvectors 1: 0.814, eigenvectors 2: 0.686—see the Supporting Information for cross-correlation plot), and so further PCA analysis was carried out on a single trajectory incorporating all the G140A/G149A and WT trajectories together (i.e., 24 RDFMD trajectories concatenated into one). This allows for direct conformational sampling comparisons to be made between the WT and mutant systems (details given later).

Analysis of the eigenvectors generated by MD simulations shows high diagonal correlation with those of the RDFMD simulations for both the WT (correlation coefficient between inner-products of the first eigenvectors: 0.870, second eigenvectors: 0.766—see the Supporting Information for cross-correlation plot) and mutant simulations (correlation coefficient between inner-products of the first eigenvectors: 0.754, second eigenvectors: 0.600). This demonstrates that the catalytic loop undergoes the same types of movement in RDFMD simulations as it does in MD simulations, thus confirming the RDFMD methodology to sample reasonable conformations of the catalytic loop. Owing to the similarity of the eigenvectors generated by MD and RDFMD simulations and to evaluate the sampling of the MD and RDFMD simulations rigorously, the trajectories were projected onto the same eigenvectors. For this purpose, the first two eigenvectors generated by the concatenated WT and mutant RDFMD trajectories were chosen. Although the first two eigenvectors of the MD and RDFMD simulations are similar, the eigenvectors of the RDFMD simulations have been generated through the concatenation of several simulations using several different starting structures, rather than based on just a single simulation, as with the MD simulations.

It is noted that an increase in motion is expected to be observed for the RDFMD simulations compared with the MD simulations as both sets of trajectories are projected onto the eigenvectors of the RDFMD simulations.

Figure 3 shows the extremes of loop motion of the WT and mutant HIV-1 IN loop as a result of the projection of the trajectories onto the first two eigenvectors. The first eigenvector shows the opening/closing gating motion of the catalytic loop toward and away from the active site (Figure 3(c),(d)), a motion thought to be associated with access to the active site, with the loop predicted to overhang the active site in the closed conformation. The second eigenvector shows the loop moving from a more compact structure which

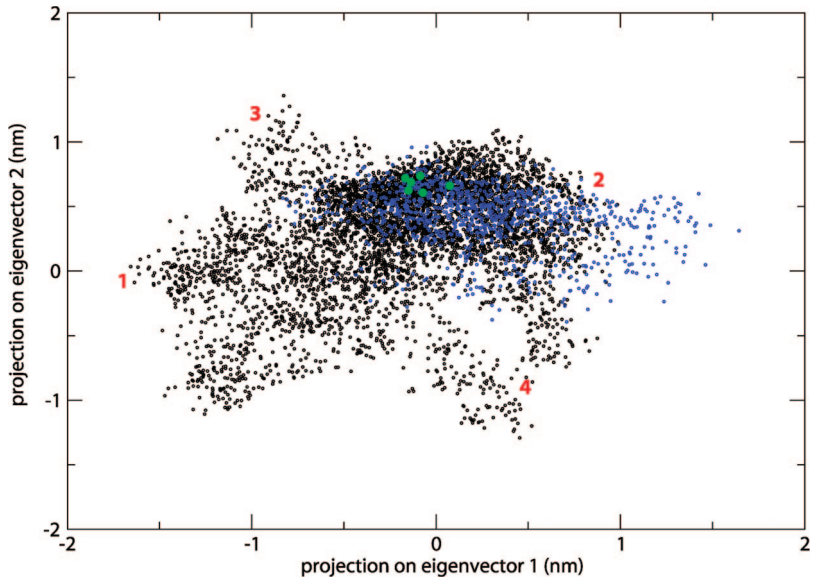

Figure 4. Sampling of the first two eigenvectors, generated from the projection of the WT RDFMD (black) and the MD (blue) simulations onto the eigenvectors of the catalytic loop created from the concatenation of all RDFMD trajectories. (The six different starting structures are represented by green circles). The numbers show the extremes of sampling. Regions marked 1 and 2 on a plot are represented by Figure 3(c),(d) (black), and regions 3 and 4 marked on a plot are represented by Figure 3(e),(f) (black).

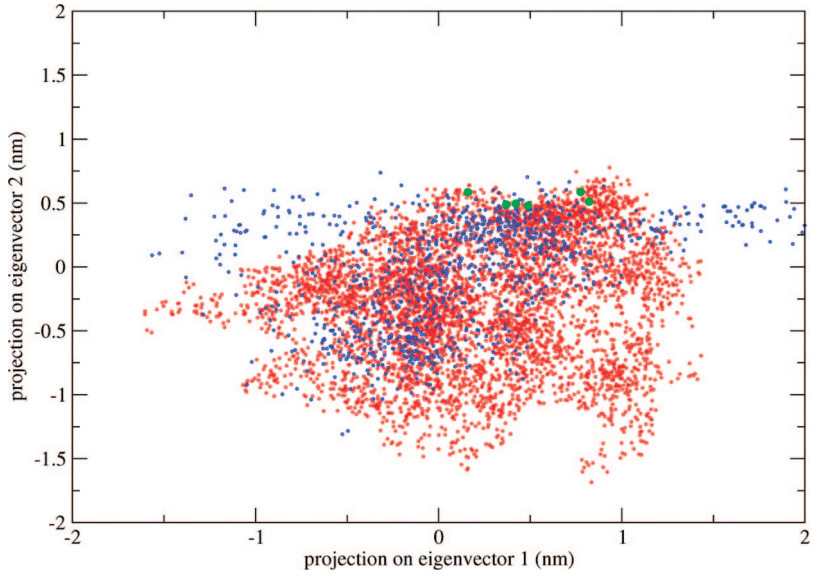

Figure 5. Sampling of the first two eigenvectors, generated from the projection of the G140A/G149A mutant RDFMD (red) and the MD (blue) simulations on the eigenvectors of the catalytic loop created from the concatenation of all RDFMD trajectories. (The six different starting structures are represented by green circles.)

leans over to one side, to a more extended conformation of the loop which spans a larger area (Figure 3(e),(f)).

3.3. Comparison of the Conformations of Catalytic Loop Sampled by the WT and G140A/G149A HIV-1 IN MD and RDFMD Simulations. The WT and G140A/G149A MD and RDFMD trajectories have been projected onto the first two eigenvectors generated from the concatentation of the 24 WT and G140A/G149A HIV-1 IN RDFMD trajectories, resulting in the two-dimensional plots shown in Figures 4 and 5. The sampling of these plots will demonstrate any correlation between these two eigenvectors and will also highlight differences in the conformations sampled by the MD and RDFMD simulations. The numbered regions marked on the two-dimensional plot (Figures 4 and 5) show the extremes of sampling of the first two eigenvectors for the WT and G140A/G149A HIV-1 IN, and the corresponding conformations are visualized in Figure 3.

Sampling along the *x*-axis of the two-dimensional plots demonstrates the sampling of eigenvector 1, with negative values corresponding to the loop moving over the active site (closed conformation) and positive values corresponding to the loop moving backward, away from the active site (open conformation, Figure 3(c),(d)). Values along the *y*-axis demonstrate the sampling of the second eigenvector, with positive values corresponding to a compact loop conformation which leans to one side, and negative values cor-

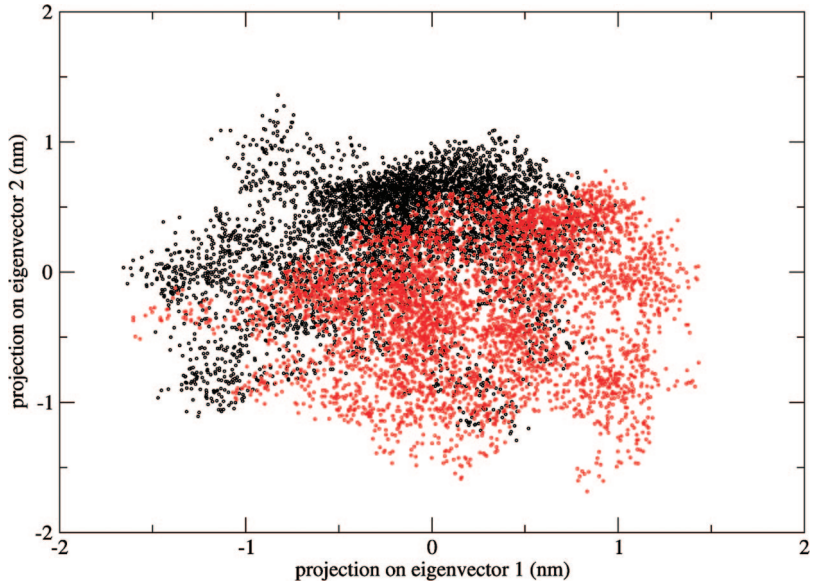

Figure 6. Projection of concatenated WT (black) and G14A/G149A (red) RDFMD simulation trajectories of the loop onto the first and second eigenvectors.

responding to the more extended catalytic loop structure which spans a larger area (Figure 3(e),(f)).

The plots clearly show the RDFMD simulations to sample a greater range of conformations of the catalytic loop compared with the MD simulations, with the sampling by the RDFMD simulations shown to largely encompass the area of sampling of the MD simulations. Owing to the larger amount of conformational space sampled by the RDFMD simulations, comparisons between the loop conformations sampled by the WT and G140A/G149A HIV-1 IN were based on the RDFMD simulations.

Figure 6 overlays the sampling of the WT and mutant RDFMD simulations, demonstrating a difference in the sampling of these two eigenvectors for the WT and G140A/G149A HIV-1 IN RDFMD simulations.

The WT RDFMD simulations show a clear main sampling area where the sampling is shown to be denser on the two-dimensional plot. This is not the case in the mutant RDFMD simulations, whose sampling distribution appears to be more diffuse with no obvious main area of sampling.

Comparison of the sampling of the WT and G140A/G149A HIV-1 IN simulations along eigenvector 1 clearly shows the mutant HIV-1 IN simulations to preferentially sample a larger proportion of open-type conformations (postive values) and fewer closed conformations (negative values) and is able to open further compared to the WT HIV-1 IN simulations (Figure 3(d)). In contrast, the loop of the WT HIV-1 IN is shown to sample significantly more conformations where the loop is closed, overhanging the active site compared with the mutant HIV-1 IN, demonstrated by the greater sampling of the negative values of eigenvector 1.

Assessment of the sampling of the second eigenvector also reveals differences between the loop conformations of the WT and G140A/G149A RDFMD simulations. The WT HIV-1 IN is shown clearly to preferentially sample conformations where the loop is in its compact conformation where it leans to one side (more positive values of eigenvector 2) and to be able to achieve conformations which are more compact compared with those sampled by the RDFMD simulations of the mutant HIV-1 IN loop (see Figure 3(e)). In contrast, the loop of the G140A/G149A HIV-1 IN demonstrates a significantly greater amount of sampling of the extended conformation (negative values of eigenvector 2).

Generally, where the WT RDFMD simulations sample the more compact conformations of the catalytic loop, where it leans to one side, the loop is more likely to be in a more upright position, intermediate between fully open or closed (i.e., the most positive/negative values of eigenvector 1 are not sampled). At these times, the mutant RDFMD HIV-1 IN loop simulations are unable to sample the compact conformations seen in the WT RDFMD simulations and are shown to sample an increased range of more extended loop conformations. Additionally, where the loop is seen to overhang the active site in the closed position (negative values of eigenvector 1), the loop is shown to more likely sample the second eigenvector corresponding to slightly extended loop conformations (Figure 3(c),(d)). This correlation of sampling indicates that, in order for the loop to move into the most closed conformation, which is postulated to be an active conformation, in closer proximity to the active site residues (D116, D64, and E152), the loop cannot exist in the very compact conformations, which are associated with the loop leaning to one side; instead the loop must be at least slightly extended.

Analysis of the ϕ and ψ torsion angles (Figure 7) of the loop hinge residues of the WT (G140/G149) and mutant (A140/A149) reveal the expected restricted dihedral sampling of the larger alanine residue, with virtually no sampling of positive ϕ values for the A140 and A149 residues. This does not appear to impact on the overall flexibility of the mutant loop, as demonstrated by RMSF (Figure 2), with the mutant enzyme demonstrating similar loop flexibility to the WT. Instead, the results indicate a change in the equilibrium of open/closed conformations sampled. Analysis of the trajectories shows the additional methyl group of the alanine compared with the glycine residue to result in steric

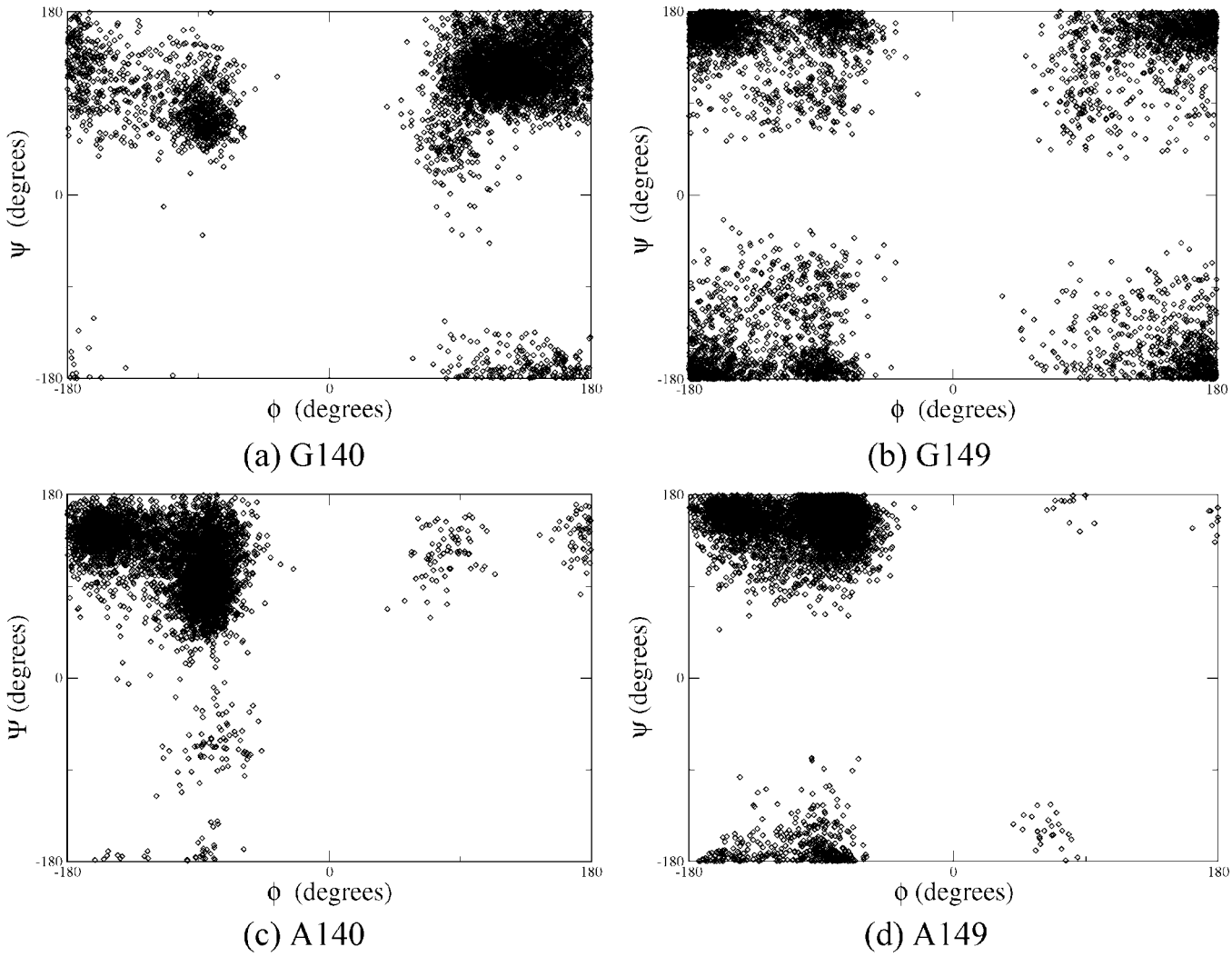

Figure 7. Sampling of ϕ - ψ dihedral angles of the two hinge residues by WT and G140A/G149A HIV-1 IN RDFMD simulations.

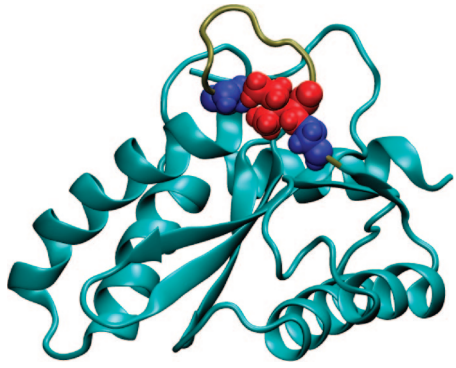

Figure 8. Compact conformation of catalytic loop sampled by WT RDFMD simulations. G140 and G149 residues highlighted in blue van der Waals representation. Ile141 represented in red van der Waals representation.

hindrance with other residues of the loop. An example is shown in Figure 8 where Gly149 and Ile141 are in close contact in the WT, allowing the loop to be in a compact conformation. In the case of the mutant catalyic loop, which is unable to sample such compact loop conformations, the residues would not be able to approach as closely owing to steric repulsion and resulting in a less compact conformation.

In summary, the loop of the G140A/G149A RDFMD simulations can open further than the WT (as also noted by Lee et al.)[43] and preferentially samples the open conformation, whereas the loop of the WT HIV-1 IN RDFMD simulations is able to close over the active site to a greater extent. The variation in the sampling of open/closed conformations seen between the WT and mutant HIV-1 IN loop may be due to the glycine to alanine mutation limiting the formation of the more closed conformations owing to steric repulsion with other residues of the loop.

3.4. Role of Tyrosine 143 (Y143). It has been predicted that the Y143 residue plays an important role in the catalytic activity of this core domain. Based on the activity of 3′−5′ exonuclease of *E. coli* polymerase I,[49] it has been suggested that this residue may be involved in the stabilization of, and in guiding, the nucleophile through a hydrogen-bond interaction, thus assisting in the catalysis of the hydrolytic and phosphoryl transfer reactions. The assumed active side chain conformation of this residue points downward toward the active site, and it has been suggested that the dynamics of the catalytic loop may play a role in the positioning of the Y143 residue. Therefore, a change in the dynamics of the loop relative to the WT HIV-1 IN would affect the catalytic

activity of the domain. The loop of the G140A/G149A HIV-1 IN is shown to be able to open to a greater extent and unable to overhang the active site as was observed in the RDFMD simulations of the WT HIV-1 IN. Therefore, the Y143 would reside further from the active site for a greater amount of time compared with the WT and reduce catalytic activity, whereas the WT samples more closed conformations, thus providing more opportunity for the catalytically important Y143 to be near the active site residues.

4. Conclusions

In this study, the RDFMD technique has been shown to efficiently enhance the range of conformations sampled for the HIV-1 IN enzyme compared with MD simulations. The results of RDFMD simulations highlight differences in the conformational dynamics of the catalytic loop between the WT and G140A/G149A HIV-1 IN which may explain the diminished disintegration activity observed in the presence of the G140A/G149A mutant. The results indicate agreement with previous suggestions proposing the importance of the Y143 residue in the catalytic mechanism and the function of the loop to position this residue in the correct orientation for the functional form.[43,50−52] However, unlike previous studies,[42,43] the mobility of the mutant loop is not reduced compared to the WT; when comparing data from several RDMFD simulations, the results indicate the mobility to be largely similar. It must be remembered, however, that the previous experimental studies[42] were carried out in the presence of cacodylate which is known to affect the conformational dynamics of the loop by preventing the binding of the essential metal ion.

The results of this study indicate the mechanism for the diminished catalytic activity could be due to a difference in the equilibrium between the open/closed conformations of the WT and G140A/G149A HIV-IN catalytic loops. As mentioned, the active conformation of the enzyme is presumed to involve the Y143 side chain being positioned close to the active site, with the hydroxyl group pointing downward. The PCA results show the G140A/G149A HIV-1 IN system to sample a larger number of conformations where the catalytic loop is open. Additionally, the loop is able to open further and is not able to close to the same extent as seen in the RDFMD simulations of the WT HIV-1 IN. The cause of the difference in sampling is postulated to be due to increased steric hindrance between loop residues in the mutant HIV-IN domain, as a result of the larger size of the alanine residue. These differences in conformational sampling of the loop may result in the decreased likelihood of the Y143 side chain being in a suitable location and conformation for the catalytic mechanism to take place in the mutant HIV-1 IN.

Owing to the sampling limitations presented by using conventional MD simulations, the RDFMD methodology has played a crucial role in this study, identifying the proposed differences in dynamics between the WT and G140A/G149A mutant HIV-1 IN.

Acknowledgment. We would like to acknowledge the contributions of S. Phillips, M. Swain, C. Edge, R. Gledhill, C. Woods, and A. Wiley for the development and improvement of the RDFMD method. This work was supported by grants from the ESPRC and BBSRC.

Supporting Information Available: Cross-correlation plots comparing the eigenvectors of WT and G140A/G149A RDFMD simulations and comparing MD and RDFMD simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Engelman, A.; Craigie, R. *J. Virol.* **1992**, *66*, 6361–6369.

(2) Engelman, A.; Bushman, F. D.; Craigie, R. *EMBO J.* **1993**, *12*, 3269–3275.

(3) van Gent, D. C.; Vink, C.; Groeneger, A. A.; Plasterk, R. H. *EMBO J.* **1993**, *12*, 3261–3267.

(4) Lee, S. P.; Xiao, J.; Knutson, J. R.; Lewis, M. S.; Han, M. K. *Biochemistry* **1997**, *36*, 173–180.

(5) Deprez, E.; Tauc, P.; Leh, H.; Mouscadet, J. F.; Auclair, C.; Brochon, J. C. *Biochemistry* **2000**, *39*, 9275–9284.

(6) Zheng, R. J.; Jenkins, T. M.; Craigie, R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13659–13664.

(7) Kahn, E.; Mack, J. P. G.; Katz, R. A.; Kulkosky, J.; Skalka, A. M. *Nucleic Acids Res.* **1991**, *19*, 851–860.

(8) Vink, C.; Groeneger, A. M.; Plasterk, R. H. *Nucleic Acids Res.* **1993**, *21*, 1419–1425.

(9) Woerner, A. M.; Marchis-Sekura, C. J. *Nucleic Acids Res.* **1993**, *21*, 3507–3511.

(10) Engelman, A.; Hickman, A. B.; Craigie, R. *J. Virol.* **1994**, *68*, 5911–5917.

(11) Puras-Lutzke, R. A.; Vink, C.; Plasterk, R. H. *Nucleic Acids Res.* **1994**, *22*, 4125–4131.

(12) Brown, P. O. In *Retroviruses*; Coffin, J. M., Hughes, H. H., Varmus, H. E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1997; pp 161−203.

(13) Chen, J. C. H.; Krucinski, J.; Miercke, L. J. W.; Finer-Moore, J. S.; Tang, A. H.; Leavitt, A. D.; Stroud, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8233–8238.

(14) Brown, P. O.; Bowerman, B.; Varmus, H. E.; Bishop, J. M. *Cell* **1987**, *49*, 347–356.

(15) Lobel, L. I.; Murphy, J. E.; Goff, S. P. *J. Virol.* **1989**, *63*, 2629–2637.

(16) Chow, S. A.; Vincent, K. A.; Eliison, V.; Brown, P. O. *Science* **1992**, *255*, 723–726.

(17) Bushman, F. D.; Engelman, A.; Palmer, I.; Wingfield, P.; Cragie, R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3428–3432.

(18) Espeseth, A. S.; Felock, P.; Wolfe, A.; Witmer, M.; Grobler, J.; Anthony, N.; Egbertson, M.; Melamed, J. Y.; Young, S.; Hamill, T.; Cole, J. L.; Hazuda, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 11244–11249.

(19) Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. *Science* **1994**, *266*, 1981–1986.

(20) Bujacz, G.; Jaskolski, M.; Alexandratos, J.; Wlodawer, A.; Merkel, G.; Katz, R. A.; Skalka, A. M. *J. Mol. Biol.* **1995**, *253*, 333–346.

(21) Bujacz, G.; Jaskolski, M.; Alexandratos, J.; Wlodawer, A.; Merkel, G.; Katz, R. A.; Skalka, A. M. *Structure* **1996**, *4*, 89–96.

(22) Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9150–9154.

(23) Eijkelenboom, A. P.; van den Ent, F. M.; Vos, A.; Doreleijers, J. F.; Hard, K.; Tullius, T. D.; Plasterk, R. H.; Kaptein, R.; Boelens, R. *Curr. Biol.* **1997**, *7*, 739–746.

(24) Eijkelenboom, A. P.; Lutzke, R. A.; Boelens, R.; Plasterk, R. H.; Kaptein, R.; Hard, K. *Nat. Struct. Mol. Biol.* **1995**, *2*, 807–810.

(25) Lodi, P. J.; Ernst, J. A.; Kuszewski, J.; Hickman, A. B.; Engelman, A.; Craigie, R.; Clore, G. M.; Groenenborn, A. M. *Biochemistry* **1995**, *34*, 9826–9833.

(26) Cai, M.; Zheng, R.; Caffrey, M.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. *Nat. Struct. Biol.* **1997**, *4*, 567–577.

(27) Berman, H. M.; Westbrook, Z.; Feng, J.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids. Res* **2000**, *28*, 235–242.

(28) Chen, Z. G. *J. Mol. Biol.* **2000**, *296*, 521–533.

(29) Chen, J. C.; Krucinski, J.; Miercke, L. J.; Finer-Moore, J. S.; Tang, A. H.; Leavitt, A. D.; Stroud, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8233–8238.

(30) Yang, Z. N.; Muserm, T. C.; Bushman, F. D.; Hyde, C. C. *J. Mol. Biol.* **2000**, *296*, 535–548.

(31) Wolfe, A. L.; Felock, P. J.; Hastings, J. C.; Blau, C.; Hazuda, D. J. *J. Virol.* **1996**, *70*, 1424–1432.

(32) Ellison, V.; Brown, P. O. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7316–7320.

(33) Vink, C.; Lutzke, R. A.; Plasterk, R. H. *Nucleic Acids Res.* **1994**, *22*, 4103–4110.

(34) Hazuda, D. J.; Felock, P. J.; Hastings, J. C.; Pramanik, B.; Wolfe, A. L. *J. Virol.* **1997**, *71*, 7005–7011.

(35) Maignan, S.; Guilloteau, J. P.; Zhou-Liu, Q.; Clement-Mella, C.; Mikol, V. *J. Mol. Biol.* **1998**, *282*, 359–368.

(36) Laboulais, C.; Deprez, E.; Leh, H.; Mouscadet, J.; Brochon, J.; Le Bret, M. *Biophys. J.* **2001**, *81*, 473–489.

(37) Lins, R. D.; Briggs, J. M.; Straatsma, T. P.; Carlson, H. A.; Greenwald, J.; Choe, S.; McCammon, J. A. *Biophys. J.* **1999**, *76*, 2999–3011.

(38) Wijitkosoom, A.; Tonmunphean, S.; Truong, T. N.; Hannongbua, S. *J. Biomol. Struct. Dyn.* **2006**, *23*, 613–624.

(39) Lins, R. D.; Adesokan, A.; Soares, T. A.; Briggs, J. M. *Pharmacol. Ther.* **2000**, *85*, 123–131.

(40) Molteni, V.; Greenwald, J.; Rhodes, D.; Hwang, Y.; Kwiatkowski, W.; Bushman, F. D.; Siegel, J. S.; Choe, S. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *57*, 536–544.

(41) Bujacz, G.; Alexandratos, J.; Qing, Z. L.; Clement-Mella, C.; Wlodawer, A. *FEBS Lett.* **1996**, *398*, 175–178.

(42) Greenwald, J.; Le, V.; Butler, S. L.; Bushman, F. D.; Choe, S. *Biochemistry* **1999**, *38*, 8892–8898.

(43) Lee, M. C.; Deng, J.; Briggs, J. M.; Duan, Y. *Biophys. J.* **2005**, *88*, 3133–3146.

(44) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161–9175.

(45) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277–9287.

(46) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiang, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(47) Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13040–13043.

(48) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714–2723.

(49) Beese, L. S.; Steitz, T. A. *EMBO J.* **1991**, *10*, 25–33.

(50) van Gent, D. C.; Groeneger, A. A.; Plasterk, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9598–9602.

(51) van Gent, D. C.; Groeneger, A. A.; Plasterk, R. H. *Nucleic Acids Res.* **1993**, *21*, 3373–3377.

(52) De Luca, L.; Vistoli, G.; Pedretti, A.; Barreca, M. L.; Chimirri, A. *Biochem. Biophys. Res. Commun.* **2005**, *336*, 1010–1016.

(53) Barreca, M.; Lee, W.; Chimirri, A.; Briggs, J. M. *Biophys. J.* **2003**, *84*, 1450–1463.

(54) Brigo, A.; Lee, W.; Mustata, G. I.; Briggs, J. M. *Biophys. J.* **2005**, *88*, 3072–3082.

(55) Brigo, A.; Lee, W.; Fogolari, F.; Mustata, G. I.; Briggs, J. M. *Proteins* **2005**, *59*, 723–741.

(56) Vriend, G. *J. Mol. Graphics* **1990**, *8*, 52–56.

(57) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1–41.

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

(59) Xiang, Z.; Honig, B. *J. Mol. Biol.* **2001**, *311*, 421–430.

(60) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. *J. Mol. Biol.* **2002**, *320*, 597–608.

(61) Jenkins, T. M.; Hickman, A. B.; Dyda, F.; Ghirlando, R.; Davies, D. R.; Craigie, R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6057–6061.

(62) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283–312.

(63) MacKerell, A. D.; Bashford, D.; Bellott, D.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joesph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(64) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(65) Wiley, A. P.; Swain, M. T.; Phillips, S. C.; Edge, C. M.; Essex, J. W. *J. Chem. Theory. Comput.* **2005**, *1*, 24–35.

(66) Wiley, A. P.; Williams, S. L.; Essex, J. W. Unpublished.

(67) Phillips, S. C.; Swain, M. T.; Wiley, A. P.; Essex, J. W.; Edge, C. M. *J. Phys. Chem. B* **2003**, *107*, 2098–2110.

(68) Phillips, S. C.; Essex, J. W.; Edge, C. M. *J. Chem. Phys.* **2000**, *112*, 2586–2597.

(69) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.

(70) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

CT800162V