

# Molecular Dynamics Simulations of the HIV-1 Integrase Dimerization Interface: Guidelines for the Design of a Novel Class of Integrase Inhibitors

Martin Sippel and Christoph A. Sotriffer\*

Institute of Pharmacy and Food Chemistry, Julius-Maximilians-University Wuerzburg, Am Hubland, D-97074 Wuerzburg, Germany

Received October 16, 2009

HIV-1 integrase (IN) is a validated target of anti-AIDS research. The classical approach of designing active-site directed ligands has largely been exploited. A promising alternative strategy to inactivate the enzyme is to prevent the formation of IN dimers. The rational design of dimerization inhibitors, however, is hampered by the lack of relevant structural data about the targeted monomeric form. Therefore, we performed molecular dynamics simulations and subsequent analyses to gain insight into the structural features of the IN catalytic-core-domain dimerization interface. As a result, the formation of a groove and a cavity along the dimerization interface of the IN monomer could be revealed. Both were shown to be suited for accommodating an inhibitory peptide. The results form a valuable basis for the design of ligands targeting the dimerization interface and, thus, of a whole new class of HIV-1 integrase inhibitors.

## INTRODUCTION

HIV-1 integrase (IN) is an essential enzyme of the HIV life cycle, catalyzing the integration of viral DNA into the host genome.<sup>1</sup> Current drugs for antiretroviral treatment address virtually all steps of the viral life cycle: Cell entry is blocked by the fusion inhibitor maraviroc and the entry inhibitor enfuvirtide; transcription is hindered by nucleoside and non-nucleoside reverse transcriptase inhibitors; and proteolysis of the viral polyprotein is addressed by numerous HIV-protease inhibitors.<sup>2,3</sup> As the first and only marketed drug directed against integration, the IN inhibitor Raltegravir (brand name Isentress) was announced in October 2007,<sup>4,5</sup> after two decades of IN research. It confirmed IN as a pivotal target in antiretroviral therapy but also underlined the difficulties associated with anti-IN drug development. Interestingly, computational methods including molecular dynamics (MD) simulations and docking studies played an early role in the development of this drug.<sup>6</sup>

The vast majority of the currently known inhibitors was developed with the intention to address the active site,<sup>7</sup> and much effort is still put into the investigation of their detailed mode of action to establish a comprehensive model for active-site inhibitor design; an example for this is given by the diketo acid derivatives.<sup>8,9</sup> However, alternative strategies for inhibition can be derived from the fact that IN exerts its function as a multimer. For the concerted integration of viral DNA into host DNA, an integrase tetramer is required, whereas integrase dimers are sufficient for the 3'-processing reaction. Higher-order oligomers are formed in the preintegration complexes.<sup>10,11</sup> Biasing the equilibrium between the different states of multimerization was shown to result in inhibition.<sup>12</sup> Thus, the IN dimerization interface is entering the focus of interest in IN research. On the one hand, small-molecule inhibitors binding to the catalytic-core-domain

dimer interface of the intact dimer have been reported;<sup>13–15</sup> for one of these inhibitors, a structural analog of chicoric acid, the mechanism of inhibition has been linked to a dynamic modulation of IN multimeric structure.<sup>15</sup> On the other hand, the dimerization of the catalytic core domain itself may be targeted, possibly resulting in a perturbed monomer–dimer equilibrium leading to catalytically inactive monomers. So far, only a few peptide inhibitors derived from the dimer interface follow this strategy.<sup>12,16,17</sup>

At first glance, addressing protein–protein interfaces with small, druglike molecules seems not feasible because of the large interaction surfaces. However, recent studies have shown that protein–protein interactions are determined by small subsets of amino acids (“functional epitopes”) rather than by the complete interaction surface.<sup>18</sup> For the design of small-molecule inhibitors, this means that it is not necessary to cover the entire protein–protein interface; instead, it is essential to mimic the functional epitope to disrupt the protein–protein interaction. These findings are supported by numerous examples of small molecules binding to protein–protein interfaces<sup>19,20</sup> and modulating protein–protein interactions among a variety of therapeutic targets in oncology and virology.<sup>21–23</sup>

For the structure-based design of such inhibitors, detailed structural knowledge of the targeted protein–protein interface is essential. In the case of HIV integrase, one obstacle to the exploitation of the interface region for inhibitor design is that little is known about the structural details of this region, in particular with respect to conformational adaptations upon dimerization. In this study, we present one of the first MD simulations which puts its focus on the IN dimerization interface. In contrast to the recent study of Hu et al.,<sup>24</sup> who used MD simulations to calculate binding energies and to propose binding locations of a large inhibitory peptide at the interface, the focus of our work is to reveal detailed interaction sites on the IN dimerization interface that can be addressed by small-molecule dimerization disruptors.

\* Corresponding author phone: +49 931 3185443; fax: +49 931 3185494; e-mail: sotriffer@uni-wuerzburg.de.

Hence, the goal of our study is to elucidate the structural features of the interface by means of MD simulations in order to derive valuable information for subsequent structure-based design efforts. The study follows the rationale that the interface conformation as seen in the (static) IN dimer crystal structure does not necessarily correspond to the conformations shown by the solvated monomer. A small molecule binding to the monomer will primarily “see” these monomer conformations and tend to stabilize them, whereas it is rather unlikely that it may be able to stabilize the same conformation as the large protein binding partner in the dimer (unless the interface is intrinsically very rigid). Accordingly, we first analyze the structural dynamics of the interface region in the monomer by means of MD simulations of the monomer (as well as of the dimer, for reference purposes). Subsequently, trajectory snapshots are selected as input structures for hot-spot analyses and docking studies of a hexameric inhibitory peptide binding to the dimerization interface.<sup>16,17</sup> The hot spots and the suggested binding mode of the peptide indicate crucial protein–ligand interactions, which serve to derive guidelines for the design of small-molecule dimerization inhibitors as a new class of IN inhibitors.

## METHODS

**Protein Preparation.** The structure of the IN catalytic-core-domain monomer used for the MD simulations corresponds to subunit A of PDB file 1QS4,<sup>25</sup> which was set up as in previous docking studies.<sup>26</sup> Five crystallographic water molecules (1035, 1229, 1234, 1263, and 1340) were retained due to their important role in stabilizing a cavity at the dimer interface (as further outlined in the Results and Discussion section). Hydrogens were added to the structure using the *t leap* module of AMBER9<sup>27</sup> after removing the ligand SCITEP from the structure but retaining the Mg<sup>2+</sup> ion observed in the active site. Despite its role in the mechanism of the catalytic reaction, a second Mg<sup>2+</sup> ion was not used in the simulations. As pointed out by Bacchi et al.,<sup>9</sup> only one Mg<sup>2+</sup> ion has been detected in the active site in crystallographic studies, and the presence of the viral DNA is believed to be critical for the binding of the second metal ion. Accordingly, substrate-free IN (as simulated in our study) should be modeled with only one metal ion bound to its active site.

Using the *t leap* module, parameters of the Amber 2003 force field<sup>28</sup> were assigned. A short energy minimization of 200 steps was performed using a generalized Born implicit solvent model.<sup>29,30</sup> The minimized protein structure was again loaded into *t leap* and solvated in a rectangular water box with a minimum distance of 8 Å between solute and box wall using pre-equilibrated TIP3P water boxes.<sup>31</sup> This gave a 65 × 65 × 50 Å simulation box with 6580 water molecules. To neutralize the protein charge, two chloride counterions were added using the *addions* command of *t leap*.

**Molecular Dynamics Simulations.** MD simulations were performed using the SANDER module of AMBER9 with the Amber 2003 force field. The minimized and solvated protein structure was gradually heated to the simulation temperature of 300 K in a two-step procedure. The first step was to disorder the solvent molecules in a constant-volume simulation, keeping the protein position fixed by applying the *ibelly* option. The temperature was increased from 100

to 300 K over 20 ps and then decreased from 300 K back to 100 K over 5 ps, employing the Berendsen weak coupling algorithm with a time constant of 0.5 ps. In the second step, the whole system (i.e., solute and solvent) was allowed to move and was heated to 300 K over 25 ps. To allow a time step of 0.002 ps, bonds involving hydrogen atoms were constrained using the SHAKE algorithm throughout the MD simulation. An 8 Å cutoff was used for the van der Waals interactions, while the electrostatics were treated by the particle mesh Ewald method.<sup>32</sup> After the heatup procedure, the constant-volume periodic boundary conditions were switched to constant pressure; isotropic position scaling was employed, with a default reference pressure of 1 bar and a pressure relaxation time of 1.0 ps. The time constant of the heat bath coupling was switched to 1.0 ps. After a 300 ps equilibration phase, MD trajectory data were collected over 13 ns for the monomer simulation. The dimer simulation carried out for comparative purposes was sampled over 6 ns. The coordinates were saved at intervals of 1 ps. Postprocessing of the trajectory was done using the *ptraj* module.

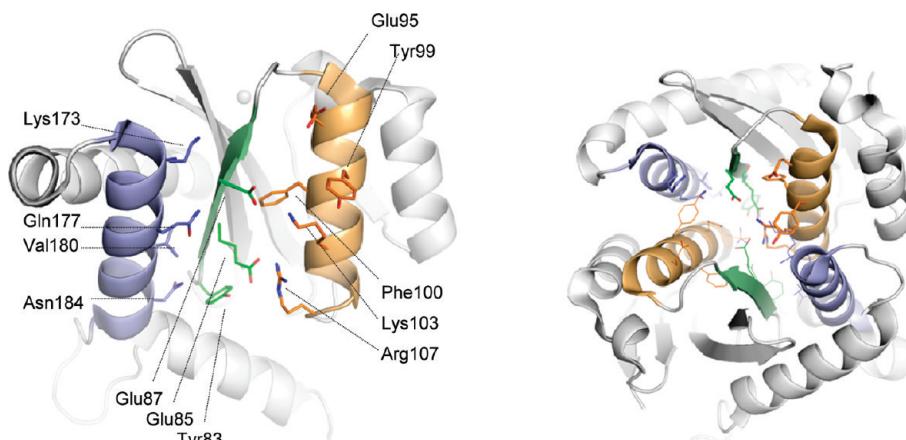
**Docking.** For the docking study, GOLD3.2<sup>33</sup> was used. The helical peptide YFLLKL was constructed using MOE.<sup>34</sup> The backbone was kept rigid in the standard-helix conformation throughout the docking runs, while the side chains were allowed to rotate. All eleven snapshots selected from the MD simulation after the 2D-rmsd analysis (described in the Results and Discussion section) were used as protein input structures for docking to identify the most favored conformer for binding the inhibitory peptide. For each protein conformer, 50 independent docking runs were performed. The binding site was defined by a sphere of a 10 Å radius centered at the amide hydrogen of Ala86. The genetic algorithm parameters were set as follows: population size, 100; number of islands, 5; selection pressure, 1.1; number of operations, 150 000; niche size, 2; migrate, 10; mutate, 95; and crossover, 95. Clustering of the resulting docking poses was done using the *rms\_analysis* utility, using a threshold of 1.5 Å.

**Hot Spot Analysis.** For the determination of preferred interaction sites (“hot spots”) within the peptide binding site, the HotSpots function of Drugscore<sup>CSD</sup><sup>35–37</sup> was employed. Sybyl atom type C.ar served as aromatic probe; C.3 served as hydrophobic probe, and N.3 served as hydrophilic nitrogen probe. The grid size was set to 97 × 97 × 64 points around the dimerization interface center, with a grid spacing of 0.5 Å. The contour plots were normalized for a more intuitive score ranging from 0 to 1 for each probe.

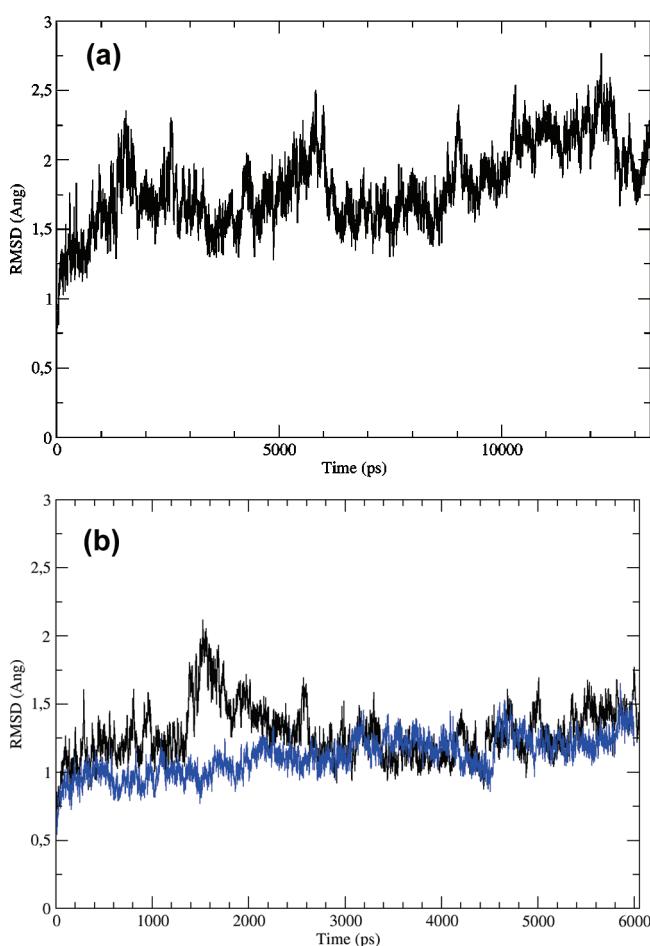
**Site Finder Analysis.** The purpose of the MOE *Site Finder* utility is to identify potential binding sites in a receptor by means of a geometric method based on Alpha Shapes. Apart from graphical output for visual analysis, the algorithm provides the number of receptor contact atoms in the putative binding sites and partitions them into hydrophobic or hydrophilic as well as backbone or side-chain atoms. Using the MD-derived snapshots as input structures, *Site Finder* was applied to retrieve the number of contact atoms within the dimer interface binding site.

## RESULTS AND DISCUSSION

The dimerization interface of the IN catalytic core domain is formed primarily by the β3-sheet (Tyr83–Ile89), the

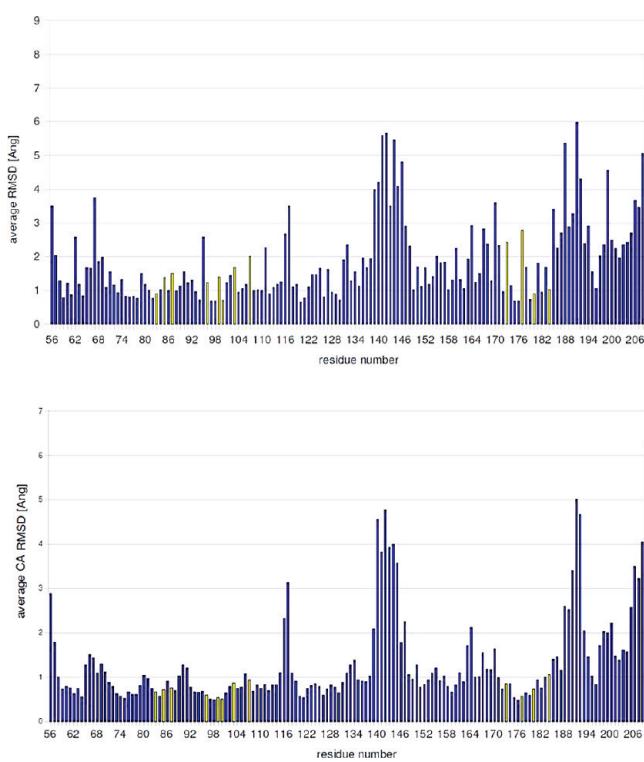


**Figure 1.** IN dimerization interface, comprising the  $\alpha_5$ -helix (blue), the  $\beta_3$ -sheet (green), and the  $\alpha_1$ -helix (wheat). Left: secondary structure elements and residues forming the dimer interface, which is located on the opposite side of the catalytic center (metal ion shown as gray sphere). Right: symmetrical IN dimer. The  $\alpha_1$ -helix of one monomer (top) interacts with the  $\alpha_5$ -helix of the other monomer (bottom) and vice versa.



**Figure 2.** (A) Average rms deviation of the C $\alpha$  atoms from the starting structure in the IN catalytic-core monomer MD as a function of simulation time. (B) Comparison of the average rms deviations from the starting structure in the MD simulations of the isolated monomer (black line) and the intact dimer (blue line). In both cases, the rmsd values were calculated for the C $\alpha$  atoms of one monomer and excluding the residues of the two exposed loop regions 140–150 and 186–196.

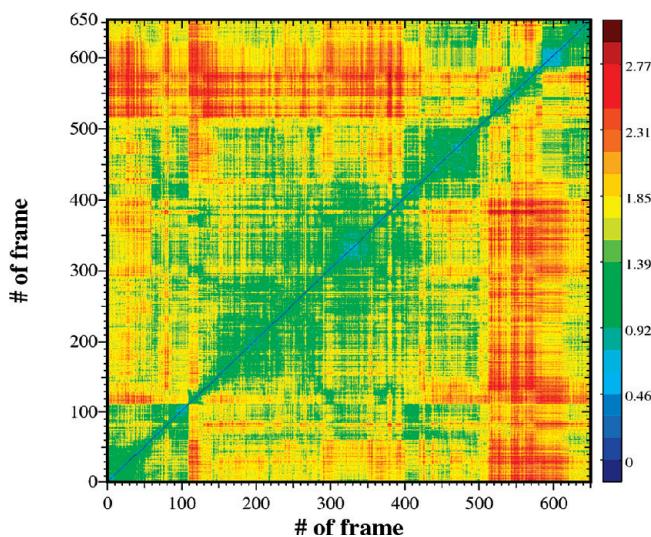
$\alpha_1$ -helix (Thr93–Arg107), and the  $\alpha_5$ -helix (His171–Lys186). Most important for the formation of the interaction surface are the residues which point toward the surface with their side chains. These are Tyr83, Glu85, and Glu87 of the  $\beta_3$ -sheet; Glu96, Tyr99, Phe100, Lys103, and Arg107 of the



**Figure 3.** Per-residue rmsd values in the IN catalytic-core monomer MD. Top: Fit and rmsd calculation based on all atoms. Bottom: Fit and rmsd calculation based on C $\alpha$  atoms. The interface residues are highlighted in yellow.

$\alpha_1$ -helix; Lys173, Gln177, Val180, and Asn184 of the  $\alpha_5$ -helix. Figure 1 provides a detailed view of this interface.

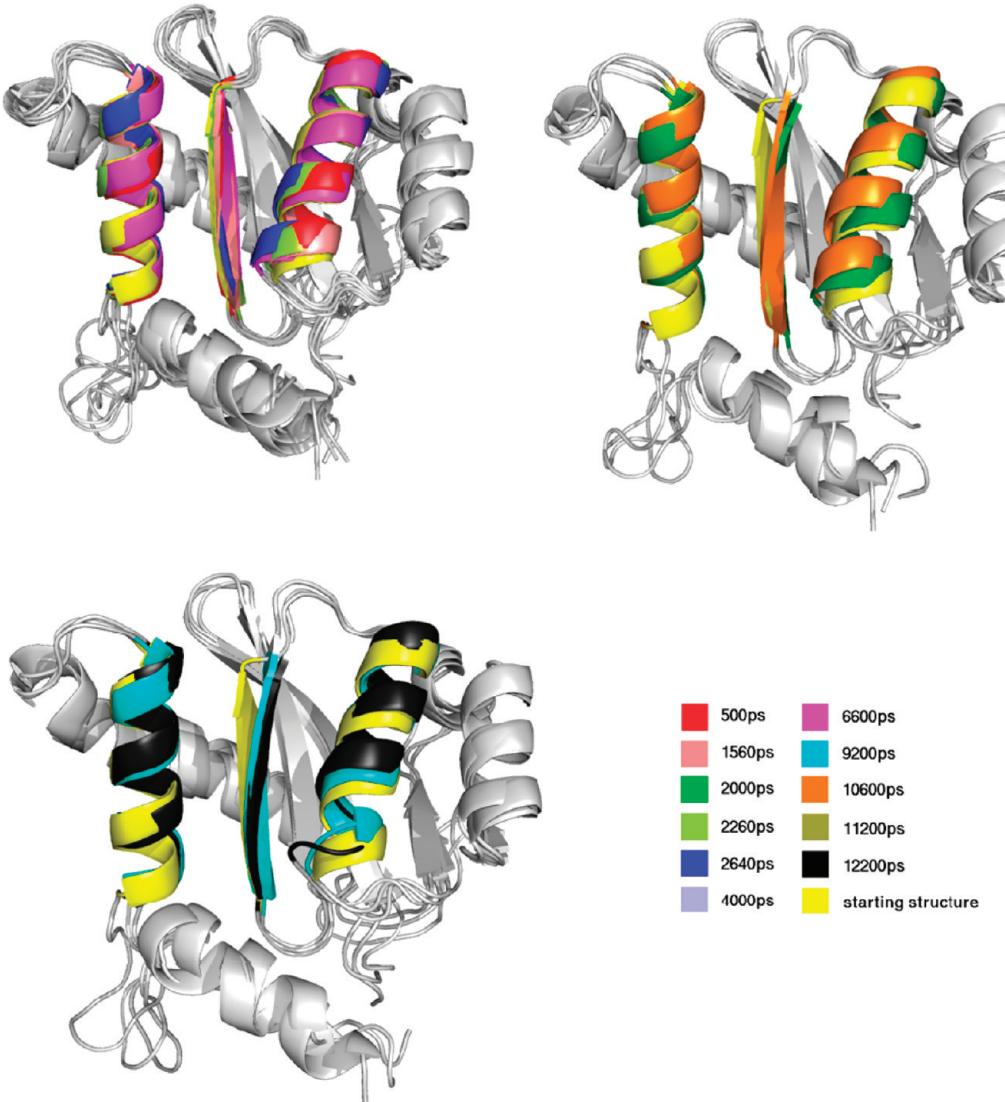
To probe the conformation space accessible to this dimer interface in solution, an MD simulation of the IN catalytic-core domain monomer was carried out. As a measure of overall stability, the root-mean-square deviation (rmsd) of the C $\alpha$  atoms was followed. Figure 2A shows that the protein fold was fairly stable during the simulation, with a mean C $\alpha$  rmsd of  $1.80 \pm 0.29$  Å (averaged over the entire trajectory). Nevertheless, significant structural changes are apparent from this diagram. The highest per-residue rmsd values (averaged over the trajectory) were observed for the flexible loop close to the active site and the terminal loop near the dimer interface, as well as for the C-terminal



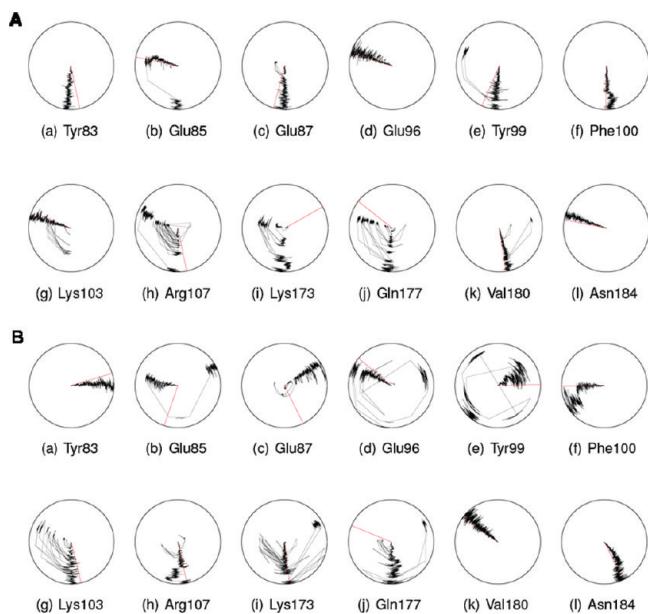
**Figure 4.** 2D-rmsd plot for the 13 ns monomer simulation, based on a superposition of all dimer interface atoms (as defined in the text) of snapshots taken every 20 ps. Each square in the plot holds the mutual rmsd value (in Å) between two snapshots, represented with the color code shown in the right column. The maximum value observed in the plot corresponds to 3.08 Å.

residues (cf. Figure 3). Excluding the residues 140–150 and 186–196 of these exposed loop regions from the rmsd calculation, a trajectory average of  $1.42 \pm 0.29$  Å was obtained (see also Figure 2B). The interface residues (as defined above) show low to moderate rmsd values, and the comparison between the all-atom data in the top diagram and the values for the C $\alpha$  atoms in the bottom diagram in Figure 3 suggests that the structural changes are mostly restricted to the side-chain level.

Because tracing possible conformational changes at the dimer interface was of pivotal interest, a 2D-rmsd plot was recorded, covering solely the dimer interface region as defined by the surface-exposed residues given above. For the 2D-rmsd calculation, every twentieth frame (snapshot) of the trajectory was used, corresponding to an interval of 20 ps. The pairwise rmsd value between all 650 frames was calculated after superimposing the aforementioned dimer interface residues. The resulting plot (Figure 4) shows rms deviations of up to 3.08 Å. One large cluster covering major parts of the simulation and two smaller clusters at the start and the end of the simulation are immediately apparent. For further studies, representatives of distinct clusters and



**Figure 5.** Secondary structure overview of the dimer interface. The colored areas correspond to  $\alpha_1$ -helix (right),  $\alpha_5$ -helix (left), and  $\beta_3$ -sheet (middle). Upper left: snapshots at  $t = 500, 1560, 2260, 2640, 4000, 6600$ , and  $11200$  ps; upper right: snapshots at  $t = 2000$  and  $10\,600$  ps; lower left: snapshots at  $t = 9200$  and  $12\,200$  ps. The starting structure has been superposed for reference purposes.



**Figure 6.** Dialsplots of the interface-residue torsion angles. The side-chain angles  $\chi_1$  and  $\chi_2$  were measured. Values in the starting structure are indicated by red lines. The plots show the time-dependent change of the dihedrals, with the center of the circle corresponding to  $t = 0$  ps. A (top):  $\chi_1$  torsions; B (bottom):  $\chi_2$  torsions.

subclusters were chosen, covering especially the beginning of the simulation to follow the most immediate structural changes upon relaxation of the interface in aqueous solution. After detailed visual inspection of the clusters and subclusters within the 2D-rmsd plot, the following eleven frames were selected: 25, 78, 100, 113, 132, 200, 330, 460, 530, 560 and 610, corresponding to simulation times of 500, 1560, 2000, 2260, 2640, 4000, 6600, 9200, 10600, 11200, and 12200 ps, respectively.

First, the secondary structure at the dimer interface was analyzed. Figure 5 shows a structural comparison of the eleven selected snapshots. While no dramatic changes in the arrangement of the helices are observed, the  $\beta$ 3-sheet shows a slight drift toward the  $\alpha$ 1-helix. The latter shows fluctuations in its helical content (cf. states at 9200 and 12 200 ps), as does the  $\alpha$ 6-helix to a greater extent (in Figure 5, the  $\alpha$ 6-helix corresponds to the uncolored helix at the southern end of the interface). Interestingly, an alteration of secondary structure elements at the dimer interface has also been observed by Circular Dichroism measurements.<sup>12</sup> Of course, the time range of feasible MD simulations does not allow one to follow major structural changes (which happen on a microsecond time scale). Nevertheless, correct tendencies can be observed. In this case, the fluctuations in the helical content of the two helices show that they are most likely liable to major structural changes (as observed in the Circular Dichroism spectra).

In the next step, the dynamics of the dimerization interface residues (i.e., Tyr83, Glu85, Glu87, Glu96, Tyr99, Phe100, Lys103, Arg107, Lys173, Gln177, Val180, and Asn184) was analyzed. Torsion-angle dialsplots were recorded for the interface residues. The  $\chi_1$  and  $\chi_2$  angles of the side chains were measured throughout the trajectory using *ptraj*. Post-processing of this data to generate the dialsplots was done using an in-house script. The results are shown in Figure 6, where the dihedral angle of the corresponding residue in the

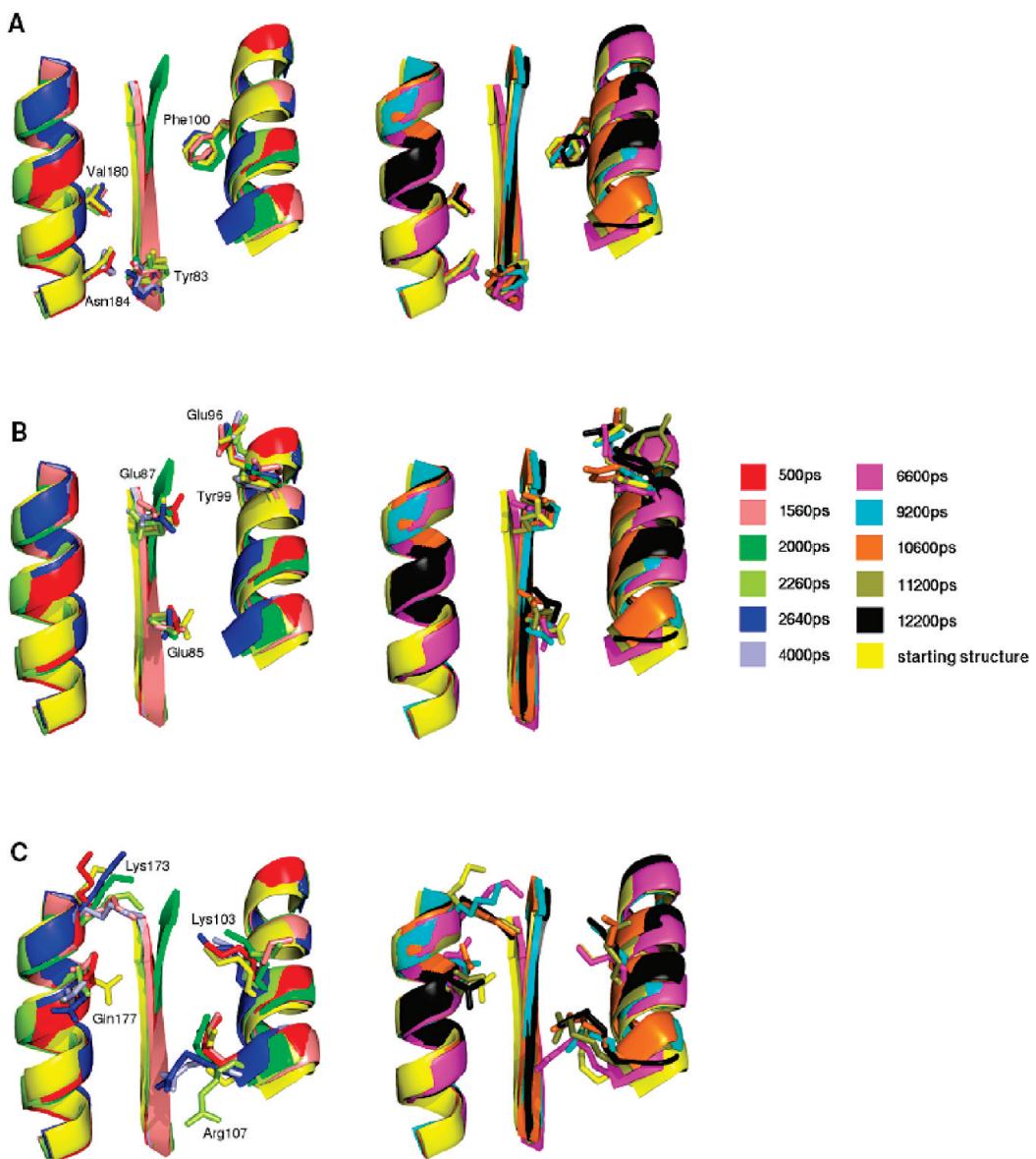
starting structure is indicated by a red line. Tyr83 keeps its  $\chi_1$  and  $\chi_2$  starting dihedrals with only minor changes and small fluctuations. For Glu85,  $\chi_1$  first oscillates around the initial value, whereas in the last third of the simulation a 90° shift is observed. This  $\chi_1$  change coincides with a shift in  $\chi_2$  of about 140°. The striking differences between the mean and the initial value (red line in Figure 6) for the  $\chi_2$  angle of Glu85, Glu87, and Gln177 indicate that the (minimized) starting conformations of these residues are not representative for the monomeric state in aqueous solution. The  $\chi_1$  dihedral of Glu96 smoothly oscillates around the starting value, whereas the  $\chi_2$  dihedral shows a second distinct side-chain orientation in the last quarter of the simulation, characterized by a 160° shift. A similar behavior is observed for the Tyr99  $\chi_1$  dihedral. Phe100 shows relatively weak changes. Lys103 has  $\chi_1$  and  $\chi_2$  values close to the starting value, intermittent by temporary shifts. In Arg107, Lys173, and Gln177,  $\chi_1$  shows interesting oscillations of ~100° during the simulation. A similar observation is made for the Lys173  $\chi_2$  dihedral. Val180 and Asn184 do not show any significant changes with respect to the starting orientation of their side chains.

Figure 7A shows the side-chain orientations of the residues found to be less flexible during the MD simulation, in particular Tyr83 (located on the  $\beta$ 3-sheet), Phe100 (located on the  $\alpha$ 1-helix), and Val180 as well as Asn184 (both located on the  $\alpha$ 5-helix). These residues are rather buried and less solvent exposed, which explains the lower degree of flexibility.

Figure 7B shows the side-chain orientations of the residues found to be moderately flexible during the MD simulation, i.e., Glu85 and Glu87 (both located on the  $\beta$ 3-sheet), as well as Glu96 and Tyr99 (both located on the  $\alpha$ 1-helix). Tyr99 oscillates weakly at the beginning of the simulation, whereas a higher degree of mobility is observed in the second half of the trajectory (see Figure 7B, right). Glu87 shows moderate flexibility, but changes in the orientation of the terminal carboxy group have a dramatic effect on a cavity formed with Lys103. The higher flexibility of Glu85, Glu87, Glu96, and Tyr99 compared to the previously mentioned interface residues is due to their more direct contact to the surrounding solvent.

Figure 7C shows the side-chain orientations of the residues found to be highly flexible during the MD simulation: Lys173 and Gln177 (both located on the  $\alpha$ 5-helix), as well as Lys103 and Arg107 (both located on the  $\alpha$ 1-helix). All four residues are highly solvent exposed. The flexibility of Lys103 has dramatic effects on a cavity formed with Glu87. During the simulation, Lys173 displays a tendency to bend toward the groove between the two helices, with the sheet forming the “bottom” of this groove. Bending of Lys173 blocks this groove, best seen in the surface representation of the dimer interface (cf. Figure 8, where the essential residues Glu87, Phe100, Lys103, and Lys173 are colored). Structural changes in this groove region are of pivotal importance, bearing in mind that this is the putative binding site for inhibitors of dimerization (based on analogy with the binding site of the  $\alpha$ 1'-helix of the second IN monomer).

The surface representation of the different snapshots illustrates the changes in the interface region, especially with respect to the groove between the two helices and the cavity formed by Glu87 and Lys103. A crucial factor influencing the groove is Lys173. In the starting structure, this residue



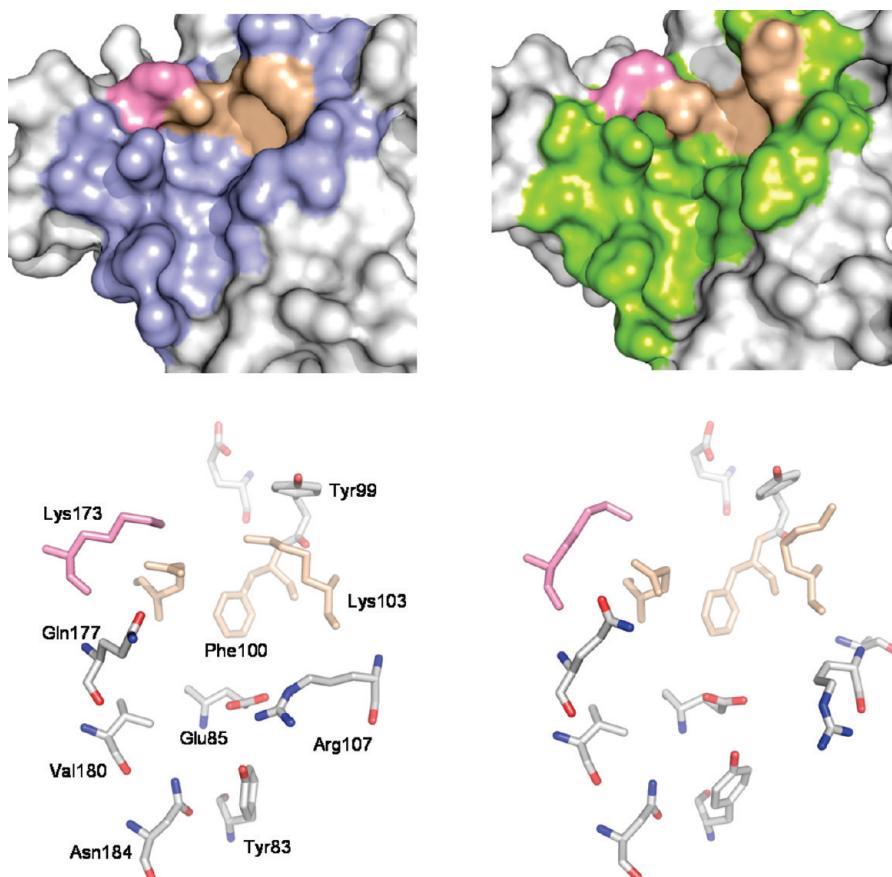
**Figure 7.** Left: Different structural states observed from the first half of the trajectory (at  $t = 500, 1560, 2000, 2260, 2640$ , and  $4000$  ps, respectively) superimposed with the starting structure. Right: Different structural states from the second half of the trajectory (at  $t = 6600, 9200, 10600, 11200$ , and  $12200$  ps, respectively) superimposed with the starting structure. A (top): Less flexible side chains; B (middle): moderately flexible side chains; C (bottom): highly flexible side chains.

is directed toward the aqueous surrounding (referred to as “unblocked” state), therefore contributing only marginally to the dimer interface. During the simulation, Lys173 bends towards the groove and narrows it (referred to as “blocked” state). In the trajectory, an oscillation between unblocked ( $500, 2000, 2640$ , and  $9200$  ps) and blocked state ( $1560, 2260, 4000, 6600, 10600, 11200$ , and  $12200$  ps) can be observed, with the blocked state prevailing at the end of the simulation.

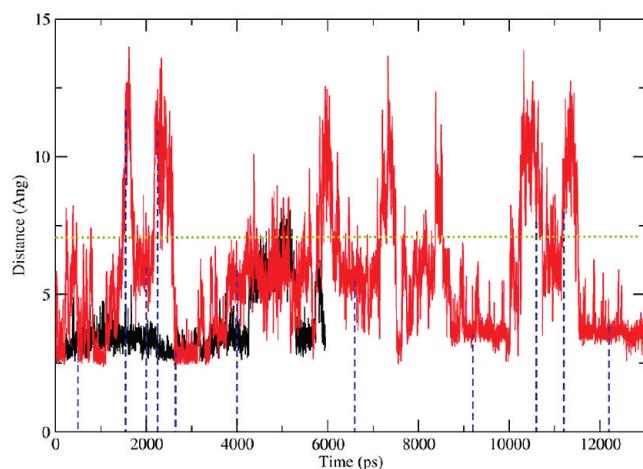
The cavity formed by Glu87 and Lys103 is distinct in the starting structure; during the MD simulation, fluctuations in the shape of the cavity are observed, best seen in the surface representation of the interface region (cf. Figure 8). As a quantitative measure of the cavity structure fluctuations, the distance between the terminal amino group of Lys103 and the terminal carboxy group of Glu87 was recorded and plotted versus simulation time (Figure 9). From this graph, the oscillation in the distance leading to a collapse and recurring formation of the cavity is clearly visible. A distinct

cavity (i.e., a relatively close contact between Glu87 and Lys103) is observed in the snapshots at  $t = 500, 2640, 4000, 6600, 9200$ , and  $12200$  ps but not in the snapshots at  $t = 1560, 2000, 2260, 10600$ , and  $11200$  ps. The distance curve in Figure 9 indicates that formation and collapse of the cavity is an oscillating process, without clear tendency toward stabilization or collapse.

For reference purposes, a 6 ns MD simulation of the dimeric IN structure was performed. The protein fold did not show significant structural deviations from the crystal structure, proving that the crystallographic structure is stable in a noncrystalline environment. In line with the expectation, structural deviations and fluctuations were observed mainly in the exposed loop regions mentioned above. Excluding these loop regions, average  $C\alpha$  atom rmsd values of  $1.23 \pm 0.14$  Å and  $1.12 \pm 0.14$  Å were observed for the two monomers in the 6 ns dimer simulation (cf. Figure 2B). Both monomeric units displayed similar structural behavior over the trajectory. The first monomer was extracted, analyzed,



**Figure 8.** Surface and stick representations of the dimer interface. Colored areas indicate surfaces of the  $\alpha$ 5-helix, the  $\alpha$ 1-helix, and the  $\beta$ -sheet. The left column represents the snapshot at 4000 ps with a blocked groove and a formed cavity, whereas the right column represents the snapshot at 2260 ps with a blocked groove and a collapsed cavity. Lys173 is colored pink, the cavity-forming residues Glu87, Phe100, and Lys103 are colored wheat.

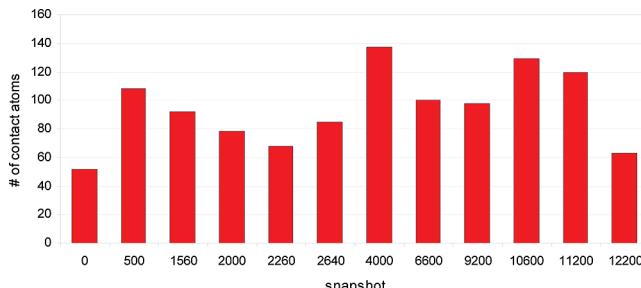


**Figure 9.** Glu87–Lys103 distance versus simulation time. Red: monomer MD simulation; black: dimer MD simulation; dotted blue lines: selected snapshots. The dotted green line indicates the formation and collapse (below and above, respectively) of the cavity.

and used as a reference for the monomer MD simulation. Looking at the Glu87–Lys103 distances obtained in the dimer MD simulation (see Figure 9, black line), the fluctuations are rather slight and the largest distance (at  $t = 5000$  ps) is only roughly one-half of the largest distance obtained in the monomer MD. In contrast to the free monomer, no collapse of the cavity formed by Glu87 and Lys103 is observed in the dimer.

In most of the IN structures, crystallographic water molecules are observed at the dimer interface. Five such molecules are found in the vicinity of Glu87 and Lys103, building a water network between Glu87 and Glu96. In a first attempt, a solute–solvent system was prepared where all crystallographic water molecules had been removed prior to the solvation process (no direct water mediation between Glu87 and Lys103 via hydrogen bonds is observed in the crystal structure). However, the *t leap* routine of AMBER9 did not place water molecules at the corresponding positions. MD simulations based on this system yielded a collapse of the cavity formed by Glu87 and Lys103 (data not shown). In the actual study, the aforementioned five water molecules have been retained. In contrast to the former MD simulation, a collapsed state of the cavity is not maintained. The water network between Glu87 and Glu96 seems to reduce the mobility of Glu87. Additionally, the water molecules placed close to Glu87 and Lys103 begin to bridge both residues in the course of the MD simulation. Both features lead to the assumption that these water molecules play a crucial role in limiting the movements of Glu87 and Lys103, thereby preventing the cavity from a permanent collapse.

In summary, the most interesting features of the dimerization interface with respect to ligand accommodation appear to be the groove and the cavity. To roughly estimate the ability of the different snapshots to accommodate a ligand, the number of potential contact atoms in the protein capable of interacting with a ligand was counted using the *Site Finder* methodology of MOE. The whole protein surface



**Figure 10.** Total number of contact atoms at the protein dimer interface, according to the *Site Finder* analysis (see text for further explanations). The snapshot at 4000 ps shows the highest count.

was scanned. In each snapshot, the dimer interface was revealed as the site with the highest contact–atom count. In Figure 10, the total number of contact atoms in the dimer interface is reported for each snapshot. The lowest number of contact atoms (52) is found for the starting structure. All simulation snapshots show higher values, with the maximum reached at snapshot 4000, corresponding to a total of 137 contact atoms. Accordingly, the protein conformation corresponding to this snapshot would appear to be most suited for accommodating potential ligands. To further confirm this assumption, all extracted snapshots were subjected to a comparative docking analysis using the peptide YFLLKL as ligand.

The hexapeptide YFLLKL was derived from a parent inhibitory dodecamer (TAYFLLKLAGRW). This dodecamer has the same amino acid sequence as the IN  $\alpha 1'$ -helix (residues 97'–108'), which binds (as part of the second monomer) to the dimer interface of the first IN monomer. It was shown<sup>12</sup> that the synthetic dodecamer competes for the same binding site, preventing dimerization of IN. Interestingly, mutation studies<sup>16</sup> showed that the lysine residue can be replaced by an isoleucine residue without affecting the affinity. This seems surprising, given the role of this lysine (Lys103') in salt-bridge formation with Glu87 across the interface. However, it has been hypothesized that this salt bridge contributes only to a minor extent to the overall dimer stability.<sup>38</sup> Truncation studies of the dodecamer showed that the minimal active structure is the hexameric peptide YFLLKL.<sup>16</sup> In analogy to the dodecamer, one can expect that this hexamer competes for the same binding site in the

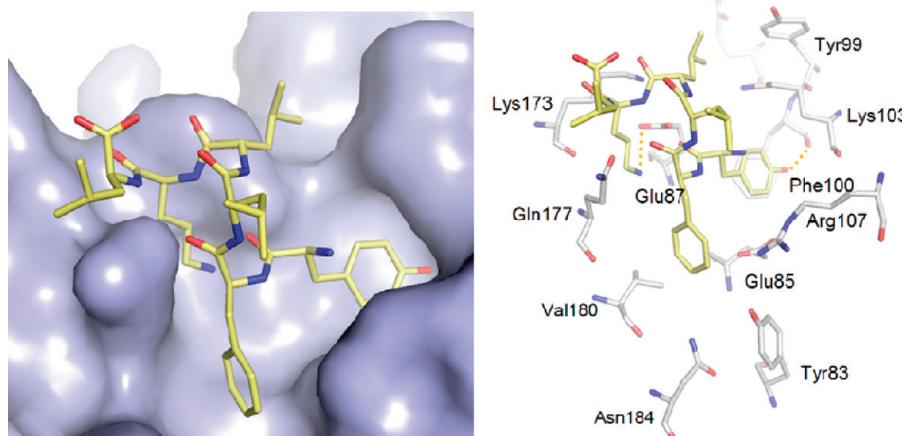
**Table 1.** Clustering and Scoring Data of the Docking Study Analyzing the Interaction of YFLLKL with the Different MD Snapshots<sup>a</sup>

	# of members in top-ranked cluster	total GOLD score	hydrogen bond score	hydrophobic score	internal score
start structure	24	13.74	0.60	39.54	-26.40
500 ps	49	-23.53	5.15	-1.71	-26.97
1560 ps	4	11.37	0.00	38.10	-26.73
2000 ps	6	12.95	0.16	13.31	-26.42
2260 ps	43	24.20	5.72	45.18	-26.70
2640 ps	1	16.19	0.00	42.98	-26.79
<b>4000 ps</b>	<b>47</b>	<b>28.34</b>	<b>3.40</b>	<b>53.10</b>	<b>-28.16</b>
6600 ps	8	13.28	3.12	36.30	-26.14
9200 ps	46	-4.63	0.63	23.57	-28.83
10600 ps	5	-4.46	0.00	22.88	-27.34
11200 ps	1	-10.50	0.00	16.44	-26.94
12200 ps	4	6.38	0.00	32.58	-26.20

<sup>a</sup> The best total score as well as very good clustering is achieved with the snapshot at 4000 ps as input structure (bold).

dimer interface region. However, in the dimeric crystal structure, this hexameric portion shows only very few contacts to the first monomer. This leads to the assumption that the inhibitory peptide YFLLKL may still bind to the same site, yet with a different binding mode which allows more favorable interactions.

Consequently, we performed docking studies of this peptide to the different MD-derived snapshots to obtain potential binding modes. The results were analyzed in terms of the scoring values and by visual inspection. Comparing the GOLDscore values from docking to different protein conformations, it is striking that the snapshot taken from the central part of the trajectory (i.e., 4000 ps) yielded the most favorable score; moreover, good clustering of the top ranked results was observed (see Table 1). This protein conformer is part of the largest cluster in the 2D-rmsd plot (see Figure 4) and, therefore, most likely corresponds to a relatively stable conformation of the IN monomer. The scores indicate that YFLLKL will preferably bind to this protein conformation, presumably stabilizing it further. The favorable accommodation of the ligand by the snapshot at 4000 ps is in fair agreement with the expectation based on the *Site Finder* analysis described above.



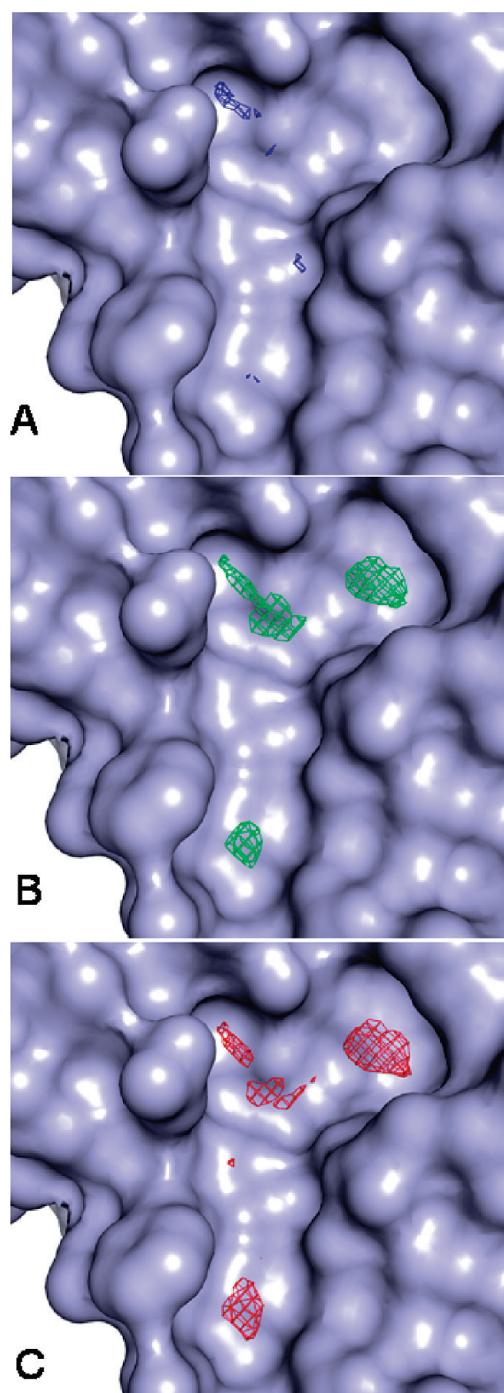
**Figure 11.** Top-ranked binding mode obtained by docking of the hexameric peptide YFLLKL to the dimerization interface (snapshot at  $t = 4000$  ps). Left: Surface representation of the protein. Right: Stick representation of the protein; the hydrogen-bond interactions described in the text are highlighted as dotted lines.

Figure 11 shows the binding mode of YFLLKL to snapshot 4000 ps. Residue Y1 is well accommodated by the cavity formed by Glu87 and Lys103. The aromatic system displays interactions with the phenyl ring of Phe100. Furthermore, the Tyr-OH group donates a hydrogen bond to the backbone oxygen of Phe100. The aromatic system of F2 is placed in the groove. The cationic residue K5 binds to the smaller cavity formed by Glu87 and Lys173, where it forms a salt bridge with Glu87.

When using AutoDock3.0<sup>39–41</sup> as the docking tool (data not shown), the aromatic residues of the hexamer were placed toward the aqueous surrounding, which is not quite likely. Furthermore, subjecting the different binding modes obtained from GOLD and AutoDock3.0 to a rescoring with SFC-score<sup>42</sup> (employing both the SFC\_290p and SFC\_290m scoring schemes) favored the binding modes from GOLD. This was further supported by the hot-spot analysis (Figure 12): In the GOLD binding mode, the favorable positions for aromatic probes in the cavity are occupied by Y1 and F2, respectively. Their prominent positions indicate a major contribution to the overall affinity, which is also in accordance with experimental data. Alanine scanning studies of the parent dodecamer TAYFLLKLAGRW showed that tyrosine and phenylalanine are essential for inhibitory activity; furthermore, hexameric peptides derived from this dodecamer were inactive when they did not include these two residues.<sup>16</sup>

The elucidation of important structural features at the dimerization interface as well as the investigation of potential binding modes of a known dimerization inhibitor yield guidelines for the design of small-molecule interface binders. Three structural features (as seen in the conformer at 4000 ps) of the interface should be addressed: the two cavities formed by Glu87/Lys103 and Glu87/Lys173, respectively, and the groove between the two helices. The hot-spot analysis (Figure 12) gives advice on where to place certain fragments. The larger cavity formed by Glu87/Lys103 with Phe100 at the bottom could be addressed by an aromatic residue; phenolic residues appear particularly favorable in order to saturate the Phe100 backbone carbonyl as a potential hydrogen-bonding partner. Aromatic or hydrophobic fragments should be placed in the groove. The smaller cavity formed by Glu87/Lys173 appears to have amphiphilic character, as favorable hot spots for a variety of probes are observed in this area. This is likely due to the presence of hydrophobic “floor” residues in this superficially polar cavity, namely Val88 and Val176, which explain the presence of favorable hydrophobic interaction sites. Hot spots for a hydrophilic probe (aliphatic amine N) were also observed in this cavity, showing a similar prevalence as the hydrophobic probe. These polar hot spots indicate possible interaction sites with the amide oxygen of Gln177 and the carboxyl function of Glu87, both located at the rim of the cavity.

On the basis of these three binding subsites, a three-point Y-shaped pharmacophore model could be created, which might serve as a query in a virtual screening campaign. For de novo design, it is conceivable to address the smaller cavity with a suitable donor group, such as a urea moiety for hydrogen bonding with the Glu87 carboxy oxygens. The larger cavity could be filled with an aminophenyl, a phenol, or similar groups, which simultaneously show hydrophobic interactions in the cavity and hydrogen bonding with the



**Figure 12.** Important subsites of the dimer binding site (snapshot at  $t = 4000$  ps): Drugscore<sup>CSD</sup> contourplots (contour level = 0.70) are shown with a nitrogen probe (A, blue), an aliphatic probe (B, green), and an aromatic probe (C, red), illustrating the Y-shaped interaction sites which have to be addressed.

Phe100 backbone oxygen. The fragments addressing the two cavities should be connected by a central linker, which should bear a third fragment to protrude into the groove. In this way, the work presented here provides a structural basis for the design of small-molecule interface binders as a potential new class of HIV integrase inhibitors.

## CONCLUSION

An inspection of the dimer interface has so far mostly been ignored in MD simulation studies of HIV-1 integrase. The

discovery of peptide inhibitors binding to this interface and disrupting the integrase multimers forms the basis for a new approach to enzyme inhibition but also calls for a deeper understanding of the structural features of this interface region. In this study, we have identified a groove which very likely accommodates the inhibitory peptide. This groove became more distinct during the MD simulation, although a highly flexible residue, Lys173, was observed to be able to obstruct this groove. Additionally, a cavity adjacent to the groove was identified and analyzed. This cavity, formed by Glu87 and Lys103, was found to be of transient nature, oscillating between formation and collapse. Both groove and cavity appear to be suitable regions for accommodating inhibitors at the dimer interface. Snapshots derived from the MD simulation were used as protein input structures in a docking study with the peptide YFLLKL to reveal its potential binding mode. The docking procedure showed that the peptidic ligand binds to a protein conformation where the groove is blocked by Lys173 and the Glu87/Lys103 cavity is formed. Additionally, a second, less distinct cavity formed by Glu87 and Lys173 is observed. These important structural features of the dimerization interface provide a guideline for pharmacophore modeling or de novo design.

## REFERENCES AND NOTES

- Pommier, Y.; Johnson, A. A.; Marchand, C. Integrase inhibitors to treat HIV/AIDS. *Drug Discovery Today* **2005**, *4*, 236–248.
- Flexner, C. HIV drug development: the next 25 years. *Nat. Rev. Drug Discov.* **2007**, *6*, 959–966.
- Hardman, J. G.; Limbird, L. E. Antiretroviral Agents. In *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 11th ed.; Gilman, A. G., Ed.; McGraw-Hill: New York, 2005; pp 1349–1380.
- Cahn, P.; Sued, O. Raltegravir: a new antiretroviral class for salvage therapy. *Lancet* **2007**, *369*, 1235–1236.
- Steigbigel, R. T.; Cooper, D. A.; Kumar, P. N. Raltegravir with optimized background therapy for resistant HIV-1 infection. *N. Engl. J. Med.* **2008**, *359*, 339–354.
- Morris, G. M. AutoDock's role in Developing the First Clinically-Approved HIV Integrase Inhibitor. <http://autodock.scripps.edu/news/autodocks-role-in-developing-the-first-clinically-approved-hiv-integrase-inhibitor> (accessed Mar 15, 2009).
- Savarino, A. A historical sketch of the discovery and development of HIV-1 integrase inhibitors. *Expert Opin. Invest. Drugs* **2006**, *15*, 1507–1522.
- Sechi, M.; Bacchi, A.; Carcelli, M.; Compari, C.; Duce, E.; Fisicaro, E.; Rogolino, D.; Gates, P.; Derudas, M.; Al-Mawsawi, L. Q.; Neamati, N. From Ligand to Complexes: Inhibition of Human Immunodeficiency Virus Type 1 Integrase by Diketo Acid Metal Complexes. *J. Med. Chem.* **2006**, *49*, 4248–4260.
- Bacchi, A.; Biemmi, M.; Carcelli, M.; Carta, F.; Carlotta, C.; Fisicaro, E.; Rogolino, D.; Sechi, M.; Sippel, M.; Sottriffer, C.; Sanchez, T. W.; Neamati, N. From Ligand to Complexes. Part 2. Remarks on Human Immunodeficiency Virus Type 1 Integrase inhibition by a Diketo Acid Metal Complexes. *J. Med. Chem.* **2008**, *51*, 7253–7264.
- Faure, A.; Calmels, C.; Desjobert, C.; Castroviejo, M.; Caumont-Sarcos, A.; Tarrago-Litvak, L.; Litvak, S.; Parissi, V. HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res.* **2005**, *33*, 977–986.
- Esposito, D.; Craigie, R. HIV integrase structure and function. *Adv. Virus Res.* **1999**, *52*, 335–350.
- Maroun, R. G.; Gayet, S.; Benleulmi, M. S.; Poroumb, H.; Zargarian, L.; Merad, H.; Leh, H.; Mouscadet, J.-F.; Troalen, F.; Fermandjian, S. Peptide inhibitors of HIV-1 integrase dissociate the enzyme oligomers. *Biochemistry* **2001**, *40*, 13840–13848.
- Moltensi, V.; Greenwald, J.; Rhodes, D.; Hwang, Y.; Kwiatkowski, W.; Bushman, F. D.; Siegel, J. S.; Choe, S. Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *57*, 536–544.
- Al-Mawsawi, L. Q.; Fikkert, V.; Dayam, R.; Witvrouw, M.; Burke, T. R., Jr.; Borchers, C. H.; Neamati, N. Discovery of a small-molecule HIV-1 integrase inhibitor-binding site. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10080–10085.
- Kessl, J. J.; Eidahl, J. O.; Shkriabai, N.; Zhao, Z.; McKee, C. J.; Hess, S.; Burke, T. R., Jr.; Kvaratskhelia, M. An allosteric mechanism for inhibiting HIV-1 integrase with a small molecule. *Mol. Pharmacol.* **2009**, *76*, 824–832.
- Li, H.-Y.; Zawahir, Z.; Song, L.-D.; Long, Y.-Q.; Neamati, N. Sequence-based design and discovery of peptide inhibitors of HIV-1 integrase: insight into the binding mode of the enzyme. *J. Med. Chem.* **2006**, *49*, 4477–4486.
- Zhao, L.; O'Reilly, M. K.; Shultz, M. D.; Chmielewski, J. Interfacial peptide inhibitors of HIV-1 integrase activity and dimerization. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1175–1177.
- Gonzalez-Ruiz, D.; Gohlke, H. Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr. Med. Chem.* **2006**, *13*, 2607–2625.
- Toogood, P. L. Inhibition of protein-protein association by small molecules: approaches and progress. *J. Med. Chem.* **2002**, *45*, 1543–1558.
- Yin, H.; Hamilton, A. D. Strategies for targeting protein-protein interactions with synthetic agents. *Angew. Chem., Int. Ed.* **2005**, *44*, 4130–4163.
- Antuch, W.; Menon, S.; Chen, Q.; Lu, Y.; Sakamuri, S.; Beck, B.; Schauer-Vukasinovic, V.; Agarwal, S.; Hess, S.; Dömling, A. Design and modular parallel synthesis of a MCR derived alpha-helix mimetic protein-protein interaction inhibitor scaffold. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1740–1743.
- Chen, L.; Yin, H.; Farooqi, B.; Sebti, S.; Hamilton, A. D.; Chen, J. p53 alpha-helix mimetics antagonize p53/MDM2 interaction and activate p53. *Mol. Cancer Ther.* **2005**, *6*, 1019–1025.
- Camarasa, M.-J.; Velazquez, S.; San-Felix, A.; Perez-Perez, M.-J.; Gago, F. Dimerization inhibitors of HIV-1 reverse transcriptase, protease and integrase: a single mode of inhibition for the three HIV enzymes. *Antiviral Res.* **2006**, *71*, 260–267.
- Hu, J. P.; Gong, X. Q.; Su, J. G.; Chen, W. Z.; Wang, C. X. Study on the molecular mechanism of inhibiting HIV-1 integrase by EBR28 peptide via molecular modeling approach. *Biophys. Chem.* **2008**, *132*, 69–80.
- Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13040–13043.
- Sottriffer, C. A.; Ni, H.; McCammon, J. A. HIV-1 integrase interactions at the active site: prediction of binding modes unaffected by crystal packing. *J. Am. Chem. Soc.* **2000**, *122*, 4109–4117.
- Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Sebra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, 2006.
- Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semi-analytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- Srinivasan, J.; Trevathan, M.; Beroza, P.; Case, D. A. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **1999**, *101*, 426–434.
- Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald-an Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
- MOE, version 2007\_09; Chemical Computing Group: Montreal, Canada, 2007.
- Velec, H. F. G.; Gohlke, H.; Klebe, G. DrugScoreCSD - Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction. *J. Med. Chem.* **2005**, *48*, 6296–6303.
- Gohlke, H.; Hendlich, M.; Klebe, G. Predicting binding modes, binding affinities and “hot spots” for protein-ligand complexes using a knowledge-based scoring function. *Persp. Drug Discovery Des.* **2000**, *20*, 115–144.

- (37) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
- (38) Al-Mawsawi, L. Q.; Hombrouck, A.; Dayam, R.; Debyser, Z.; Neamati, N. Four-tiered interaction at the dimeric interface of HIV-1 integrase critical for DNA integration and viral infectivity. *Virology* **2008**, *377*, 355–363.
- (39) Goodsell, D. S.; Olson, A. J. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins* **1990**, *8*, 195–202.
- (40) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed Automated Docking of Flexible Ligands to Proteins: Parallel Applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293–304.
- (41) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- (42) Sottriffer, C. A.; Sanschagrin, P.; Matter, H.; Klebe, G. SFCscore: Scoring functions for affinity prediction of protein-ligand complexes. *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 395–418.

CI900403S