Comparative Molecular Dynamics Simulations of HIV-1 Integrase and the T66I/M154I Mutant: Binding Modes and Drug Resistance to a Diketo Acid Inhibitor

Alessandro Brigo, 1,2† Keun Woo Lee, 2‡ Federico Fogolari, Gabriela Iurcu Mustata, 28 and James M. Briggs 2*

HIV-1 IN is an essential enzyme for viral replication and an interesting target for the design of new pharmaceuticals for use in multidrug therapy of AIDS. L-731,988 is one of the most active molecules of the class of β -diketo acids. Individual and combined mutations of HIV-1 IN at residues T66, S153, and M154 confer important degrees of resistance to one or more inhibitors belonging to this class. In an effort to understand the molecular mechanism of the resistance of T66I/M154I IN to the inhibitor L-731,988 and its specific binding modes, we have carried out docking studies, explicit solvent MD simulations, and binding free energy calculations. The inhibitor was docked against different protein conformations chosen from prior MD trajectories, resulting in 2 major orientations within the active site. MD simulations have been carried out for the T66I/M154I DM IN, DM IN in complex with L-731,988 in 2 different orientations, and 1QS4 IN in complex with L-731,988. The results of these simulations show a similar dynamical behavior between T66I/M154I IN alone and in complex with L-731,988, while significant differences are observed in the mobility of the IN catalytic loop (residues 138-149). Water molecules bridging the inhibitor to residues from the active site have been identified, and residue Gln62 has been found to play an important role in the interactions between the inhibitor and the protein. This work provides information about the binding modes of L-731,988, as well as insight into the mechanism of inhibitor-resistance in HIV-1 integrase. Proteins 2005;59:723-741. © 2005 Wiley-Liss, Inc.

Key words: HIV-1 integrase; automated docking; double mutant integrase T66I/M154I; diketo acid HIV-1 integrase inhibitors; binding free energy calculation; MM/PBSA methodology

INTRODUCTION

HIV-1 is the etiological agent of AIDS. The HIV genome encodes 3 enzymes: reverse transcriptase, protease, and integrase. Only the first 2 enzymes have been successfully exploited as targets for antiviral drug therapy. The emergence of strains resistant to currently available reverse

transcriptase and protease inhibitors has led to the necessity to study new targets for antiretroviral therapy. An essential step in HIV replication is the integration of the transcribed double-stranded DNA into the host chromosomes carried out by the IN protein. Integration of the double-stranded viral DNA genome occurs in 2 subsequent reactions: In the first step, termed 3'-processing, an activated water molecule attacks at the 3'-ends of the viral DNA, removing a terminal dinucleotide from each 3'-end; in the second step, called "strand transfer," each exposed viral DNA 3'-OH ribose is activated for nucleophilic attack on the opposite strands of the host DNA becoming covalently attached to it. Divalent metal ions such as Mg²⁺ or Mn²⁺ are required for both of the reactions and for the formation of the IN complex with viral DNA.

HIV-1 IN is a 32-kDa enzyme that comprises 3 structurally and functionally distinct domains, all of them required for full catalytic activity. The N-terminal domain (residues 1–50) contains a conserved HHCC motif, which

Abbreviations: AIDS, acquired immunodeficiency syndrome; CVFF, Consistent Valence Force Field; DM, double mutant; HIV-1, human immunodeficiency virus type 1; IN, integrase; ; IN DM, 1QS4 with double resistance mutation (T66I/M1541); LGA, Lamarckian Genetic Algorithm; MD, molecular dynamics; MM/PBSA, molecular mechanics Poisson—Boltzmann solvent-accessible surface area; NPT, consistent number of particles, pressure and temperature; PBE, Poisson—Boltzmann equation; PDB, Protein Data Bank; PME, particle mesh Ewald; 1QS4, HIV-1 integrase catalytic domain with double solubility mutation F185K/W131E; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; 3D, three-dimensional; TIP3P, Transferable Intermolecular Potential Three Point Model; vdW, van der Waals.

¹Dipartimento di Scienze Farmaceutiche, Università degli Studi di Padova, Padova, Italy

²Department of Biology and Biochemistry, University of Houston, Houston, Texas

³Dipartimento Scientifico-Tecnologico, Università degli Studi di Verona, Verona, Italy

The Supplementary Materials referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat

Grant sponsor: NIH Program on Structural Biology of AIDS Related Proteins; Grant number: GM56553. Grant sponsor: Robert A. Welch Foundation; Grant number: E-1497.

[†]Current address: The Genetics Company, Inc., Wagistrasse 27, 8952 Schlieren, Zurich Area, Switzerland (www.the-genetics.com)

[‡]Current address: Department of Biochemistry, Gyeongsang National University, Gazwa-dong, Jinju, Kyeongnam-do, 660-701, Korea [§]Current address: Emisphere Technologies, Inc., Tarrytown, NY

<sup>10591
*</sup>Correspondence to: James M. Briggs, Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001.

E-mail: jbriggs@uh.edu Received 22 July 2004; Accepted 6 December 2004

Published online 6 April 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20447

binds a structural Zn2+ ion.8-10 The N-terminal domain promotes multimerization of the enzyme and enhances its catalytic activity. 10 The catalytic domain (residues 50-212) contains a conserved D,D-35-E motif, common to all retroviruses, retrotransposons, and some bacterial transposases. 11-13 This motif is represented in IN by the residues Asp64, Asp116, and Glu152. These residues are all crucial for the catalytic activity, and it has been proposed that 2 divalent metal ion cofactors are coordinated by these residues.^{5,14-17} Finally, the C-terminal domain (residues 212–288) appears to have a nonspecific DNA binding activity. ^{15,18,19} All of the isolated domains form homodimers in solution, and their 3D structures are publicly available. 20-25 There is no structural information about how the domains are organized in the active complex with DNA, but it seems that IN is likely to function as an octamer.²⁶ The most complete crystal structure of HIV-1 IN revealed that a divalent metal ion bound to the active site interacts with residues Asp64 and Asp116. The active site loop (residues 140–145) is located close to the active site, and mutations that constrain the conformational flexibility of this particular region have demonstrated that its mobility is important for the catalytic activity of IN,²⁷ and that it is involved in the stabilization of the substrate during the reaction with the DNA-substrate. 28,29

A large number of classes of compounds have been identified as active either against the 3'-processing or the strand transfer reactions. 30-33 Recently, a series of compounds, typified by an aryl β-diketo motif, has been identified by Merck.³⁴ These compounds prevent HIV-1 replication in cells by selectively inhibiting the strand transfer process. These molecules represent one of the most promising classes of IN inhibitors in terms of potency and selectivity of their inhibitory activity, with L-731,988 and L-708,906 (molecules 1 and 2, Fig. 1) representing 2 of the most active compounds. Single or multiple mutations of HIV-1 IN at residues T66, S153, and M154 confer degrees of resistance to one or more compounds belonging to this class.³⁴ However, structural information is not available for the mutated enzyme, nor for the binding modes of diketo acids within the active site of the IN catalytic domain. The only available X-ray structure of IN in complex with an inhibitor in the active site contains 5CITEP, the novel Shionogi inhibitor (molecule 3, in Fig. 1).35 This small molecule contains similar chemical features to the compounds belonging to the diketo acids class: a β-diketo motif in the center of the molecule; an aromatic group on the left portion of the molecule; and an acidic group (either a carboxylic group or an ionizable heteroaryl ring) on the right portion of the molecule. These similarities in the molecular architecture between the Shionogi and Merck inhibitors have led to important insights into the binding modes of these molecules, even though the orientation of 5CITEP within the active site seems to be affected by crystal packing, and its activity is significantly lower than that of L-731,988. Molecular docking calculations, 36-38 MD simulations, 39,40 and experimental stud $ies^{34,38,41}$ have been used in an attempt to identify the type of interactions that take place between the HIV-1 IN

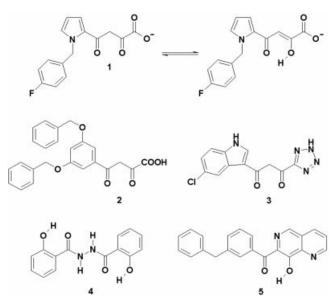


Fig. 1. Structures of HIV-1 integrase inhibitors. L-731,988 and its tautomeric form (1), L-708,906 (2), 5CITEP (3), salicylhydrazide derivative (4), 8-hydroxy-[1,6]-naphthyridine (5).

enzyme and different inhibitors. The latest hypotheses regarding the interactions between the protein and diketo acid inhibitors point toward a metal ion chelation mechanism: The β-diketo motifs, as well as other functional groups (i.e., salicylhydrazide derivatives; compound 4 in Fig. 1) of the inhibitors, interact with the divalent IN metal ion cofactor, preventing them from participating in the catalytic activity of the enzyme. 33,38,41,42 Furthermore, a recent report⁴³ has suggested that novel bioisosteric diketo acid equivalents containing an 8-hydroxy-[1,6]naphthyridine moiety (compound 5 in Fig. 1), inhibit HIV-1 IN via a mechanism similar to that of the diketo acids. The formation of an IN-inhibitor complex capable of constraining the flexibility of the loop region containing residues 138-149 is another hypothesis that has been proposed to explain the decrease in the enzymatic activity^{27,40} induced by the compounds belonging to this class.

Our aim in the present work is to study, at an atomic level, the mechanism of HIV-1 IN T66I/M154I inhibitor resistance against the Merck compound L-731,988. The resistance to this compound, based on mutations in the enzyme target (i.e., IN), may be justified in 2 different ways: First, the inhibitor may no longer be able to bind within the active site after the protein mutations; second, the inhibitor may still be able to bind within the active site of the protein, but with an orientation that allows the DM IN to maintain its catalytic activity.

Because of the lack of structural information concerning both the DM HIV-1 IN and the binding mode of the compound L-731,988 in the enzyme, we used a comparative computational approach to perform our study. We have carried out 2 MD simulations on DM IN in complex with L-731,988, with 2 different orientations within the active site. Two orientations of the inhibitor have been determined by molecular docking calculations performed against several protein conformations, randomly extracted

from prior MD trajectories obtained from the T66I/M154I IN catalytic domain without the inhibitor in complex (Brigo et al., manuscript in preparation). The results of the aforementioned MD simulations of the DM IN with no inhibitor have been compared to those obtained by the MD simulations carried out for the same system in the presence of the inhibitor. Moreover, in order to compare the dynamic behavior of the DM IN/L-731,988 complex with the nonresistant enzyme in complex with the same inhibitor, we have also performed an MD simulation of the catalytic domain of the 1QS435 IN PDB structure (see Methods section). The orientation of the inhibitor was also determined from docking calculations, which are also contrasted against the conformation of the 5CITEP/1QS4 complex. Two orientations of the inhibitor L-731,988 have been identified within the IN active site. One orientation is very similar to that determined for 5CITEP in the crystal structure of the catalytic domain of IN.35 The other orientation facilitates the interaction of the L-731,988 β-diketo motif with the divalent cofactor Mg²⁺. The IN (1QS4 and DM) conformations that favor one or the other inhibitor orientation are discussed in the Results section.

We show that the dynamic behavior of HIV-1 T66I/M154I IN is not affected by the presence of the inhibitor, regardless of its orientation within the active site. Conversely, an apparently important constraint in the catalytic loop region (residues 138–149) occurs in the 1QS4 enzyme when L-731,988 is in complex. Furthermore, water molecules involved in the stabilization of protein—inhibitor interactions have been identified.

In order to gain additional insight into the binding modes of this inhibitor in the active site of 1QS4 and DM IN, we have performed binding free energy calculations, using the computed MD trajectories, with the MM/PBSA methodology. We observed that the binding free energy between the HIV-1 IN 1QS4 and L-731,988 is more favorable than that calculated for the 2 MD simulations carried out for the inhibitor-resistant DM enzyme, in agreement with the experimental results. 34,35

METHODS

Docking

Docking studies were performed with AutoDock v3.0.⁵⁰ The program performs automated docking of flexible ligands to an enzyme receptor by rapid energy evaluation achieved by precalculating atomic affinity potentials for each atom type in the substrate molecule. This program uses a grid-based procedure (AutoDock module AutoGrid3.0) in such a way that the protein is embedded in a 3D grid and a probe atom is placed at each grid point. The interaction energy of this single atom with the protein is assigned to the grid point. The LGA was employed to find the best orientations, conformations, and binding positions of the ligand.

The following equation is used to model the free energy of binding, adding to the classical molecular mechanics equation the entropy contribution:

$$\Delta G = \Delta E_{vdw} + \Delta E_{hbond} + \Delta E_{elec} + \Delta E_{conf} + \Delta G_{tors} + \Delta G_{sol}$$

 ΔG_{tors} models the restriction of internal rotors, and global rotation and translation; ΔG_{sol} models desolvation.

The automated docking of the ligand L-731,988 against the DM and the 1QS4 HIV-1 IN catalytic domains has been carried out using a standard protocol with an initial population of 50 randomly placed individuals, a maximum number of energy evaluations of $1.5\times10^6,$ and a maximum of 27,000 generations. The mutation rate was set to 0.02 with Cauchy distribution parameters $\alpha=1$ and $\beta=0.$ The crossover rate was set to 0.80 and the elitism to 1. The local search was performed with the Pseudo-Solis and Wets algorithm using a maximum of 300 iterations per local search. The probability of performing a local search on an individual in the population was 0.06, and the step size, $\rho,$ of the local search space to sample was 1. The maximum number of consecutive successes or failures before doubling or halving ρ was 4.

We carried out 50 independent docking runs for each chosen DM IN and 1QS4 IN conformation. Results differing by less than 0.15 nm in positional RMSD were clustered together and ranked according to free energy of binding.

Ligand setup

The structure of the Merck IN inhibitor (L-731,988) was generated with Insight II. Since this compound is a 1,3-diketone, we had to take into account its keto–enolic tautomeric equilibrium, which, in solution, favors the enolic form because of the formation of a strong intramolecular hydrogen bond. The carboxylic group of the ligand is deprotonated in water because of its low pKa value and it has been consequently modeled in the inhibitor. The ligand was next minimized with DISCOVER using the CVFF force field through 200 steps using the steepest descent algorithm.

Atomic charges were assigned using the Gasteiger–Marsili formalism,⁵² and the AutoTors procedure implemented in AutoDock Tools⁵³ was used to define 4 flexible torsions to be taken into account during the docking calculations. Finally, nonpolar hydrogen atoms in the ligand were merged.

Protein setup

In order to take into account protein flexibility during the docking studies, we used different conformations of the protein extracted from MD simulations. The HIV-1 1QS4 IN catalytic domain conformations used for the docking studies were chosen at 200-ps intervals from a prior 1.8-ns MD simulation⁴⁰ carried out for the 1QS4 crystallographic structure³⁵ in complex with the Shionogi inhibitor (5CITEP). The 1QS4 structure (subunit A, residues 56-140, 145-209) was completed from residues 140-145, as previously described. 40 Note that this structure (1QS4) contains 2 mutations that confer suitable solubility properties to the catalytic domain (F185K/W131E) for crystallographic studies. Since there is no indication that these mutations affect the catalytic activity, we maintained these mutations in our simulations. Therefore, what we call a drug-resistant "double mutant" is really a quadruple

mutant, although only the DM affects drug efficacy. When we refer to the "1QS4 IN" in this article, we are referring to the F185K/W131E IN that is reported in the 1QS4 crystallographic structure. This gave us 9 different conformations of the protein from which we have removed the inhibitor 5CITEP. The 9 snapshots were minimized with DISCOVER 51 using the CVFF force field through 200 steps of steepest descent.

The HIV-1 DM IN conformations were also chosen at 200 ps intervals from two 2-ns MD trajectories carried out for the system with no inhibitor in complex. These 2 trajectories were taken from a previous study (Brigo et al., manuscript in preparation) in which we carried out a total of 4 MD simulations on the DM enzyme starting from different conformations. The purpose of that study was to determine if there were significant differences in the DM IN dynamical behavior with respect to the 1QS4 IN. Given the importance of the catalytic loop comprising residues 138-149,²⁷ we chose the 2 trajectories that showed the highest transition in the catalytic loop in order to take into account the highest number of the loop conformations. We obtained 20 snapshots of the DM IN, which were used for our docking studies. All of these conformations were minimized with DISCOVER⁵¹ using the CVFF force field through 200 steps of steepest descent. During this minimization, a cutoff of 1.0 nm and a distance-dependent dielectric constant of $1 \times r$ have been employed in order to reduce artifacts resulting from missing solvent.

These structures were set up for docking studies using the utilities implemented in AutoDock Tools. Partial charges were assigned using Kollman united atom charge-type nonpolar hydrogen atoms, and lone pairs were merged; last, solvation parameters were added. Histidine protonation states were modeled as reported in previously carried out MD simulations (A. Brigo et al., manuscript in preparation). The grid maps, representing the protein in the docking process, were calculated with AutoGrid3.0, with the grid box centered on the IN metal cofactor (Mg²⁺). The grid box dimensions were set to $3.0 \times 3.0 \times 3.0$ nm, with a spacing of 0.0375 nm between the grid points, in order to include a larger portion of the IN active site. In this way, we take into account all of the protein regions that may be involved in interactions with L-731,988.

Molecular Dynamics Simulations

From the molecular docking results, we chose 3 most favorable IN complexes with compound L-731,988 (i.e., the ligand orientation in the active site that led to the highest number of conformations in 1 cluster at the lowest binding free energy). Those systems (protein and ligand) were used as starting points for our MD simulations. We carried out 1 MD simulation on the HIV-1 1QS4 IN catalytic domain in complex with L-731,988 and 2 MD simulations on HIV-1 DM IN catalytic domain in complex with L-731,988 oriented in 2 different ways within the active site. The details of the ligand setup for the MD calculations, including atomic net charge calculations along with inhibitor parameters, are provided in the Supplementary Material.

The systems (DM IN and 1QS4 in complex with L-731,988) were solvated in a cubic box (with 7.2-nm edges) of pre-equilibrated waters. Water molecules that were found to be within 0.26 nm of any atom in the solute were removed. The system was neutralized by replacing 1 solvent water molecule with 1 chloride ion. The counterion was placed in proximity of the region of the protein surface at the most favorable electrostatic potential, calculated using the particle-based method implemented in the GRO-MACS package v3.0.5.54 The chloride ion was placed in the proximity of protein regions not important for the catalytic activity and far away from the active site. The protein and solvent interact via the CHARMM27 force field, where all protein atoms are explicitly represented,55 and where water is characterized by the TIP3P model. 56,57 The hydrogen atoms of the protein were placed using the HBUILD routine implemented in CHARMM.⁵⁸ Amino acid ionization states were chosen based on an earlier Poisson-Boltzmann-based protein pKa prediction made for this system.⁵⁹ The SHAKE algorithm,⁶⁰ with a tolerance of 10⁻⁷ nm was applied to constrain all bonds involving hydrogen atoms in all simulations.

The system involving 1QS4 IN finally contained 2407 solute atoms (protein, ligand, Mg2+ ion, chloride ion) and 34,572 solvent atoms. The systems involving DM IN contained 2414 solute atoms (protein, ligand, Mg2+ ion, chloride ion), and between 34,509 and 34,576 solvent atoms. The difference in the total number of water molecules was due to the different conformations of the protein and ligand orientations that require a slightly different number of waters to be properly solvated. The temperature was kept constant by coupling to a heat bath through the Berendsen algorithm.⁶¹ The pressure was adjusted by isotropic position scaling using a Berendsenlike algorithm. To avoid edge effects, periodic boundary conditions were applied to all the simulations. Atomic coordinates were saved for later analysis every 0.5 ps. A 1.0-nm short-range cutoff was used for all nonbonded interactions, and long-range electrostatic interactions were treated by PME method, 62 with a grid size of 7.2 nm \times 7.2 $nm \times 7.2 nm$.

Before the MD runs were performed, the systems were energy minimized by 200 steps of steepest descent, keeping the solvent and the chloride ion fixed in order to minimize the energy of interaction between the IN and L-731,988. Next, 200 steps of steepest descent were run, keeping the solute (protein, ligand, Mg ion and chloride ion) fixed, followed by 1000 steps of conjugate gradient with no constraint. The systems were all heated up to 300 K over a period of 30 ps, with a 2-fs timestep and preliminarily equilibrated for 70 ps, with the velocities being reassigned in the system every 0.1 ps. The systems were next equilibrated with a 2-fs timestep for 50 ps, at which point the energy of the system achieved stability. Production runs were performed at 300 K and carried out under NPT conditions for 2 ns, with a 2 fs timestep.

All MD calculations were performed using the NAMD program, ⁶³ and the analyses of the trajectories were

performed using the GROMACS software package 54 and CHARMM version $27b2.^{58}$

Essential dynamics 64,65 is a technique that enables the identification of the correlated motions of a protein during a trajectory generated by an MD simulation. This type of analysis also shows the concerted atomic motions that take place in different parts of the protein structure during the trajectory. After removal of the rotational motions, a covariance matrix is constructed. Diagonalization of the matrix leads to a set of eigenvectors/eigenvalues. Each eigenvector represents a single direction in a multidimensional space, while the eigenvalue is the amplitude of the motion along the eigenvector. Only a few eigenvectors show large eigenvalues, which means that the essential/ correlated motions of a protein occur along only a few spacial directions, while along the other directions, there are only small Gaussian (harmonic) fluctuations. The projection of the displacements on each eigenvector shows the width of the essential space explored by the system as a function of time.

MM/PBSA Methodology

The free energy corresponding to each MD snapshot (taken at 10-ps intervals) has been estimated using the MM/PBSA methodology. 44-49 In this approach, only solute degrees of freedom are explicitly represented, so that the potential of mean force W is written as the sum of solute energy term $U(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_n)$ and a solvation free energy term, which can be further split in a polar (electrostatic) and a nonpolar (hydrophobic) term:

$$W = U(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n) + \Delta G^{polar} + \Delta G^{nonpolar}. \tag{1}$$

The solute energy may be used to estimate the solute enthalpic component of the free energy corresponding to each conformation. This choice is roughly equivalent to the assumption that each conformation is a sample from an ensemble of conformations possessing the same potential of mean force. The entropic part is more difficult to estimate. The method proposed by Kollman et al.⁴⁹ involves normal mode analysis of the conformation, which, in turn, requires extensive structure minimization that is known to introduce artifacts. We have neglected this contribution altogether, assuming that this contribution will be similar for all conformations. This assumption is reasonable for compact structures.

The solute energy term $U(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n)$ has been computed using CHARMM version c27b2, a classic and well-tested molecular mechanics force field. All structures have been energy minimized with 200 steepest descent and 200 conjugate gradient minimization steps, in order to remove high-energy spots. During this minimization, a cutoff of 1 nm and a distance-dependent dielectric constant of $1 \times r$ have been employed in order to reduce artifacts ensuing from missing solvent. After minimization, the energy was evaluated with no cutoff and with a dielectric constant of 1.0 in order to be consistent with the PBSA solvation energy estimation. ΔG^{polar} has been computed according to the Poisson–Boltzmann theoretical framework $^{66-69}$ as the difference in free energy for the hypothetical charging

process of the solute in vacuo and in an ionic solvent. In order to remove self-energy contributions, which are strongly dependent on grid mesh and positioning, 2 identical computations are performed, differing only in the solvent dielectric constant and ionic strength (80 and 100 mM, respectively, for solvent calculation, and 1 and 0 mM, respectively, for the in vacuo calculation). The choice of 1.0 for interior dielectric constant requires a word of explanation. In most studies involving the MM/PBSA approach or some of its variants, the interior dielectric constant is assumed to be in the range 2-4, typical of organic molecules. In applications exploring MD simulations within the MM/PBSA approach, it was found that substantially higher dielectric constant values (4.0-17.0) must be employed in order to achieve reasonable MD trajectories, thus posing serious problems of consistency between the MM and PB energy estimations. 44 Ramachandran maps generated with dielectric constant values other than 1.0 do not show the typical low-energy regions corresponding to representative backbone conformations. While we suspect that many problems in the MM/PBSA MD simulations are due to the numerical approximations involved in the PBE solution, we preferred to adopt the value of 1.0 for the dielectric constant in order to be more consistent with the MM energy estimation and correctly estimating for model compounds, the energy of backbone conformations. No significant dependence of the results on the dielectric constant has been found (see Results section). The following protocol has been used: First the linearized PBE is solved on a large grid, with 40³ grid points, with 0.25 nm spacing between the grid points, centered on the protein in order to obtain boundary conditions for the following focusing steps. Then, the same equation is solved on smaller and finer grids (333 points at a grid spacing of 0.045 nm) centered on each amino acid's center of geometry. In each of the focusing steps, the potential at each atomic position of the amino acid is stored and used later to estimate the electrostatic free energy of the protein, which is expressed for the linear PBE as

$$\Delta G = \sum_{i} \frac{q_i \phi_i}{2}.$$
 (2)

The linearized PBE is appropriate for moderately charged biomolecules. AGnonpolar is taken to be proportional to the solvent-accessible surface area A (i.e., $\Delta G^{nonpolar} = \gamma A$). Although a large range of values have been used in the literature for the surface tension coefficient γ , we have used a value of 20 kcal mol⁻¹ Å⁻², which is lower than the 50 kcal mol⁻¹ Å⁻² value proposed by Nicholls et al. To proteins. The rationale for this choice is that in the absence of solvent molecules, unbalanced intramolecular van der Waals forces provide a strong tendency toward collapse. The free energy of binding (not entailing translational and rotational entropy restrictions, which should, however, be almost equal among the 3 complexes considered here) was estimated as follows:

$$egin{aligned} \Delta G^{bind} &= U^{inter} + \Delta G^{solv}(complex) - \Delta G^{solv}(protein) \ &- \Delta G^{solv}(inhibitor) \end{aligned}$$

TABLE I. Results of 50 Independent Docking Runs Computed for the Inhibitor L-731,988 Onto Different Conformations of 1QS4 and T66I/M154I IN

	Total No. of Clusters	No. of Members in Selected Clusters ^b	$\Delta G_{ m bind}$ (kcal/mol)	Protein-ligand Interactions		
				Hydrogen bonds	$ m vdW^c$ interactions	Hydrophobic contacts
1QS4 IN conformations	s^a					
1 (0 ps)	6	40	-10.68	Q148, N155	D64, D116	Q62, H114, T115, G140, I151,
2(800ps)	4	38	-8.13	Q148	D64, D116, E152, N155, K159	Q62, H114, T115, F139, G140
$3(1400\mathrm{ps})$	9	33	-7.64	_	D116, E152	Q62, H114, T115, G140, I141, I151
4 (1600 ps)	8	26	-7.81	_	D64, D116, N155, K159	Q62, H114, T115, F139, G140, I141, I151
T66I/M154I IN conform	nationsa					
$5 (200 \mathrm{ps}, \mathrm{T1^d}) \\ \mathrm{RMSD^f: 0.11 nm}$	17	31	-8.89	N155	E152, K159	Q62, G140, F139, I141,
$6 (1000 \mathrm{ps, T1^d}) \ \mathrm{RMSD^f: 0.27 nm}$	16	27	-8.61	N155, K159	H67, E92, N117	I151
$7 (2000 \text{ ps}, \text{T1}^{d})$ RMSD ^f : 0.22 nm	24	19	-7.05	H67	D92, N120, Q143, N155, K159, Mg ²⁺	C65, I66, F121, V150
$8 (1200 \text{ ps, } T2^{e})$ RMSD ^f : 0.20 nm	19	24	-9.18	Q62, N155	D116, K156, K159	T115, I151
9 (2000 ps, T2 ^e) RMSD ^f : 0.26 nm	10	41	-9.88	Q62, E152	D64, K156	T115, D116 F139, G140, I141, I151

^aIN snapshots that gave the most significant results in the docking studies.

where U^{inter} is protein–drug interaction energy computed using CHARMM and ΔG^{solv} is the solvation energy estimated using PBSA methodology. ΔG^{solv} is computed for the complex and for isolated protein and drug excised from the complex without any further optimization.

RESULTS

Automated Docking

Docking of the L-731,988 compound (1 in Fig. 1), modeled in its keto-enolic form (as explained in the Methods section), with the negatively charged carboxylate group exhibited a consistent set of binding modes for all of the 1QS4 and DM IN conformations used in this study. In general, every investigation carried out by 50 independent docking runs converged to a small number of different orientations within the active site of the protein. We docked L-731,988 into 29 snapshots of IN, taken at 200 ps intervals from prior trajectories generated for both systems. A prior 1.8-ns MD simulation carried out for the 1QS4 IN in complex with the Shionogi inhibitor (5CITEP)⁴⁰ gave us 9 different conformations. HIV-1 DM IN conformations were also chosen at 200-ps intervals from two 2-ns-long MD trajectories that were computed on the DM system with no inhibitor (Brigo et al., manuscript in preparation). In Table I, the results from the docking calculations are shown for the most representative protein conformations (9 out of a total of 29 investigated). By "the most representative protein conformations," we mean those that gave very similar clustering results and a similar average ligand orientation. In general, the first clusters ranked according to the best binding free energy are also the most populated (1 cluster contains ligand conformations differing by less than 0.15 nm in RMSD), showing a reasonable consistency of the searching algorithm. Our results show (see Table I) that the binding free energy is negative (i.e., favorable) for all of the examined protein conformations, as well as having a good clustering of the ligand conformations. As we mentioned in the Introduction, there is only 1 crystal structure of IN showing the orientation of 5CITEP (compound 3 in Fig. 1) in the active site (1QS4)35; therefore, we took the relative position of this compound as a reference point and we compared it with the ligand orientations obtained from our docking calculations. Figure 2 shows the most frequently identified orientation of the ligand within the 1QS4 IN catalytic site: It has an RMSD value (calculated on all non-H atoms) of 0.234 nm with respect to the relative position of 5CITEP found in the IN crystal structure.³⁵ Figures 3 and 4 show the 2 main relative positions of the inhibitor L-731,988 within the DM IN active site. As shown in Figure 2, the best orientation identified by AutoDock⁵⁰ for L-731,988 is slightly rotated with respect to 5CITEP (in the 1QS4 crystal structure), with the fluoro-benzyl ring interacting with the side-chain of Asp116.

^bSelected cluster represents the cluster with the highest number of members and the most favorable binding free energy.

^cvan der Waals interactions.

^{d,e}Trajectories obtained by MD simulations carried out for DM IN, with no inhibitor in complex.

^fBackbone RMSD calculated using 1QS4 IN 0-ps conformation as reference.

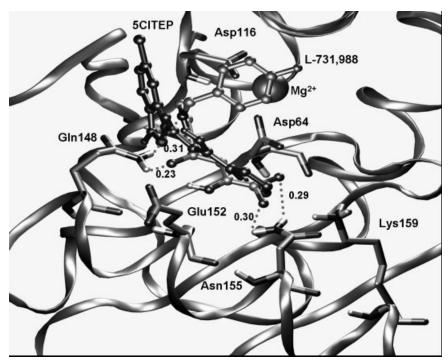


Fig. 2. Most favorable relative orientation of the inhibitor L-731,988 within the 1QS4 IN active site obtained by docking calculations. The compound 5CITEP is displayed in dark gray ball-and-stick representation, and overlaid onto the protein with the same orientation adopted in the crystal structure. ⁷³ The thick dashed lines indicate the most important interactions and relative distances expressed in nanometers between the inhibitor and the 1QS4 IN residues.

The keto-enolic group points out of the active site (toward bulk solvent) and makes no interactions with either the metal ion or Glu152. On the other hand, the keto oxygen of L-731,988 is hydrogen-bonded to the side-chain of Gln148. Gln62 and His114 are involved in hydrophobic contacts with the carbon atoms of the pyrrolic ring, while the Asn155 side-chain is hydrogen-bonded to the carboxylate group of the inhibitor. The orientations of L-731,988 in the 1QS4 IN active site found in our study are all characterized by relatively small adjustments in their positions with respect to the orientation that 5CITEP adopted in the molecular docking calculations performed in 2000 by Sotriffer et al. 36 and corroborated by the published crystal structure (1QS4), 35 maintaining the major ligand-protein interactions described above. In all of our results on the 1QS4 enzyme, the carboxylic group of L-731,988 adopts the same relative position of the tetrazole ring (both of them are deprotonated) as that of 5CITEP in the crystal structure. At the same time, the pyrrole ring of the Merck inhibitor occupies the same region of the active site of IN as the chloro-indole ring of the 5CITEP inhibitor in the crystal structure, while the fluoro-benzyl ring of L-731,988 shows a larger local RMSD with respect to the chloroindole ring of 5CITEP. Conversely, DM IN conformations allow 2 main ligand orientations. The first one (see Fig. 3) is close to the 5CITEP orientation identified in the IN crystal structure with an RMSD value of 0.238 nm. This value is in perfect agreement with the one found in the case of 1QS4 IN, thus demonstrating that L-731,988 may adopt an orientation similar to the one adopted by 5CITEP

for both the 1QS4 and DM enzymes. In this case, the fluoro-benzyl group is not interacting with the Asp116 side-chain as in the 1QS4 complex, but it is involved in hydrophobic contacts with Ile141 and Ile151. Moreover, the nitrogen of the pyrrolic ring is hydrogen-bonded to the Gln62 side-chain, and the carboxylate of the ligand is hydrogen-bonded to the Asn155. The keto-enolic motif is pointing outward from the active site and, contrary to the 1QS4 case, it is hydrogen-bonded to the side-chain of Glu152. In Figure 4, the second allowed orientation of L-731,988 in the DM IN active site is displayed. The RMSD value between this ligand orientation and the position adopted by 5CITEP in the crystal structure is 0.746 nm. The keto-enolic moiety interacts with the Mg²⁺ ion complexed between the residues Asp64 and Asp116. This orientation is interesting, since it would be in agreement with the hypothesis that the mechanism of inhibition of this compound is due to chelation of the divalent metal ion cofactor by the keto-enolic group. 38,41,42 The carboxylate group of the ligand interacts via a hydrogen bond with His67, while the pyrrole ring is involved in hydrophobic interactions with the side-chain of Cys65. The difference in the DM IN conformation that may lead to the inhibitor orientation displayed in Figure 4 is given by a differing orientation of the side-chain of Gln62. In fact, when the side-chain of Gln62 is capable of forming a hydrogen bond with the pyrrole nitrogen of the inhibitor, the predominant ligand orientation is similar to the one adopted by 5CITEP in the 1QS4 crystal structure. On the contrary, when the side-chain of Gln62 is too distant to form a hydrogen bond,



Fig. 3. Most favorable relative orientation of the L-731,988 inhibitor within the T66l/M154l DM IN active site obtained by docking calculations. The compound 5CITEP (ball-and-stick representation, dark gray) has the same orientation as the one adopted in the crystal structure.⁷³ The thick dashed lines indicate the most important interactions and relative distances expressed in nanometers between the inhibitor and the DM IN residues.

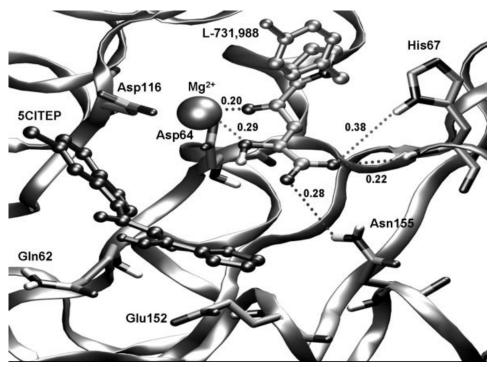


Fig. 4. The second most favorable orientation of the L-731,988 inhibitor within the T66I/M154I DM IN active site obtained by docking calculations. The compound 5CITEP (ball-and-stick representation, dark gray) has the same orientation as the one adopted in the crystal structure. The thick dashed lines indicate the most important interactions and relative distances expressed in nanometers between the inhibitor and the DM IN residues.

the ligand adopts an orientation that allows the direct interaction between the ${\rm Mg}^{2^+}$ ion and the keto–enolic moiety (Fig. 4). It is noteworthy that this type of binding mode was identified only with 1 conformation of DM IN (Table I, DM IN conformation 7) and it showed a less favorable binding free energy, as well as a higher number of clusters (i.e., more diffuse binding) with respect to all of the other docking studies.

Similar docking solutions are found for different protein conformations. In particular, DM IN conformation 5 (Table I) has the lowest RMSD (calculated using the 0 ps 1QS4 backbone conformation as the reference) of 0.11 nm, and the docking solutions are perfectly comparable with those found for the other DM IN conformations (with an RMSD from the reference structure of up to 0.27 nm) in terms of ligand orientation, number of clusters, and binding free energy (Table I).

In summary, the most favorable orientations of the inhibitor L-731,988 have been found in the center of 1QS4 IN, as well as in the DM IN active site, in a location very close to the one adopted by 5CITEP in the crystal structure. This demonstrates that the analogies in the structures between 5CITEP (compound 3 in Fig. 1) and L-731,988 (compound 1 in Fig. 1) are sufficient to lead to similar complexes, regardless of the mutated residues and of the different starting protein conformations. This affinity has been further confirmed by additional docking studies carried out on different DM IN conformations with the diketo acid inhibitor L-708,90634 (compound 2 in Fig. 1). The low binding affinity complexes given by this ligand are correlated to those exhibited by L-731,988 in the DM IN conformations (see Supplementary Material) in particular with reference to the binding free energy and the docking pose.

Molecular Dynamics Simulations

In order to compare the dynamical behavior of 1QS4 IN and the T66I/M154I DM when the inhibitor L-731,988 is bound in the active site, we performed 3 MD simulations: 1 on the 1QS4 IN with the ligand complexed in an orientation comparable to the one that 5CITEP adopted in the crystal structure, and 2 on the DM IN with the ligand complexed in 2 different orientations within the active site. The 3 starting complexes have been chosen from the docking results described above. The starting point for the 2-ns MD simulation carried out for the 1QS4 IN catalytic domain with L-731,988 in the active site was the docked complex obtained from conformation 1, as indicated in Table I. This conformation led to the best clustering (40 members out of 50) and the best binding free energy (-10.68 kcal/mol) among all of the 9 docking studies performed on the 1QS4 enzyme.

The starting points for the 2 MD simulations carried out for the T66I/M154I DM IN catalytic domain interacting with L-731,988 were the docked complexes obtained from DM conformations 7 and 9 in Table I. Conformation 7 (Table I) led to only 1 complex that allowed the inhibitor to interact with the Mg²⁺ ion through its keto-enolic motif; this type of complex is in agreement with the hypothesis

that IN inhibition occurs by means of divalent cofactor chelation. 38,41,42 Conformation 9 (Table I) is the one that showed the best binding free energy (-9.88 kcal/mol), as well as the best clustering (i.e., the highest number of members in the selected cluster; 41 members) and the lowest number of clusters identified (10 clusters). The orientation of the inhibitor found in this complex is similar to the one adopted by 5CITEP in the crystal structure.

In one of our studies (Brigo et al., manuscript in preparation), we addressed the issue of the convergence of the results coming from MD simulations carried out for the same system starting from different conformational states. We showed that 4 MD simulations carried out for the DM IN catalytic domain starting from 4 different conformations led to perfectly comparable results in terms of sampled conformational space. Moreover, we demonstrated that 2-ns simulations are sufficient for achieving convergence for a system of this size (i.e., 154 residues) and of well-defined secondary structure (Brigo et al., manuscript in preparation).

We carried out the MD simulations for the systems described above to contribute to a comparative study, in order to gain insight into the following issues: (1) dynamical behavior of the T66I/M154I DM IN with and without (Brigo et al., manuscript in preparation) the inhibitor L-731,988; (2) dynamical behavior of the protein (1QS4 and DM) during the 3 MD simulations involving the presence of the ligand adopting different orientations within the IN active site; and (3) dynamical behavior of the inhibitor L-731,988 when it is in complex with 1QS4 and DM IN, in order to determine the most plausible modes of binding of this ligand as a function of time, taking into account protein flexibility.

T66I/M154I double mutant HIV-1 integrase

In the present work, we compare the results from the MD simulations carried out for the DM IN with the inhibitor L-731,988 in complex, with 2 different orientations adopted within the active site and on the same system without any ligand (Brigo et al., manuscript in preparation). In our previous work, we carried out 4 MD simulations on the system without inhibitor. We chose a representative trajectory (designated as unbound hereafter) to be compared with those computed in the presence of the ligand. This unbound simulation (unbound) showed the most stable RMSD for the protein backbone atoms and a relatively lower flexibility (with respect to the other 3 MD trajectories) in the region comprised by residues 138-149. Figure 5 shows the backbone atom positional RMSD with respect to each equilibrated conformation for the DM IN with (designated as ion-bound-DM and bound-DM) and without (unbound) L-731,988 in complex. In the bound-DM simulation, we can see that the RMSD is very similar to the one in the unbound simulation (Fig. 5), almost perfectly overlapping throughout the entire length of the simulation. The RMSD achieves stability after about 250 ps and slightly increases up to 0.30 nm. This behavior is due to the high mobility of the most flexible regions of the protein that cause structural changes during the

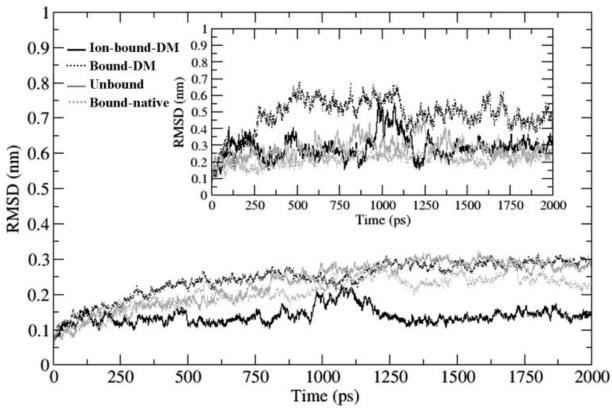


Fig. 5. RMSD of DM (T66I/M154I) HIV-1 IN and 1QS4 backbone atoms with respect to the equilibrated conformation as a function of time: lon-bound-DM stands for MD simulation carried out for DM IN in complex with the L-731,988 in the orientation that allows the interaction between the keto-enolic group of the inhibitor and the enzyme Mg²⁺ cofactor (black line); Bound-DM stands for MD simulation carried out for DM IN in complex with L-731,988 in a relative orientation similar to the one adopted by 5CITEP in the IN crystal structure (black dashed line); Unbound stands for MD simulation carried out for DM IN with no inhibitor in complex (gray line); Bound-native stands for MD simulation carried out for IN 1QS4 in complex with L-731,988 in a relative orientation similar as the one adopted by 5CITEP in the IN crystal structure (gray dashed line). The inset shows the RMSD calculated for the catalytic loop (residues 138–149).

entire length of the simulations designated as bound-DM and unbound. From the analysis of the RMSD fluctuations per residue of the backbone atoms (Fig. 6), we can see that the T66I/M154 IN has 3 highly flexible regions. These regions comprise residues 66–74, 138–149, and 185–195. The same features are present for all 3 simulations, and the region of the catalytic loop (residues 138–149) shows high flexibility in all of the simulations regardless of the presence or absence of the inhibitor. It is noteworthy that the presence of the inhibitor in the active site (which is close to the catalytic loop) of DM IN does not affect the mobility of the catalytic loop (residues 138–149) despite the fact that the 2 different binding modes investigated herein are rather different.

In order to obtain insight into the essential space sampled by DM IN during the trajectories generated from the MD simulations, as well as the concerted motions of the protein, we performed essential dynamics analyses. ^{64,65} The eigenvalues identified for eigenvector 1 (typically representing the essential motion occurring with the highest amplitude) in the 3 analyzed trajectories (ionbound-DM, bound-DM, and unbound) have approximately the same value (data not shown), showing that the properties of the motion amplitudes along the specified eigenvec-

tor are almost identical regardless of the presence of the inhibitor in the active site. By projecting the trajectories along each of the most important eigenvectors (Fig. 7), it is possible to represent the essential space explored by the protein along each direction. Note that the protein showed significant essential motions only along 4 out of 462 eigenvectors. From Figure 7, we can see that the amount of essential space explored by DM IN is almost identical along the 4 most important eigenvectors. The pathway followed is very consistent for all of the 3 trajectories (ion-bound-DM, bound-DM, and unbound) and after 2 ns of MD simulation, the explored essential space converges to the same point along all 4 eigenvectors. The C_a displacement along the 4 main eigenvectors (Fig. 8) can provide information about the concerted motions of the different parts of the protein along each direction. The catalytic loop (residues 138-149) shows high mobility for all of the simulations along the 4 main eigenvectors. This demonstrates that the presence of the inhibitor does not affect the essential space accessible by the catalytic loop of DM-IN.

DM IN and 1QS4 IN in complex with L-731,988

The results obtained from the analysis of the MD simulation carried out for the 1QS4 IN in complex with the

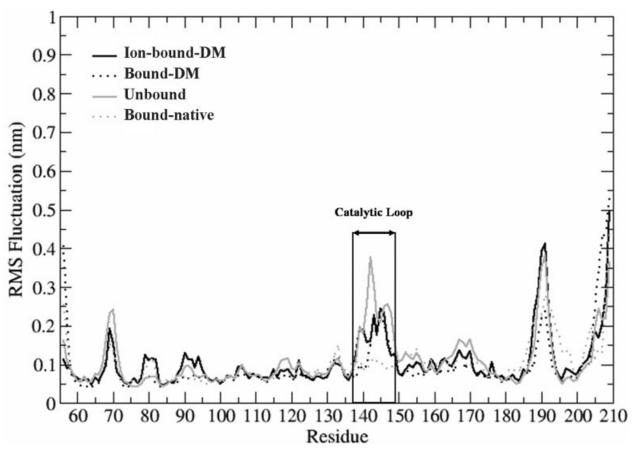


Fig. 6. Backbone atom root-mean-square fluctuation per residue for the 3 trajectories computed on the T66I/M154I double mutant IN: Ion-bound-DM, MD simulation carried out for DM IN in complex with the L-731,988 in such an orientation that allows the interaction between the keto-enolic group of the inhibitor and the enzyme Mg²⁺ cofactor (black line); Bound-DM, MD simulation carried out for DM IN in complex with L-731,988 in a relative orientation similar to that found for the compound 5CITEP in the IN crystal structure (black dashed line); Unbound, MD simulation carried out for DM IN with no inhibitor in complex (gray line); Bound-native, MD simulation carried out for IN 1QS4 in complex with L-731,988 in a relative orientation similar to the one adopted by 5CITEP in the IN crystal structure (gray dashed line).

L-731,988 (designated here as bound-native) have been compared to the results obtained from the other 2 MD simulations carried out for the DM IN complexed with the same inhibitor. As we can see from Figure 5, the RMSD of the backbone atoms of 1QS4 IN, calculated with respect to its equilibrated conformation (i.e., 0 ps of production run), achieves stability after 250 ps, and its profile is fairly stable during the whole simulation. After 1120 ps, the RMSD of 1QS4 IN undergoes a small increase up to about 0.25 nm, a value that is maintained until the end of the simulation. The overall RMSD for the bound-native simulation is lower than the backbone RMSD calculated for the simulation bound-DM. The RMSFs per residue (Fig. 6) show a good overlap among the 3 MD simulations (ionbound-DM, bound-DM, and bound-native), with the consistent identification of the most flexible protein regions. The catalytic loop region comprised of residues 138-149 for the 1QS4 IN displays very low mobility. However, the simulation carried out in 1999 by Lins et al. 59 on 1QS4 IN with no inhibitor in complex showed very high flexibility in this region. The fact that the catalytic loop is constrained only in the simulation of the bound-native system demonstrates that the presence of the inhibitor in the 1QS4 IN active site affects the catalytic loop degrees of freedom. The RMSD of the backbone atoms of the catalytic loop region (residues 138–149) is clearly more stable and lower (see inset, Fig. 5) for the bound-native system (1QS4 IN) than for the other simulations. It achieves stability very quickly and maintains a constant value comprised between 0.20 nm and 0.25 nm for the entire time of the simulation. Moreover, essential dynamics analysis confirmed the constraint of the catalytic loop (residues 138–149) in the MD simulation. The C_{α} displacement along the 4 main eigenvectors shows that the 1QS4 IN catalytic loop is one of the most stable parts of the protein, while in the case of the DM IN simulations, it displays high mobility along eigenvectors 1, 2, and 3 (Fig. 8).

L-731,988 binding modes

The drug resistance showed by DM IN against the inhibitor L-731,988³⁴ may be explained by the different types of interactions that occur between the ligand and the protein, and sometimes mediated by water molecules. The RMSDs of all non-hydrogen atoms of the inhibitor with respect to each equilibrated orientation within the active site (i.e., 0 ps of production run) have been calculated for

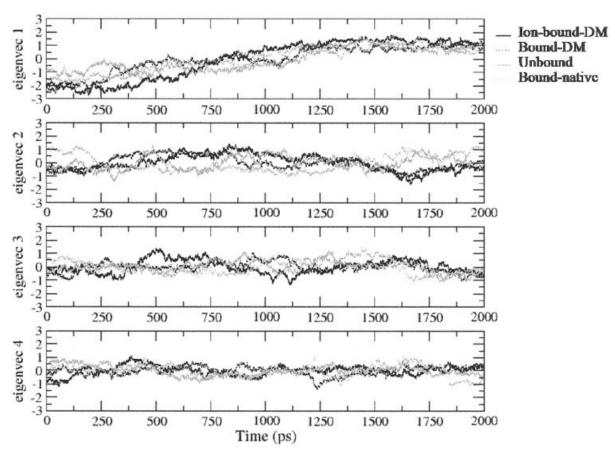


Fig. 7. Projections on the 4 selected eigenvectors of the MD simulations carried out for DM IN (Ion-bound-DM, black line; Bound-DM, black dashed line; Unbound, gray line) and 1QS4 IN (Bound-native, gray dashed line).

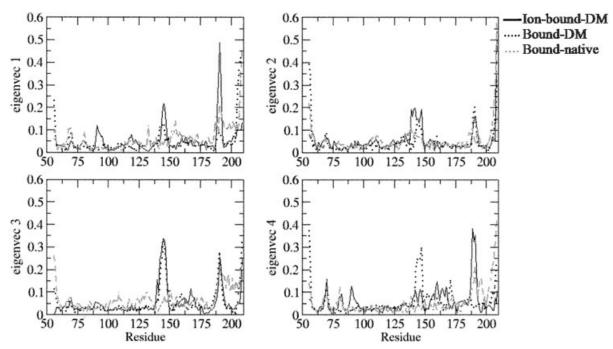


Fig. 8. Residue displacement calculated for each C_{α} atom along the 4 selected eigenvectors for the MD simulations carried out for DM IN (lon-bound-DM, black line; Bound-DM, black dashed line) and 1QS4 IN (Bound-native, gray dashed line).

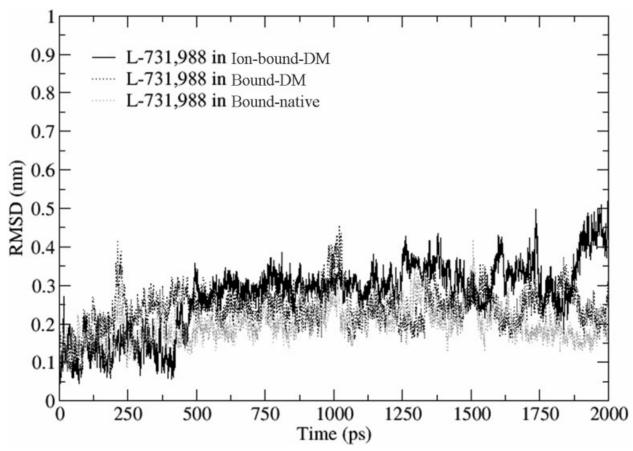


Fig. 9. RMSD from the equilibrated structure measured for all non-H atoms of the L-731,988 ligand. The least-square fit has been performed on the backbone atoms of the protein, taking into account the rotations and translations of L-731,988. Ion-bound-DM, black line; Bound-DM, black dashed line; Bound-native, gray dashed line.

simulations ion-bound-DM, bound-DM and bound-native (Fig. 9). As can be seen, the inhibitor never exhibits very high mobility within the active site of the protein. The RMSD value fluctuates between 0.20 nm and 0.30 nm, achieving a maximum of 0.5 nm for simulation ion-bound-DM. This indicates that the orientations generated from the docking studies are fairly well preserved during the entire length of the simulations. The pathway followed by the inhibitor in the simulations bound-DM and boundnative (these are the 2 simulations in which the inhibitor adopts the same relative orientation within the active site) is very similar, revealing wide overlapping regions during the MD simulations. In the simulation ion-bound-DM, the inhibitor maintains a stable RMSD until 500 ps, thereafter increasing suddenly from 0.15 nm up to 0.30 nm. At the beginning of simulation, the keto-enolic motif of the ligand interacts with the Mg²⁺ ion (Figs. 4 and 10); the carboxylate group is hydrogen-bonded to the nitrogen atom of His67 side-chain, and the pyrrolic ring is involved in hydrophobic interactions with Cys65. These interactions are conserved for about 500 ps, until the ligand undergoes a rotation that leads to the loss of the contacts with the Mg²⁺ ion for the remainder of the simulation (Fig. 10). During the last 1500 ps, the keto-enolic motif of L-731,988 forms a hydrogen bond with His67, and the carboxylate group interacts with Asn155 and Lys159.

At the beginning of the bound-DM simulation (i.e., MD) simulation carried out for DM IN in complex with L-731,988 in a relative orientation similar to that found for the compound 5CITEP in the IN crystal structure), the Gln62 side-chain is hydrogen-bonded to the pyrrolic ring of the ligand, as well as to the sp2 oxygen of the keto-enolic motif. Moreover, the residue Asp116 is interacting with the fluoro-benzyl group, and the side-chain of Glu152 is also hydrogen-bonded to the keto-enolic motif (Fig. 3). The pyrrolic ring of L-731,988 is the most stable functional group (RMSD fluctuating between 0.12 nm and 0.15 nm) as it is hydrogen-bonded to the side-chain of Gln62 for the entire simulation. The fluoro-benzyl group also exhibits small RMSD values (i.e., 0.2 nm) and it maintains the interaction with residue Asp116 for the entire simulation. The keto-enolic motif (including the carboxylate group) displays the highest RMSD, achieving a value of 0.60 nm after 800 ps of simulation. This is due to an internal rotation of the inhibitor after 800 ps that led to the interaction between the carboxylate group of the ligand with the Mg²⁺ ion and the loss of the hydrogen bond between the side-chain of Glu152 and the keto-enolic

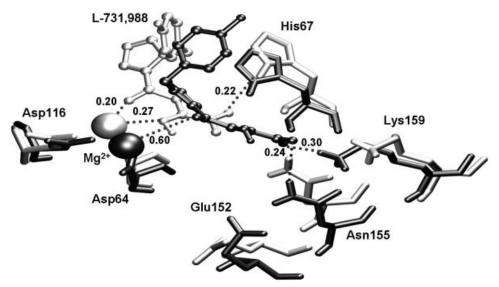


Fig. 10. Representation of the rotation experienced by L-731,988 during the ion-bound-DM MD simulation. The starting conformation for the protein and the ligand are colored in light gray. The final conformation for the protein and the ligand after 2 ns of simulation is colored dark gray. The dashed thick lines indicate the most important interactions and relative distances expressed in nanometers between the inhibitor and the DM IN residues.

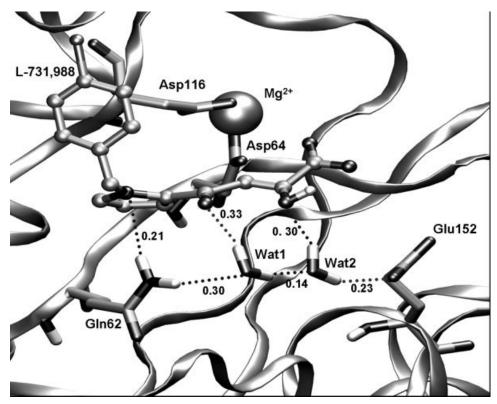


Fig. 11. Representation of the most important interactions between DM IN and L-731,988 (ball-and-stick representation) occurring after 800 ps of the bound-DM MD simulation.⁷⁴ The dashed thick lines indicate the most important interactions and relative distances expressed in nanometers between the inhibitor and the DM IN residues.

motif (Fig. 11). The analysis of the water molecules that may be directly involved in the stabilization of the protein—inhibitor interaction during the bound DM simulation has revealed the constant presence of 2 water molecules (see

Fig. 11) that connect, by means of hydrogen bonds, the side-chain of Gln62 with the backbone nitrogen of Glu152. Furthermore, 1 of the 2 water molecules is involved in bridging the sp2 oxygen of the inhibitor with the side-

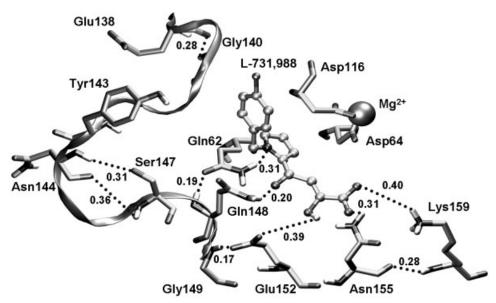


Fig. 12. Hydrogen bond network occurring between 1QS4 IN and the L-731,988 inhibitor during the bound-native MD simulation. The catalytic loop (residues 138–149) is represented as ribbon.⁷³ Distances are expressed in nanometers.

chain of Gln62. These waters play a significant role in the stabilization of the interaction between L-731,988 and DM IN, and in hampering the ligand from moving within the active site. More importantly, the inhibitor does not interact with any residues of the catalytic loop (residues 138–149), which maintains its high mobility and its capability of facilitating enzymatic function.

In the bound-native simulation, the 3 functional groups of the ligand exhibit a similar RMSD to that calculated from simulation bound-DM (data not shown). The pyrrolic ring is the most stable functional group, maintaining a 0.10-nm RMSD for the entire simulation. This group is constantly hydrogen-bonded to the side-chain of Gln62 and is involved in hydrophobic interactions with Asp116. The fluoro-benzyl group is constantly involved in hydrophobic interactions with the side-chains of His114 and Asp116, and the keto-enolic motif, together with the carboxylate group, undergoes a larger RMSD, fluctuating between 0.10 nm and 0.50 nm. The sp2 oxygen of the keto-enolic motif is hydrogen-bonded to the Gln148 side-chain for the entire length of the trajectory. We have pointed out that the catalytic loop (residues 138-149) of the 1QS4 IN is constrained when L-731,988 is in complex (Figs. 6 and 9). Gln148 is part of this loop and is constantly hydrogenbonded to the ligand (Figs. 2 and 12). Moreover, the carbonyl oxygen of the side-chain of Gln62 is hydrogenbonded to the backbone nitrogen of Gln148. The side-chain of Ser147 is also hydrogen-bonded to the backbone carbonyl oxygen of Asn144 (Fig. 12). The catalytic residue Glu152 is hydrogen-bonded to Gly149 and to the enolic group of the inhibitor at the same time. This hydrogen bond network utilizes the ligand as a bridge between the flexible loop and the inner portions of IN active site (Fig, 12). Moreover, the side-chain of Tyr143, which is also a residue in the active site loop and is proposed to play a role in the catalytic activity, ^{29,59} is always pointing away from the active site in all the simulations carried out for the DM IN. Conversely, in the case of bound-native simulation, these specific interactions will cause the catalytic loop to adopt a constrained conformation that leads the Tyr143 side-chain to point toward the active site (Fig. 12). This orientation might either be unfavorable for the formation of multimeric active complex of IN or may constitute a hindrance that disallows the substrate to bind to or interact with residues of the active site.

Binding Free Energy Calculations

For every snapshot taken from the MD simulations, the free energy of binding (not entailing translational and rotational entropy) was computed according to the MM/ PBSA methodology. The results for the 3 different molecular models are reported in Figure 13. The snapshots taken from the bound-native simulation show that this system has the lowest average free energy of all 3 complexes (ion-bound-DM, bound-DM, and bound-native DM) with a value of -23.0 ± 5.8 kcal/mol. The bound-DM simulation has a binding free energy of -15.7 ± 7.4 kcal/mol, which is higher than the value calculated for the bound-native. Overall, these results are in agreement with the experimental results published by Hazuda et al.,34 who reported that L-731,988 is 15 times less active against the DM IN than in the wild-type. Our results correctly predict that the interaction of the inhibitor with 1QS4 IN is very favorable and that, in fact, the catalytic activity is inhibited. Interestingly, the predicted binding free energy for the 2 DM IN complexes demonstrates that the interaction is less favorable in both cases. In particular, the ion-bound-DM simulation shows a totally unfavorable free energy (15.9 \pm 14.0 kcal/mol), confirming further that the best relative orienta-

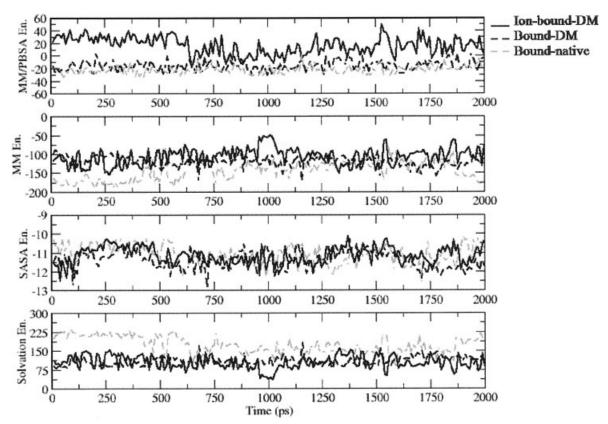


Fig. 13. MM/PBSA energy (calculated on every 10 ps snapshot of the MD simulations of the 3 complexes) plotted versus time. Ion-bound-DM, black line; Bound-DM, black dashed line; Bound-native, gray dashed line.

tion of L-731,988 within IN active site is the one similar to the position found for the 5CITEP in the crystal structure.

The dependence of these results on the dielectric constant values has been tested. The choice of dielectric constant is not free of constraints, because force field parameters are consistent with a dielectric constant 1.0. The solvation energy component of the computed binding free energy scales approximately with dielectric constant. In particular, for the complexes studied here, we have seen that the solvation free energy of binding scales almost exactly with dielectric constant. Therefore, a test on the influence of the dielectric constant on computed free energies has been carried out as follows: Both the Coulombic and electrostatic solvation energy components have been divided by the inverse of a set of dielectric constant values. For all tested dielectric constant values (from 1.0 to virtually infinite), the conclusions concerning the ranking of MM/PBSA free energy of binding is maintained. This is due to the fact that for the bound-native and bound-DM simulations, similar electrostatic binding energies are computed, which are quite lower than those calculated for the ion-bound-DM simulation.

The effect of increasing the dielectric constant is to lower the binding free energy for all complexes, because for all of them, electrostatics (Coulombic interactions and solvation) opposes the binding. For dielectric constant 1.0, the average electrostatic component of the binding free energy is 16.0 kcal/mol for the bound-native simulation, 40.6

kcal/mol for ion-bound DM simulation, and 17.2 kcal/mol for the bound-DM simulation. The free energy difference between the 3 complexes is reduced by increasing the dielectric constant by an amount proportional to the figures reported above.

Although the ranking of free energy of binding of the 3 complexes studied is not influenced by the dielectric constant choice, the computed values of binding free energy are, however, dependent on the chosen dielectric constant.

DISCUSSION

In the present work, we have carried out computational studies on HIV-1 1QS4 IN and T66I/M154I DM in complex with the Merck inhibitor L-731,988. Because of the lack of structural information concerning the IN-ligand complex, we performed automated docking calculations in order to find the most favorable ligand orientations, taking into account protein flexibility. The results showed a complete consistency in the relative orientation adopted by the ligand within both 1QS4 IN and DM. The ligand position that we found is in good agreement with that seen for 5CITEP in the only available IN crystal structure with an inhibitor in complex. 35 Both 5CITEP and L-731,988 belong to the diketo acid inhibitor class, and their functional groups (i.e., tetrazole ring and carboxylic group, as well as chloro-indole ring and fluoro-benzyl-pyrrole ring, respectively) are bioisosteres (see compounds 1 and 3 in Fig. 1).

As we mentioned before, the relative orientations of the ligand were very similar for most of the 1QS4 IN and DM conformations taken into account in the docking studies, with the inhibitor substructures occupying the same regions within the binding pockets. Despite the similar relative position, the interactions that occur between the 1QS4 and DM IN showed significant differences (Table I). The side-chain of Gln148 in 1QS4 IN is hydrogen-bonded to the keto-enolic motif of the ligand in 7 conformations out of 9 studied, while in DM IN, Gln62, Asn155, and Lys159 are the residues forming hydrogen bonds with the ligand in 14 conformations out of 20 analyzed. Interestingly, we found a completely different orientation of the inhibitor when the side-chain of Gln62 points out from the active site in 5 of the DM IN conformations. This orientation allows for the direct interaction of the keto-enolic motif with the Mg2+ ion, as well as a hydrogen bond between the carboxylate group and His67.

We carried out a comparative MD study in order to gain insight into the possible differential dynamical behavior of 1QS4 and DM IN when they are in complex with the inhibitor. By comparing the simulation carried out for the DM IN with no inhibitor in complex (Brigo et al., manuscript in preparation) with those carried out for the same mutated enzyme with L-731,988 complexed in 2 different orientations, we observed an almost identical dynamical behavior for the protein. The RMSF for the 2 MD simulations carried out for the DM IN identified the same most flexible regions, in particular the protein portion comprised of residues 138-149, whose flexibility has already been shown to be important for catalytic activity. The concerted motions and the essential space accessible by the system identified by essential dynamics analysis have also been found to be comparable for the DM IN, regardless of either the presence of the inhibitor or the adopted orientation at the beginning of the simulations.

The analysis of the MD simulation carried out for 1QS4 IN in complex with L-731,988 revealed similar dynamical features to those found for DM IN, except for the catalytic loop region, which appeared to be constrained during the entire simulation. The inhibitor L-731,988 is directly hydrogen-bonded to Gln148, which is part of the catalytic loop only in the 1QS4 IN simulation, also favoring a hydrogen bond network (Fig. 12) that maintains the stability of this protein region. Moreover, the side-chain of Tyr143 points toward the active site in the 1QS4 IN simulation, while it points outward in the DM IN simulations, with and without the inhibitor in complex (Brigo et al., manuscript in preparation). This Tyr143 side-chain orientation, together with the catalytic loop rigidity, may explain the biological activity of the inhibitors belonging to this class, which act in a way that can either render difficult the interaction of the enzyme with the biological substrate or the formation of the active IN multimeric complex. The inhibitor, in simulation bound-DM (i.e., MD simulation carried out for DM IN in complex with L-731,988 in a relative orientation similar to that found for the compound 5CITEP in the IN crystal structure) was slightly

more buried in the active site than in the complex with 1QS4 IN. This position allowed the inhibitor to have conservative direct hydrogen bonds with Gln62 that kept the ligand position stable during the simulation. Moreover, 2 water molecules have been identified as playing a role in the ligand stabilization, anchoring it to residues Gln62 and Glu152. However, no direct interaction has been found between L-731,988 and the catalytic loop residues. These characteristics confer high flexibility to the catalytic loop (region comprising residues 138–149) and may guarantee enough space within the active site to accommodate the biological substrate. It is worth noting that the slight difference in the binding modes of L-731,988 within DM IN and the 1QS4 active site is sufficient to give different binding affinities through specific interactions that, in the case of the bound-native complex, are inducing a strong restraint of the flexibility of the catalytic loop. The T66I/M154I mutations are therefore conferring drug resistance by inducing the modification of Gln148 side-chain orientation that renders impossible the hydrogen bond interactions between L-731,988 and Gln148 itself. Moreover, a different Gln148 side-chain orientation seems to be directly involved in maintaining Tyr143 side-chain in an orientation potentially more suitable for the catalytic activity.

An interesting ligand orientation, in accordance with the most recent hypotheses of diketo acid mechanism of action, 38,41,42 has been found in our docking studies. This ligand orientation has been investigated by the ionbound-DM MD simulation. The results show that the contacts between the keto-enolic motif of the inhibitor and the divalent metal cofactor last for only 500 ps of MD simulation, leading to the formation of a hydrogen bond between His67 and the same ligand motif. This evidence goes against the latest proposed mechanism of IN inhibition, 38,41,42 meaning that the orientation we found may be either totally unfavorable or requires the presence of the second metal ion (that we did not model, since no 3D structure of HIV-1 IN shows both ions) to be strongly stabilized and, consequently, would make this relative orientation favorable. The binding free energy values calculated with the MM/PBSA methodology also support the hypothesis that this is not a likely binding mode. L-731,988 interacted favorably only with 1QS4 IN (boundnative simulation, -23.0 ± 5.8 kcal/mol), while the interactions in the 2 DM IN complexes were significantly less favorable, particularly for the ion-bound-DM. We pointed out that the relative orientation adopted by the ligand in simulation bound-DM was very similar to that adopted in the native-bound, but it revealed different specific interactions with the protein residues that may render the formation of the complex possible.

The discrepancy found between the results given by the scoring function implemented in AutoDock and the values of binding free energy calculated with MM/PBSA methodology is due to the different methodology used by the 2 approaches. AutoDock uses a semiempirical approach, where scaling factors of various terms have been optimized to fit binding energies for approximately 30 compounds. In

this sense, it is a semiempirical method. The MM/PBSA methodology does not use any adjustable parameter but employs previously determined force field parameters and standard values for representing the solvent as a continuum. In this sense, it may be considered a first principles methodology. It is, however, reassuring that similar values for the binding energies are found (10–15 kcal/mol should be subtracted from MM/PBSA to take into account global and local rotation and translation restrictions).⁷⁴

In this article, we have gained insight into the binding modes of the ligand L-731,988 with 1QS4 and DM IN, as well as into the dynamical behavior of the 2 enzymes. We provide further evidence that the higher rigidity of the catalytic loop and a more widely occupied active site of 1QS4 IN, with respect to the DM protein, can be used to explain the diketo-acid compound mechanism of IN inhibition.

ACKNOWLEDGMENTS

Our thanks to NRAC for a grant of supercomputer time (to J. M. Briggs). Additional computing, data storage, and visualization resources were provided by the Institute for Molecular Design (IMD) and the Texas Learning and Computation Center at the University of Houston (UH). Gratitude is also expressed to Accelrys, Inc. for software licenses provided to the IMD at UH. A. Brigo thanks Dr. Maria Letizia Barreca for valuable discussions. F. Fologan thanks Drs. G. Tecchiolli and P. Zuccato of Exadron, the HPC Division of the Eurotech Group, for providing hardware and technical assistance.

REFERENCES

- Asante-Appiah E, Skalka AM. Molecular mechanisms in retrovirus DNA integration. Antiviral Res 1997;36:139–156.
- 2. Hindmarsh P, Leis J. Retroviral DNA integration. Microbiol Mol Biol Rev 1999;63:836–843.
- Brown PO. Integration. In: Coffin JM, Hughes SH & Varmus HE, Editors. Retroviruses. Plainview, NY: Cold Spring Harbor Laboratory Press: 1997. p 161–203.
- Ellison V, Brown PO. A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro. Proc Natl Acad Sci USA 1994;91:7316–7320.
- Wolfe AL, Felock PJ, Hastings JC, Uncapher Blau C, Hazuda DJ.
 The role of manganese in promoting multimerization and assembly of human immunodeficiency virus type 1 integrase as a catalytically active complex on immobilized long terminal repeat substrates. J Virol 1996;70:1424–1432.
- Engelman A, Bushman FD, Craigie R. Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. EMBO J 1993;12:3269– 3275.
- Johnson MS, McClure MA, Feng DF, Gray J, Doolittle RF. Computer analysis of retroviral pol genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. Proc Natl Acad Sci USA 1986;83:7648-7652.
- Haugan IR, Nilsen BM, Worland S, Olsen L, Helland DE. Characterization of the DNA-binding activity of HIV-1 integrase using a filter binding essay. Biochem Biophys Res Commun 1995;217:802–810.
- Lee SP, Han MK. Zinc stimulates Mg²⁺-dependent 3'-processing activity of human immunodeficiency virus type 1 integrase in vitro. Biochemistry 1996;35:3837–3844.
- Zheng R, Jenkins TM., Craigie R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. Proc Natl Acad Sci USA 1996;93:13659– 13664.
- 11. Engelman A, Craigie R. Identification of conserved amino acid

- residue critical for human immunodeficiency virus type 1 integrase function in vitro. J Virol 1992;66:6361-6369.
- Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol Cell Biol 1992;12:2331–2338.
- 13. Polard P, Chandler M. Bacterial transposases and retroviral integrases. Mol Microbiol 1995;15:13–23.
- Ellison V, Brown PO. A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro. Proc Natl Acad Sci USA 1994:91:7316-7320.
- Vink C, Lutzke RA, Plasterk RH. Formation of a stable complex between the human immunodeficiency virus integrase protein and viral DNA. Nucleic Acid Res 1994;22:4103–4110.
- Hazuda DJ., Felock PJ, Hastings JC, Pramanik B, Wolfe AL. Differential divalent cation requirements uncouple the assembly and catalytic reactions of human immunodeficiency virus type 1 integrase. J Virol 1997;71:7005–7011.
- Beese LS, Steitz TA. Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. EMBO J 1991;10:25–33.
- Woerner AM, Marcus-Sekura CJ. Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. Nucleic Acid Res 1993;21:3507–3511.
- Engelman A, Hickman AB, Craigie R. The core and carboxyterminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. J Virol 1994;68:5911–5917.
- Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to the other polynucleotidyl transferases. Science 1994;266:1981–1986.
- Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, Davies DR. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. Proc Natl Acad Sci USA 1998;95:9150-9154.
- Maignan S, Guilloteau JP., ZhoupLiu Q, Clement-Mella C, Mikol V. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. J Mol Biol 1998;282: 359-368.
- Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, Clore GM, Gronenborn AM. Solution structure of the DNA binding domain of HIV-1 integrase. Biochemistry 1995;34: 9826-9833.
- Eijkelenboom AP, Lutzke RA, Boelens R, Plasterk RH, Kaptein R, Hard K. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. Nat Struct Biol 1995:2:807–810.
- Cai ML, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. Nat Struct Biol 1997;4:839–840.
- 26. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y, De Clercq Y, Debyser Z. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J Biol Chem 2003;278:372–381.
- Greenwald J, Le V, Butler SL, Bushman FD, Choe S. The mobility
 of an HIV-1 integrase active site loop is correlated with catalytic
 activity. Biochemistry 1999;38:8892

 –8898.
- Heuer TS, Brown PO. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. Biochemistry 1997; 36:10655-10665
- Esposito D., Craigie R. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. EMBO J 1998;17:5832–5843.
- Fesen MR, Kohn KW, Leteurtre F, Pommier Y. Inhibitors of human immunodeficiency virus integrase. Proc Natl Acad Sci USA 1993;90:2399-2403.
- Fesen M, Pommier Y, Leteurtre F, Hiroguchi S, Yung J, Kohn KW. Inhibition of HIV-1 integrase by flavones, caffeic acid phenethyl ester (CAPE) and related compounds. Biochem Pharmacol 1994;48: 595–608.
- 32. Robinson WE, Reinecke MG, Abdel-Malek S, Jia Q, Chow SA. Inhibitors of HIV-1 replication that inhibit HIV integrase. Proc Natl Acad Sci USA 1996;93:6926-6331.
- 33. Neamati N, Hong H, Owen JM, Sunder S, Winslow HE, Chris-

- tensen JL, Zhao H, Burke TR Jr, Milne WA, Pommier Y. Salicylhydrazide-containing inhibitors of HIV-1 integrase: implication for a selective chelation in the integrase active site. J Med Chem 1998:41:3202–3209.
- 34. Hazuda DJ, Felock P, Witmer M, Wolfe A, Stillmock K, Grobler JA, Espeseth A, Gabryelski L, Schleif W, Blau C, Miller MD. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. Science 2000;287:646-650.
- 35. Goldgur Y, Craigie R, Cohen GH, Fujiwara T, Yoshinaga T, Fujishita T, Sugimoto H, Endo T, Murai H, Davies DR. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. Proc Natl Acad Sci USA 1999:96:13040-13043.
- 36. Sotriffer CA, Ni H, McCammon JA. HIV-1 integrase inhibitor interactions at the active site: prediction of binding modes unaffected by crystal packing. J Am Chem Soc 2000;122:6136-6137.
- Sotriffer CA, Ni H, McCammon JA. Active site binding modes of HIV-1 integrase inhibitors. J Med Chem 2000;43:4109-4117.
- Pais GCG, Zhang X, Marchand C, Neamati N, Cowansage K, Svarovskaia ES, Pathak VK, Tang Y, Nicklaus M, Pommier Y, Burke TR Jr. Structure activity of 3-aryl-1,3-diketo-containing compounds as HIV-1 integrase inhibitors. J Med Chem 2002;45: 3184-3194.
- Ni H, Sotriffer CA, McCammon JA. Ordered water and ligand mobility in the HIV-1 integrase-5CITEP complex: A molecular dynamics study. J Med Chem 2001;44:3043–3047.
- Barreca ML, Lee KW, Chimirri A, Briggs JM. Molecular dynamics studies of the 1QS4 and double mutant HIV-1 integrase complexed with the 5CITEP inhibitor: mechanism for inhibition and drug resistance. Biophys J 2003;84:1450–1463.
- 41. Grobler JA, Stillmock K, Hu B, Witmer M, Felock P, Espeseth AS, Wolfe A, Egbertson M, Bourgeois M, Melamed J, Wai JS., Young S, Vacca J. Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. Proc Natl Acad Sci USA 2002;99:6661–6666.
- Neamati N, Lin Z, Karki RG, Orr A, Cowansage K, Strumberg D, Pais GCG, Voigt JH, Nicklaus MC, Winslow HE, Zhao H, Turpin JA, Yi J, Skalka AM, Burke TR, Pommier Y. Metal-dependent inhibition of HIV-1 integrase. J Med Chem 2002;45:5661–5670.
- 43. Zhuang L, Wai JS, Embrey MW, Fisher TE, Egbertson MS, Payne LS, Guare JP Jr, Vacca JP, Hazuda DJ, Felock PJ, Wolfe AL, Stillmock KA, Witmer MV, Moyer G, Schleif WA, Gabryelski LJ, Leonard YM, Lynch JJ Jr, Michelson SR, Young SD. Design and synthesis of 8-hydroxy-[1,6]naphthyridines as novel inhibitors of HIV-1 integrase in vitro and in infected cells. J Med Chem 2003;46:453-456.
- 44. Fogolari F, Esposito G, Viglino P, Molinari H. Molecular mechanics and dynamics of biomolecules using a solvent continuum model. J Comp Chem 2001;22:1830–1842.
- Honig B, Nicholls A. Classical electrostatic in biology and chemistry. Science 1995;268:1144–1149.
- Davis ME, McCammon JA. Electrostatics in biomolecular structure and dynamics. Chem Rev 1990;90:509

 –521.
- Fogolari F, Brigo A, Molinari H. The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology. J Mol Recognit 2002;15:377–392.
- 48. Baginski M, Fogolari F, Briggs JM. Electrostatic and nonelectrostatic contributions to the binding free energies of anthracycline antibiotics to DNA. J Mol Biol 1997;274:253–267.
- 49. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE III. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–897.
- Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 1998;19:1639–1662.
- 51. ÎnsightII. San Diego, CA: Accelrys Inc.; 2000.
- 52. Gasteiger J, Marsili M. Iterative partial equalization of orbital

- electro negativity—a rapid access to atomic charges. Tetrahedron 1980;36;3219-3228.
- 53. Goodsell DS, Morris GM, Olson AJ. Docking of flexible ligands: applications of AutoDock. J Mol Recognit 1996;9:1–5.
- 54. van der Spoel D, van Drunen R, Berendsen HJC. Groningen machine for chemical simulations. Department of Biophysical Chemistry, BIOSON Research Institute Nijenborgh 4 NL-9717 AG Groningen, 1994. E-mail to gromacs@chem.rug.nl
- 55. MacKerell AD, Bashford D, Bellot M, Dunbrack RL, Evanseck JD, Field MJ, MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586-3616.
- Jorgensen WL, Chandrasekar J, Madura J, Impey R, Klein M. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926-935.
- Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. J Chem Phys 1996;105:1902–1921.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. A program for macromolecular energy minimization and dynamics calculations. J Comp Chem 1983;4:187–217.
- Lins RD, Briggs JM, Straatsma TP, Carlson HA, Greenwald J, Choe S, McCammon JA. Molecular dynamics studies on the HIV-1 integrase catalytic core domain. Biophys J 1999;76:2999–3011.
- Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comp Phys 1977;23:327–341.
- Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
- Essmann U, Perera L, Berkowitz ML, Darden J, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys 1995;103; 8577–8593
- Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: greater scalability for parallel molecular dynamics. J Comp Phys 1999:151:283–312.
- Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins 1993;17:412

 –425.
- 65. van Alten DMF, Amadei A, Linssen ABM, Eijsink VGH, Vriend G, Berendsen HJC. The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. Proteins 1995;22:45–54.
- Fogolari F, Briggs JM. On the variational approach to the Poisson– Boltzmann free energies. Chem Phys Lett 1997;281:135–139.
- Sharp KA, Honig B. Calculating total electrostatic energies with the non-linear Poisson–Boltzmann equation. J Phys Chem 1990; 94:7684–7692.
- Marcus RA. Calculation of thermodynamic properties of polyelectrolytes. J Chem Phys 1955;23:1057–1068.
- Zhou HX. Macromolecular electrostatic energy within the nonlinear Poisson-Boltzmann equation. J Chem Phys 1994;100:3152

 3162
- Fogolari F, Elcock AH, Esposito G, Viglino P, Briggs JM, McCammon JA. Electrostatic effects in homeodomain–DNA interactions. J Mol Biol 1997;267:368–381.
- Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.
- 72. Fogolari F, Brigo A, Molinari H. A protocol for MM/PBSA molecular dynamics simulations of proteins. Biophys J 2003;85:159–166.
- Humphrey W, Dalke A, Schulten K. VMD—Visual Molecular Dynamics. J Mol Graph 1996;14:33–38.
- 74. Finkelstein AV, Janin J. The price of lost freedom: entropy of bimolecular complex formation. Protein Eng 1989;3:1-3.