JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

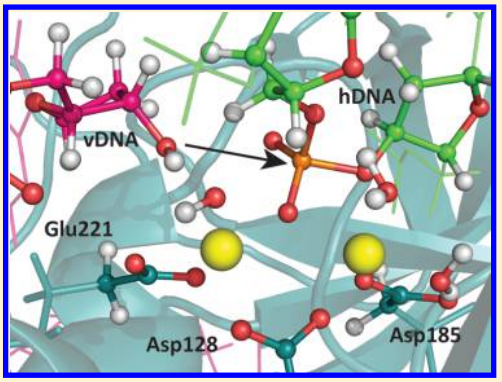
Catalytic Mechanism of Retroviral Integrase for the Strand Transfer Reaction Explored by QM/MM Calculations

Ana R. Araújo, António J. M. Ribeiro, Pedro A. Fernandes, and Maria J. Ramos*

REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal

Supporting Information

ABSTRACT: Integrase (IN) is one of the three fundamental enzymes for the HIV life cycle. It irreversibly inserts the viral DNA into the host DNA, infecting the host cells. Although there are 37 compounds currently used in the HIV-1 antiretroviral therapy, only three have IN as a target. Lack of structural and mechanistic information on IN greatly contributes to such a small number. Prototype Foamy Virus (PFV) IN has an enzymatic activity remarkably similar to HIV IN and is considered a model system to study the catalytic mechanism of HIV IN. Recently, the crystal structure of the PFV intasome became available, which allowed us to perform accurate high-level quantum mechanics/molecular mechanics (QM/MM) calculations to determine the strand transfer reaction mechanism followed by IN. We describe here, for the first time with atomic detail, the integration of a viral genome into the DNA of a host cell. We found that the strand transfer reaction mechanism has three distinct steps: deprotonation and activation of the nucleophile; S_N2 transesterification involving a pentacoordinated transition state; and protonation of the leaving group. The chemical steps have a limiting potential activation energy of 14.8 kcal/mol at the MPWB1K/6-311++G(2d,2p) level of theory, which is consistent with the upper limit established experimentally (25.1 kcal/mol) associated with the product release. This work improves the mechanistic knowledge on the IN chemistry and provides accurate structures of all the intermediates and transition states, which can be used as templates for the discovery of new IN inhibitors.



1. INTRODUCTION

Integrase (IN) is a fundamental enzyme to the Human Immunodeficiency Virus (HIV) life cycle that mediates the integration of viral DNA (vDNA) into the host DNA, a process that is crucial for the retroviral replication. After the integration process, the vDNA is irreversibly inserted into the chromosomes of the infected cells: it becomes impossible to completely eliminate the viral infection.[1] Up until now, the Food and Drug Administration (FDA) approved 37 compounds/drugs used in the highly active antiretroviral therapy (HAART),[2,3] from which only three target IN − Raltegravir, Elvitegravir, and Dolutegravir.[4−8] The lack of structural and catalytic information about HIV IN made it difficult to develop new drugs that target that enzyme.[9−13]

The integration process occurs in two catalytic steps: the 3′-end processing reaction and the strand transfer reaction. In the first step, IN removes two or three nucleotides from one or both 3′ ends of the long terminal repeats (LTRs) of the vDNA, exposing the 3′-hydroxyl groups. In the second step, IN catalyzes the attack of the 3′-hydroxyls groups of the processed vDNA to two phosphate groups of the host cell DNA, which results in the integration of the vDNA into the host DNA.[14−17] Both reactions occur in the same catalytic center, although the first one takes place in the cytoplasm, whereas the strand transfer reaction only happens after the translocation of the IN:vDNA complex to the nucleus of the host cell.[18]

Mutagenesis studies showed that two aspartates and one glutamate,[19] known as the D,DX_{35}E motif, and two magnesium ions are required for both IN reactions. The D,DX_{35}E motif is characteristic of the transposase family, which catalyzes the movement of DNA fragments from one part of the genome to another, using a single active site.[20] For this reason, the catalytic mechanism of IN should be akin to similar enzymatic reactions (breaking and formation of phosphodiester bonds) that occur in similar active sites like those of ribonuclease H1[21] and transposase Tn5.[22] The catalytic mechanism of HIV-IN for the 3′-end processing of vDNA was also recently described.[23]

The measured experimental turnover for the strand transfer reaction of HIV-IN is $7.6 \pm 1.4 \times 10^{-4}$ min^{-1}. This slow turnover is associated with the dissociation of the products, rather than with the chemical step. For this reason, the turnover gives us an upper limit for the reaction barrier (25.1 kcal/mol) but not its exact size.[24] Despite this wealth of information, what happens in atomistic scale during the strand transfer reaction is still not fully understood.

Received: July 3, 2014
Published: October 29, 2014

   dx.doi.org/10.1021/ct500570g | J. Chem. Theory Comput. 2014, 10, 5458−5466

Several structures of the isolated domains of HIV-1 were solved, but the exact three-dimensional structure of the full monomer is not known. It has been difficult to crystallize the entire HIV-IN protein due to poor solubility and interdomain flexibility problems.[25−31] However, prototype foamy virus integrase (PFV IN) has high solubility and uses small vDNA substrates for concerted integration assays, which makes it easier to study.[32] PFV IN has only 15% sequence similarity with HIV-1 IN, but the similarity increases to 22% when considering only the catalytic core domain. Furthermore, their enzymatic activity is remarkably similar.[33] Thus, PFV IN is considered a model system to study the catalytic mechanism of HIV-1 IN as well as the interaction with HIV-1 IN inhibitors.[32,34,35]

Retroviral IN proteins have three common domains: a N-terminal domain (NTD), the catalytic core domain (CCD), which includes the active site, and the C-terminal domain (CTD).[19,34,36,37] The NTD contains a characteristic zinc-binding HHCC motif and the CCD harbors the D,DX$_{35}$E motif and a pair of divalent metal cations (usually Mg^{2+} or Mn^{2+}), essential for both enzymatic activities.[17,19,38,39] The crystal structure of the prototype foamy virus (PFV) intasome (IN complexed with two strands of vDNA) revealed a fourth domain called N-terminal extension domain (NED).[34] The two reactions catalyzed by IN are supposed to occur via bimolecular nucleophilic substitution (S_N2), similar to the reactions catalyzed by metal-dependent nucleotidyl transferases and some nucleases.[17,21,40,41] The two metal ion cofactors seem to have two main roles: to maintain the reacting groups in an adequate conformation and to stabilize the postulated pentacoordinated transition state.[42−44] The X-ray crystal structure of PFV intasome was initially solved at 2.8 Å[34] and refined at a resolution of 2.0 Å,[45] revealing a tetramer of IN with a pair of synapsed viral DNA ends. As predicted on the basis of the partial structures,[30,46] the tetramer is composed of a dimer of dimers, where the inner subunits are responsible for the catalysis of the strand transfer reaction and the outer subunit seems to give support at the structure of the inner pair of working monomers.[47,48] Recently, the crystal structures of all PFV IN-DNA complexes observed in the retroviral integration became available. These structures finally showed the exact positions of the two metal ions and of the chemical reacting groups that take part in this reaction.[49]

Here, we establish the reaction mechanism of strand transfer catalyzed by PFV IN with atomic detail, by using high-level quantum mechanics/molecular mechanics (QM/MM) calculations.[50] These computational methods have been widely used in the past to elucidate the catalytic mechanism of many enzymes.[51−57] The ONIOM scheme was applied on the crystallographic structure of the target capture complex, in which the intasome of PFV is bound to the host DNA.[49] The QM layer was treated with the MPWB1K density functional[58] with a triple-ζ basis sets, and the MM layer was treated with the Amber force field. The entire model has more than 20000 atoms.

This work shows all the intermediates and transition states (TS) structures along the integration of the linear vDNA into the host cell chromosome as well as the respective activation and reaction energies. These TS structures can be used as a template for the screening of new IN inhibitors, and the disclosure of integration mechanism may help to solve some of the problems associated with HIV retroviral therapy.

2. METHODS

2.1. The Model. The starting point for all computational studies was the crystal structure of the target capture complex of PFV integrase containing all the components necessary for strand transfer in the reactants state, with a resolution of 3.02 Å (Protein Data Bank code: 4E7K).[49] The homodimer structure of the enzyme was constructed with the Protein Interfaces, Surfaces and Assemblies' service (PISA),[59] with the rotation and translation matrices proposed in the PDB file (citar artigo X-ray). The structure of each monomer comprises the four domains of PFV IN − NED, NTD, CCD, and CTD − and two extra CCD domains from another monomer (Figure 1A). The

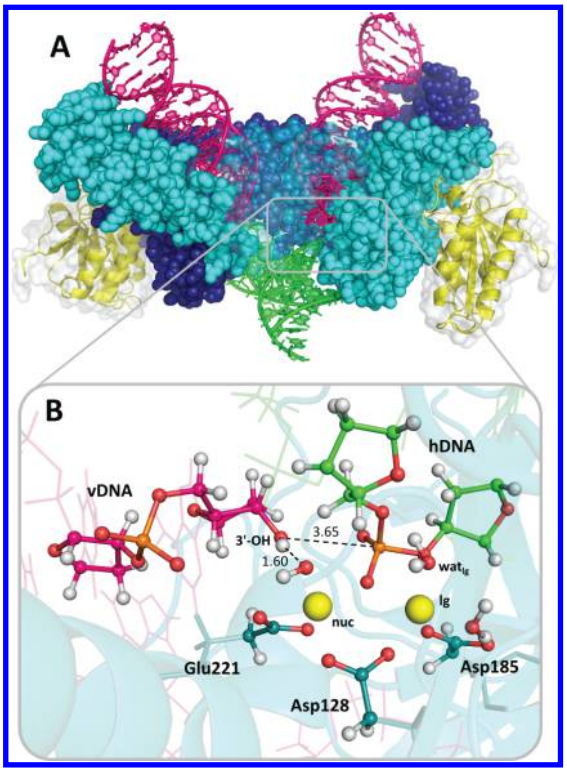


Figure 1. Model used in the QM/MM calculations. (A) The protein atoms of the two monomers are represented as blue and cyan spheres. The extra CCD domains of the other two monomers are represented in yellow cartoon. The vDNA is represented in pink cartoon and the hDNA in green cartoon. (B) Detail of the QM layer of the model. The protein portion of the QM layer is colored by atom type and depicted in ball and stick format. The Mg^{2+} ions are represented as yellow spheres. The QM portion of the DNA is represented in ball and stick as well but with the carbon atoms colored light green (hDNA) or pink (vDNA). The MM layer is represented in transparent dark green cartoon (enzyme) and pink and green sticks (vDNA and hDNA, respectively). Relevant interatomic distances (in Angstrom) are included.

structure of the NED domain lacks the first 8 amino acid residues, but they are not important for the catalytic mechanism.[32,60] The model also includes two double strands of viral DNA (vDNA) and a double strand of human DNA (hDNA) (Figure 1A). We replaced the manganese ions with magnesium ions, the natural cofactor.

The protonation state of every residue was checked with the H++ server[61] (Table S1). Hydrogen atoms and 55 counterions (Na^+) were added with the X-leap program[62] in the more electrostatically favorable positions, which were mainly near the

dx.doi.org/10.1021/ct500570g | J. Chem. Theory Comput. 2014, 10, 5458−5466

DNA phosphate groups. The water coordination spheres of magnesium ions (12 water molecules) were added with GaussView[63] in order to build an octahedral coordination sphere, typically observed in Mg^{2+} ions.

In order to stabilize the model and release bad contacts from the crystallographic structure, three sets of 250 steps of steepest descent energy minimization were performed with AMBER default parameters. In the first one only water molecules were minimized; in the second one, hydrogen atoms; and in the last one, all the atoms in the system were allowed to move. The minimization procedure was performed with the AMBER 9 simulation package,[62] using the ff03 force field for protein and nucleic acids atoms and the zinc AMBER force field (ZAFF) library[64] for the zinc ions and ligands in the N-terminal domains (parameters files are included in the SI).

2.2. QM/MM Calculations. The quantum mechanics/molecular mechanics (QM/MM) calculations performed to determine the potential energy surface (PES) of the strand transfer reaction were done with the Gaussian 09 program,[65] within the ONIOM scheme.[66−68] We used the previously optimized structure of the IN:DNA complex to build the QM/MM model, in which we included the entire enzyme:hDNA:vDNA complex and the 55 counterions. The system contains a total of 20538 atoms divided in two layers. The QM layer has 86 atoms and includes the acetate portion of the side chain of the three catalytic residues (Asp128, Asp185, and Glu221), the two magnesium ions, the metal-bound water molecules (two water molecules and one hydroxide), and a "portion" of the substrate that includes the scissile phosphate group, the previous 3′ phosphate group, and the four ribose rings connected to those. Details can be observed in Figure 1B, and the coordinates of the system can be consulted in Structure S1 of the Supporting Information. The MM layer contains all the atoms of the system. For completing the valence of the truncated bonds between the two regions we used hydrogen atoms as link atoms.

The QM high layer was optimized with Gaussian 09 standard convergence criteria at the density theory level (DFT level). We used the B3LYP functional and the 6-31g(d) basis set as implemented in Gaussian 09. The MM low layer was treated with the AMBER force field[69] as implemented in Gaussian 09. Parameters for the zinc coordination sphere were added from the (ZAFF) library.[64] The electrostatic interaction between the two layers was treated with the electrostatic embedding method, which includes the low layer point charges in the high layer Hamiltonian.

Several linear scans were performed using the proper coordinates for each reaction in order to obtain transition states, intermediates, and product geometries. We used step values between 0.05 and 0.01 Å to locate all the stationary points. To obtain the first TS we started from the reactants structure and shortened the distance between the hydrogen atom of the 5′ terminal hydroxyl group of vDNA and the oxygen of the hydroxide ion with −0.05 Å increments in order to deprotonate the nucleophile. To obtain the second TS, which corresponds to the integration reaction, we started from the first intermediate structure and reduced the distance between the oxygen atom of the 5′ terminal hydroxyl group of vDNA and the phosphorus atom of the hDNA, with increments between −0.05 and −0.01 Å. Starting with the structure of the second intermediate we obtained the third TS by shortening the distance between the oxygen atom of the leaving group and a hydrogen atom of a $Mg^{2+}{}_{lg}$-bound water molecule with −0.05 Å increments in order to protonate the leaving group. Apart from the scanned coordinates, all atoms of the system were free to move in these calculations.

Single-point energy calculations were performed on the optimized geometries of reactants, intermediates and products and on the higher energy structures of the scans that correspond to the TS geometries (the coordinates of reactants, intermediates, T,S and Products are in the Supporting Information, Structures S1−S7). It was not possible to optimize the TS structures freely due to the large number of atoms in the system, which makes the calculation of the Hessian matrix unfeasible. The QM layer in the single point calculations was described at the MPWB1*K*/6-311++G(2d,2p) level. This choice was based on a benchmarking study of DFT functionals, which showed that this combination of functional and basis sets yields energies comparable to CCSD(T)/CBS calculations in the description of the hydrolysis of phosphodiester bonds.[58] Point charges were calculated using the Mulliken scheme. The system was divided in five parts (nucleophile, phosphate group, leaving group, hydroxide of nucleophile, and water of leaving group) and the atoms that belong to each one are shown in Figure 1 of the Supporting Information. To make sure that we were not incurring in error regarding dispersion, we also performed single-point energy calculations with the M06 and M062X functionals as well as calculating the Grimme D3 Dispersion correction term for all functionals.[70] The differences between these methods were not meaningful (Table S2), and therefore we chose the best candidate to describe the reaction mechanism, i.e. the MPWB1K functional.

3. RESULTS AND DISCUSSION

IN (Integrase) belongs to the large superfamily of polynucleotidyl transferases, which comprises important nucleic acid-processing enzymes such as nucleases and transposases.[36] The enzymes in this family catalyze two types of reaction: the 3′-end processing reaction, in which they catalyze the hydrolysis of a terminal phosphate group of a nucleic acid that results in the formation of a 3′OH group; and the strand-transfer reaction where the 3′OH group formed in the 3′-end processing reaction acts as a nucleophile and attacks the phosphate group of the target DNA molecule. These reactions are dependent on divalent metals ions, usually Mg^{2+}, which are essential to coordinate and activate the nucleophiles and to stabilize the transition states and the leaving groups.[42,43,71]

Here, we focused on the strand transfer reaction, which begins with the deprotonation and the activation of the nucleophile (the 3′OH group) and finishes with the protonation of the leaving group (a negatively charged deoxyribose group from the hDNA). The phosphoryl transfer reaction step was found to be of S_N2 type, in agreement with other enzymes that catalyze similar reactions.[44] In the next sections, we discuss the three steps of this mechanism. We will use the notation (*a*;*b*;*c*), where *a*, *b*, and *c* refer to interatomic distances (in Å) or charges (Mulliken point charges) at the reactants, TSs, and products involved in each step. The activation and reaction energies were calculated at the MPWB1K-D3/6-311++G(2d,2p):Amber//B3LYP/6-31G-(d):Amber level using the model represented in Figure 1A.

3.1. The Reactants. In the reactants conformation, the magnesium ion bonded to the nucleophile ($Mg^{2+}{}_{nuc}$) is coordinated to Glu221, Asp128, a water molecule, the 3′OH group of the vDNA, and the phosphate group of the hDNA. The leaving group magnesium ion ($Mg^{2+}{}_{lg}$) is coordinated to

dx.doi.org/10.1021/ct500570g | J. Chem. Theory Comput. 2014, 10, 5458−5466

Asp128, Asp185, two water molecules, and the phosphate group of hDNA by two oxygen atoms (Figure 1B).

We tested three different reactants structures concerning three possible bases that can activate the 3′OH group, by accepting its proton: the 3′-end phosphate group of the vDNA (Figure 2A), the carboxylate group of Glu221 (Figure 2B), or a metal-bound hydroxide (Figure 2C). In the first hypothesis, the phosphate should accept the proton of the 3′OH group through a bridging water. In exploratory PES scans we found that this mechanism is unfavorable. The potential activation energy increased excessively and continuously up to 40 kcal/mol, and the proton transfer did not take place (Figure S2). In the second hypothesis, an oxygen atom of the Glu221 carboxylate should accept the proton of the 3′OH group, also through a bridging water molecule. We tested this hypothesis, but we found that the conformation with a protonated Glu221 does not correspond to a stationary point on the potential energy surface, probably because both oxygen atoms of the Glu221 carboxylate are coordinating the Mg^{2+}_{nuc}.

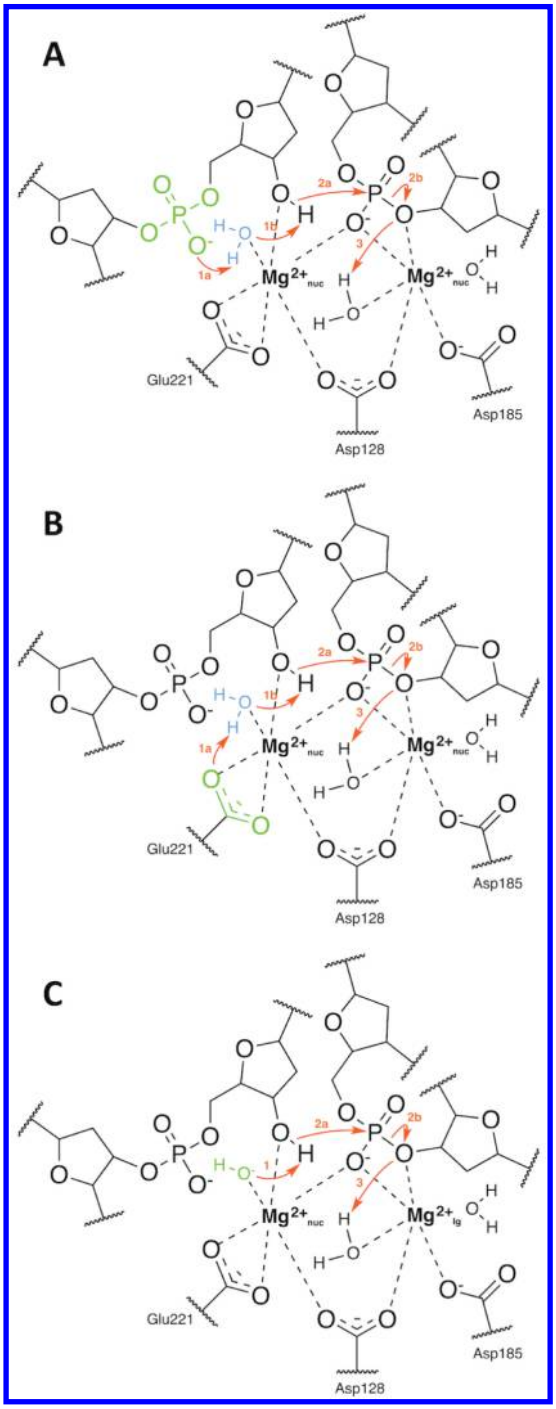


Figure 2. Schematic representation of the three possible mechanistic approaches for the deprotonation of the nucleophile: (A) the 3′-end phosphate group of the vDNA; (B) the carboxylate of Glu221; and (C) a metal-bound hydroxide. The bases that should deprotonate the nucleophile in each mechanism are represented in green. The bridging water molecules are represented in blue.

In the last hypothesis, which proved to be the correct one, the base that deprotonates and activates the nucleophile is a metal-bound hydroxide (Figure 1B). Similar findings were observed in a previous study on the HIV-1 IN 3′-end processing reaction, in which a hydroxide coordinated to the Mg^{2+}_{nuc} deprotonates efficiently the nucleophilic water molecule but a vDNA phosphate and an active site carboxylate do not.[23] Since the strand transfer reaction is almost symmetrical to the 3′-end processing reaction, and occurs within an identical chemical environment, it is likely that both reactions share a similar reaction path.

In the optimized geometry of the reactants, the hydroxide oxygen atom is at 1.66 Å from the 3′OH proton, and both are adequately orientated for the proton transfer. The phosphate oxygen atom of the hDNA is positioned between the two Mg^{2+} ions, and the oxygen of the 3′OH group is 3.65 Å away from the phosphorus atom that will be attacked in the integration step. The two Mg^{2+} ions have an octahedral coordination sphere. The Mg^{2+}_{nuc} is 2.86 Å from the 3′OH, 1.97 Å from the hydroxide, 2.14 Å from the scissile phosphate group, and 3.60 Å from the Mg^{2+}_{lg}, which are the most relevant interactions for the strand transfer reaction.

3.2. Transition State 1 - Deprotonation and Activation of Nucleophile. The first step of the strand transfer reaction consists in the deprotonation and activation of the nucleophile, the 3′OH group of the vDNA. In this step the metal-bound hydroxide receives the proton of the nucleophile. The negatively charged 3′-oxoanion is now an ideal nucleophile. At the transition state the proton is halfway between the reactants and the first intermediate structure (INT1); it is 1.20 Å away from both the 3′-oxoanion and the hydroxide (Figure 3A). In the INT1 structure, the active nucleophile is 1.65 Å away from the recently protonated water molecule (distance between the 3′-oxoanion and the proton) (Figure 3B). This reaction has a potential activation energy of 4.4 kcal/mol and a potential reaction energy of −0.5 kcal/mol (Figure 4). The distance between the oxygen of 3′OH and the phosphorus of the scissile phosphate group decreased slightly (3.65, 3.32, 3.28).

During this step, the Mg^{2+}_{nuc} coordination sphere suffered some adjustments that reflect the new charge distribution. The distance between the oxygen of 3′OH and Mg^{2+}_{nuc} decreased (2.86, 2.19, 2.04) due to the anionic character of the 3′oxoanion (0.681, 0.300, −0.438) and the distance between the oxygen from the hydroxide group and the Mg^{2+}_{nuc} increased (1.97, 2.12, 2.37), since the hydroxide group became a neutral water molecule, presenting a more positive charge (−0.532, −0.127, 0.098). The distance between the two metal ions increased a little (3.60, 3.85, 3.93).

3.3. Transition State 2 – Transesterification Reaction. The second step is the main strand transfer reaction, in which

dx.doi.org/10.1021/ct500570g | J. Chem. Theory Comput. 2014, 10, 5458−5466

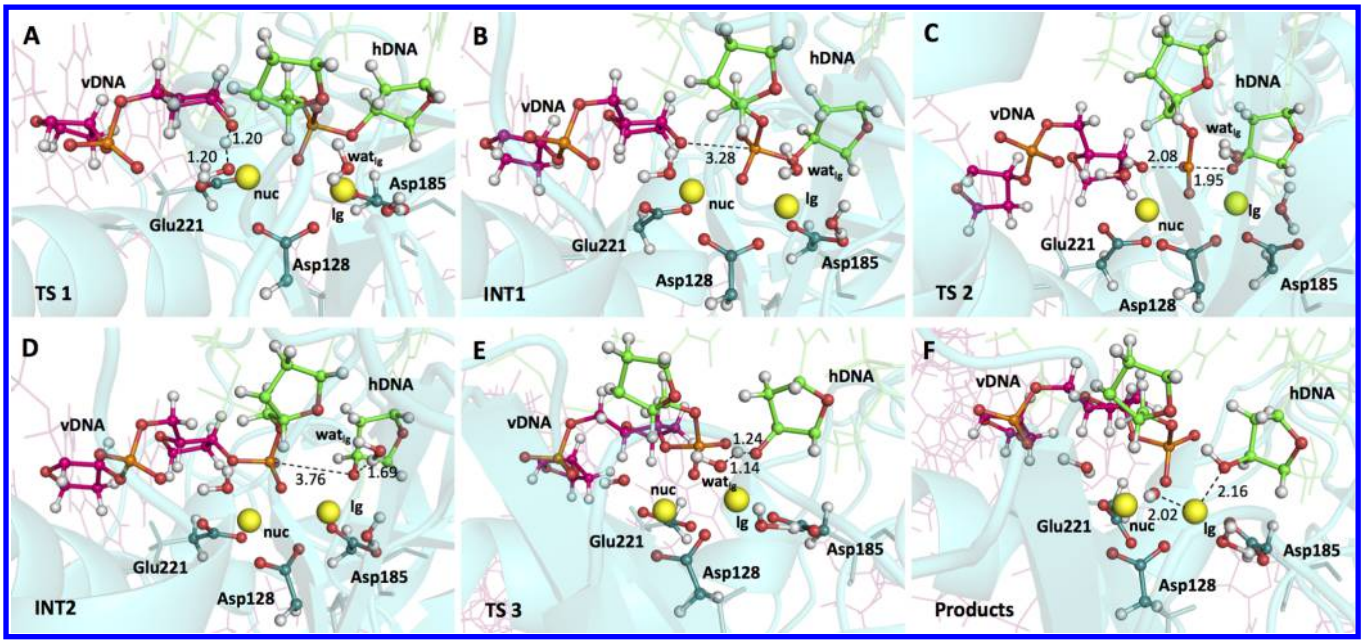


Figure 3. Stationary points for the strand transfer mechanism (reactants are shown in Figure 1B). Relevant interatomic distances (in Angstrom) are included. The representation includes all atoms in the high level QM layer, and the rest of the enzyme and viral and host DNA are represented in transparency (MM layer). The QM layer is colored by element and depicted in ball and stick format. The Mg^{2+} ions are represented in yellow spheres, and the pink and green carbons in sticks belong to vDNA and hDNA, respectively.

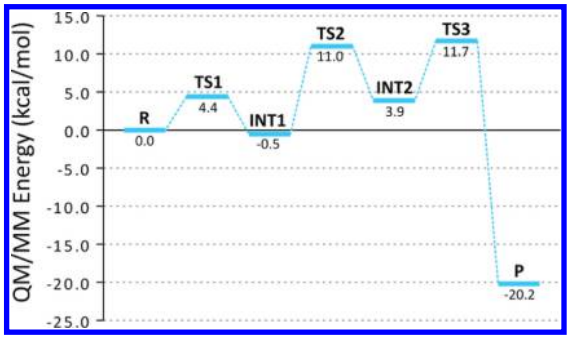


Figure 4. Most favorable pathway for the strand transfer reaction. The QM/MM energies (reactants are taken as zero) were obtained by single-point calculations at the MPWB1K-D3D3/6-311++G-(2d,2p):Amber//B3LYP/6-31G(d) level.

IN integrates the vDNA into the hDNA. The nucleophilic 3′-oxoanion of vDNA attacks the phosphorus atom of the phosphate scissile group, while the opposite O5′-P bond is broken. A pentacoordinated phosphate intermediate, which was initially proposed to exist in this kind of reactions, is never formed.[44] Instead, such structure corresponded to the transition state for the transesterification reaction (TS2). In the TS2, the phosphorus atom is halfway from the 3′-oxoanion of the vDNA and the oxygen of the leaving group of hDNA (2.08 and 1.95 Å, respectively) (Figure 3C).

The activation energy of this step is 11.0 kcal/mol and the reaction energy is 4.4 kcal/mol (Figure 4). At the end of this reaction step, the hybrid DNA strand of vDNA/hDNA is already formed (Figure 3D). The phosphorus atom is now 3.76 Å away from the leaving group, while the negatively charged leaving group (0.275, 0.176, −0,825) is closer to the $Mg^{2+}{}_{lg}$ (2.28, 2.13, 1.95). As before, some rearrangements occurred in the coordination sphere of both Mg^{2+} ions. The oxygen atom of the integrated phosphate group increased its distance to $Mg^{2+}{}_{lg}$ and became closer to $Mg^{2+}{}_{nuc}$ (2.17, 2.11, 2.31 from $Mg^{2+}{}_{lg}$ and 2.19, 2.02, 2.06 from $Mg^{2+}{}_{nuc}$).

Since this step of the reaction is the one that leads to the viral DNA integration, we used it to investigate further the catalytic role of the two Mg^{2+} ions and surrounding amino acid residues. The role of surrounding amino acids is fairly limited. Apart from the aspartate and glutamate residues coordinating the metal ions no other amino acid is a good candidate for catalytic roles. In the optimized structure of INT1, for example, there are no amino acid residues in a radius of 5 Å of the scissile phosphate. We can conclude from this simple observation that the amino acid residues in the surroundings of the active site have only a structural function, by guaranteeing that the structure of the active site remains stable. We studied the influence of the Mg^{2+} ions in this reaction by calculating the energy of the second transition state with and without the magnesium ions, through single point energy calculations. Without the magnesium ions, the activation energy of this step increases 3.2 kcal/mol. This result means that the Mg^{2+} ions have a contribution to catalysis that corresponds roughly to 2 orders of magnitude of the k_{cat}. However, since experimental data shows that the lack of divalent metal ions have a greater detrimental effect on catalysis, one must conclude that their structural role is even more important. Without the ions, the active site will not be able to adopt the reactive conformation.

The distance between the two Mg^{2+} ions decreased during this reaction step (3.93, 3.79, 3.49). Asp128 suffered a rotation, and now the same oxygen atom coordinates both $Mg^{2+}{}_{nuc}$ and $Mg^{2+}{}_{lg}$. Apart from this, a water molecule left the coordination sphere of $Mg^{2+}{}_{lg}$ and got to be 3.95 Å from it. This water molecule (wat_{lg}) has one of its protons in line with the 3′-oxoanion of the leaving group at 1.69 Å. This structure is thus prepared for the last step of the integration reaction that corresponds to the protonation of the leaving group. This step is necessary for the removal of the leaving group from the coordination sphere of the $Mg^{2+}{}_{lg}$ ion.[44]

dx.doi.org/10.1021/ct500570g | J. Chem. Theory Comput. 2014, 10, 5458−5466

3.4. Transition State 3 - Protonation of the Leaving Group. The third and last step of the strand transfer reaction consists in the protonation of the leaving group by a water molecule, in order to facilitate the departure from the active site of IN.

In this last step, the wat_{lg} gives a proton to the 3′-oxoanion of the leaving group. In the TS structure TS3, the proton is halfway between the oxygen atom of the wat_{lg} (1.14 Å) and the 3′-oxoanion (1.24 Å) (Figure 3E). This step has a barrier of 11.7 kcal/mol and a reaction potential energy of −24.1 kcal/mol (Figure 4). The high potential energy difference between the INT2 and the products includes the potential energy from the chemical reaction but also from a conformational adjustment.

At the end of this step, the leaving group is further away from the $Mg^{2+}{}_{lg}$ ion (1.95, 2.02, 2.16) and is now ready to leave the active site of IN (Figure 3F). On the other hand, the hydroxide that resulted from this last step is closer to $Mg^{2+}{}_{lg}$ (3.95, 2.17, 2.02) and is coordinating both Mg^{2+} ions. As the $Mg^{2+}{}_{nuc}$ ion had already six coordinating atoms, Glu221 suffered a carboxylate shift: it is now bonded to $Mg^{2+}{}_{nuc}$ through a single oxygen atom.

3.5. The Kinetics of the Chemical Reaction. The integration mechanism of vDNA in the host DNA is a crucial step in the HIV-1 life cycle. The turnovers of both the 3′ processing and the integration reactions, needed to complete a successful integration, have been determined and are both very slow: 0.002 min^{-1} for the 3′-end processing reaction and 0.0008 min^{-1} for the strand transfer reaction.[71,72] These turnover numbers are much lower than the usual enzyme kinetics, because they are not related with the chemical step but instead with the exit of the DNA products. Therefore, these values define only upper limits for the activation energies. They correspond to a barrier of 22.1 kcal/mol for the 3′-end processing reaction and a barrier of 25.1 kcal/mol for the strand transfer reaction, according to transition state theory.[73]

It was recently shown by theoretical methods that the barrier of the 3′-end processing reaction in HIV-1 integrase is 15.4 kcal/mol,[23] which is consistent with the expected enzyme kinetics. As the integration reaction itself is very similar to the 3′-end processing reaction we expected similar activation energies.

Before comparing the two reactions, however, we must make some corrections to the strand transfer barrier, due to the presence of a magnesium-bound hydroxide ion in the structure of the reactants. The concentration of hydroxide is much lower than the concentration of water in the cytoplasm, and so there is an entropic cost associated with the presence of a hydroxide group in the active center. On the other hand, the hydroxide ion is a much more favorable metal ligand than water.

The entropic cost can be translated into a free energy increase of 12.1 kcal/mol (Section S1). The free energy gain associated with exchanging a water molecule by a bulk solvent hydroxide in the active site (with the water molecule leaving to bulk solution) amounts to −9.0 kcal/mol, as measured by thermodynamic integration calculations for HIV-1 integrase, which is a similar system.[23] The sum of these two contributions is +3.1 kcal/mol. We need to add this value to the TS3 potential energy (11.7 kcal/mol) to obtain the value of the potential activation energy of the transesterification reaction at physiological conditions. The overall potential activation energy for the strand transfer reaction is then 14.8 kcal/mol, which is consistent with the value obtained for the 3′-end processing reaction[23] and is lower than the experimental threshold of 25.1 kcal/mol measured for strand transfer reaction.[24] The calculated potential reaction energy of −20.2 kcal/mol (Figure 4) shows that this reaction is exothermic and irreversible, as expected for phosphoester reactions and compatible with the irreversibility of the integration of the vDNA into the hDNA.

These energies are still missing a few components: the zero point energy, the entropy of reaction, and the thermal energy (all calculations were done at 0 K). These quantities can be calculated approximately from a frequency calculation on the QM/MM model and from basic statistical mechanics, but this calculation is too expensive, in computational terms for the size of our model. Nevertheless, we found these contributions to be very small in a model system for phosphoester reactions as well as in many other enzymatic reactions in the literature, where they typically amount to 1−3 kcal/mol. Their size should be similar for the current system.[58] Up until now, our results are the best quantitative answer for the kinetics and reaction mechanism of the chemistry of the strand transfer reaction.

3.6. Mechanism of Enzymes Involved in Making and Breaking Phosphodiester Bonds. Enzymes that catalyze phosphoester bonds formation or hydrolysis have similar active site geometries. Most of these enzymes work through a two-Mg^{2+}-ion catalysis like RNase H, DNA polymerase I of *E. coli*, Tn5.[44,74] The two Mg^{2+} are located on each side of the scissile phosphate group where one of them helps to stabilize and activate the nucleophile and the other one stabilizes the transition state and the leaving group.[75] We observed the same for the IN strand transfer reaction studied in this paper: the $Mg^{2+}{}_{nuc}$ helps to stabilize the nucleophile and the activation of the 3′OH which will then attack the phosphorus of the scissile phosphate group; and the $Mg^{2+}{}_{lg}$ helps to stabilize the second transition state as well as the leaving group after the its protonation. The formation of a pentacoordinated transition state and the lack of a pentacoordinated intermediate is also a feature of these kinds of phosphoester reactions.

In a recent paper about the mechanism of HIV-1 reverse transcriptase described with QM/MM simulations it was suggested that the base that deprotonates and activate the nucleophile is a Mg^{2+} bound aspartate. On the other hand, in HIV-1 Integrase, we found that is a hydroxide molecule that act as a base.[76] A mechanism similar to the one proposed for reverse transcriptase was also tested in our model, when we hypothesized that an oxygen atom from Glu221 should accept the proton from the 3′OH. However, we were not able to optimize this reactant structure because the two atoms from Glu221 carboxylate are part of the Mg^{2+}nuc coordination sphere and cannot hold and additional bond to a proton. We think that, although the general mechanism of bimetallic phosphoester enzymes is always the same, enzymes can use different bases for the activation of the nucleophile. Since the activation of the nucleophile is not the limiting step, the overall velocity of reactions should not be affected by these differences. Indeed, the total barrier described for the reaction catalyzed by HIV-1 reverse transcriptase is 18.4 kcal, which is similar to the barrier we propose for the mechanism of the HIV-1 IN strand transfer reaction.[76]

4. CONCLUSION

In this work, we describe the strand transfer mechanism of PFV IN at an atomic-level. To the best of our knowledge this is the first time that an integration reaction catalyzed by an integrase is clarified with atomic detail. The reaction proceeds through

dx.doi.org/10.1021/ct500570g | J. Chem. Theory Comput. 2014, 10, 5458−5466

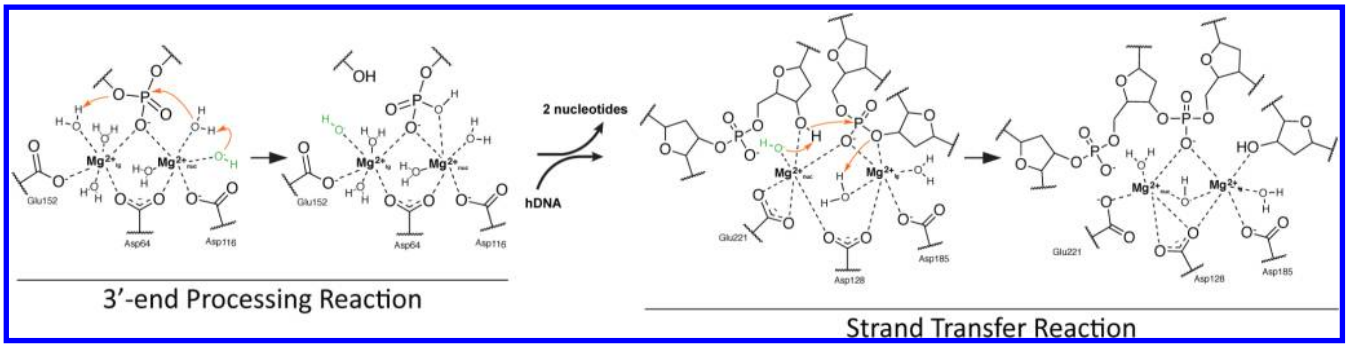


Figure 5. The two catalytic reactions needed for the integration of the vDNA into the hDNA, as catalyzed by IN - 3′end processing reaction and strand transfer reaction. Here are represented the QM layer used in each of the studies performed in order to dissect the catalytic mechanism of IN.[23]

three mechanistic steps. The first one corresponds to the activation of the nucleophile, a 3′-oxoanion, which is deprotonated by a hydroxide ion. The second step is the nucleophilic attack to the scissile phosphate of the hDNA by the 3′-oxoanion of vDNA. The formation of the hybrid DNA strand of vDNA and hDNA is concluded at this step. The third step of the mechanism consists of the protonation of the leaving group by a water molecule. In the end, this reaction originates a hybrid DNA strand that includes the integrated vDNA into the host DNA.

The calculated potential activation energy is 14.8 kcal/mol at the MPWB1K-D3/6-311++G(2d,2p):Amber level of theory. This value is consistent with both typical enzymatic kinetics and the upper limit established by the experimental turnover number (25.1 kcal/mol).[24] The total potential reaction energy is −20.2 kcal/mol.

It is now possible to understand the overall picture of how IN performs its catalytic activity, at an atomic level, since both steps needed to integrate a strand of vDNA into the hDNA are known. The 3′end processing reaction was described in HIV IN,[23] and here the strand transfer reaction is described in PFV IN. Figure 5 compiles the information about the two mechanisms. The overall process is an elegant showcase of how enzymes achieve complex tasks with simplicity. Within the same catalytic center, IN catalyzes two highly symmetrical reactions. The same solution is applied for the deprotonation of the nucleophile and protonation of the leaving group, in both reactions. The formation and breaking of bonds involving the phosphate group also occur through similar transition states.

As PFV IN is very similar to HIV IN, this QM/MM study is a valid contribution to drug design research against HIV-1 virus. This work provides structures of all the intermediates and transition states of the strand transfer reaction, which can prove very helpful in the discovery of new inhibitors. We and others can now use these structures to design more effective and targeted drugs that will block the action of HIV IN. This work fulfills important parts of the knowledge gap that is delaying the discovery of new drugs for HIV-1 IN. Furthermore, PFV is a promising vector to be used in gene therapy. This work might also be of help in designing better, more effective, and secure vectors.

■ ASSOCIATED CONTENT

Supporting Information

Table with the pK_a results for all the ionizable residues as obtained by the H++ server; xyz files of all stationary states of the structure used in QM/MM calculations; zinc parameters files; table with the activation and reaction energies for all the stationary points described in four different functionals and the 6-311++g(2d,2p) base with and without Grimme D3 dispersion correction term; figure that shows the potential energy surface scan of the phosphate acting as a base hypothesis tested for strand transfer reaction; calculation of the cost of having a hydroxide molecule in the structure of the reactants instead of a water molecule. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mjramos@fc.up.pt.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has been financed by the program FEDER/COMPETE and by the Fundação para a Ciência e Tecnologia (EXCL/QEQ-COM/0394/2012).

■ ABBREVIATIONS

IN, integrase; vDNA, viral DNA; hDNA, human DNA

■ REFERENCES

(1) Long, Y. Q.; Huang, S. X.; Zawahir, Z.; Xu, Z. L.; Li, H.; Sanchez, T. W.; Zhi, Y.; De Houwer, S.; Christ, F.; Debyser, Z.; Neamati, N. *J. Med. Chem.* **2013**, *56*, 5601−5612.

(2) U.S. Food and Drug Administration - Consumer Information by Audience: Antiretroviral drugs used in the treatment of HIV infection. http://www.fda.gov/ForConsumers/byAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm118915.htm (accessed March 4, 2014).

(3) Sechi, M.; Rizzi, G.; Bacchi, A.; Carcelli, M.; Rogolino, D.; Pala, N.; Sanchez, T. W.; Taheri, L.; Dayam, R.; Neamati, N. *Bioorg. Med. Chem.* **2009**, *17*, 2925−2935.

(4) Murray, J. M.; Emery, S.; Kelleher, A. D.; Law, M.; Chen, J.; Hazuda, D. J.; Nguyen, B.-Y. T.; Teppler, H.; Cooper, D. A. *AIDS* **2007**, *21*, 2315−2321.

(5) Summa, V.; Petrocchi, A.; Bonelli, F.; Crescenzi, B.; Donghi, M.; Ferrara, M.; Fiore, F.; Gardelli, C.; Gonzalez Paz, O.; Hazuda, D. J.; Jones, P.; Kinzel, O.; Laufer, R.; Monteagudo, E.; Muraglia, E.; Nizi, E.; Orvieto, F.; Pace, P.; Pescatore, G.; Scarpelli, R.; Stillmock, K.; Witmer, M. V.; Rowley, M. *J. Med. Chem.* **2008**, *51*, 5843−5855.

(6) Marchand, C.; Maddali, K.; Metifiot, M.; Pommier, Y. *Curr. Top. Med. Chem.* **2009**, *9*, 1016−1037.

(7) Sax, P. E.; DeJesus, E.; Mills, A.; Zolopa, A.; Cohen, C.; Wohl, D.; Gallant, J. E.; Liu, H. C.; Zhong, L.; Yale, K.; White, K.; Kearney, B. P.; Szwarcberg, J.; Quirk, E.; Cheng, A. K. *Lancet* **2012**, *379*, 2439−2448.

(8) Cahn, P.; Pozniak, A. L.; Mingrone, H.; Shuldyakov, A.; Brites, C.; Andrade-Villanueva, J. F.; Richmond, G.; Buendia, C. B.; Fourie, J.;

Ramgopal, M.; Hagins, D.; Felizarta, F.; Madruga, J.; Reuter, T.; Newman, T.; Small, C. B.; Lombaard, J.; Grinsztejn, B.; Dorey, D.; Underwood, M.; Griffith, S.; Min, S. *Lancet* **2013**, *382*, 700−708.
(9) Johnson, A. A.; Marchand, C.; Pommier, Y. *Curr. Top. Med. Chem.* **2004**, *4*, 1059−1077.
(10) Pommier, Y.; Johnson, A. A.; Marchand, C. *Nat. Rev. Drug Discovery* **2005**, *4*, 236−248.
(11) Dayam, R.; Al-Mawsawi, L. Q.; Neamati, N. *Drugs R&D* **2007**, *8*, 155−168.
(12) Blanco, J. L.; Varghese, V.; Rhee, S. Y.; Gatell, J. M.; Shafer, R. W. *J. Infect. Dis.* **2011**, *203*, 1204−1214.
(13) Malet, I.; Calvez, V.; Marcelin, A. G. *Curr. Opin. Virol.* **2012**, *2*, 580−587.
(14) Chiu, T. K.; Davies, D. R. *Curr. Top. Med. Chem.* **2004**, *4*, 965−977.
(15) Asante-Appiah, E.; Skalka, A. M. *Antiviral Res.* **1997**, *36*, 139−156.
(16) Delelis, O.; Carayon, K.; Saib, A.; Deprez, E.; Mouscadet, J. F. *Retrovirology* **2008**, *5*, 114.
(17) Engelman, A.; Mizuuchi, K.; Craigie, R. *Cell* **1991**, *67*, 1211−1221.
(18) Poeschla, E. M. *Cell Mol. Life Sci.* **2008**, *65*, 1403−1424.
(19) Engelman, A.; Craigie, R. *J. Virol.* **1992**, *66*, 6361−6369.
(20) Polard, P.; Chandler, M. *Mol. Microbiol.* **1995**, *15*, 13−23.
(21) Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W. *Cell* **2005**, *121*, 1005−1016.
(22) Davies, D.; Goryshin, I.; Reznikoff, W.; Rayment, I. *Science* **2000**, *289*, 77−85.
(23) Ribeiro, A.; Ramos, M.; Fernandes, P. *J. Am. Chem. Soc.* **2012**, *134*, 13436−13447.
(24) Dicker, I.; Terry, B.; Lin, Z.; Li, Z.; Bollini, S.; Samanta, H.; Gali, V.; Walker, M.; Krystal, M. *J. Biol. Chem.* **2008**, *283*, 23599−23609.
(25) Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9150−9154.
(26) Maignan, S.; Guilloteau, J. P.; Zhou-Liu, Q.; Clement-Mella, C.; Mikol, V. *J. Mol. Biol.* **1998**, *282*, 359−368.
(27) Cai, M.; Zheng, R.; Caffrey, M.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. *Nat. Struct. Biol.* **1997**, *4*, 567−577.
(28) Chen, J. C.; Krucinski, J.; Miercke, L. J.; Finer-Moore, J. S.; Tang, A. H.; Leavitt, A. D.; Stroud, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8233−8238.
(29) Lodi, P. J.; Ernst, J. A.; Kuszewski, J.; Hickman, A. B.; Engelman, A.; Craigie, R.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1995**, *34*, 9826−9833.
(30) Wang, J. Y.; Ling, H.; Yang, W.; Craigie, R. *EMBO J.* **2001**, *20*, 7333−7343.
(31) Eijkelenboom, A. P.; Sprangers, R.; Hard, K.; Puras Lutzke, R. A.; Plasterk, R. H.; Boelens, R.; Kaptein, R. *Proteins* **1999**, *36*, 556.
(32) Valkov, E.; Gupta, S.; Hare, S.; Helander, A.; Roversi, P.; McClure, M.; Cherepanov, P. *Nucleic Acids Res.* **2009**, *37*, 243−255.
(33) Hossain, M.; Ali, M.; Shin, C.-G. *Viruses* **2013**, *5*, 1850−1866.
(34) Hare, S.; Gupta, S.; Valkov, E.; Engelman, A.; Cherepanov, P. *Nature* **2010**, *464*, 232−236.
(35) Hare, S.; Smith, S.; Métifiot, M.; Jaxa-Chamiec, A.; Pommier, Y.; Hughes, S.; Cherepanov, P. *Mol. Pharmacol.* **2011**, *80*, 565−572.
(36) Li, X.; Krishnan, L.; Cherepanov, P.; Engelman, A. *Virology* **2011**, *411*, 194−205.
(37) Kulkosky, J.; Jones, K. S.; Katz, R. A.; Mack, J. P.; Skalka, A. M. *Mol. Cell. Biol.* **1992**, *12*, 2331−2338.
(38) Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. *Science* **1994**, *266*, 1981−1986.
(39) Zheng, R.; Jenkins, T. M.; Craigie, R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13659−13664.
(40) Kennedy, A. K.; Haniford, D. B.; Mizuuchi, K. *Cell* **2000**, *101*, 295−305.
(41) Mizuuchi, K.; Adzuma, K. *Cell* **1991**, *66*, 129−140.
(42) Harding, M. M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 678−682.
(43) Nowotny, M.; Yang, W. *EMBO J.* **2006**, *25*, 1924−1933.
(44) Yang, W.; Lee, J.; Nowotny, M. *Mol. Cell* **2006**, *22*, 5−13.
(45) Hare, S.; Vos, A.; Clayton, R.; Thuring, J.; Cummings, M.; Cherepanov, P. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20057−20062.
(46) Hare, S.; Di Nunzio, F.; Labeja, A.; Wang, J.; Engelman, A.; Cherepanov, P. *PLoS Pathog.* **2009**, *5*, e1000515.
(47) Cherepanov, P.; Maertens, G.; Hare, S. *Curr. Opin. Struct. Biol.* **2011**, *21*, 249−256.
(48) Krishnan, L.; Engelman, A. *J. Biol. Chem.* **2012**, *287*, 40858−40866.
(49) Hare, S.; Maertens, G.; Cherepanov, P. *EMBO J.* **2012**, *31*, 3020−3028.
(50) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(51) Alberto, M. E.; Marino, T.; Ramos, M. J.; Russo, N. *J. Chem. Theory Comput.* **2010**, *6*, 2424−2433.
(52) Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. *Chem. - Eur. J.* **2007**, *13*, 8507−8515.
(53) Himo, F. *Theor. Chem. Acc.* **2006**, *116*, 232−240.
(54) Leopoldini, M.; Marino, T.; Michelini, M.; Rivalta, I.; Russo, N.; Sicilia, E.; Toscano, M. *Theor. Chem. Acc.* **2007**, *117*, 765−779.
(55) Ramos, M. J.; Fernandes, P. A. *Acc. Chem. Res.* **2008**, *41*, 689−698.
(56) Rothlisberger, U.; Carloni, P.; Doclo, K.; Parrinello, M. *J. Biol. Inorg. Chem.* **2000**, *5*, 236−250.
(57) Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 11139−11149.
(58) Ribeiro, A. J. M.; Ramos, M. J.; Fernandes, P. A. *J. Chem. Theory Comput.* **2010**, *6*, 2281−2292.
(59) Krissinel, E.; Henrick, K. *J. Mol. Biol.* **2007**, *372*, 774−797.
(60) Pahl, A.; Flügel, R. M. *J. Biol. Chem.* **1995**, *270*, 2957−2966.
(61) Gordon, J.; Myers, J.; Folta, T.; Shoja, V.; Heath, L.; Onufriev, A. *Nucleic Acids Res.* **2005**, *33*, 71.
(62) 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.; Seabra, 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.
(63) Dennington, R. K. T.; Millam, J. *GaussView, Version 5*; Semichem Inc.: Shawnee Mission, KS, 2009.
(64) Peters, M.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M.; Merz, K. *J. Chem. Theory Comput.* **2010**, *6*, 2935−2947.
(65) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.02*; Gaussian Inc.: Wallingford, CT, 2009.
(66) Dapprich, S.; Komáromi, I.; Byun, K.; Morokuma, K.; Frisch, M. *J. Mol. Struct.: THEOCHEM* **1999**, *461−462*, 1−21.
(67) Feliu, M.; Keiji, M. *J. Comput. Chem.* **1995**, *16*, 1170−1179.
(68) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357−−19363.
(69) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(70) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(71) Lee, S.; Kim, H.; Censullo, M.; Han, M. *Biochemistry* **1995**, *34*, 10205−10214.
(72) Tramontano, E.; Colla, P.; Cheng, Y. *Biochemistry* **1998**, *37*, 7237−7243.
(73) Laidler, K. J. *Theories of Chemical Reaction Rates*; McGraw-Hill: New York, 1969.
(74) Steitz, T. A.; Steitz, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6498−502.
(75) Nowotny, M. *EMBO Rep.* **2009**, *10*, 144−51.
(76) Rungrotmongkol, T.; Mulholland, A. J.; Hannongbua, S. *MedChemComm* **2014**, *5*, 593−596.