31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Study on the Binding Mode of the Integrase with DNA via Steered Molecular Dynamics Simulation

Cun Xin Wang, Ming Liu, Ping Li, Xiao Jing Cong, Jian Jun Tan, Wei Zu Chen,

Abstract—**Human immunodeficiency virus type 1 (HIV-1) integrase (IN) is an essential enzyme in the lifecycle of this virus and also an important target for the study of anti-HIV drugs. In the current work, a model for the active site of IN and viral DNA was built by combining experimental data with the results of steered molecular dynamics simulation. The model was then taken into a series automatic molecular docking calculations with two groups of inhibitors. According to the results of molecular docking, the inhibitors of the second group share a similar binding model with those of the first group, though they have no common scaffold. The newly built model of the IN-DNA complex is helpful for our subsequent research on the design of IN inhibitors.**

I. INTRODUCTION

THE pol gene of HIV-1 encodes three enzymes that are essential for the virus: protease (PR), reverse transcriptase (RT) and integrase (IN). Inhibitors targeting at RT and PR have become successful drugs in the fight against AIDS, however, efforts toward development of IN inhibitors have been hampered by the absence of crystal structures of the full-length enzyme and by uncertainties in the understanding of the biochemical mechanism of integration. HIV-1 IN is composed of 288 residues (32 kD). This enzyme can be divided into three distinct functional domains: the N-terminal domain (residue 1-49), the catalytic core domain (residue 50-212), and the C-terminal domain (residue 213-288). The catalytic core domain contains three highly conserved residues, i.e. D64, D116 and E152, which coordinate divalent cations such as Mg^{2+} or Mn^{2+}[1-3]. HIV-1 IN mediates the insertion of viral DNA into the host genome. This process occurs through two separate steps, i.e. the 3'-processing and the strand transfer (ST) reactions.

Although the crystal structure of the full length IN is not available yet, the three domains of IN have been determined separately by means of X-ray crystallography and NMR spectroscopy[4-8]. To facilitate the development of IN inhibitors, many works are concentrated on the modeling of IN-DNA complex in the last decade. Luca et al. modeled a full length IN dimer, and then obtained the IN-DNA complex through a modified DNA docking approach[9]. Wang et al. modeled a full length IN tetramer and docked a 27bp DNA to the tetramer through rigid body docking[10]. Additionally, Karki and Tang et al. manually assembled three IN tetramer-DNA complex as models for virtual drug screening[11]. Although these models provide useful information for the research of the mechanism of IN and rational drug design, the modifications for the modeling method are still highly required. In the recent two years, it was found that the crystal structure of bacterial transposase Tn5 can be used as a useful surrogate model for IN in an attempt to address the potential binding modes of IN strand transfer inhibitors[12]. Based on this background, Chen et al. constructed and validated an IN core-DNA complex by superimposing the active site of IN and Tn5[13]. According to this model, it was found that the DNA plays a role in the binding of IN inhibitors.

In the current study, we built an IN-DNA complex through a similar method used by Chen[13] and then the molecular dynamics (MD) and steered molecular dynamics (SMD) simulations were performed to modify the model and investigate the interactions between IN and DNA. Finally, two groups of IN inhibitors were automatically docked to the modified model for the sake of testing this model. The inhibitory mechanism of some IN inhibitors was also investigated.

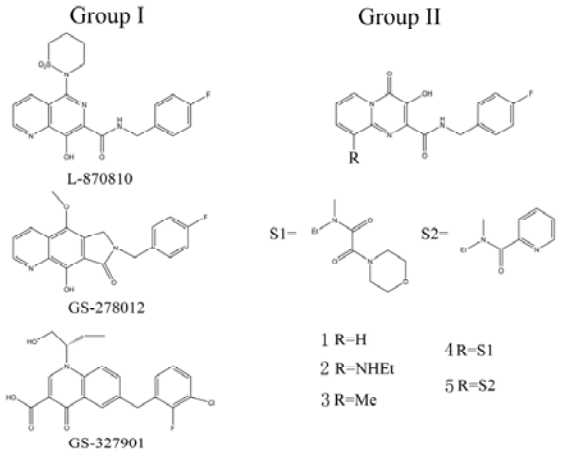

Figure 1. The two groups of IN inhibitors used in the molecular docking tests.

Manuscript received April 3, 2009. This work was supported in part by National Natural Science Foundation of China (No.30670497), the Beijing Natural Science Foundation (No.5072002 and No.7082006), and the National Basic Research Program of China (No.2009CB930203).

Cun Xin Wang is with the College of Life Science and Bioengineering, Beijing University of Technology, 100124 CHINA (corresponding author to provide phone: +86 10 67392724; fax: +86 10 67392837; e-mail: cxwang@bjut.edu.cn).

Ming Liu, Ping Li, Xiao Jing Cong, Jian Jun Tan, and Wei Zu Chen are with the College of Life Science and Bioengineering, Beijing University of Technology, 100124 CHINA (e-mail: ice_flame@yahoo.cn).

II. MATERIALS AND METHODS

A. Building the DNA-IN Complex

The crystal structures of the IN core and the Tn5 was obtained form the Protein Data Bank[14] with the PDB entry of 1EX4[8] and IMUH[15], respectively. Firstly, the core domain of Tn5 was superimposed onto that of the IN on the basis of the approach addressed by Chen[13]. After the superimposition, the DNA was right located around the active site of the IN and the three manganese coordinating residues (D97, D188 and E326) of Tn5 align directly with the corresponding magnesium residues of IN (D64, D116 and E152). The coordinates of the DNA and Mn^{2+} ion in the Tn5-DNA complex were then extracted and copied to the PDB file of the IN core. The Mn^{2+} ion (suited between D64 and E152) was then substituted by a Mg^{2+} ion. The second Mg^{2+} (suited between D64 and D116) ion was modeled according to 1QS4[16].

In order to simulate the 3'processed DNA, we manually deleted the last two nucleotide in the 3' terminal nearby the active site of IN and then substituted the two new terminal bases with C and A. The missing functional loop on each monomer was built with Sybyl 7.3. The modeled loops were then minimized by 1000 steps with the steepest descent method followed by 1000 steps of the conjugate gradient energy minimization. Finally, the processed structure was checked by Ramachandran map. Most of the residues were located at the most allowed regions.

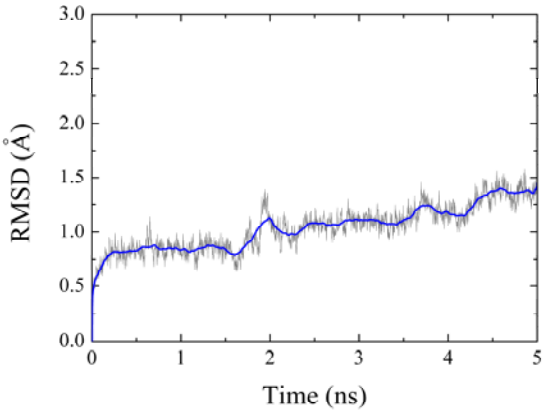

Figure 2. The C_α RMSD plot of the IN core domain and the IN core without loops.

B. Molecular Dynamics Simulation

After the preparation of the IN-DNA complex, we intended to modify it with MD simulation. The complex was first merged into a TIP3P[17] water box and then some counter ions were added to hold the system in electric neutrality. All MD simulations were performed with the NAMD 2.6[18] and the AMBER all-atom force field[19] was used. During the simulations, all bond lengths were constrained employing the SHAKE algorithm[20] and the integration time step was set to 2 fs. Firstly, the two systems were respectively minimized by 20,000 steps with solutes constrained, followed by 20,000 steps minimizations without any constraint. Then the minimized systems were slowly heated form 0 up to 300 K within 120 ps with the backbone (IN and DNA) and the Mg^{2+} ion constrained. Finally, the non-constrained simulations were performed by 5 ns at a constant temperature of 300 K and a constant pressure of 1 atm.

C. Steered Molecular Dynamics Simulation

Steered molecular dynamics (SMD) simulation induces movements and conformational changes of biomolecules on time scales accessible to molecular dynamics simulations. Previous study has proved that SMD can be employed as a useful approach in attempt to accommodate the flexibility of biomacromolecule in molecular docking[21]. To modify the IN-DNA complex model, the SMD simulation was carried out and the interaction energies between IN and DNA were monitored. Firstly, the DNA was manually placed 4 Å away from its original position along the connecting line of the centroids of IN and DNA. Secondly, the DNA was slowly pulled back toward its original position by a virtual spring. The pulling speed was set to 1Å/ns and the force constant of the virtual spring was 300 pN/Å. The whole simulation time was 4 ns.

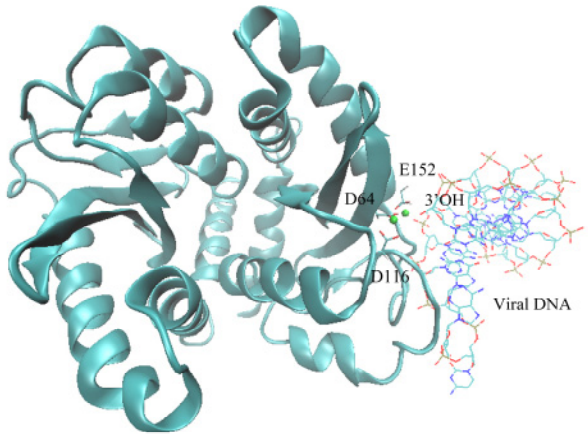

Figure 3. The Model of the IN-DNA complex.

D. Molecular Docking

To validate the SMD modified IN-DNA complex model, three well-tested IN inhibitors were selected for the following molecular docking calculations. Additionally, a group of newly reported IN inhibitors[22] were also prepared to explore their possible binding mode and inhibitory mechanism. All inhibitors were generated and minimized with Sybyl 7.3. The inhibitors used in molecular docking are depicted in Fig. 1. Group I contains three IN inhibitors with different scaffolds, whereas group II is constituted by five IN inhibitors sharing a common scaffold.

The molecular dockings calculations were performed with the AutoDock 4.0 package[23, 24]. All single bonds of the ligand were treated as rotatable in all docking tests. The center of the grid box was set to the centroid of the three catalytic residues (DDE). The box size was set to $60\times60\times60Å^3$ with grid spacing 0.375 Å in each dimension.

The box size is large enough for the free rotation of all inhibitors. All other docking parameters were set to default. Each docking calculation generated 100 complex structures. The binding mode for each inhibitor was selected according to the predicted binding energy, the clustering analysis as well as the known experimental information.

III. Results and Discussions

A. Analyses of the MD Trajectories of the IN-DNA Complex Simulation

The root mean square deviation (RMSD) for C_α atoms of the IN core was calculated as a measurement of the stability of this model. The RMSD versus simulation time was plotted in Fig. 2. Based on this plot, it is found that the IN core becomes stable within the first 0.5 ns and then fluctuates around 0.7 Å from 0.5 to 1.8 ns. From 2 to 4 ns, the RMSD fluctuates around 1.2 Å. After 4 ns, the RMSD raises up to 1.3 Å. The average structure calculated from 0.5 to 4 ns was taken as the starting structure for the following SMD simulation. The determined structure was shown in Fig. 3. According to this figure, the three Mg^{2+} coordinating residues chelate the two Mg^{2+} ions with their negatively charged side chains. The average distance between the carboxylate O atoms of the three residues and the corresponding Mg^{2+} ions are less then 2.5 Å. The adenosine phosphate of the conserved CA terminal coordinates a Mg^{2+} ion with its 3'OH and the average coordinating distance is less than 3 Å.

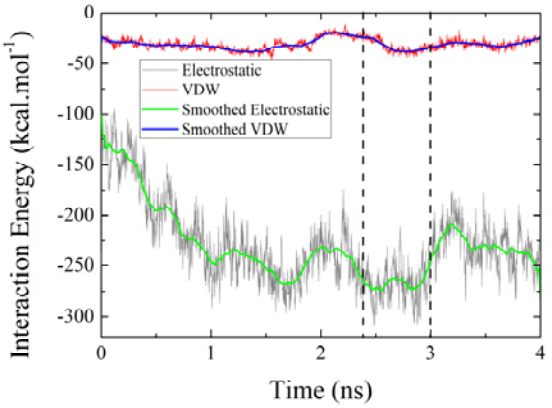

Figure 4. The Electrostatic and VDW interaction energies between residues in the IN-DNA interface.

B. The Modifications of the IN-DNA Complex Model with SMD

In order to evaluate the SMD result, the electrostatic interaction energy and the van der Waals (VDW) energy between the interface residues of the IN and the interface nucleotides of the viral DNA are calculated, and the profiles are depicted in Fig. 4. Obviously, the VDW energy does not change much during the whole simulation. On the contrary, the electrostatic interaction energy decreases dramatically during the first 1.5 ns. The electrostatic energy is much smaller than VDW energy, suggesting the interactions between IN and DNA mainly depend on electrostatic forces. Besides, both the electrostatic energy and the VDW energy reach minima around 2.4 to 3 ns, suggesting the IN-DNA interface formed within this time range is energetically favored. Therefore, the average structure from 2.4 to 3 ns was taken as the final model for the following molecular docking tests. The differences between the modified model and the previous model mainly reside in the non-interfacial nucleotides of the viral DNA. Additionally, the positively charged side chains of K156 and K159 directly point to the viral DNA, thus result in a slight rearrangement in the direction of the DNA. This rearrangement makes the modified complex model conform to the previous photo-crosslinking experiment[25].

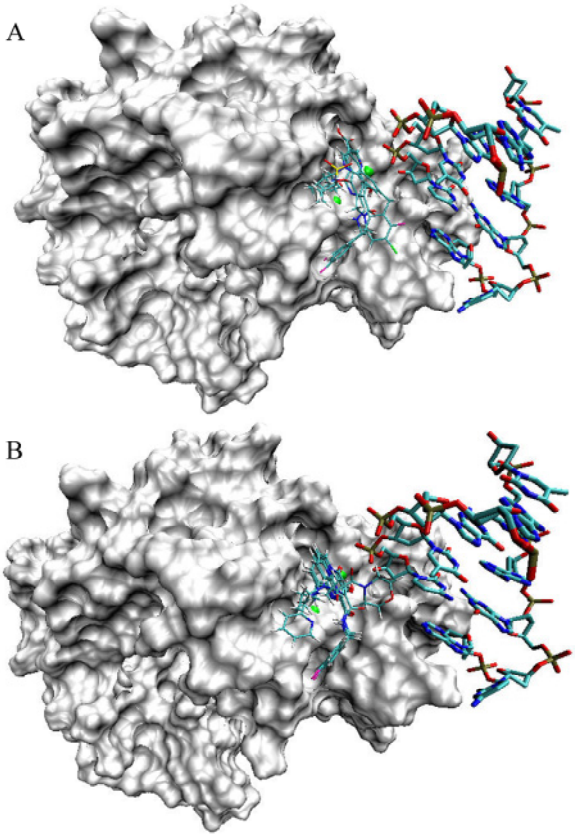

Figure 5. Binding modes for the IN inhibitors of group I (A) and II (B) with IN and viral DNA.

C. Analyses of the Binding Modes of the Inhibitors

The docking results of the two groups of IN inhibitors are illustrated in Fig 5A and B, respectively. According to Fig. 5A, the three different inhibitors share a similar binding mode when interacting with IN and DNA. Basically, the three inhibitors coordinate one or two Mg^{2+} ions with their diketo acid like moieties, and their substituted benzyl moieties insert deep into a pocket composed of E92, T93, G94, T97, T115, D116 and F121. This binding mode is roughly consistent with previous work[13], hence suggesting the validity of our model and indicating the modification by SMD is helpful. As for the five inhibitors of group II, they also coordinate the Mg^{2+} ions and make hydrophobic interactions with the pocket

addressed above, resembling the binding mode of the inhibitors of group I. In other words, the two group inhibitors share a common binding mode when interacting IN and the viral DNA. Therefore, the working mechanism of the inhibitors of group II may be similar with that of the inhibitor of group I.

IV. Conclusions

In the current study, an IN-DNA complex model was built by superimposing the Tn5-DNA complex onto an IN core domain. The model was then modified by MD and SMD simulations. During the SMD simulation, the electrostatic and VDW interaction energies between IN and DNA were calculated in order to determine the final model. The final model was basically in accordance with previous experimental information. Finally, two groups of IN inhibitors were automatically docked to the determined model. According to the docking results, it is found that the two groups of inhibitors share a common binding mode: coordinating the Mg^{2+} ions with their specific functional moieties and make hydrophobic interactions with a deep pocket, suggesting the five novel IN inhibitors may share a similar working mechanism with the inhibitors of group I. These findings are helpful for the modification and drug design of IN inhibitors.

Acknowledgment

Ming Liu thanks all colleagues' helpful discussions.

References

[1] A. Engelman and R. Craigie, "Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro," J. Virol., vol. 66, pp. 6361-6369, 1992.

[2] J. Kulkosky, K. S. Jones, R. A. Katz, and J. P. G. Mack, "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., vol. 12, pp. 2331-2338, 1992.

[3] P. Polard and M. Chandler, "Bacterial transposases and retroviral integrases," Mol. Microbiol., vol. 15, pp. 13-23, 1995.

[4] F. Dyda, A. B. Hickman, T. M. Jenkins, A. Engelman, R. Craigie, and D. R. Davies, "Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases," Science, vol. 266, pp. 1981-1986, 1994.

[5] M. Cai, R. Zheng, M. Caffrey, R. Craigie, G. M. Clore, and A. M. Gronenborn, "Solution structure of the N-terminal zinc binding domain of HIV-1 integrase," Nat Struct Mol Biol, vol. 4, pp. 567-577, 1997.

[6] S. Maignan, J.-P. Guilloteau, Q. Zhou-Liu, C. Clément-Mella, and V. Mikol, "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., vol. 282, pp. 359-368, 1998.

[7] Y. Goldgur, F. Dyda, A. B. Hickman, T. M. Jenkins, R. Craigie, and D. R. Davies, "Three new structures of the core domain of HIV-1 integrase: An active site that binds magnesium," Proc. Natl. Acad. Sci. U. S. A., vol. 95, pp. 9150-9154, 1998.

[8] J. C. H. Chen, J. Krucinski, L. J. W. Miercke, J. S. Finer-Moore, A. H. Tang, A. D. Leavitt, and R. M. Stroud, "Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding," Proc. Natl. Acad. Sci. U. S. A., vol. 97, pp. 8233-8238, 2000.

[9] L. De Luca, A. Pedretti, G. Vistoli, M. Letizia Barreca, L. Villa, P. Monforte, and A. Chimirri, "Model of full-length HIV-1 integrase complexed with viral DNA as template for anti-HIV drug design," Biochem Biophys Res Commun. , vol. 310, pp. 1083-1088, 2003.

[10] L. D. Wang, C. L. Liu, W. Z. Chen, and C. X. Wang, "Constructing HIV-1 integrase tetramer and exploring influences of metal ions on forming integrase-DNA complex," Biochem Biophys Res Commun. , vol. 337, pp. 313-319, 2005.

[11] R. G. Karki, Y. Tang, T. R. Burke, and M. C. Nicklaus, "Model of full-length HIV-1 integrase complexed with viral DNA as template for anti-HIV drug design," Journal of Computer-Aided Molecular Design, vol. 18, pp. 739-760, 2004.

[12] M. L. Barreca, F. Ortuso, N. Iraci, L. De Luca, S. Alcaro, and A. Chimirri, "Tn5 transposase as a useful platform to simulate HIV-1 integrase inhibitor binding mode," Biochem Biophys Res Commun., vol. 363, pp. 554-560, 2007.

[13] X. Chen, M. Tsiang, F. Yu, M. Hung, G. S. Jones, A. Zeynalzadegan, X. Qi, H. Jin, C. U. Kim, S. Swaminathan, and J. M. Chen, "Modeling, analysis, and validation of a novel HIV integrase structure provide insights into the binding modes of potent integrase inhibitors," J. Mol. Biol., vol. 380, pp. 504-519, 2008.

[14] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," Nucl. Acids Res., vol. 28, pp. 235-242, 2000.

[15] D. R. Davies, I. Y. Goryshin, W. S. Reznikoff, and I. Rayment, "Three-dimensional structure of the Tn5 synaptic complex transposition intermediate," Science, vol. 289, pp. 77-85, 2000.

[16] Y. Goldgur, R. Craigie, G. H. Cohen, T. Fujiwara, T. Yoshinaga, T. Fujishita, H. Sugimoto, T. Endo, H. Murai, and D. R. Davies, "Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design," Proc. Natl. Acad. Sci. U. S. A., vol. 96, pp. 13040-13043, 1999.

[17] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of Simple Potential Functions for Simulating Liquid Water," J. Chem. Phys., vol. 79, pp. 926-935, 1983.

[18] L. Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten, "NAMD2: Greater Scalability for Parallel Molecular Dynamics," J. Comput. Phys., vol. 151, pp. 283-312, 1999.

[19] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," J. Am. Chem. Soc., vol. 117, pp. 5179-5197, 1995.

[20] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes," J. Comput. Phys., vol. 23, pp. 327-341, 1977.

[21] A. Heifetz, S. Pal, and G. R. Smith, "Protein-protein docking: Progress in CAPRI rounds 6-12 using a combination of methods: The introduction of steered solvated molecular dynamics," Proteins: Structure, Function, and Bioinformatics, vol. 69, pp. 816-822, 2007.

[22] M. Donghi, O. D. Kinzel, and V. Summa, "3-Hydroxy-4-oxo-4H-pyrido[1,2-a]pyrimidine-2-carboxylates--A new class of HIV-1 integrase inhibitors," Bioorganic & Medicinal Chemistry Letters, vol. 19, pp. 1930-1934, 2009.

[23] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson, "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function," J. Comput. Chem., vol. 19, pp. 1639-1662, 1998.

[24] R. Huey, G. M. Morris, A. J. Olson, and D. S. Goodsell, "A semiempirical free energy force field with charge-based desolvation," J. Comput. Chem., vol. 28, pp. 1145-1152, 2007.

[25] T. M. Jenkins, D. Esposito, A. Engelman, and R. Craigie, "Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking," EMBO J., vol. 16, pp. 6849-6859, 1997.