Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 475–477
Bioorganic & Medicinal Chemistry Letters

Docking dinucleotides to HIV-1 integrase carboxyl-terminal domain to find possible DNA binding sites

Hai Mei Zhu, Wei Zu Chen and Cun Xin Wang*

College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100022, China

Received 26 July 2004; revised 2 October 2004; accepted 4 October 2004
Available online 25 November 2004

Abstract—HIV-1 integrase mediates a processed viral DNA into the host genome DNA. The binding modes between HIV-1 integrase and viral/host DNA have not been clarified until now. The C-terminal domain of integrase has been found to be a DNA-binding domain. In this work, we have explored the possible DNA binding sites of dimeric C-terminal domain by docking dinucleotides to HIV-1 integrase. The docking results suggest that two symmetrical DNA-binding sites are likely located on the outside surface of the dimeric C-terminal domain and not located in the groove. Those sites are in agreement with the experimental data.

1. Introduction

HIV-1 integrase (IN) is a 32-kDa viral protein, which is responsible for the insertion of a DNA copy of HIV-1 genome into the host genome.[1,2] The binding modes between IN and viral/host DNA were explored by some groups using docking approach. Perryman and McCammon[3] docked eight different dinucleotides to the core domain of IN and developed a model of the binding sites for viral and human DNA. Adesokan et al.[4] predicted interactions between HIV-1 IN and a double-strand viral DNA by fast molecular docking. De Luca et al.[2] constructed a full-length IN and docked with a 27-bp viral DNA.

Due to the computational time consume, Perryman and McCammon[3] and Adesokan et al.[4] just studied the binding modes between DNA and the core domain of HIV-1 IN. The C-terminal domain of IN was also taken into account as a DNA-binding domain.[5] It is found that there is a nonspecific but strong DNA binding activity in the C-terminal domain similar to the full-length IN.[5,6] Therefore, it is necessary to study the DNA binding sites of HIV-1 C-terminal domain. In addition, the multimerization of the protein is needed for IN activity.[7] The dimerization of C-terminal domain has been shown to be very important for correct multimerization of IN.[8] In the constructed full-length IN,[2] the two C-terminal domains were found to be two monomers and could not form a dimer. In our work, the dimeric C-terminal domain was used to explore the possible DNA binding sites. The isolated C-terminal domain forms a homodimer in solution and both subunits of the dimer are folded in a SH3-like manner.[9,10] The groove between the subunits forms a large cleft that is proposed to bind DNA.[11] The second mode suggests that DNA could bind at the surface outside the groove, which located in both sides of the dimeric C-terminal domain.[1,2,11] Since there is no experimental structural information for the C-terminal domain of HIV-1 IN complexed with DNA, the aim of this work is to explore the possible DNA-binding sites by docking dinucleotides to the dimeric C-terminal domain. It is proposed that the favorable positions of dinucleotides will give insight pertaining to where DNA binds to dimeric C-terminal domain.

2. Methods

The coordinates of receptor were chosen from PDB data of the first NMR structure for dimeric C-terminal domain (PDB code 1QMC)[11], in which there is no any metal ion. The AMBER united charges[12] were assigned on the receptor with the Biopolymer module of the SYBYL program package[13]. A $9 \times 9 \times 9$ nm grid box with a grid spacing of 0.075 nm was generated for the receptor. The six dinucleotides, such as d(pCpA), d(pGpT),

Keywords: HIV-1 integrase; C-terminal domain; Dinucleotides; DNA binding site.
* Corresponding author. Tel.: +86 10 67392724; fax: +86 10 67392837; e-mail: cxwang@bjut.edu.cn

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.10.003

d(pApG), d(pCpT), d(pTpA), and d(pGpC), were constructed with the SYBYL package. Then they were minimized using the Tripos force field and assigned charges using the Gasteiger–Hückel method.

The AutoDock 3.0[14] was used to perform docking calculations. It allows automated docking of fully flexible ligands to the rigid enzyme receptor. The Lamarckian genetic algorithm (LGA)[14] was used to find the appropriate binding positions, orientations, and conformations of the ligands. We started the global optimization with a population of 100 randomly positioned individuals. The maximum number of energy evaluations was increased to 2.5×10^6. The maximum number of generations in the LGA algorithm was 27,000. The Solis & Wets local search was performed with a maximum number of 3000. During each docking experiment, 100 runs were carried out, and 10 experiments were executed for each ligand. The resulting data were taken from docking experiments in which the lowest total docking energy was obtained. Finally, we got 6000 different conformations of ligands.

3. Results and discussion

We found that the 6000 conformational structures are all located at the symmetry outside of the dimeric C-terminal domain. The 60 top ligand conformations (10 dominant conformations for each dinucleotides) with the final docked energy that refers to the lowest energy were selected from the 6000 ligand conformations for result analysis and figure plotting. We did not specify the energy range to choose conformations. As to each dinucleotide, among the 10 selected conformations, the difference between the highest and the lowest conformational energy is about 1.67 kcal/mol. The 60 top ligand binding conformations of dimeric C-terminal domain was shown in Figure 1. We can see that the ligand conformations are not located in the groove produced by the two monomers of the C-terminal domain but distributed symmetrically in the outside of the dimeric C-terminal domain. This result supports the second DNA-binding model of the HIV-1 IN C-terminal domain mentioned above. The dinucleotides are much shorter than DNA substrate and they were docked flexibly to the dimeric C-terminal domain. Therefore, our results suggest that the DNA-binding site is not located in the groove region.

In order to see clearly how the dinucleotides bind to the C-terminal domain, we show only the positions of the phosphate atoms on the docked dinucleotides (Fig. 2). On both the outsides of the dimeric C-terminal domain, the phosphate atoms are located in a similar manner. It is found that the distribution manners of the phosphate atoms on the peripheral surface of each C-terminal domains looks somewhat similar to a double-strand DNA substrate. This strongly recommends that these two regions on the surface are the DNA-binding sites.

The orientation of DNA is consistent with the model obtained by docking the constructed full-length IN to the viral DNA.[2] But we cannot distinguish whether they are the human or viral DNA binding sites. The C-terminal domain could be suggested as a responsible target for binding the human DNA due to cation-dependent nonspecific DNA binding activity.[15] However, studies based on IN chimeras[16] and UV cross-linking studies[17] have demonstrated that residues in the C-terminal domain may be also concerned with the binding of viral DNA.

The dinucleotides formed hydrogen bonds with residues R228, K244, E246, R262, R263, K264, K266, I267, and R269 on the C-terminal domain. Most of these residues show positive charges on the surface, providing a favorable binding surface for negatively charged dinucleotides. Residues E246, R262, R263, and K264 have been revealed to involve in DNA binding by mutational

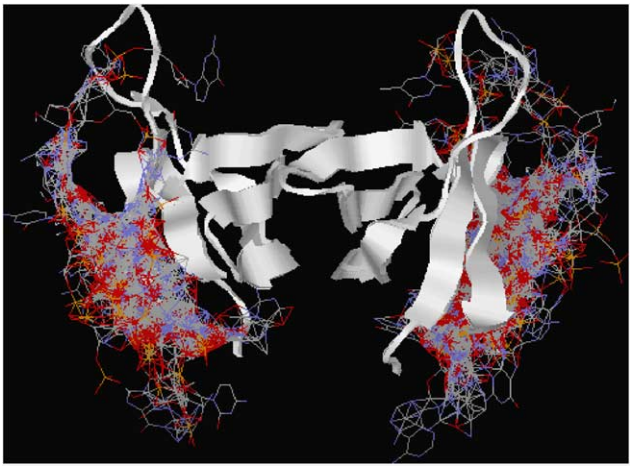

Figure 1. The dimeric C-terminal domain of HIV-1 IN binding ligands dinucleotides. The dimeric C-terminal domain was shown as the white cartoon mode. The 60 ligand conformations of the six dinucleotides were shown as the color lines. All of these conformations distributed symmetrically in the outside of the dimeric C-terminal domain.

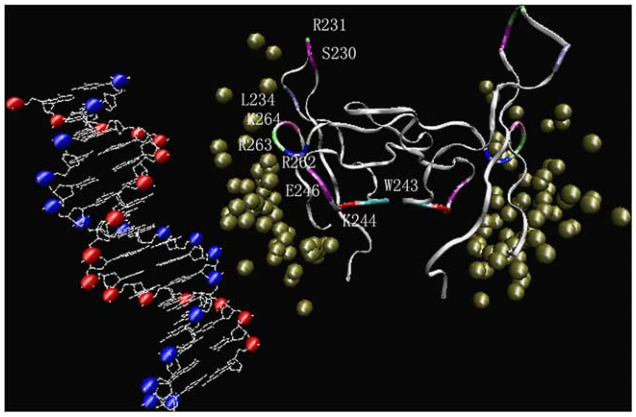

Figure 2. The dimeric C-terminal domain (white ribbons) and the positions of the phosphate atoms of the docked dinucleotides (yellow balls) near the ribbons are shown in this picture. The distributing manner of the phosphate atoms looks like B-form DNA shown on the left. In the B-form DNA, the red balls are the Cα atoms in one chain of DNA, blue balls are the Cα atoms in another chain of DNA. All hydrogen atoms of the DNA chains are not shown.

analysis[7,8,11,15] and protein footprinting.[1] In the model developed by De Luca et al.[2] the viral DNA formed direct interactions with S230, R231, W243, K244, K263, and K264 in the C-terminal domain.[2] According to our docking results, it is difficult for a double-strand DNA to form interactions with W243 and K244 at the same time because the hydrophobic residue W243 is buried in the interface of the C-terminal domain dimer, whereas K244 is pointed to the outside. The mutation of R231A showed reduced DNA-binding activity in the context of the full-length protein,[5] but it is not distinguished whether this reduction is due to a loss of DNA binding or intra- or intermolecular interactions.[8] The photo cross-linking studies suggested that S230 might make hydrogen bonds with the viral DNA, so that the mutation in S230 has not been investigated. In addition, the mutation on L234A has been found to reduce the DNA binding affinity of the dimeric C-terminal domain.[11] The residue L234 is located far away from the groove. But if the dimeric C-terminal domain binds with DNA at its outside surface, the residue L234 could be easily contacted with DNA.

4. Conclusion

Although the function of the C-terminal domain in full-length IN has not been clearly understood, it is suggested to be the DNA-binding domain. Experiments revealed that IN is functional in a multimer, but the exact number of the monomer is still not clear. The C-terminal domain is a dimer in solution, which is important for correct multimerization of IN. Analyses of mutations of the residues on the C-terminal domain have suggested two possible modes for binding DNA. By docking dinucleotides to the dimeric C-terminal domain, we have found that one particular binding mode between the C-terminal domain and DNA is much more probable: the DNA binding sites are likely located symmetrically on the outside surface of the dimeric C-terminal domain. These results are in agreement with the protein footprinting and mutation experimental data. In our docking simulation, we did not consider the flexibility of the receptor. The work on conformational change of HIV-1 integrase dimeric C-terminal domain during the process of binding with all dinucleotides and the DNA segment is currently underway.

Acknowledgements

We thank Professor Olsen for providing us with the AutoDock package. This work was supported in part by the Beijing Natural Science Foundation (Grant Nos. 5032002 and 5042003) and the Foundation from Beijing City Education Committee (No. KZ200410005002).

References and notes

1. Dirac, A. M. G.; Kjems, J. *Eur. J. Biochem.* **2001**, *268*, 743–752.
2. De Luca, L.; Pedretti, A.; Vistoli, G.; Barreca, M. L.; Villa, L.; Monforte, P.; Chimirria, A. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 1083–1088.
3. Perryman, A. L.; McCammon, J. A. *J. Med. Chem.* **2002**, *45*, 5624–5627.
4. Adesokan, A. A.; Roberts, V. A.; Lee, K. W.; Lins, R. D.; Briggs, J. M. *J. Med. Chem.* **2004**, *47*, 821–828.
5. Vink, C.; Oude Groeneger, A. M.; Plasterk, R. H. *Nucleic Acids Res.* **1993**, *21*, 1419–1425.
6. Engelman, A.; Craigie, R. *J. Virol.* **1992**, *66*, 6361–6369.
7. Heuer, T. S.; Brown, P. O. *Biochemistry* **1998**, *37*, 6667–6678.
8. Lutzke, R. A.; Plasterk, R. H. *J. Virol.* **1998**, *72*, 4841–4848.
9. Eijkelenboom, A. P.; Lutzke, R. A.; Boelens, R.; Plasterk, R. H.; Kaptein, R.; Hard, K. *Nature Struct. Biol.* **1995**, *2*, 807–810.
10. 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.
11. Eijkelenboom, A. P.; Sprangers, R.; Hard, K.; Puras Lutzke, R. A.; Plasterk, R. H.; Boelens, R.; Kaptein, R. *Proteins* **1999**, *36*, 556–564.
12. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
13. SYBYL 6.5 Manual. Tripos Inc., 1699 South Hanley Rd, St Louis, MO, USA. 1999.
14. Morris, G. M.; Goodsell, D. S.; Hakkiday, R. S.; Ruth, H.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
15. Lutzke, R. A.; Vink, C.; Plasterk, R. H. *Nucleic Acids Res.* **1994**, *22*, 4125–4131.
16. Katzman, M.; Sudol, M. *J. Virol.* **1998**, *72*, 1744–1753.
17. Heuer, T. S.; Brown, P. O. *Biochemistry* **1997**, *36*, 10655–10665.