AutoDocking Dinucleotides to the HIV-1 Integrase Core Domain: Exploring Possible Binding Sites for Viral and Genomic DNA

Alexander L. Perryman*,† and
J. Andrew McCammon†,‡

Howard Hughes Medical Institute, Department of Chemistry & Biochemistry, and Department of Pharmacology, University of California at San Diego, La Jolla, California 92093-0365

Received June 10, 2002

Abstract: To understand the binding of both viral and human DNA to HIV-1 integrase, fully flexible dinucleotides were docked onto the core domain of integrase. AutoDocking did identify sites on integrase where favorable interactions with nucleotides can occur, and those sites were in agreement with recently published protein fingerprinting data. By analyzing the phosphates of the docked dinucleotides, we developed a model indicating where the viral cDNA and human DNA bind to the integrase core domain.

Although there are good inhibitors of HIV protease and of HIV reverse transcriptase, there are currently no suitable inhibitors of HIV integrase. Integrase is the HIV enzyme that catalyzes the 3′ processing of the viral cDNA, and then it covalently attaches that processed viral cDNA to human genomic DNA. Some compounds developed to inhibit integrase only bind to the integrase−DNA complex and not to integrase by itself.[1] Thus, information concerning the atomic structure of the complex of integrase with DNA (especially the viral cDNA) is essential to enabling the structure-based drug design efforts to proceed.

Because one cannot currently dock large pieces of flexible, double-stranded DNA (dsDNA) to integrase, fully flexible dinucleotides were used in the AutoDocking studies discussed herein. AutoDock3.0.5 was chosen because it utilizes a fully flexible ligand in its docking algorithm (although it is still docked to a rigid protein) and because it has been shown to successfully reproduce many crystal structure complexes.[2] Although this approach does not necessarily place the dinucleotides into the precise position and conformation that would give the global minimum for each complex structure, characterizing the many local minima for such complexes was the goal. AutoDocking did identify sites on integrase where favorable interactions with nucleotides can occur, and those sites were in agreement with recently published protein fingerprinting data.[8] It is proposed that the convergent positions of the phosphates of the AutoDocked dinucleotides will give insight pertaining to where the viral cDNA and where the human DNA bind to the HIV integrase core domain.

* To whom correspondence should be addressed. Phone: 858-534-2798. FaxL 858-534-0006. E-mail: aperryma@mccammon.ucsd.edu.
† Howard Hughes Medical Institute and Department of Pharmacology.
‡ Department of Chemistry & Biochemistry.

The docking studies discussed within were performed on only the core domain of HIV-1 integrase. There is no publicly available structure of the full-length integrase monomer. We are aware that crystal structures of both the N-terminal and core domains or of both the C-terminal and core domains of HIV-1 integrase are available; however, we chose to focus on just the core domain because of computational and time constraints. Each job of AutoDocking a fully flexible dinucleotide (which has 12−14 active torsions) to just the core domain consumed between 560 and 960 h (i.e., 1 job composed of only 10 runs took 23−40 days), and owing to the nature of the genetic algorithm utilized, each job must be run continuously on a single processor. If these docking studies were attempted on a two or three domain structure (instead of just on the core domain), then the amount of computer time required for each job would have been substantially greater. More importantly, the flexible linkages between the core and both the N- and C-terminal domains introduce additional complications in the analysis of possible modes of interactions with DNA in the full integrase−DNA complex. Although there are crystal structures of two of the three domains of integrase, the inherent flexibility of the regions that connect the three domains in combination with the nature of the crystallization environment would most likely cause artificial interdomain interactions to occur in those two domain crystal structures of integrase. The way in which all three domains of integrase interact is likely affected by the presence of all three domains and by the presence of double-stranded DNA. Thus, because there is no reliable information concerning the structure of the full-length integrase, this initial study had to be restricted to focus solely on the largest structural unit for which there is reliable structural information, i.e., a single domain of integrase. This search was limited to just the core domain because the core domain is the region that performs the actual catalysis, which means it is the domain on which subsequent drug design studies will be focused. The positions of the phosphates of the docked dinucleotides will be used to guide the subsequent manual docking of much larger stretches of double-stranded DNA in the future.

To optimize the run parameters used within AutoDock and to verify that those parameters could be used to successfully dock a flexible dinucleotide to a nucleic-acid-modifying enzyme, the complex of barnase (a G-specific endonuclease) with a d(GpC) dinucleotide was used as a control. The crystal structure 1RNB.pdb of Barnase complexed to that dinucleotide was used in the control studies.[3] We did successfully reproduce the structure of the complex of Barnase with a fully flexible GC (as judged by superimposing our docked results with the crystal structure of the complex), which can be seen in the Supporting Information. Importantly, the docked phosphate of the fully flexible GC superimposed exactly on the phosphate from the crystal structure's GC (see figures in the Supporting Information). Thus, the control trial was a success.

10.1021/jm025554m CCC: $22.00
Published on Web 11/14/2002

The ligand and protein preparation procedures and the docking protocols that were tested on the barnase control and then applied to the AutoDocking of dinucleotides to the HIV-1 integrase core domain are listed in the Supporting Information. A summary of the optimized AutoDocking run parameters is as follows: the maximum number of energy evaluations was increased to 9×10^{10} per run (ga_num_evals); the maximum number of generations in the Lamarckian genetic algorithm was increased to 90 000 (ga_num_generations); the maximum number of iterations in the pseudo-Solis-and-Wets-minimization/local search was increased to 3000 (sw_max_its). All other run parameters were maintained at their default settings. The exact same preparatory and AutoDocking processes were applied to both the barnase control and the integrase experiments. All ligands were constructed, Gasteiger−Marsili charges were added with SYBYL, and all images were generated with InsightII.[4,5] Before the barnase or integrase proteins were used in the AutoDocking, the polar hydrogens and Kollman united atom charges were added with SYBYL, and a quick minimization was done with AMBER6.[6]

For the docking of fully flexible dinucleotides to integrase, the crystal structure 1QS4.pdb of the core domain of HIV-1 integrase (residues C56−Q209) was used.[7] The A chain was kept, while the B and C chains were deleted. The Shionogi inhibitor within the active site and all the waters were also deleted. However, the active site magnesium ion was maintained throughout the entire process (with a charge of +2). The eight different fully flexible 5′-phosphorylated dinucleotides that were docked to the entire surface of the integrase core domain are as follows: d(pApC), d(pCpIsoA)[9], d(pTpC), d(pGpA), d(pApA), d(pApG), d(pGpT), and d(pCpA). Those eight dinucleotides were built in the extended conformation, but the docking of a d(pApC) that started in the B-form conformation was also performed and yielded no major differences. There were 10 runs for each system using our optimized run parameters (8 dinucleotides in extended conformations and 1 dinucleotide that began as B-form × 10 runs/dinucleotide = 90 runs), and there were 4 additional jobs of 5 runs each (2 for CisoA and 2 for AC × 5 runs = 20 runs) that used a more rigorous search with a larger population size (75 or 100 members instead of the default value of 50) and an increased local search frequency (0.15 or 0.12 instead of the default value of 0.06). Only 5 runs per job were done for the more rigorous search protocol because each job consumed twice as much CPU time. But the runs involving the more rigorous search protocol produced results similar to results from the runs using just the values described in the previous paragraph. A total of 110 different runs of AutoDocking fully flexible dinucleotides to the entire surface of the integrase core domain were attempted in this trial. Of those 110 trials, 1 quit early when the wall-time of 1000 hours expired.

Of all the dinucleotides docked to the integrase core, $^{99}/_{109}$ of the docked dinucleotides were near residues that are known from a recent protein fingerprinting study to interact with DNA.[8] For all the figures involving integrase, the red residues are residues whose cleavage was protected by the presence of dsDNA, and the residues of the catalytic triad are shown in green. Thus, the red and green residues are known to interact with DNA. Most of those 99 dinucleotides that docked "near" residues known to interact with DNA were actually in contact with the red residue. However, some of the dinucleotides that docked to the region on the right of Figure 1 were within a cleft with two walls composed of DNA-interacting residues, but they were not in actual contact with the red residue.

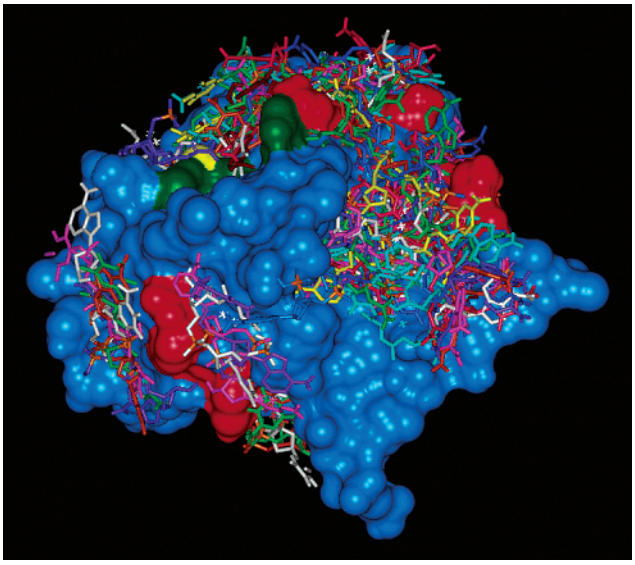

Figure 1. Ninety-nine AutoDocked dinucleotides are shown on the Connolly solvent-accessible surface of the integrase core domain. The active site Mg ion is shown as the yellow region of the integrase surface, while the catalytic triad is displayed in green. A total of 99 out of 109 runs docked the dinucleotides near residues known experimentally to interact with dsDNA (the red and green parts of the surface), as judged by a recent protein fingerprinting study.[8]

Of the 109 completed runs, 99 docked the dinucleotide near a residue known to interact with DNA (i.e., they docked near a red or green residue. This supports our initial hypothesis that dinucleotides in silico will dock to the surface of a protein in a manner similar to the way in which large pieces of dsDNA have been observed to bind in vitro. And this agreement occurred even though the AutoDocking studies were performed on just the core domain of integrase, while the fingerprinting studies were performed on the full, 3 domain integrase. Thus, even though we docked to only the core domain, the results do correlate well with the full integrase−DNA complex as characterized experimentally in a biologically relevant context. The other 10 runs docked to the flat surface where dimerization between core domains occurs in the crystal structures. Perhaps the results of those 10 runs could be used to provide information pertaining to the design of an inhibitor that could help prevent the multimerization process.

When all 99 of the docked dinucleotides are displayed upon the core domain's surface in Figure 1, a chaotic picture is created from which useful information might be difficult to extract. However, if one shows only the phosphate positions of the docked dinucleotides, the picture becomes much clearer. As can be seen by the overlapping orangish-brown X groups (the phosphates of the docked dinucleotides) in Figure 2, several different runs placed the phosphates in the same general posi-

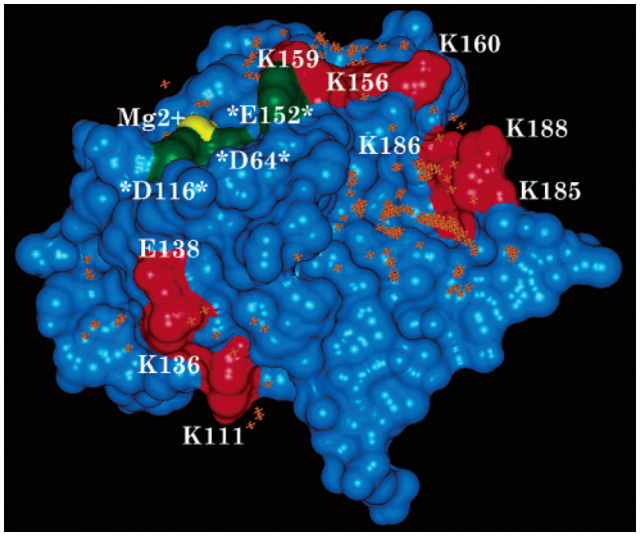

Figure 2. Phosphates of the docked dinucleotides are shown as organish-brown X groups on the Connolly surface of the integrase core domain. The same coloring scheme is used as in the other figures (the red and green residues are known to interact with DNA, and the Mg ion is shown as the yellow region of the integrase surface). The DNA-interacting residues are labeled, and the catalytic triad residues have asterisks (*) surrounding their labels. The region on the left has a discrete pattern of a few locations where favorable interactions with phosphate groups can occur. However, the region on the right showed a convergence of many overlapping sites where phosphate groups can interact favorably.

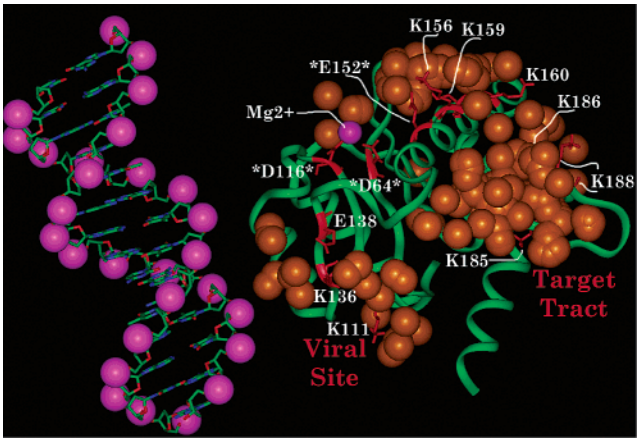

Figure 3. Integrase core domain is shown as green ribbons, and the positions of the phosphates of the docked dinucleotides are shown as brown CPK spheres on that ribbon. A random B-form dsDNA is shown on the left to indicate the relative scale, and the dsDNA's phosphates are displayed as magenta CPK spheres. The same orientation of the integrase core that is shown in the previous figures was used here. The residues of the catalytic triad have asterisks (*) surrounding their labels.

tion. Although the region on the left has a discrete pattern of a few docked phosphate groups, many of the docked phosphate groups on the right half of the figure overlap substantially.

The "target tract" will be defined as the region on the right of Figure 3 (and of all the other integrase figures) where there is a dense packing of many different but overlapping sites that can interact favorably with phosphate groups (shown as brown CPK spheres), which would allow the accommodation of many different sequences/structures and motions of dsDNA in this

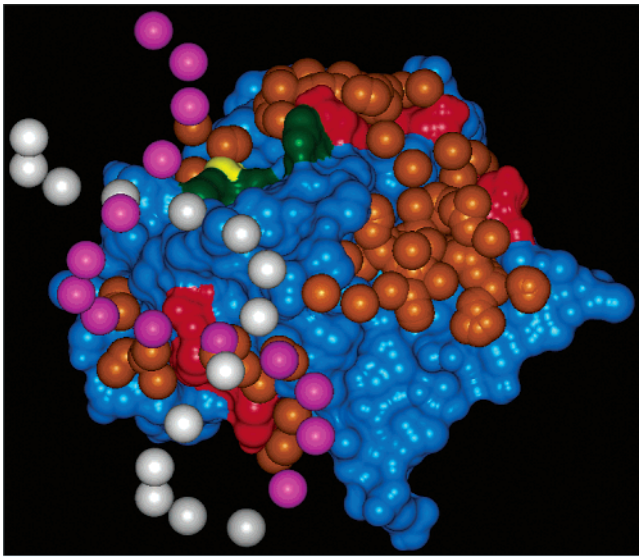

Figure 4. Connolly surface of integrase is shown with the phosphates of the docked dinucleotides portrayed as brown CPK spheres. The dsDNA helix from Figure 3 was simply translated to the right, its bases and sugars were deleted, and the front strand became white. The phosphates of the back strand of the dsDNA helix are shown as magenta CPK spheres. The residues of the catalytic triad are green, and the active site Mg is yellow. The pattern of the docked phosphates in the viral site correlates well to the locations of the phosphates in the back strand of a B-form dsDNA helix.

region. Because integrase attaches the viral cDNA to the human genome in a nonsequence-specific manner, this "target tract" is probably where integrase binds to human DNA. Conversely, the "viral site" on the left of Figure 3 has a much more specific pattern of locations where favorable interactions with phosphates can occur.

The "specificity" we describe is specificity related to the number of sites in that region where favorable interactions with phosphate groups can occur. Conveniently, GT (part of the viral LTR) docked to the viral site with a much better average final docked energy than the average of all the dinucleotides that docked to the viral site (−9.45 for the 3 GT runs vs −7.63 for all 26 runs). But we do not claim that our results prove specificity concerning which sequence of DNA will bind best to particular regions of integrase.

We use the argument that because integrase does bind to and preferentially cleave the sequence of the viral LTR, sequence specificity in binding interactions with viral cDNA must exist. Similarly, because integrase attaches that viral cDNA to virtually any sequence within the human genome, the binding interactions with the human target DNA must not be governed by sequence-specific effects. Our results showed that the region we have proposed as the "viral site" has a much more specific pattern of a small number of discrete sites (leading up to the active site) where favorable interactions with phosphate groups can occur; thus, fewer sequences would likely be able to deform suitably to enable strong binding with that region. The sequence determines the relative deformability of the DNA, which governs the positions that the phosphate groups could assume when bound upon the integrase surface. Because there is a more discrete pattern of locations within this site where favorable interactions with phosphate groups can occur, we propose that the

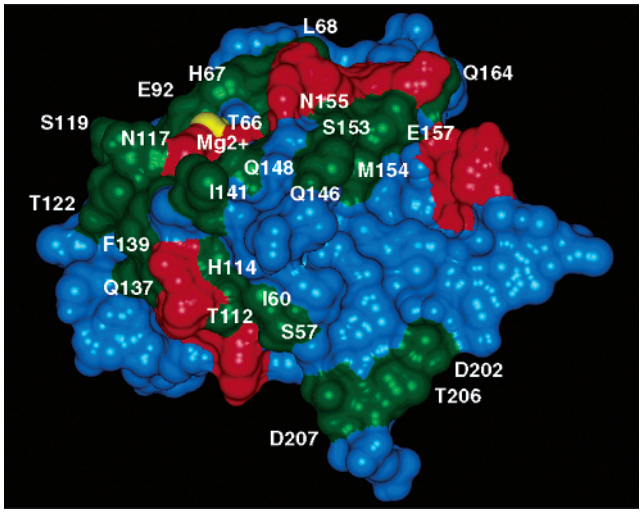

Figure 5. Residues suggested for sequential Cys-scanning mutagenesis and crystallization with a thiol-containing DNA are green and labeled.

sequence-specific interactions with DNA are more likely to occur in this region, which is why we propose this region as the binding site for the viral cDNA. But we make no claim about the specific sequence that would bind best to that region; we simply state that the region we call the "viral site" is more likely to be involved in sequence-specific interactions with DNA. Conversely, the region we have proposed as the "target tract" has a dense cluster of many overlapping sites where favorable interactions with phosphate groups can occur. Because many different positions and motions of phosphate groups can easily be accommodated within this region, many different sequences of double-stranded DNA should be able to interact within this tract. Thus, we propose that this region is more likely to be involved in interactions with DNA that are not guided by the specific sequences involved, which is why we propose that the human DNA binds within this "target tract."

This discrete pattern of docked phosphate positions within the "viral site" is quite similar to the organization of the phosphate positions of the back strand of a regular B-form dsDNA helix. If the viral cDNA does bind in this region in the manner that we suggest, then that line of red residues (known to be protected from cleavage by the presence of dsDNA[8]) would align with the viral cDNA's helical axis, and they would be the residues most protected from cleavage in that region (see Figure 4).

To summarize, presented here is a study that utilized the convergent positions of the phosphates of fully flexible, AutoDocked dinucleotides to suggest the regions on the HIV-1 integrase core domain's surface where the viral cDNA binds versus where the human genomic DNA binds.

The model developed here suggests a variety of experiments. For example, the introduction of solvent-exposed cysteine residues (i.e., focused Cys scanning mutagenesis) within the target tract or along the viral site might facilitate the crystallization of the integrase core domain (or the full integrase monomer) with a dsDNA helix containing an engineered sulfur on its surface (and perhaps also with a nonhydrolyzable linkage after the conserved CA). The thiol-containing dsDNA could then cross-link to the integrase surface, and that tethering might allow crystallization of the complex to proceed. See Figure 5 for details on the specific residues one might wish to mutate to Cys.

Acknowledgment. We thank Dr. Christoph Sotriffer and Dr. Garrett Morris for helpful advice concerning AutoDock. We thank Dr. Robert Konecny for compiling the programs and for keeping the KeckII cluster running for over 40 continuous days. We also thank Dr. Jung-Hsin Lin, Dr. Chung Wong, Dr. Haihong Ni, and Dr. Thomas P. Quinn. This work was supported, in part, by grants to J.A.M from NIH, NSF, and NPACI/SDSC. Additional support has been provided by the W. M. Keck Foundation, the National Biomedical Computational Resource, and Accelrys, Inc. A.L.P. is an HHMI Predoctoral Fellow.

Supporting Information Available: Detailed modeling protocols and parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hazuda, D. J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J. A.; Espeseth, A.; Gabryelski, L.; Schleif, W.; Blau, C.; Miller, M. D. Inhibitors of Strand Transfer That Prevent Integration and Inhibit HIV-1 Replication in Cells. *Science* **2000**, *287*, 646−650.

(2) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(3) Baudet, S.; Janin, J. Crystal Structure of a Barnase−d(GpC) Complex at 1.9 Å Resolution. *J. Mol. Biol.* **1991**, *219*, 123−132.

(4) *SYBYL*, version 6.7.2; Tripos, Inc.: St. Louis, MO, 2001.

(5) *InsightII*; Accelrys Inc.: San Diego, CA, 2000.

(6) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, C.; Weiner, P.; Kollman, P. A. *AMBER 6*; University of California, San Francisco, 1999.

(7) Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. Structure of the HIV-1 Integrase Catalytic Domain Complexed with an Inhibitor: A Platform for Antiviral Drug Design. *PNAS* **1999**, *96*, 13040−13043.

(8) Dirac, A. M. G.; Kjems, J. Mapping DNA-Binding Sites of HIV-1 Integrase by Protein Fingerprinting. *Eur. J. Biochem.* **2001**, *268*, 743−751.

(9) Taktakishvili, M.; Neamati, N.; Pommier, Y.; Pal, S.; Nair, V. Recognition and Inhibition of HIV Integrase by Novel Dinucleotides. *J. Am. Chem. Soc.* **2000**, *122*, 5671−5677.

JM025554M