Supporting Information

Christoph A. Sotriffer, Haihong Ni, J. Andrew McCammon

HIV-1 integrase inhibitor interactions at the active site: prediction of binding modes unaffected by crystal packing

Journal of the American Chemical Society

Docking. Docking was performed with version 3.0 of the program AutoDock[1], using the new empirical free energy function and the Lamarckian genetic algorithm. A standard protocol was applied, with an initial population of 50 randomly placed individuals, a maximum number of 1.5×10^6 energy evaluations, a mutation rate of 0.02, a crossover rate of 0.80, and an elitism value of 1. Proportional selection was used, where the average of the worst energy was calculated over a window of the previous 10 generations. For the local search the so-called pseudo-Solis and Wets algorithm was applied using a maximum of 300 iterations per local search. The probability of performing local search on an individual in the population was 0.06, and the maximum number of consecutive successes or failures before doubling or halving the local search step size was 4. 50 independent docking runs were carried out for each ligand. Results differing from each other by less than 1.0 Å in positional root mean square deviation (rmsd) were clustered together and represented by the result with the most favorable free energy of binding.

Ligand setup. Special care was required for the setup and modeling of 5CITEP, as mesomeric and tautomeric effects need to be considered. These considerations were necessary since the crystal structure of the compound in the integrase complex obtained at 2.1 Å resolution[2] can obviously not provide all required details. The compound is a 1,3-diketone, with two heterocyclic aromatic substituents. Compounds of this class generally exist in a keto-enol tautomeric equilibrium, where the enol form is often dominant. In crystal structures of many different 1,3-diaryl-1,3-propanediones generally the enol form is found, and due to the very strong intramolecular hydrogen bond it may also be predominant in solution[3]. The compound was therefore modeled as keto-enol. In principle, two forms of enols are possible, depending on the position of the enolic proton. As discussed by Bertolasi *et.al.*[3], an electronegativity rule can be proposed according to which the proton resides preferably on the carbonyl connected to the more electronegative group. For 5CITEP the proton has therefore been chosen to reside on the oxygen adjacent to the tetrazole substituent. The tetrazole part deserves special attention itself. In its neutral form, tetrazole exists in two tautomers, depending on the position of the hydrogen atom. However, since tetrazole has a pKa of about 5, the most relevant form of the tetrazole substituent in 5CITEP should be deprotonated and the compound was therefore modeled accordingly.

The structure of the ligand generated according to these considerations was optimized with the CHARMm® force field implemented in QUANTA® (these programs are distributed by Molecular Simulations Incorporated). Atomic charges were assigned using the Gasteiger-Marsili formalism[4], which is the type of atomic charges used in calibrating the AutoDock empirical free energy function[1].

To get an estimate for the torsion barriers in 5CITEP at the bonds connecting the keto-enol group with the tetrazole and the chloroindole ring, respectively, fully relaxed single bond torsional potential were evaluated by *ab initio* calculations at the HF/6-31G* level in increments of 15°. These calculations were done with Gaussian98[5]. Based on the *ab initio* results and a comparison with the conformation found in the crystal structure of the complex, the tetrazolium was kept fixed and planar with respect to the keto-enol group, while the bond between keto-enol group and chloroindole was allowed to rotate in the docking process, but constrained to values within $\sim \pm 45°$ relative to the planar minimum conformation.

Protein setup. For the IN catalytic core domain the structure obtained by X-ray analysis of the complex with 5CITEP was used (PDB 1QS4)[2]. For the purpose of docking, subunit A was selected, which is the only monomer in the asymmetric unit where the position of the ligand could be determined. The missing residues at position 141-144 in this subunit were incorporated from monomer B of the IN structure PDB 1BIS[6] after superposition of the backbones of residues 135-140 and 145-150.

The structure was setup for docking as follows: polar hydrogens were added using the PROTONATE utility (written by D.A. Case, K. Cross, and G.P. Gippert, and distributed with AutoDock[1] and AMBER[7]). Histidine protonation states were modeled as in related MD

simulation studies[8]. In order to optimize the inserted loop residues, the structure was subjected to a short energy minimization using the SANDER program of AMBER5.0[7] and the AMBER united atom force field[9], in accordance with the type of force field and protein charges of the AutoDock empirical free energy function. Solvation parameters were added to the final protein file using the ADDSOL utility of AutoDock3.0. For docking to the dimer, where one inhibitor molecule was kept in the active site, solvation parameters had to be added for the inhibitor as well; these were taken from parameters of corresponding atoms in amino acid side chains.

The grid maps representing the protein in the actual docking process were calculated with AutoGrid. The grids (one for each atom type in the ligand, plus one for electrostatic interactions) were centered on the active site(s), but chosen to be sufficiently large to include not only the active site, but also significant portions of the surrounding surface. The dimensions of the grids were thus 30 Å × 30 Å × 30 Å, with a spacing of 0.375 Å between the grid points.

References

1. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

2. Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13040–13043.

3. Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *J. Am. Chem. Soc.* **1991**, *113*, 4917–4925.

4. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219–3228.

5. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98.* Gaussian, Inc., Pittsburgh PA, 1998.

6. Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davies, D. R. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9150–9154.

7. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 5.* University of California, San Francisco, 1997.

8. Lins, R. D.; Briggs, J. M.; Straatsma, T. P.; Carlson, H. A.; Greenwald, J.; Choe, S.; McCammon, J. A. *Biophys. J.* **1999**, *76*, 2999–3011.

9. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.