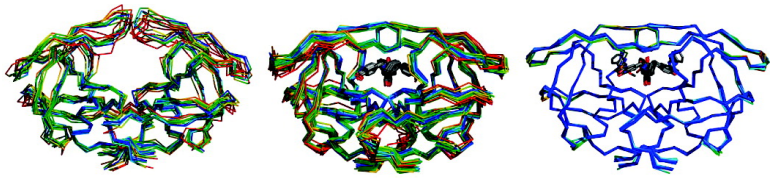
Article

Exploring Experimental Sources of Multiple Protein Conformations in Structure-Based Drug Design

Kelly L. Damm, and Heather A. Carlson

J. Am. Chem. Soc., **2007**, 129 (26), 8225-8235• DOI: 10.1021/ja0709728 • Publication Date (Web): 08 June 2007



MD Snapshots-Apo Form | **NMR Ensemble-Bound Form** | **Crystal Structures-Bound Form**

More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 18 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 06/08/2007

Exploring Experimental Sources of Multiple Protein Conformations in Structure-Based Drug Design

Kelly L. Damm and Heather A. Carlson*

Contribution from the Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1065

Received February 16, 2007; E-mail: carlsonh@umich.edu

Abstract: Receptor flexibility must be incorporated into structure-based drug design in order to portray a more accurate representation of a protein in solution. Our approach is to generate pharmacophore models based on multiple conformations of a protein and is very similar to solvent mapping of hot spots. Previously, we had success using computer-generated conformations of apo human immunodeficiency virus-1 protease (HIV-1p). Here, we examine the use of an NMR ensemble versus a collection of crystal structures, and we compare back to our previous study based on computer-generated conformations. To our knowledge, this is the first direct comparison of an NMR ensemble and a collection of crystal structures to incorporate protein flexibility in structure-based drug design. To provide an accurate comparison between the experimental sources, we used bound structures for our multiple protein structure (MPS) pharmacophore models. The models from an NMR ensemble and a collection of crystal structures were both able to discriminate known HIV-1p inhibitors from decoy molecules and displayed superior performance over models created from single conformations of the protein. Although the active-site conformations were already predefined by bound ligands, the use of MPS allows us to overcome the cross-docking problem and generate a model that does not simply reproduce the chemical characteristics of a specific ligand class. We show that there is more structural variation between 28 structures in an NMR ensemble than 90 crystal structures bound to a variety of ligands. MPS models from both sources performed well, but the model determined using the NMR ensemble appeared to be the most general yet accurate representation of the active site. This work encourages the use of NMR models in structure-based design.

Introduction

Structure-based drug design (SBDD) has emerged as a very important tool in drug discovery research.[1,2] Advances in computational power have increased the practicality of using molecular modeling techniques in the drug development process, although the tradeoff between speed and accuracy still exists. Accounting for the conformational changes that can occur within the binding site of proteins has increased the difficulty of SBDD. The protein in solution exists as an ensemble of energetically accessible conformations and is best described when all states are represented.[3−5] Upon ligand binding, further conformational changes in the receptor can be induced. Although ligand flexibility can be accurately reproduced, replicating the innumerable degrees of freedom of a protein is impractical.

Much progress has been made in developing clever, computationally feasible methods that simulate the inherent flexibility of a ligand−receptor system using both experimentally and theoretically determined structures. Several reviews have been published in this area of research.[6−9] In general, there are four techniques that are most often employed. The original method is termed "soft docking" and involves relaxing the criterion used to model steric fit, allowing for overlap of the protein and ligand surfaces.[10,11] A second method utilizes a single representative structure, the average of a collection of conformational states.[12,13] A third way is to generate receptor conformations "on the fly" such that side chains are allowed to move to accommodate ligand binding using a predetermined rotamer library to define acceptable, alternative conformations.[14,15] A final approach is to use an ensemble of protein structures.[13,16−19] Models can be

(1) Hardy, L. W.; Malikayil, A. *Curr. Drug Discovery* **2003**, December, 15−20.
(2) Jorgensen, W. L. *Science* **2004**, *303*, 1813−1818.
(3) Carlson, H. A.; McCammon, J. A. *Mol. Pharmacol.* **2000**, *57*, 213−218.
(4) Luque, I.; Freire, E. *Proteins: Struct., Funct., Bioinf.* **2000**, Suppl. 4, 63−71.
(5) Ma, B.; Shatsky, M.; Wolfson, H. J.; Nussinov, R. *Protein Sci.* **2002**, *11*, 184−197.
(6) Carlson, H. A. *Curr. Pharm. Des.* **2002**, *8*, 1571−1578.
(7) Teague, S. J. *Nat. Rev. Drug Discovery* **2003**, *2*, 527−541.
(8) Teodoro, M. L.; Kavraki, L. E. *Curr. Pharm. Des.* **2003**, *9*, 1635−1648.
(9) Wong, C. F.; McCammon, J. A. *Annu. Rev. Pharmacol. Toxicol.* **2003**, *43*, 31−45.
(10) Jiang, F.; Kim, S. H. *J. Mol. Biol.* **1991**, *219*, 79−102.
(11) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. *J. Med. Chem.* **2004**, *47*, 5076−5084.
(12) Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M. *J. Mol. Biol.* **1997**, *266*, 424−440.
(13) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. *J. Med. Chem.* **2004**, *47*, 45−55.
(14) Frimurer, T. M.; Peters, G. H.; Iversen, L. F.; Andersen, H. S.; Moller, N. P.; Olsen, O. H. *Biophys. J.* **2003**, *84*, 2273−2281.
(15) Kallblad, P.; Dean, P. M. *J. Mol. Biol.* **2003**, *326*, 1651−1665.
(16) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. *J. Med. Chem.* **2000**, *43*, 2100−2114.
(17) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. *J. Mol. Biol.* **2001**, *308*, 377−395.
(18) Cavasotto, C. N.; Abagyan, R. A. *J. Mol. Biol.* **2004**, *337*, 209−225.

generated by overlaying the different conformations in the ensemble, or each structure can be considered separately.

Previously, Carlson et al. developed a robust receptor-based pharmacophore method based on an ensemble of unbound protein structures to account for the inherent flexibility of HIV-1 integrase.[16] The multiple protein structure (MPS) method was then extended to develop pharmacophore models from three unbound structures of human immunodeficiency virus-1 protease (HIV-1p).[20,21] In both studies, molecular dynamics (MD) simulations were used to generate the multiple conformations. The MPS pharmacophore models outperformed those generated from static models and were successful at predicting known inhibitors from druglike noninhibitors. It is of particular interest that the models created using the unbound HIV-1p were able to identify ligand conformations found in cocrystal structures.[20]

A difficulty that arises when working with multiple structures is deciding which receptor conformations are the most appropriate to use. A further issue is the source of the structures; are structures generated from MD simulations or solved using NMR or X-ray crystallography more suitable? Previously, NMR structures have been shown to sample more conformational space than MD simulations and account for additional protein flexibility.[22] However, multiple groups have demonstrated that dynamics simulations provide complete sampling of the multiple flap conformations of HIV-1p.[23,24]

Crystal structures are thought to provide a more accurate depiction of a protein despite the fact that NMR structures are solved in a more biologically relevant environment.[25−27] This may be due to the fact that X-ray crystallography generally provides a greater amount of high-quality experimental data than NMR spectroscopy, which can be assessed using standard quality control measurements. Good agreement is usually seen in the protein backbones of crystal structures versus NMR structures, and the conformational sampling is focused on loop regions and side chains.[25,28] Two independent groups found that crystal and NMR structures often provide complementary structural information and should be used in conjunction with one another as methods to solve protein structures.[29,30] It is also known that the choice of protein structure can heavily influence the outcome of a simulation; different conformations do perform better than others in virtual-screening applications.[31,32] A recent review summarizes the use of crystal structures in SBDD and discusses the associated limitations.[33]

Crystal structures provide only a static snapshot of the dynamic structure of a protein, and bound structures can lead to the "cross-docking" problem. The binding site is already predefined for the cocrystallized ligand and may not fit other conformationally diverse structures. Nonetheless, using a collection of crystal structures bound to a variety of ligand classes offers an ensemble of conformations and can elucidate structural changes that occur upon ligand binding.[34] Limitations exist such that most systems rarely have a large number of crystal structures solved in complex with many diverse ligands. Also, crystal structure conformations can be influenced by crystallization conditions such as crystal packing effects, pH, buffers, and temperature and may not be a fully correct representation of the structure in solution. Finally, flexible regions may be ill-defined due to a lack of electron density.

Conversely, the use of NMR spectroscopy as a method of three-dimensional structure determination provides an ensemble of conformations found in solution. The ensemble is comprised of low-energy structures that satisfy acceptance criteria based on the experimental data. Each conformation alone can be thought of as a static snapshot; however, they provide a dynamic representation of the protein when used as a collection. As with crystallography, experimental conditions may influence the determined conformations. Also, the structural variability may not be a result of true motion in the protein but rather due to insufficient experimental data.[35]

In the literature, almost all studies use crystal structures in SBDD, both collections and single, static conformations. There are a few occurrences where NMR ensembles are also employed. For example, Knegtel et al. used NMR ensembles to examine ras p21 and uteroglobin.[12] Additionally, Huang and Zou found that ensemble docking to NMR structures of HIV-1p resulted in the identification of more known inhibitors than docking to single, static crystal structures (91% vs 66%, respectively).[36] Furthermore, it is common to utilize information from NMR such as NOE-derived distance constraints and torsion angle constraints to aid in both protein−protein and ligand−protein docking.[37−39] Fragment-based screening through NMR or "SAR by NMR" has also been widely used in drug discovery for the past 10 years.[40−42] However, no one has compared the use of NMR structures to collections of crystal structures.

We are now interested in expanding our MPS technique to incorporate experimental structures from either an NMR ensemble or a collection of crystal structures. There are few examples where a diverse set of experimentally determined structures is available, but one such case is HIV-1p; structures are available from both NMR and X-ray crystallography. HIV-1p is a key drug target because it is a viral enzyme critical to

(19) Fernandes, M. X.; Kairys, V.; Gilson, M. K. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1961−1970.
(20) Meagher, K. L.; Carlson, H. A. *J. Am. Chem. Soc.* **2004**, *126*, 13276−13281.
(21) Meagher, K. L.; Lerner, M. G.; Carlson, H. A. *J. Med. Chem.* **2006**, *49*, 3478−3484.
(22) Philippopoulos, M.; Lim, C. *Proteins: Struct., Funct., Bioinf.* **1999**, *36*, 87−110.
(23) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915−920.
(24) Toth, G.; Borics, A. *J. Mol. Graphics Modell.* **2006**, *24*, 465−474.
(25) Garbuzynskiy, S. O.; Melnik, B. S.; Lobanov, M. Y.; Finkelstein, A. V.; Galzitskaya, O. V. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 139−147.
(26) Lee, M. R.; Kollman, P. A. *Structure* **2001**, *9*, 905−916.
(27) Spronk, C. A. E. M.; Linge, J. P.; Hilbers, C. W.; Vuister, G. W. *J. Biomol. NMR* **2002**, *22*, 281−289.
(28) Billeter, M. Q. *Rev. Biophys.* **1992**, *25*, 325−377.
(29) Snyder, D. A.; Chen, Y.; Denissova, N. G.; Acton, T.; Aramini, J. M.; Ciano, M.; Karlin, R.; Liu, J.; Manor, P.; Rajan, P. A.; Rossi, P.; Swapna, G. V.; Xiao, R.; Rost, B.; Hunt, J.; Montelione, G. T. *J. Am. Chem. Soc.* **2005**, *127*, 16505−16511.
(30) Yee, A. A.; Savchenko, A.; Ignachenko, A.; Lukin, J.; Xu, X.; Skarina, T.; Evdokimova, E.; Liu, C. S.; Semesi, A.; Guido, V.; Edwards, A. M.; Arrowsmith, C. H. *J. Am. Chem. Soc.* **2005**, *127*, 16512−16517.
(31) Thomas, M. P.; McInnes, C.; Fischer, P. M. *J. Med. Chem.* **2006**, *49*, 92−104.
(32) Subramanian, J.; Sharma, S.; B-Rao, C. *J. Med. Chem.* **2006**, *49*, 5434−5441.
(33) Davis, A. M.; Teague, S. J.; Kleywegt, G. J. *Angew. Chem., Int. Ed.* **2003**, *42*, 2718−2736.
(34) Barril, X.; Morley, S. D. *J. Med. Chem.* **2005**, *48*, 4432−4443.
(35) Spronk, C. A. E. M.; Nabuurs, S. B.; Bonvin, A. M.; Krieger, E.; Vuister, G. W.; Vriend, G. *J. Biomol. NMR* **2003**, *25*, 225−234.
(36) Huang, S-Y.; Zou, X. *Protein Sci.* **2007**, *16*, 43−51.
(37) Fesik, S. W. *J. Biomol. NMR* **1993**, *3*, 261−269.
(38) McCoy, M. A.; Wyss, D. F. *J. Am. Chem. Soc.* **2002**, *124*, 2104−2105.
(39) Zabell, A. P. R.; Post, C. B. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 295−307.
(40) Hicks, R. P. *Curr. Med. Chem.* **2001**, *8*, 627−640.
(41) Hajduk, P. J. *Mol. Interventions* **2006**, *6*, 266−272.
(42) Zartler, E. R.; Shapiro, M. J. *Curr. Pharm. Des.* **2006**, *12*, 3963−3972.

continuing the life cycle of HIV, the retrovirus responsible for acquired immunodeficiency syndrome.[43] Upon ligand binding, multiple conformational changes occur in the protease such that there is an inward rotation of each monomer and the flaps that cover the active site assume a closed conformation (5−7 Å shift from apo form).[44,45] Although there are nine marketed therapies that target HIV-1p,[46] this is still a very active area of research due to the associated toxicity, poor pharmacokinetic properties, and resistance that has developed to the existing drugs.

We focus on comparing the use of two protein collections in our MPS method, an NMR ensemble of HIV-1p with a bound cyclic urea (cu) inhibitor and multiple unique crystal structures with cu inhibitors. The location and chemical characteristics of the pharmacophore elements are consistent between the models; however, additional elements exist in the cu-crystal model. Interestingly, even when the protease is in a bound conformation, the features of our previous model generated from apo HIV-1p[21] are still reproduced. In an effort to incorporate the most structural data, we also create a model from 90 crystal structures of susceptible HIV-1p. We show that the structural variation between the collection is very small, resulting in a very similar model to the cu-crystal model. We are also able to demonstrate that models generated from protein ensembles are more successful at discriminating between known HIV-1p inhibitors and inactive druglike molecules than are models from a single "average" structure. Erickson et al. have shown that "average" structures of HIV-1p also perform poorly when docking a ligand into its binding site with a successful docking rate of only 32.5%.[13] To our knowledge, this is the first time a direct comparison of NMR ensembles and crystal collections was made using the same protein in SBDD.

Methods

Protein Preparation. A cu-bound NMR structure (PDB ID: 1BVE)[47] comprised of 28 distinct models was downloaded from the Protein Data Bank (PDB)[48] along with the restrained minimized average NMR structure (PDB ID: 1BVG).[47] The Binding MOAD database[49] was used to obtain 174 bound crystal structures, all having a resolution of ≤2.5 Å. Any structure with a mutation known to confer resistance or known to alter the biological activity of the protein (i.e., A25N) was discarded, resulting in a collection of only drug-susceptible, active strains. Because of the ambiguity in the data, any structure with residues in multiple orientations in the active-site region (defined as any residue within 10 Å of the active-site center) was removed. Structure 1AID was also discarded from this study as an outlier due to the unusual conformation of the flap region,[50] resulting in a final set of 90 structures. Of the 90 structures, 10 are drug-susceptible, active strains bound to unique cu ligands and were used as a collection to provide a direct comparison to the NMR ensemble: 1AJX,[51] 1DMP,[52] 1HVR,[53] 1HWR,[44] 1PRO,[54] 1QBR,[55] 1QBS,[56] 1QBT,[55] 1QBU,[55] and 1T7K.[57] The structures of the cu ligands and inhibition constants are provided in the Supporting Information along with the PDB IDs of the entire crystal collection.

All NMR and crystal structures were first prepared by using MolProbity[58] to check the side-chain orientations, and histidine tautomers were checked by hand. Next, ligands and solvent ions were removed from each structure. Any hydrogen atom was stripped from the crystal structures then added with xleap in the AMBER6[59] suite and minimized to convergence with 10 000 steps of conjugate gradient energy minimization using Sander Classic. This ensured uniform setup of all structures, whether from NMR or crystallographic sources.

MUSIC Simulation. The active site of each NMR and crystal structure was flooded with 500 small molecule probes using a 12 Å radius sphere to define the initial placement. Benzene, ethane, and methanol were utilized as the probes. The sphere was centered at the midpoint of the active site to ensure complete random sampling throughout the entire binding cavity. Each structure was then used in a multiunit search for interacting conformers (MUSIC) simulation with the BOSS program,[60] using the OPLS force field[61] and holding the protein atoms fixed. The small molecule probes were minimized via a low-temperature Monte Carlo sampling, revealing energetically favorable regions of the active-site surface for each chemical functionality. Benzene probes elucidate aromatic and hydrophobic interactions, ethane probes clarify general hydrophobic interactions from aromatic, and methanol probes demonstrate hydrogen-bond donating and accepting sites. Probes do not interact with other probes, but the full interaction energy is calculated with the protein atoms. Further details describing the MUSIC simulation have been previously published.[16]

Pharmacophore Elements. Each structure was then examined to determine clusters, regions where multiple probes had minimized to the same location on the protein surface. This was done both manually and using an autoclustering method based on our in-house Jarvis−Patrick codes. Any cluster within 9.5 Å of the catalytic aspartic acid residues 25 and 25′ was investigated, and if eight probes were present, the cluster was represented by its "parent", the lowest-energy probe calculated in the MUSIC simulation.

An average structure was calculated for each protein set: the NMR ensemble, all-crystal collection (90 structures), and cu-crystal collection (10 structures). Each set of structures was superimposed to a reference protein, the structure in the ensemble with the smallest rmsd to the calculated average structure, using a Gaussian-weighted rmsd (wrmsd) alignment,[62] setting the scaling factor equal to 2 Å^2. The overlaid sets were then used to determine "cluster of clusters" or consensus clusters of the probe molecules. A consensus cluster is defined by having parent probes from ≥50% of the protein conformations. For example, the NMR

(43) Babine, R. E.; Bender, S. L. *Chem. Rev.* **1997**, *97*, 1359−1472.
(44) Ala, P. J.; Huston, E. E.; Klabe, R. M.; Jadhav, P. K.; Lam, P. Y.; Chang, C. H. *Biochemistry* **1998**, *37*, 15042−15049.
(45) Wlodawer, A.; Gustchina, A. *Biochim. Biophys. Acta* **2000**, *1477*, 16−34.
(46) Eder, J.; Hommel, U.; Cumin, F.; Martoglio, B.; Gerhartz, B. *Curr. Pharm. Des.* **2007**, *13*, 271−285.
(47) Yamazaki, T.; Hinck, A. P.; Wang, Y. X.; Nicholson, L. K.; Torchia, D. A.; Wingfield, P.; Stahl, S. J.; Kaufman, J. D.; Chang, C. H.; Domaille, P. J.; Lam, P. Y. *Protein Sci.* **1996**, *5*, 495−506.
(48) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(49) Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. *Proteins: Struct., Funct., Bioinf.* **2006**, *60*, 333−340.
(50) Rutenber, E.; Fauman, E. B.; Keenan, R. J.; Fong, S.; Furth, P. S.; Ortiz de Montellano, P. R.; Meng, E.; Kuntz, I. D.; DeCamp, D. L.; Salto, R.; Rose, J. R.; Craik, C. S.; Stroud, R. M. *J. Biol. Chem.* **1993**, *238*, 15343−15346.
(51) Backbro, K.; Lowgren, S.; Osterlund, K.; Atepo, J.; Unge, T.; Hulten, J.; Bonham, N. M.; Schaal, W.; Karlen, A.; Hallberg, A. *J. Med. Chem.* **1997**, *40*, 898−902.
(52) Hodge, C. N.; et al. *Chem. Biol.* **1996**, *3*, 301−314.
(53) Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C-H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. *Science* **1994**, *263*, 380−384.
(54) Sham, H. L.; et al. *J. Med. Chem.* **1996**, *39*, 392−397.
(55) Jadhav, P. K.; Ala, P.; Woerner, F. J.; Chang, C. H.; Garber, S. S.; Anton, E. D.; Bacheler, L. T. *J. Med. Chem.* **1997**, *40*, 181−191.
(56) Lam, P. Y.; et al. *J. Med. Chem.* **1996**, *39*, 3514−3525.
(57) Huang, P. P.; Randolph, J. T.; Klein, L. L.; Vasavanonda, S.; Dekhtyar, T.; Stoll, V. S.; Kempf, D. J. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4075−4078.
(58) Lovell, S. C.; Davis, I. W.; Arendall, W. B., III; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins: Struct., Funct., Bioinf.* **2003**, *50*, 437−450.
(59) Case, D. A.; et al. *AMBER 6*; University of California San Francisco: San Francisco, CA, 1996.
(60) Jorgensen, W. L. *BOSS*, version 4.2; Yale University: New Haven, CT, 2000.
(61) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(62) Damm, K. L.; Carlson, H. A. *Biophys. J.* **2006**, *90*, 4558−4573.

ensemble contains 28 structures; hence, there must be 14 parents in close proximity to be a consensus cluster. Only probes found in energetically favorable regions, conserved throughout the ensemble, will remain as a consensus cluster. Thus, protein flexibility is implicitly accounted for by focusing chemical requirements on the rigid, unforgiving regions of the binding site and allowing chemical and steric flexibility in the mobile regions.

The consensus clusters were then represented as spherical pharmacophore elements. The center of each pharmacophore element was defined by the average position of the benzene centroid, the midpoint of the carbon−carbon bond for ethane, and the oxygen atom of the methanol probe. The radius was based on the rmsd of the probe positions. Overlapping benzene and ethane clusters were combined and termed aromatic/hydrophobic elements. Individual benzene elements were labeled aromatic, but extraneous ethane clusters were discarded. Methanol elements were classified as a hydrogen-bond donor, acceptor, or doneptor (donor and acceptor). Two excluded volumes were defined by the average position of the Cγ of each catalytic aspartic acid residue and used to represent the bottom of the active site. The radii of the excluded volumes were set to 1.5 Å, the approximate length of a Cγ−Oδ bond. A more detailed description of the MPS method can be found elsewhere.[20]

Pharmacophore Model Evaluation. The resulting pharmacophore models were screened against databases of compounds with pregenerated multiple conformers (maximum number of conformations was 300) using the search option within the Pharmacophore query editor of Molecular Operating Environment (MOE).[63] This is simply a fit/no-fit comparison based on the geometry of each conformer's chemical features and the physical arrangement of the pharmacophore elements. It is not a docking calculation based on scoring functions.

Three previously created databases of compounds were used. The first database consists of 89 diverse known HIV-1p inhibitors taken from the PDB and the literature. Two databases of noninhibitors from the Comprehensive Medicinal Chemistry Index[64,65] were used as decoys. The first noninhibitor database is comprised of 85 ligands[20] identified by filtering based on size and chemistry comparable to that of known protease inhibitors, whereas the second is more general and contains 2322 druglike ligands of very diverse sizes and chemical characteristics.[66] The full set created for use in our previous work contained 2324 compounds, but for this work it was appropriate to remove the two known HIV-1p inhibitors. The preparation and composition of these data sets has been described previously.[20,66] The stringency of the pharmacophore model was examined by varying the required number of pharmacophore elements that must be matched by enabling the partial match option in the Pharmacophore query editor of MOE and also by varying the radii of the elements.

The performance of the models was evaluated by comparing the percentage of identified known inhibitors (true positives) versus the percentage of druglike noninhibitors identified (false positives). The database screening results are presented as receiver operator characteristic (ROC) curves, where the optimal model would lie at the (0, 100) point predicting 100% of true positives and 0% of false positives. The models were also qualitatively compared back to the cu ligands.

Results and Discussion

Structural Comparison of the Protein Conformations. A sequence comparison was made of the 10 cu-crystal structure sequences and that of the NMR ensemble. The sequences differ at three amino acid positions: 3, 37, and 95. However, none of the mutations confer resistance or alter the biological activity of HIV-1p. The 28 NMR models and 10 cu-crystal structures were compared by aligning their Cα coordinates to their respective average structure using a Gaussian-weighted alignment.[62] The superposition of the 28 NMR models is shown in Figure 1A along with the bound cu ligands, and the 10 crystal structures and their unique cu ligands are given in Figure 1B. The majority of the variation between the NMR backbones is in the "elbows" of the flaps and in the "cheek" region, highlighted by arrows in Figure 1A, whereas the active site appears quite rigid. A detailed analysis of the NMR ensemble is provided by Yamazaki et al.[47] The backbones of the cu-crystal structures show much less deviation.

The rmsd was calculated between each wrmsd-aligned structure and its reference structure. For the NMR ensemble, the Cα rmsd ranges from 0.65 to 1.71 Å with the average being 0.92 Å. The cu-crystal collection had much lower rmsd values and also a smaller range, 0.26−0.80 Å and an average of 0.43 Å. (The rmsd values calculated from a wrmsd alignment are higher than those calculated from a standard rmsd alignment.[62] This is because a wrmsd alignment sacrifices the fit in the flexible regions to better align the rigid core.) The point here is not to compare the literal rmsd values per se but rather to evaluate the range of the values to illustrate the conformation variation between the NMR ensemble and cu-crystal collection. This analysis demonstrates that there is a greater variation between the Cα coordinates of the NMR ensemble than the cu-crystal collection. Other groups have also found that the variation between the active sites of different crystal structures is usually small, 0.3−0.8 Å rmsd.[50,67]

Across all NMR and crystal conformations, the cu ligands are in relatively the same conformation and position in the active site of the protease, the urea oxygen accepting a hydrogen bond from the protease flaps and the diols off the seven-membered ring donating hydrogen bonds to the 25/25′ aspartic acids. The position of the urea oxygen shows more spread across the 28 NMR structures than within the crystal structures. The side chains of the cu inhibitors occupy their complementary S1/S1′ and S2/S2′ substrate recognition sites. The cu ligands bound in crystal structures 1QBR,[55] 1QBT,[55] and 1QBU[55] have larger side chains and also hydrogen bond with the flap residue Gly 48/48′.

Pharmacophore Model Comparison. The NMR pharmacophore model maintains the *C*2 symmetry of the protease and has eight sites: two hydrogen-bond donor elements near the catalytic aspartic acid residues 25/25′, two aromatic/hydrophobic elements that anchor the hydrophobic regions near the active-site center, and four aromatic/hydrophobic elements that occupy the S1/S1′ and S2/S2′ pockets of the active site. The chemical characteristics of the NMR pharmacophore elements differ slightly from the chemical features of the bound cu ligand. The hydrogen-bond donor elements are slightly displaced from the location of the hydroxyl groups that extend below the seven-membered ring. The two interior aromatic/hydrophobic elements are located between the cu ligand side chains and represent the hydrophobic features of the central scaffold, whereas the four exterior aromatic/hydrophobic elements complement the chemical features of the ligand side chains. Figure 2A presents the MPS model based on the NMR structures in relation to the HIV-

(63) *Molecular Operating Environment*; Chemical Computing Group Inc.: Montreal, Canada, 2001.
(64) *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press: Oxford, 1990; Vols. 1−6.
(65) *Comprehensive Medicinal Chemistry Database*; MDL Information Systems, Inc.: San Leandro, CA, 2003.
(66) Bowman, A. L.; Lerner, M. G.; Carlson, H. A. *J. Am. Chem. Soc.* **2007**, *129*, 3634−3640.
(67) Erickson, J. W. *Perspect. Drug Discovery Des.* **1993**, *1*, 109−128.

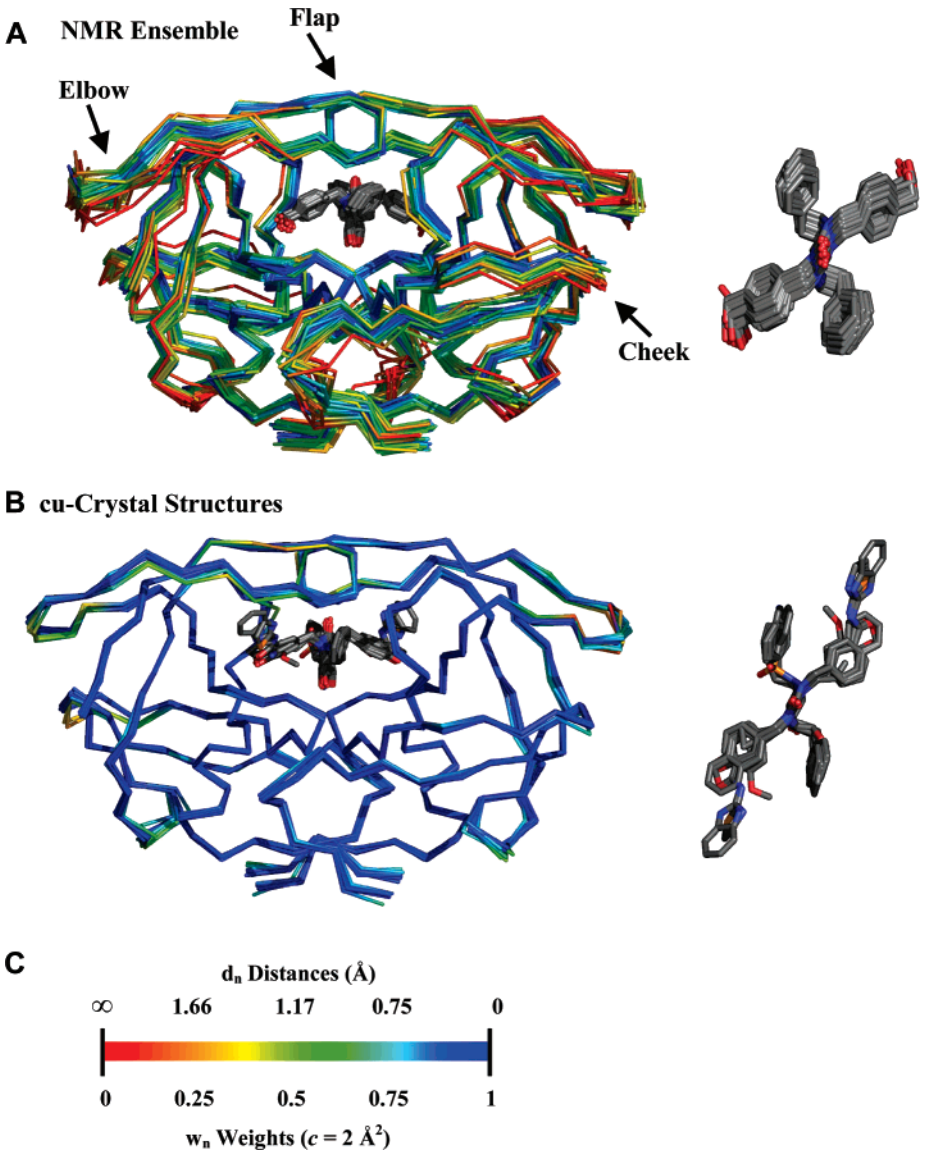


Figure 1. (A) Gaussian-weighted overlay of 28 models in the NMR ensemble along with all cu ligands (front view). The corresponding cu ligands are also shown using a top view for clarity. The regions of the protein with high backbone deviations are highlighted with an arrow. (B) Gaussian-weighted overlay of 10 crystal structures bound to unique cu ligands (front view). A top view of the 10 ligands is also shown. (C) The scale shows how smaller deviations (blue) are more heavily weighted in the wrmsd fit, $c = 2$ Å^2. Deviations over 2.45 Å have weights under 5% (red).

1p structure, and Figure 2B shows the model superimposed with the 28 cu ligands.

In order to assess how the models are improved through the use of MPS, we also generated a "static" model from a single structure. The pharmacophore element centers were defined in the same manner as previously described; however, the static model is based on the probes docked into one structure, rather than the parent probes across MPS. The pharmacophore model generated from the average NMR structure maintained the features of the model created from the NMR ensemble, but the radii are much smaller for most elements. This is most likely because without the consensus clustering step, the positions of the probes come from one structure only, even if it is an average. The spread between the probes docked within one structure is usually much smaller than the spread between the parent probes across many conformations. There are also two additional donor sites occupying the S3/S3′ subsites. Details of this static model are presented in the Supporting Information. Though the model is inferior to MPS models, it does show improvement over our earlier static model based on an apo crystal structure.[20]

The cu-crystal pharmacophore model also maintains the *C*2 symmetry of the protease and is shown in Figure 3A. However, it contains 11 elements: a hydrogen-bond acceptor element near the tips of the protease flaps, two hydrogen-bond donor elements near the catalytic aspartic acid residues 25/25′, two aromatic/hydrophobic elements near the core of the cu ligands, four aromatic/hydrophobic elements that occupy the S1/S1′ and S2/S2′ pockets of the active site, and two aromatic sites in the S3/S3′ pockets. The crystal model overlaid with the 10 unique cu ligands is given in Figure 3B.

The most interesting feature of this model is the hydrogen-bond acceptor element that perfectly overlays with urea oxygen of the 10 cu ligands. The urea oxygen is known to displace a structural water molecule that coordinates substrates/inhibitors to the tips of the protease flaps. The structural water is a key difference between mammalian and HIV proteases, and this displacement may be one reason why cu ligands are very

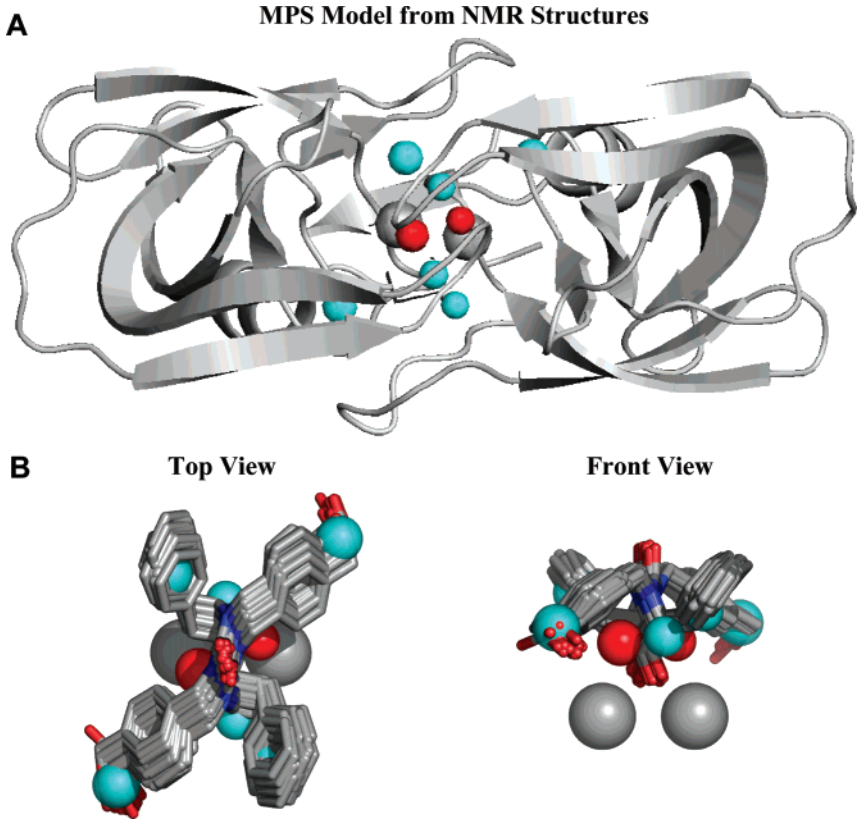


Figure 2. (A) Pharmacophore model (radii of 1 × rmsd) generated using 28 NMR structures. Elements are color-coded according to chemical functionality: red, hydrogen-bond donor; cyan, aromatic/hydrophobic. Top view of the protease backbone is shown in gray, as are the excluded volumes. (B) Pharmacophore model superimposed with 28 cu ligands colored in gray. Both top and front views are shown.

selective for HIV proteases.[55,68] Similar to the NMR models, the hydrogen-bond donor elements at the bottom of the pocket are slightly higher than the ligand diols. The four aromatic/hydrophobic elements complement the chemical features of some cu ligands but do not agree with others. The additional two aromatic sites at the S3/S3′ subsites fall at the edge of the aromatic rings in the cu ligands.

The most significant difference between the NMR and cu-crystal models is the hydrogen-bond acceptor element in the crystal model. This site was not occupied by probes in any of the 28 NMR structures or the average structure. Yamazaki et al. state that the flap tips (residues 48−51) are dynamic in solution and exhibit motion on a nanosecond time scale, whereas in crystal structures the flap tips are well-ordered.[47] The conformational variation across an NMR ensemble can be due to two things: proteins dynamics or an under-resolved structure from lack of experimental data. The HIV-1p NMR structure solved by Yamazaki et al. is regarded as a high-quality ensemble, and hence, the variation is thought to be from the dynamics of the structure.

There are also two additional aromatic sites in the cu-crystal model that are not found in the NMR model. These sites are located in the S3/S3′ subsites located at the solvent interface and are known to have broad substrate specificity.[68] The NMR and cu-crystal models are compared to the substrate recognition motifs of the HIV-1p active site in Figure 4, parts A and B, respectively. In the NMR structures, the arginine 8/8′ side chains are pushed out from the active site in variable locations. For this reason, there was more spread in the probes across the multiple conformations. The high flexibility of the arginine 8/8′ side chains that is seen in the NMR structures was also observed in the conformations sampled by MD simulations used to create previous pharmacophore models.[20,21] In both the NMR and cu-crystal structures, additional hydrogen-bond doneptor probes were observed between residues arginine 8/8′ and aspartic acid 29/29′ in the S3/S3′ pockets. However, these doneptor sites fall outside the 9.5 Å cutoff; consequently, they were not included in the pharmacophore models. A few of the larger inhibitors seen in the collection of 90 crystal structures have features that hydrogen bond to arginine 8/8′ or aspartic acid 29/29′. Nonetheless, there are many smaller ligands that maintain an extremely high potency (nanomolar to picomolar) without complementing this region, so it is appropriate that these sites were not included as an essential feature. Accordingly, ligands with the hydrogen-bonding feature will be accepted by the model, but it will not be required for identification as a potential inhibitor of HIV-1p.

Eight elements of the cu-crystal model were common to the NMR model: two hydrogen-bond donor elements and six aromatic/hydrophobic sites. The position and radii of these eight elements are very similar with the exception that the radii of the hydrogen-bond donor elements are slightly smaller for the cu-crystal model. The location and chemical character of the 8-site NMR model is highly consistent with pharmacophore models generated from MD simulations of apo HIV-1p (apo-MD model). A representative apo-MD model based on data from our previous work[21] is provided in Figure 5A. However, in the apo-MD model, several pharmacophore elements were aromatic,

(68) Wlodawer, A.; Erickson, J. W. *Annu. Rev. Biochem.* **1993**, *62*, 543−585.

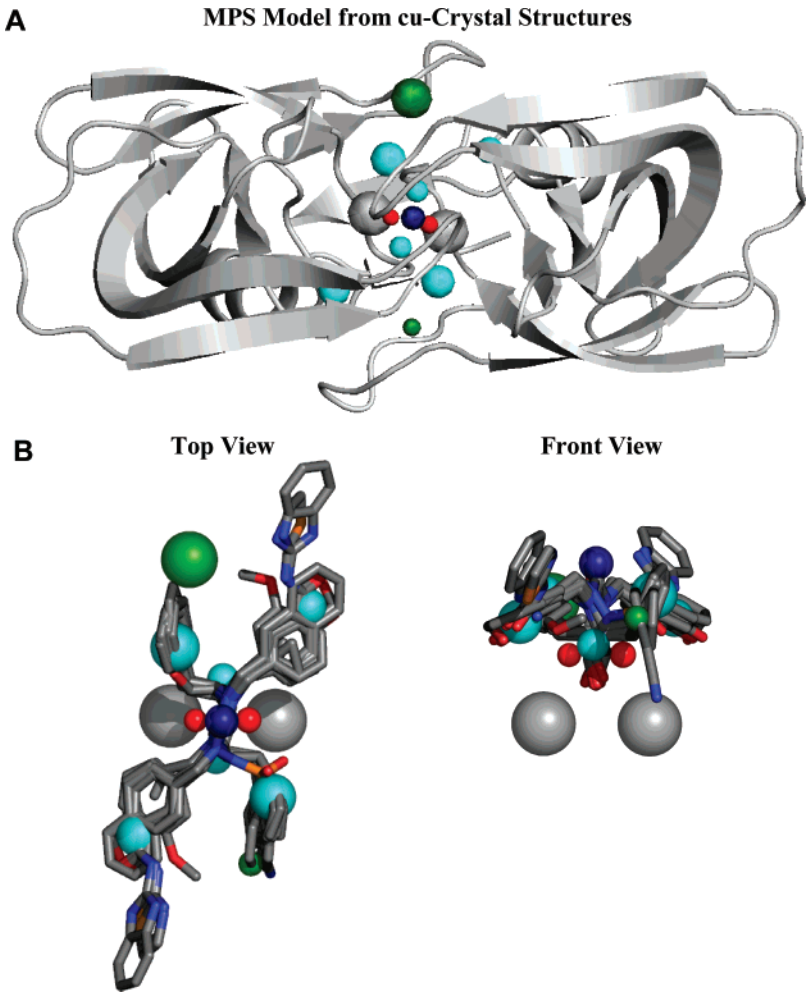


Figure 3. (A) Pharmacophore model (radii of 1 × rmsd) generated using 10 cu-crystal structures. Elements are color-coded according to chemical functionality: red, hydrogen-bond donor; blue, hydrogen-bond acceptor; cyan, aromatic/hydrophobic; green, aromatic. Top view of the protease backbone is shown in gray and so are the excluded volumes. (B) Pharmacophore model superimposed with 10 unique cu ligands colored in gray. Both top and front views are shown.

but all of the similar elements in the NMR model are aromatic/hydrophobic. The elements of the apo-MD model are also spread further apart due to the larger active-site cavity in the semiopen conformation than bound form. The cu-crystal model clearly differs from the apo model by the additional hydrogen-bond acceptor and two aromatic sites.

Additionally, we observed that the range of Cα rmsd for the apo-MD ensemble (11 structures) is similar to that of the NMR (28 structures): 0.94−1.50 versus 0.65−1.71 Å, respectively. The Gaussian-weighted superposition of the 11 structures from the apo-MD ensemble is shown in Figure 5B. The overlay of the MD ensemble clearly displays more movement in the bottom of the active site and, as one would expect due to the apo conformation, in the flap region than the NMR ensemble. It is interesting to find better agreement between bound HIV-1p conformations from NMR and the apo HIV-1p conformations from MD, rather than agreement to other bound conformations from X-ray crystallography.

Evaluation of Pharmacophore Models. The NMR and cu-crystal models were screened against a database of known HIV-1p inhibitors and two decoy data sets using the search option within the Pharmacophore query editor in MOE, while varying the stringency of the search (i.e., enabling a partial match in the pharmacophore search). Each ligand is described as a set of "annotation points" based on its chemistry and position in space. The ligand annotation points are then mapped to the pharmacophore elements and identified as a "hit" only if each of the required elements is satisfied. Hence, this is a binary fit or no-fit method, where all identified ligands are considered compatible with the pharmacophore model. The resulting data is presented as ROC curves. The best models identify the greatest number of true positives and the least number of false positives; consequently the optimal pharmacophore model is defined by having the smallest distance from (0, 100). The raw data used to generate the ROC curves of the NMR and cu-crystal pharmacophore screens is available as Supporting Information.

The performance of the pharmacophore models at discriminating known inhibitors versus a database of noninhibitors with similar size and chemistry is shown in Figure 6, parts A and B. The MPS models from NMR and cu crystals were both very successful at differentiating between the two populations. The optimal NMR model (7/8 sites, 2 × rmsd) identifies 89.9% of the true positives and only 10.6% of the false positives; the point on the ROC curve is highlighted in Figure 6A. The optimal cu-crystal model (9/11 sites, 3 × rmsd) also identifies the same number of true positives, 89.9%, but hits less false positive than the NMR model, 7.1% (point highlighted in Figure 6B).

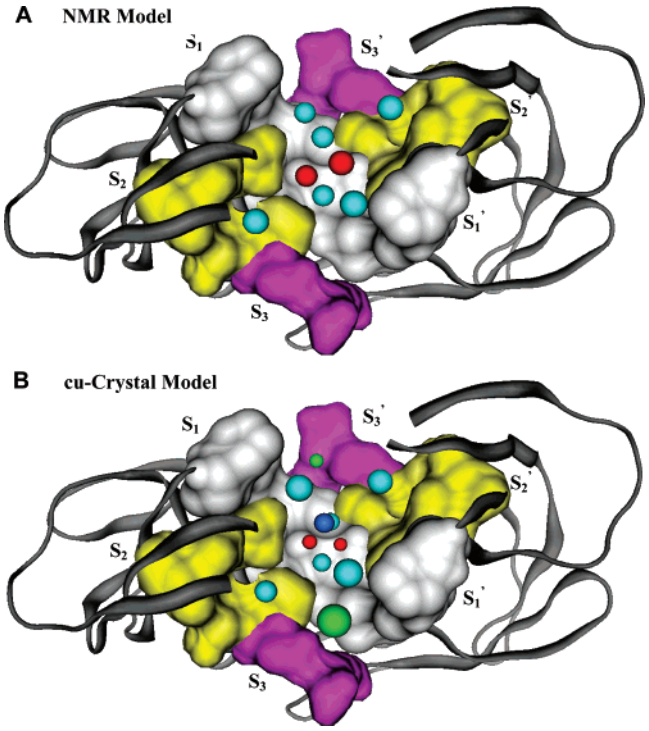


Figure 4. Comparison of known HIV-1p substrate recognition pockets with MPS pharmacophore models (radii of 1 × rmsd): white, S1/S1′ pocket; yellow, S2/S2′ pocket; purple, S3/S3′ pocket. Elements are color-coded according to chemical functionality: red, hydrogen-bond donor; blue, hydrogen-bond acceptor; cyan, aromatic/hydrophobic; green, aromatic. Flap residues 46/46′−54/54′ are removed for clarity. (A) NMR model. (B) cu-crystal structure model.

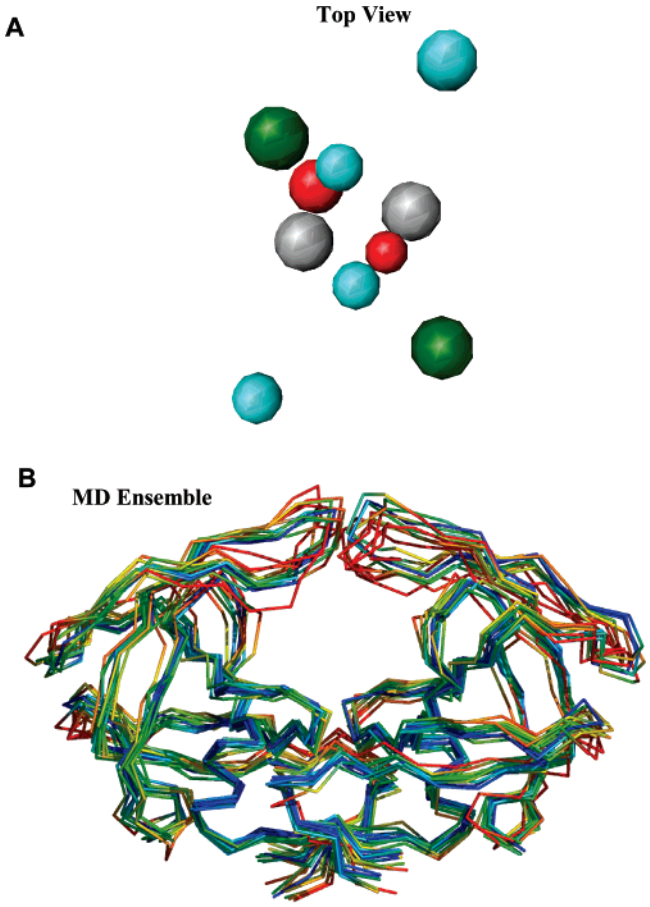


Figure 5. (A) Top view of an MPS pharmacophore model (radii of 1 × rmsd) created using 11 structures generated from a 3 ns MD simulation of apo HIV-1p (ref 21). Elements are color-coded according to chemical functionality: red, hydrogen-bond donor; cyan, aromatic/hydrophobic; green, aromatic. Excluded volumes are shown in gray. (B) Gaussian-weighted overlay of the 11 snapshots (front view). The color code of the weights is the same as in Figure 1C, and the view is comparable to Figure 1, parts A and B.

However, these results may be misleading. The 11-site cu-crystal model shows the best performance when 9 of the 11 sites are required. This demonstrates that the 11 sites are too specific; less essential features of the active site were selected out from using multiple cu-crystal structures, unlike with the NMR ensemble. If the three "extra" sites unique to the cu-crystal model are dropped, the performance is nearly identical to the optimal 11-site model. The best, "core 8-site" cu-crystal model (8/8 sites, 3 × rmsd) identifies 88.8% of the true positives and only 10.6% of the false positives. The extra sites do not significantly improve the performance of the model, which indicates that the hits from the cu-crystal model are really using the elements in common with the NMR model and apo-MD model. Extraneous sites which do not improve the performance of the models are problematic for database screening and undesirable for the MPS technique.

The optimal NMR performance is comparable to the optimal 11-site cu-crystal model but with fewer sites. Additionally, the 8/8 site, 3 × rmsd, NMR model performed quite similarly to the optimal NMR model (7/8 sites, 2 × rmsd); the number of true positives identified remained the same, while only identifying two additional false compounds. Therefore, all of the sites in the NMR model appear to encode useful information. The 11/11 cu-crystal models demonstrate mediocre performance; the best model (4 × rmsd) has a larger false positive hit rate and identifies only 65.2% of the true positives. The reduced amount of conformational sampling of the protein had to be overcome by significantly increasing the scaling factor for radii.

We use multiple structures to determine the most essential features that are conserved across different receptor conformations. Overall, the NMR model is more general, and the features do not simply reproduce the chemical characteristics of the bound cu ligand. In a recent study by our group using a different protein target, dihydrofolate reductase (DHFR), crystal structures were also employed as MPS and very minor conformational changes were observed between the collections.[66] The minimal conformational variation between the structures resulted in relatively small radii of the elements. Hence, the radii of the pharmacophore models had to be multiplied by 4 or 5 × rmsd for optimal performance.

We anticipated that the use of a more general model will be beneficial when searching large databases for novel compounds from new chemical space. We compared the performance of the pharmacophore models at discriminating known HIV-1p inhibitors from a large, general data set of 2322 decoy compounds. Again, both the NMR and crystal models display excellent performance at selecting out the known inhibitors, Figure 6, parts C and D, respectively. The NMR model again performs very well when 7/8 or 8/8 sites are required. Both 7/8, 2 × rmsd and 8/8, 3 × rmsd identified 89.9% of the true positives and only 2.8% and 4.1% of the false positives, respectively; the points on the ROC curve are highlighted in Figure 6C. Once more, the optimal cu-crystal model required 9/11 sites (2.7 × rmsd) to perform similarly to the optimal NMR

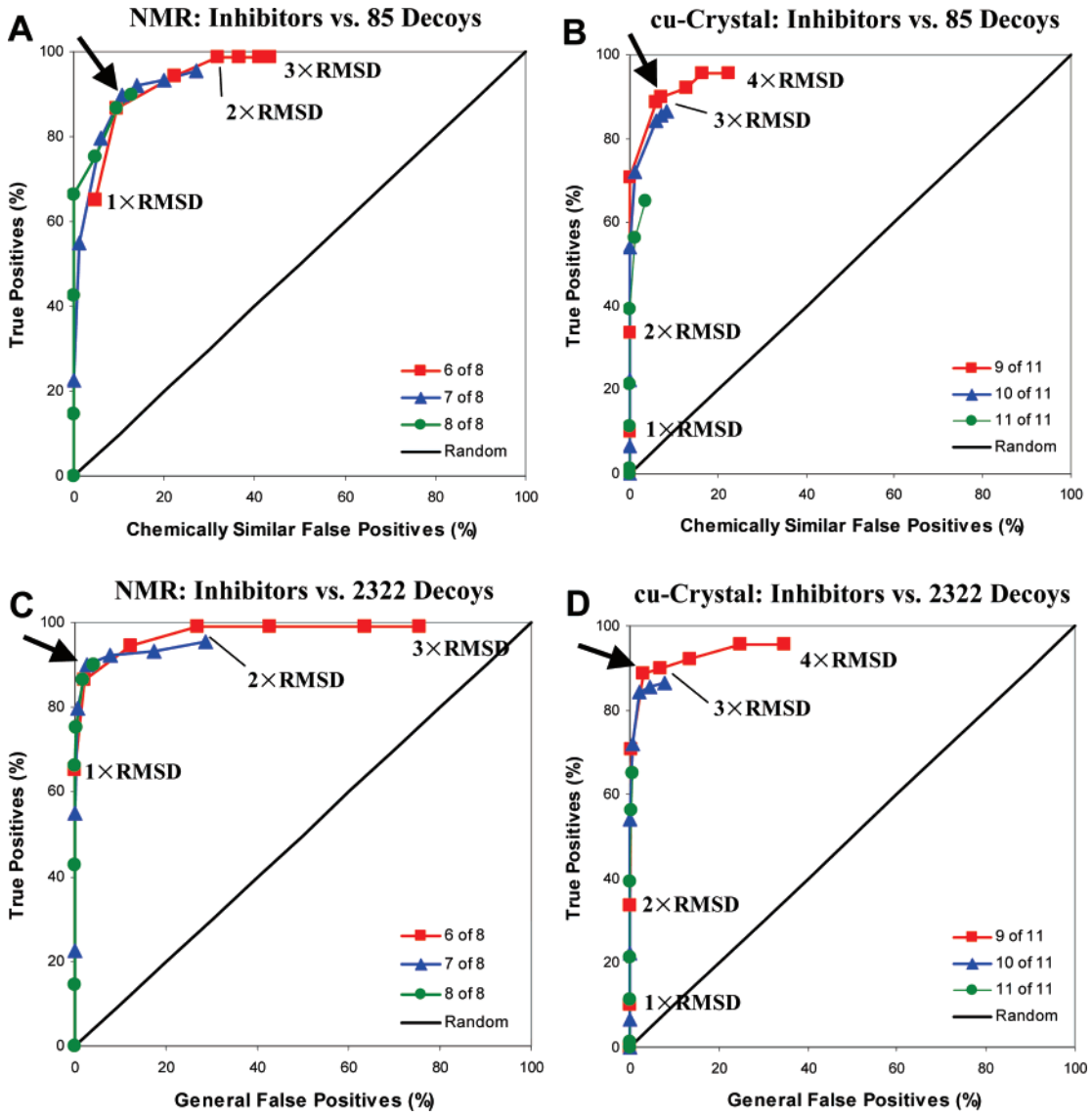


Figure 6. Receiver operator characteristic (ROC) curves generated from screening a database of 89 known HIV-1p inhibitors against a set of 85 chemically similar known inactives and 2322 general decoy compounds. Each series represents a different stringency in the screen (i.e., 6 of 8 elements are required as a hit, 7 of 8 elements are required as a hit, etc.) Points in the series are increasing radii values from 1 to 3 × rmsd for the NMR model and 1−4 × rmsd for the cu-crystal model. The radii are labeled on the 6 of 8 models based on NMR and the 9 of 11 models based on cu crystals. The optimal pharmacophore models are highlighted by an arrow. (A) MPS NMR pharmacophore models, 89 known inhibitors vs 85 decoy compounds (optimal: 7/8, 2.0 × rmsd). (B) MPS cu-crystal pharmacophore models, 89 known inhibitors vs 85 decoy compounds (optimal: 9/11, 3.0 × rmsd). (C) MPS NMR pharmacophore models, 89 known inhibitors vs 2322 general molecules (optimal: 7/8, 2.0 × rmsd). (D) MPS cu-crystal pharmacophore models, 89 known inhibitors vs 2322 general molecules (optimal: 9/11, 2.7 × rmsd).

model, identifying 88.8% of the true positives and 3% of the false positives (point highlighted in Figure 6D). However, only the NMR model was able to identify almost 100% of the true positives. The presence of extraneous sites may explain why the cu-crystal models miss identifying several of the known inhibitors even with the most generous criteria.

Several aspects of the model's performance confirm patterns we observed with the apo-MD models in our previous studies.[20,21] First, all sets of ROC curves are very steep at the beginning indicating the potential for models with smaller radii to be used in virtual-screening applications. When screening large databases of compounds, the number of true positives can be sacrificed to reduce the amount of false positives. Second, larger radii are needed when more elements are required. Third, among the false positives identified by the pharmacophore models are renin inhibitors, transition-state mimics of peptide cleavage, and small hydrophobic signaling peptides. This is not surprising since renin is a homologous aspartic protease, the function of HIV-1p is to cleave peptides, and the substrates are hydrophobic regions of proteins. Similar classes were identified from both decoy databases, but as one would expect due to the size of the general database (2322), additional classes were also seen. This list includes macrocyclics (another HIV-1p inhibitor class), β-lactams, tetracenes, and other polycyclic systems. However, for brevity we are providing only the structures of the identified false positives from the database of 85 chemically similar compounds in the Supporting Information.

Effect of the Structure Number in Ensemble. Only 10 structures were used to generate the cu-crystal model, but the NMR ensemble contains 28 conformations. We were concerned that the larger number of structures in the NMR ensemble may bias the model for better performance. To ensure a fair

comparison between the NMR and crystal structures, we also generated an additional MPS pharmacophore model from 90 crystal structures (all-crystal model). The 90 structures are bound to a variety of ligand classes. Once again, there is little backbone variation between the 90 structures; the Cα rmsd values range from 0.12 to 0.71 Å. An overlay of the structures using the Cα coordinates is provided in Supporting Information. Zoete et al. also found minimal variation between 73 HIV-1p backbones bound to different ligands.[69] Moreover, we observed that adding 35 structures with resistant mutations did not provide any additional conformational variation (125 structures total, data not shown).

We also calculated the rmsd for each heavy atom of the protein active site (defined as any atom within 10 Å of the active-site center) using 1PRO as the reference structure. This was chosen because 1PRO is the representative structure for the HIV-1p family in the Binding MOAD database[49] as it has the tightest-bound inhibitor. Moreover, it is also bound to a cu ligand, making it an appropriate choice for comparing the cu-crystal structures to the larger set of 90 structures. The rmsd values ranged from 0.16 to 1.80 Å, with an average of 0.51 ± 0.37 Å.

The small conformational variation between the crystal structures does not appear to be an effect of crystal structure refinement. It is common practice to use a previously solved crystal structure when determining the coordinates of another. However, our inspection of the electron density maps showed the structures to be of high quality with well-resolved density defining the coordinates. Crystal packing effects are known to be important for conformations of HIV-1p, but there is no evidence to suggest that is the cause of the limited sampling. Most likely, the resulting conformations are influenced by a variety of factors including the conditions used in X-ray crystallography such as temperature and pH. For example, low temperatures are typical for growing crystals and may not provide enough thermal energy for a protein to overcome the barrier to sample conformations outside of a particular local minimum. In the case of HIV-1p, apo crystal structures are found in the semiopen conformation, whereas bound structures exist in the closed state. A 1.3 Å apo crystal structure of a highly mutated HIV-1p strain was recently solved in a novel open conformation (PDB ID: 1TW7),[70] but the open state was later shown to be caused by crystal packing effects.[71]

Cross-docking studies in the literature demonstrate how different HIV-1p structures perform poorly when trying to dock ligands from other crystal structures.[72,73] However, there are also cross-docking examples where HIV-1p performs quite well.[74] Furthermore, there are examples in the literature where HIV-1p is able to reproduce docking poses of its own ligands successfully[72] and also unsuccessfully.[73] We propose that the difficulties in those studies arise from the ligands of HIV-1p, not the structures of the proteins. The majority of the bound ligands are large, flexible peptides. It is well-known that many of the docking programs have difficulty with ligands that have many rotatable bonds. The different studies in the literature used different routines for sampling ligands, and this could be the real source of poor cross-docking results. This argument supports our structural analysis of HIV-1p; there is very little variation between the crystal structures.

The resulting all-crystal pharmacophore model is very similar to the cu-crystal model. The only exception is two elements that are aromatic in the all-crystal model, rather than aromatic/hydrophobic; the model is provided in the Supporting Information. The inclusion of more structures appears to cause the two elements to become less general. The sphere centers and radii are nearly identical between the cu-crystal and all-crystal models, apart from the aromatic sites flanking the solvent-exposed region of the binding site. In the all-crystal model, they better replicate the *C*2 symmetry of the protein. The model performance does change slightly; the optimal model now requires even more elements to be dropped: 8 out of 11 elements (8/11 sites, 2.7 × rmsd). Furthermore, it identifies less of the true positives (86.5% compared to 89.9%) and more of the false positives (14.1% compared to 7.1%). The raw data from the pharmacophore screen is available in the Supporting Information. Again, as in the case of the average NMR model, the loss of the hydrophobic character of the aromatic elements in the S1/S1′ pocket does seem to negatively affect the performance of the model; it appears that an aromatic/hydrophobic element truly provides a more accurate representation of the active-site pockets.

As previously mentioned, the range of Cα rmsd values for the apo-MD ensemble (11 structures) is comparable to that of the NMR (28 structures), 0.94−1.50 versus 0.65−1.71 Å, respectively, and that the MPS models are nearly identical. However, the range of Cα rmsd values for both the cu-crystal (10 structures) and all-crystal (90 structures) collections is much smaller, 0.26−0.80 and 0.12−0.71 Å, respectively. Both models based on crystal structures have three additional sites. We strive to generate pharmacophore models from an ensemble that represents an appropriate sampling of conformational space. It appears that both NMR and MD ensembles can account for more accessible conformations than bound protein−ligand crystals structures, even those bound to a set of diverse ligands. We stress that crystal structures are very useful in many other SBDD applications, but we believe that bound HIV-1p crystal structures do not provide a complete sampling of receptor conformations and NMR models can have definite advantages when trying to represent the protein's flexibility.

Conclusion

Incorporating protein flexibility into SBDD is necessary to simulate a more accurate representation of a protein in solution. By looking for favorable interaction regions across multiple conformations of a protein, we can determine the most essential and conserved features of the active site. We are able to show that the MPS method can be extended to include the use of experimental structures as a source of multiple conformations. The use of experimentally determined structures is attractive over generating conformations from an MD simulation in order to reduce the amount of time required to develop an MPS model. Additionally, to our knowledge this is the first direct comparison

(69) Zoete, V.; Michielin, O.; Karplus, M. *J. Mol. Biol.* **2002**, *315*, 21−52.
(70) Martin, P.; Vickrey, J. F.; Proteasa, G.; Jimenez, Y. L.; Wawrzak, Z.; Winters, M. A.; Merigan, T. C.; Kovari, L. C. *Structure* **2005**, *13*, 1887−1895.
(71) Layten, M.; Hornak, V.; Simmerling, C. *J. Am. Chem. Soc.* **2006**, *128*, 13360−13361.
(72) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., III. *J. Med. Chem.* **2004**, *47*, 3032−3047.
(73) Perola, E.; Walters, W. P.; Charifson, P. S. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 235−249.
(74) Joseph-McCarthy, D.; Alvarez, J. C. *Proteins: Struct., Funct., Bioinf.* **2003**, *51*, 189−202.

of NMR ensembles and crystal collections for incorporating receptor flexibility in SBDD.

The MPS pharmacophore models generated from an NMR ensemble and collections of crystal structures were able to discriminate known HIV-1p inhibitors from druglike decoys and showed better performance than a model previously created using apo HIV-1p structures. They also showed superior performance over a model created from the average NMR structure. The average NMR model contained additional elements and lost important chemical characteristics that appeared to diminish the performance of the model, but the use of MPS identified the most important, chemically relevant features. The use of an average structure from multiple receptor conformations is an alternate method that has been proposed for incorporating protein flexibility in SBDD, but we find that ensembles of structures is a superior approach.

The present results are strong support for the use of NMR ensembles in SBDD. *The NMR model revealed only the most essential features of the binding site.* Instead, the collection of crystal structures identified three additional, and less essential, elements. These were highly related to chemical features specific to the class of cu ligands. In order to achieve a reasonable performance, additional elements had to be dropped or the radii had to be multiplied by large scaling factors. The NMR model did not simply reproduce its bound ligand. It could be used in its entirety (8/8 sites, 3 $\times$ rmsd) with exceptional performance for discriminating true inhibitors from decoy molecules. The performance improved slightly with 7/8 sites, 2 $\times$ rmsd models, which is in good agreement with the parameters previously suggested for MPS based on MD (generally, $n - 1$ of n features and radii of $\sim$2 $\times$ rmsd). Furthermore, the NMR ensemble samples a greater amount of conformational space than the crystal collection and is comparable to the amount of sampling seen in a 3 ns MD simulation of apo HIV-1p.[21]

Overall, we recommend NMR structures over crystal structures for incorporating protein flexibility into SBDD studies of HIV-1p. By no means are the crystal structures inaccurate; instead, there is simply too little variation between the different structures, even when bound to a variety of ligand classes. In fact, this analysis strongly suggests that the difficulties seen in cross-docking studies of HIV-1p do not arise from the protein structures themselves. Most likely, the difficulty comes from the ligands which inhibit HIV-1p. Many routines employed to generate ligand conformations have difficultly with large, flexible compounds, and this could be the cause of the inconsistencies in the cross-docking results.[72−74]

Acknowledgment. This work has been supported by the National Institutes of Health (GM65372), a Beckman Young Investigator Award, an NSF CAREER Award (MCB 0546073), and the Pharmacological Sciences Training Program (GM07767 NIGMS). K.L.D. is also grateful for receiving a Rackham Predoctoral Fellowship and a fellowship from the American Foundation for Pharmaceutical Education. We thank OpenEye for their generous donation of the OMEGA[75] software that was used to generate conformations of the large test set of druglike compounds (described in ref 67) and the Chemical Computing Group for generously providing MOE.[63] The authors would also like to thank K. Meagher and O. Tsodikov for insightful discussions, M. Lerner for the python scripts used in automating the MPS method, and A. Bailey for maintaining the linux clusters. PyMOL[76] was used for various visualization purposes and the creation of figures for this paper.

Supporting Information Available: Complete refs 52, 54, 56, and 59, 10 cu-crystal structures and their inhibition constants, 90 crystal structure PDB IDs with the complete citations, coordinates for the NMR, average NMR, cu-crystal, and all-crystal model pharmacophore models, figures of the average NMR structure pharmacophore model, static crystal pharmacophore model,[20] all-crystal pharmacophore model, raw data from the pharmacophore screens, Gaussian-weighted overlay of 90 crystal structures, and structures of the false positive hits from a chemically similar database of 85 compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0709728

(75) *OMEGA*; OpenEye Science Software: Santa Fe, NM, 2001.
(76) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific LLC: San Carlos, CA, 2002; http://www.pymol.org.