Refining the Multiple Protein Structure Pharmacophore Method: Consistency across Three Independent HIV-1 Protease Models

Kristin L. Meagher,[†] Michael G. Lerner,[‡] and Heather A. Carlson*[,†,‡]

Department of Medicinal Chemistry, College of Pharmacy, 428 Church St., University of Michigan, Ann Arbor, Michigan 48109-1065, and Biophysics Research Division, 930 North University, University of Michigan, Ann Arbor, Michigan 48109-1055

Received August 1, 2005

Developing methods to incorporate protein flexibility into structure-based drug design is an important challenge. Our approach uses multiple protein structures (MPS) to create a receptor-based pharmacophore model of the desired target. We have previously demonstrated the success of the method by applying it to human immunodeficiency virus-1 protease (HIV-1p). Our models, based on an apo structure, discriminated known HIV-1p inhibitors from druglike inactive compounds and also accurately identified bound conformations of known inhibitors. Here, we test the method by applying it to all three unbound crystal structures of HIV-1p. We have also improved our method with denser probe mapping of the binding site and refined our selection criteria for pharmacophore elements. Our improved protocol has led to the development of a consistent 8-site pharmacophore model for HIV-1p, which is independent of starting structure, and a robust MPS pharmacophore method that is more amenable to automation.

Introduction

Virtual screening and structure-based drug design (SBDD) have become important tools in the pharmaceutical industry.[1,2] An important challenge in SBDD is the incorporation of both ligand and receptor flexibility. Protein targets are dynamic and often cannot be adequately represented by a single, static conformation such as a crystal structure. Instead, an ensemble of receptor conformations representing the accessible conformations in the solution phase may be more useful for SBDD. This is an active area of research, and many approaches have been taken to solve this difficult problem.[3−6]

Previously, Carlson et al. introduced a novel method using multiple protein structures (MPS) to create receptor-based pharmacophore models which accounted for the protein flexibility of human immunodeficiency virus-1 (HIV-1) integrase.[7,8] Small molecule probes were used to map complementary chemical functionalities onto the HIV-1 integrase active site. Areas with a consensus of probe molecules over several protein structures were represented as pharmacophore elements, a "dynamic" pharmacophore model. We have further developed this method by creating pharmacophore models of the unbound HIV-1 protease (HIV-1p).[9] Using MPS from a molecular dynamics (MD) simulation, we identified pharmacophore models which successfully discriminated known HIV-1p inhibitors from druglike noninhibitors. In addition, our pharmacophore models were able to accurately predict binding modes of known inhibitors, starting with an unbound structure. It is rare to have such success with an apo structure.

Other groups have also applied similar techniques that map probes onto MPS. Briggs and co-workers have explored improved techniques for mapping binding sites in alanine racemase[10] and HIV-1 integrase.[11] Schechner and co-workers have used a consensus probe approach with CHARMM to map the ATP binding site of DNA gyrase B.[12] As methods based on MPS gain broader use and exposure for drug discovery,[7−12] it is important to demonstrate the consistency and reproducibility of such techniques. Therefore, we have applied the MPS pharmacophore method to the other two unbound HIV-1p structures available in the Protein Data Bank[13] (PDB). We demonstrate that the MPS pharmacophore method gives consistent, high-performing pharmacophore models using any of the unbound HIV-1p structures available in the PDB. We have also developed a method for placing the small molecule probes more densely into the target active site.[14] Here, we have also modified the parameters to incorporate this improved flooding technique into the MPS pharmacophore method.

Computational Methods

The unliganded HIV-1p structures were obtained from the PDB (accession codes 1HHP,[15] 3HVP,[16] and 3PHV[17]). Symmetry operations were used to generate the homo dimer for each system. The simulation details have been previously described for 1HHP and the same procedure was used for 3HVP and 3PHV.[9,18] Both catalytic aspartate residues were deprotonated in the calculations. The AMBER6[19] suite of programs was used along with the AMBER94 force field.[20] Each of the proteins was solvated in a TIP3P[21] solvent box and gradually heated from 50 to 298 K over 50 ps. Equilibration was performed for 200 ps with the protein restrained to allow the water to optimally complement the protein, followed by an additional 200 ps of all-atom equilibration. The sampling phase of the MD simulations was 3 ns.

Conformational snapshots were taken after the final equilibration and every 100 ps throughout the simulations. Each snapshot of the protein was flooded with 500 small molecule probes using an optimized probe-placement method.[14] The probes were placed into an 18-Å sphere encompassing the entire active site. MUSIC[8] calculations using the BOSS[22] program were used to minimize the small molecule probes to the protein surface. Probe−probe interactions were ignored, allowing the probes to cluster against the protein surface and reveal favorable interactions within the active site. Benzene, ethane, and methanol were used as probes to map aromatic, hydrophobic, and hydrogen-bond interactions, respectively.

For each protein snapshot, clusters of probes within 10 Å of the catalytic aspartate residues were analyzed. A cluster was defined as containing at least 8 closely packed probes for benzene and ethane and 12 probes for methanol. Each cluster is represented by

* To whom correspondence should be addressed at the College of Pharmacy. Phone: 734-615-6841. Fax: 734-763-2022. E-mail: carlsonh@umich.edu.

[†] College of Pharmacy.

[‡] Biophysics Research Division.

10.1021/jm050755m CCC: $33.50
Published on Web 05/18/2006

a single "parent" probe which is defined as the probe in the cluster with the lowest interaction energy with the protein. Protein snapshots along with the parent probes were overlaid by Cα coordinates, using a standard root mean square deviation (rmsd) alignment, and consensus clusters determined. A total of 11 snapshots were used for all models (1, 2, and 3 ns); the snapshots were taken every 100, 200, or 300 ps, respectively. A consensus cluster was required to contain at least 6 parent probes with representation from the beginning, middle, and end of the simulation length.

The consensus clusters were represented as pharmacophore elements. Each element was centered at the average position of the key chemical functionality within the consensus cluster (centroids of benzenes, oxygens of methanols, and midpoints of the C−C bonds in ethanes). The radii of the pharmacophore elements were based on the rmsd spread of the parent probes in the consensus clusters as previously described.[9] Ethane clusters were used to clarify aromatic from hydrophobic interactions, and overlapping benzene and ethane elements were combined and relabeled as possessing either aromatic or hydrophobic character. Excluded volumes were centered at the average position of the Cγ of the two catalytic aspartate residues and simply assigned a radius of 1.5 Å. Pharmacophore elements within 9 Å of the excluded volumes were included in the final models.

Pharmacophore models were created using 1, 2, or 3 ns of the MD trajectory for each structure investigated (1HHP, 3HVP, 3PHV). In a manner analogous to our previous work, the stringency of the pharmacophore query was investigated and the radii of the pharmacophore elements were explored. The pharmacophore models were searched against two datasets (included in the Supporting Information): a set of known HIV-1p inhibitors (89 entries) and a set of druglike noninhibitors (85 entries) taken from ref 23. Each dataset contained pregenerated ligand conformers to account for ligand flexibility. The conformations were calculated using MOE[24] and the MMFF force field.[25] The creation and composition of these datasets has been described elsewhere.[9] The pharmacophore-searching applications in MOE[24] were used to screen the pharmacophores against the two ligand datasets.

Results and Discussion

Extending the Method to Similar Starting Structures. To demonstrate the consistency and reliability of the MPS pharmacophore method, we have created models of HIV-1p using different starting structures. We have chosen the three unliganded, complete HIV-1p structures available in the PDB (accession codes: 1HHP,[15] 3HVP,[16] and 3PHV[17]). The three structures were crystallized independently and have approximately 0.5 Å Cα rmsd. The 1HHP and 3PHV structures have identical sequences, but the 3HVP sequence differs in 5 residues, none of which are biologically significant or confer resistance. This is a simple and straightforward way to demonstrate the robust nature of the MPS pharmacophore method. We expect to identify similar pharmacophore models, despite the minor differences inherent to independently solved structures.

Using three starting structures and three independent MD trajectories also allows us to further generalize the selection parameters for determining parent and consensus clusters. New criteria were necessary to treat the clusters obtained with the denser probe packing as described below. Using three independent starting structures to determine these criteria ensures that the method gives consistent results—demonstrating the reproducibility and robust nature of the MPS pharmacophore method.

Improved Probe Placement. In the extension of the MPS pharmacophore method to additional systems, it was recognized that the original routine for placing probe molecules into the active-site cavity could be improved. Initial placement of the probes was too sparse in systems with smaller, more occluded binding sites. To overcome this limitation, a python script was created to work with the PyMOL[26] program and densely flood binding sites within a user-defined sphere.[14] The user can also define a cutoff to the protein that controls how close the probes are initially placed against the protein surface. Using this method, significantly more probes are placed in the active-site cavity, and both positions and orientations of the probes are more adequately sampled. With a higher density of probes, all accessible local minima within the active site are sampled. This more thorough mapping of the binding site helped us to identify additional consensus clusters for the MPS pharmacophore models. There are also more probes in each individual cluster, making the most important interactions within the active site easier to identify.

Refining Selection Parameters for the Elements in the Pharmacophore Models. Because a higher number of probes were initially placed in the active-site cavity and more probes were present in each individual cluster, more stringent clustering criteria were needed. In selecting clustering parameters, we did not fit model performance to a particular set of equations; rather, the optimal parameters were chosen through an iterative procedure of varying parameter choices and evaluating the effect on model performance. Various criteria for clustering probes were evaluated, including the number of probes required to determine a significant cluster, the number of probes required in each consensus cluster, the stringency of pharmacophore query, and the radius size of the pharmacophore elements.

When choosing which clusters were most significant in an individual snapshot, we found that the optimal cutoff was 8 or more benzene/ethane probes in a cluster and 12 or more probes for methanol. We investigated several values for this parameter including 5, 8, 10, and 12 probes required for a significant cluster. The difference in requirements for methanol and the hydrophobic probes can be attributed to the difference in their interactions energies with the protein. This is an increase in stringency compared to our previous application to HIV-1p where we required 5 copies of the probe to identify a cluster as significant.

More restrictive criteria were also used in determining the consensus clusters from the overlaid snapshots. Originally, consensus was defined as parent probes from at least 5 of the 11 snapshots making the same interaction. We varied this parameter requiring 5, 6, or 7 probes to identify a consensus cluster. Using the improved probe placement, we found that requiring at least 6 parents was optimal. This ensures that only the most important interactions are represented by the consensus clusters. The requirement for representation among the beginning, middle, and end of the MD simulation was also maintained, again ensuring we are identifying general interactions rather than localized ones.

Using these stricter criteria reduces the subjectivity inherent in the pharmacophore method. The studies presented here were done "by hand" by the same authors as in the previous study.[9] This allows for consistency in comparing back to our previous work. These more stringent "by hand" models are also important for guiding our automation efforts. The most significant clusters are more obvious due to the larger number of probes/cluster, allowing the user to rely less on chemical intuition. Our group is in the process of creating automated techniques for identifying consensus clusters.[14] Well-defined algorithms and mathematical descriptions will also reduce the subjectivity and ensure that two users will create the exact same MPS pharmacophore models.

Creation of Consistent Pharmacophore Models. Pharmacophore models were created for the three HIV-1p structures using 1, 2, or 3 ns of the MD trajectory. The models generally

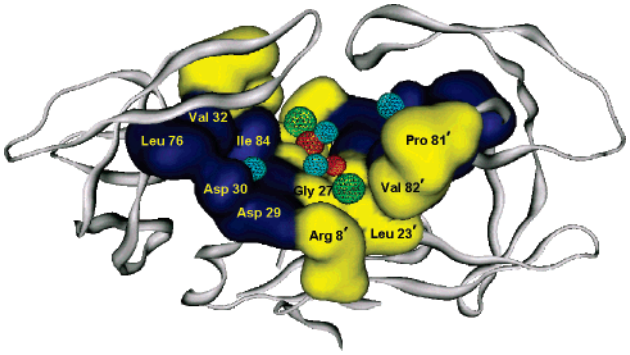

Figure 1. Comparison of known substrate recognition elements with the 1HHP-3ns pharmacophore model: yellow, residues making the S1, S1′ pockets; blue, residues making the S2, S2′ pockets. Key residues forming these pockets are labeled. Radii of the pharmacophore elements are shown as 1 × rmsd of the consensus probes, and the elements are colored according to chemical functionality: red, hydrogen-bond donating; green, aromatic; cyan, aromatic or hydrophobic.

have 8 sites; 6 of the sites are in excellent agreement with our earlier models: two hydrogen-bond donor sites near the catalytic aspartates and four aromatic/hydrophobic sites around the periphery of the active site. The new 8-site models contain two additional aromatic/hydrophobic sites near the center of the active site. These two interior aromatic/hydrophobic sites were not identified as consensus clusters in our earlier study of HIV-1p, but these interior elements had been observed in individual snapshots, just not with enough consistency to warrant inclusion in the consensus pharmacophore models. By improving the probe mapping within the active site, we have more thoroughly defined the interaction surface between the protein and small molecule probes—resulting in the inclusion of additional pharmacophore elements. The 8-site model is an enhancement of our previous 6-site model, and the pharmacophore elements are an excellent match with known substrate recognition motifs of HIV-1p (Figure 1).[27] Specifically, the pharmacophore elements in the northwest and southeast quadrants interact with protein residues Arg 8, 8′, Leu 23, 23′, and Val 82, 82′ corresponding to the S1 and S1′ recognition pockets. Residues Asp 29, 29′, Asp 30, 30′, and Ala 28, 28′ are part of the S2, S2′ pockets and contact the northeast and southwest pharmacophore elements. The interior, hydrogen-bonding elements of the pharmacophore interact with the two catalytic aspartate residues (25, 25′), and the hydrophobic elements contact Asp 25, 25′, Gly 27, 27′, and Ala 28, 28′.

Figure 2 presents the MPS pharmacophore models based on 1, 2, and 3 ns of MD simulation from 1HHP, 3PHV, and 3HVP. Coordinates and radii of the pharmacophore elements in all models are provided as Supporting Information. A consistent pharmacophore model is observed across the different simulation lengths and also across the three independent starting positions. There are minor differences among the models, but the overall trend reveals a consistent, symmetric 8-site pharmacophore model. Small variations were expected in the location and radii of elements and the inclusion of hydrophobic character to an aromatic site (i.e., green versus cyan elements in Figure 2). More significant differences typically arise when one of the 8 sites does not contain at least 6 parents, and so that site is not represented in the final consensus model. On occasion, additional elements were identified because more weakly occupied local minima happened to provide parent probes from 6 snapshots. These extra sites are generally seen in models with less conformational sampling like the 1HHP-1ns and 3PHV-2ns models. We were surprised to see that the 3HVP-3ns model contained an extra site, particularly when the 3HVP-1ns and 3HVP-2ns models contained only the common 8-site pharmacophore elements. However, the discussion below shows how testing the models against the datasets of known inhibitors and noninhibitors revealed that the inclusion of an extra site (or the absence of one of the 8 common elements) did not significantly degrade their selectivity and performance.

Due to the limited sampling inherent to MD, our models are influenced by some specific side-chain conformations sampled. The most variable pharmacophore sites are the northwest and southeast sites (Figure 2). These sites are flanked by solvent-exposed arginine residues (Arg 8, 8′). A high degree of flexibility is expected for a solvent-exposed, charged side chain, and this arginine flips freely during our simulations, pointing out into solution or into the active-site cavity. Furthermore, crystal structures of HIV-1p show that Arg 8 and 8′ adopt

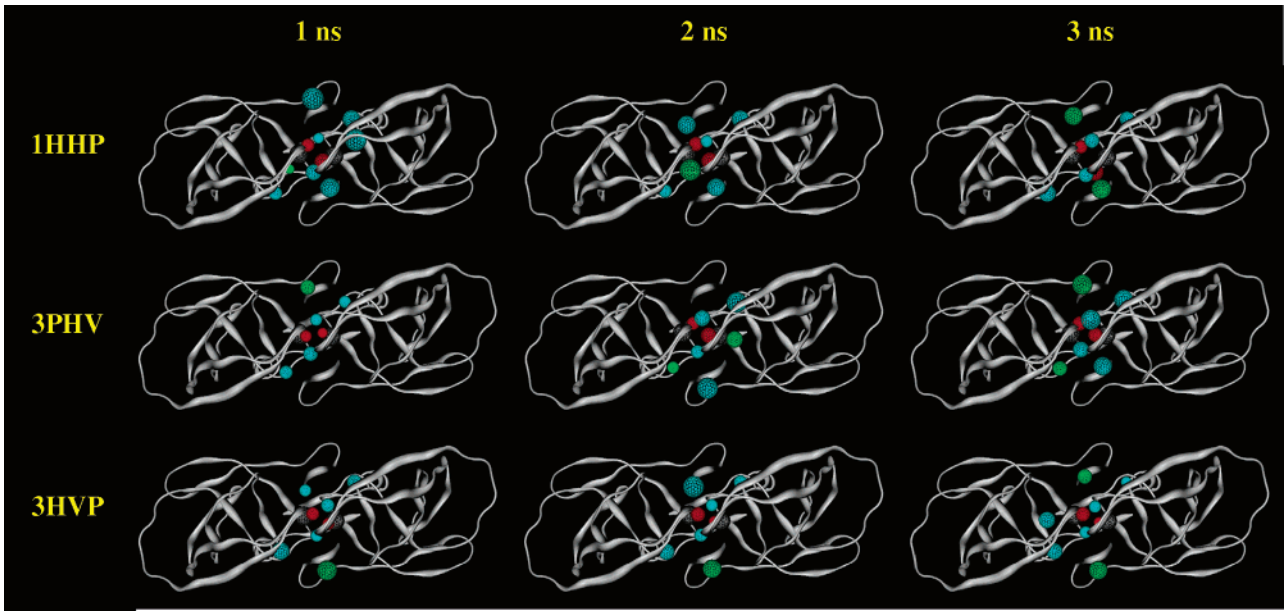

Figure 2. Pharmacophore models for 1HHP, 3PHV, and 3HVP created from 1, 2, or 3 ns simulation lengths. Radii of the elements are shown as 1 × rmsd of the consensus probes and colored according to chemical functionality: red, hydrogen-bond donating; green, aromatic; cyan, aromatic or hydrophobic; gray, excluded volume.

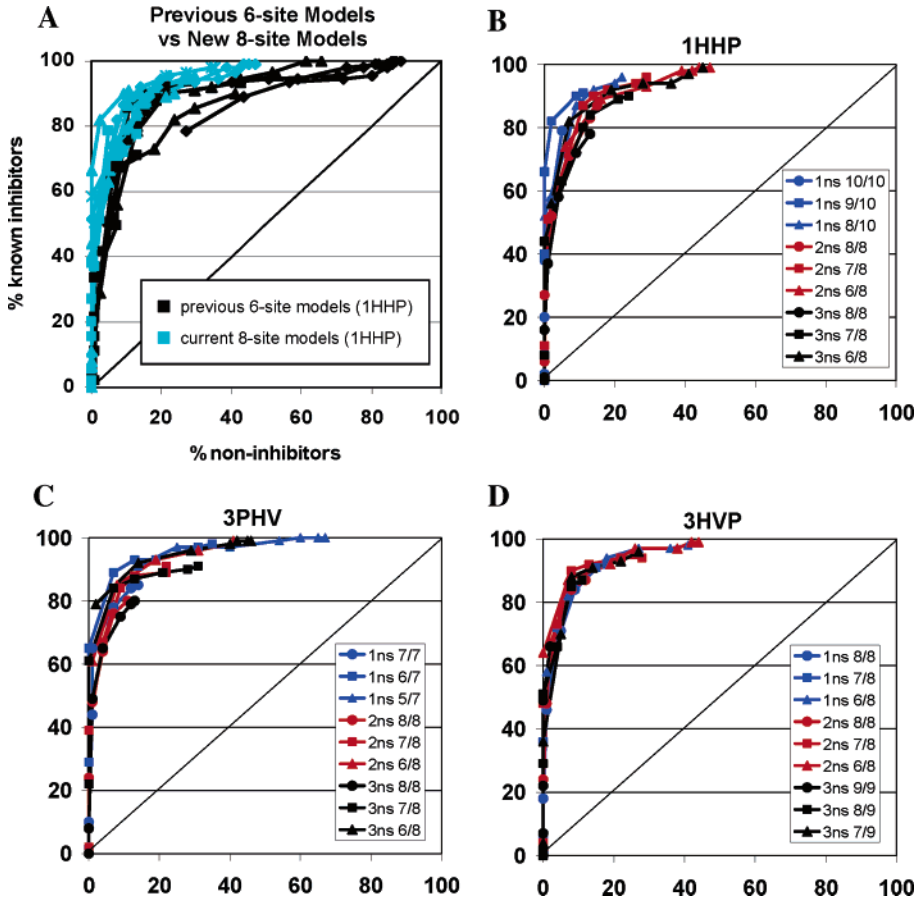

Figure 3. ROC curves for the 1HHP, 3PHV, and 3HVP pharmacophore models. The percentage of known HIV-1p inhibitors identified by the model is plotted against the percentage of compounds from the Comprehensive Medicinal Chemistry Index (false positives). (A) Comparison of previous screening results for a 6-site model of 1HHP (black) with results for the 1HHP from the current work (cyan). The 1, 2, and 3 ns models for 1HHP are shown simultaneously. (B) Details of the 1HHP results: blue, 1s; red, 2s; black, 3 ns. (C) 3PHV. (D) 3HVP.

different orientations in response to different bound ligands. This flexibility influences the placement of the minimized probes and, hence, affects the definition of the pharmacophore element. When the arginine points into solution, it can pull the pharmacophore element outside the 9 Å cutoff (3PHV-1ns and -2ns), and if it predominately points into the cavity, the element is found closer to the catalytic residues. Additionally, if the residue samples a wide enough set of conformations, the probe positions become too disordered and no consensus is observed.

Testing the Pharmacophore Models. The data from our pharmacophore screens are presented as receiver operator characteristic (ROC) curves in which the percent of known inhibitors identified by the model is plotted against the percentage of false positives predicted. In this representation of the data, a model with no selectivity would lie somewhere on the line with slope equal to 1, and the ideal model would lie in the upper left corner at the point (0, 100).

As we plot each model's ability to hit true inhibitors versus false positives, we use the "distance" from ideal to determine overall performance (i.e., distance = $[(\%\text{ known inactives})^2 + (100\% - \%\text{ known inhibitors})^2]^{1/2}$). This metric for performance allows us to objectively compare models from different structures and models that contain a different number of pharmacophore elements. It also gives us a means to evaluate the statistical significance of our findings through cross-validation. Each dataset was divided into 10 random subsets, and each subset was sequentially removed from the analysis to evaluate its contribution to the performance of the model (details are provided in the Supporting Information). The standard deviation obtained from the cross-validation is very low, meaning that the performance is not grossly dependent on the nature of the test set of inhibitors and noninhibitors.

The raw data used to generate the ROC plots is provided as Supporting Information. As in our previous study, the number of elements required for a match were varied (i.e., 8 of 8 sites versus 7 of 8 sites versus 6 of 8 sites) and the radii of the elements were varied (1 × rmsd, $1^1/_3$ × rmsd, $1^2/_3$ × rmsd, 2 × rmsd, $2^1/_3$ × rmsd, $2^2/_3$ × rmsd, and 3 × rmsd).[9] Using smaller radii and requiring more elements results in highly selective models that identify almost no false positives but at the expense of many known inhibitors. Models with large radii and fewer required elements have the least selective performance.

In assessing data across all three starting structures (1HHP, 3PHV, and 3HVP) and all three time lengths investigated (1, 2, and 3 ns), we found that the overall optimum protocol requires radii for the elements to be 2.3 × rmsd and $n - 1$ of n pharmacophore elements to be satisfied for a hit. Specific details comparing the best individual models are presented below.

Figure 3A provides a comparison of the performance of our current 1HHP models to our previous 6-site models also generated from the same 1HHP snapshots. All of the new models show as good or better performance than our previous results as determined by a leftward shift on the ROC plot. Due to the greater specificity of an 8-site versus 6-site pharmacophore model, we see a steeper initial slope, indicating that the models are identifying fewer false positives with the most restrictive models.

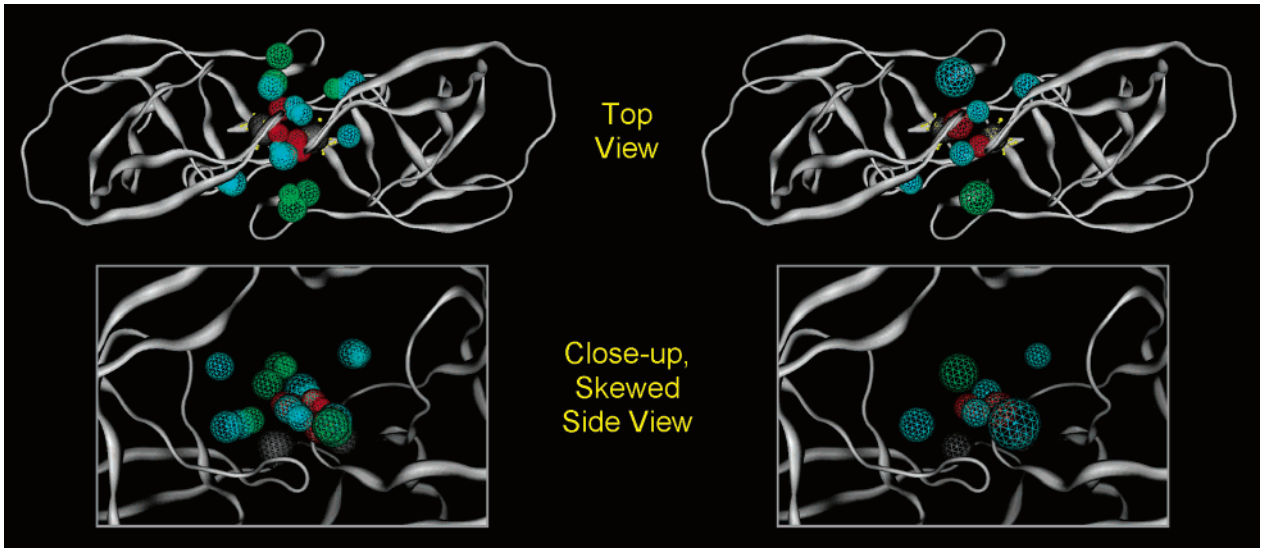

Figure 4. Creation of the "consensus of consensus" model: (left) overlay of the 3-ns pharmacophore models from 1HHP, 3PHV, and 3HVP; (right) resulting consensus of consensus pharmacophore model. Pharmacophore elements are colored according to chemical functionality: red, hydrogen-bond donating; green, aromatic; cyan, aromatic or hydrophobic; gray, excluded volume.

We were pleased to find that the best pharmacophore models from all 3 MD simulations were similar to our previous findings (Figure 3). Again, the optimal radii were approximately 2 × rmsd and tight stringency was also required. The optimal models required a molecule to match 7 of the 8 sites to be counted as a "hit" and predicted 84−92% of the known inhibitors correctly with a false positive rate of 7−15%. The false positives identified were examined and found to contain the same renin inhibitors, small hydrophobic peptides, and peptidomimetics that were observed in our previous work.[9] Of course, renin inhibitors and peptide mimics should be identified as potential inhibitors by our models.

In this study, we do not see the same improvement with increasing simulation length as previously reported. Previously, the longer simulations sampled additional protein conformations and also probe positions. By improving the initial probe placement, we are able to identify consensus with less MD sampling—which can significantly speed up the MPS procedure. We should note that other protein systems that we are investigating still show improvement with longer MD simulations,[14] and this characteristic may be system dependent. Using multiple protein structures is still necessary, as pharmacophore models derived from a single static structure are inconsistent across different starting structures and are not reasonable for drug discovery (see Supporting Information).

We have analyzed the subtle differences between the individual pharmacophore models. The 1HHP-1ns model actually has 10 sites. Despite the inherent difficulty of a molecule fitting all 10 sites, the model shows good performance—the optimal model identifies 92% known inhibitors with a false positive rate of only 14% (8/10 sites, $2^1/_3$ × rmsd). In the 1HHP-2ns and 1HHP-3ns models, the two extraneous sites "disappear" as greater sampling of the protein is achieved. Interestingly, the 3PHV-1ns pharmacophore model has only 7 sites; it is missing the hydrophobic site corresponding to the S1 recognition pocket (southeast element) that is influenced by the Arg 8 conformation. Conversely, the 3PHV-2ns model is missing the northwest site which again is influenced by the mobility of the Arg 8′ residue. However, the 3PHV model does converge to 8 sites by 3 ns. In the case of 3HVP, it is the 3 ns model that has an extra site while the 1 and 2 ns models fit the common 8-site template. Despite these minor differences, all of the pharmacophore models show good performance and successfully discriminate known HIV-1p inhibitors from inactives (Figure 3B−D).

A significant goal of this work was to develop parameters for the MPS method that are robust and consistent among similar starting structures. The current MPS pharmacophore method is very successful but quite time-consuming and labor intensive. The user must identify clusters from each MD snapshot and then combine the resulting parent probes. Next, the user must identify which parents are in consensus to create the MPS pharmacophore elements. To improve the method for use by the entire scientific community, we are in the process of automating the method of pharmacophore determination.[14] However, an automated method lacks a user's chemical intuition and may miss one pharmacophore element or add an extraneous element. We have shown that our models are robust enough to provide good performance, even when they deviate from the ideal.

Creating a Consensus of Consensus Model. By combining all the conformational information gained through the three independent MD simulations of the three crystal structures, we created a "consensus of consensus" pharmacophore model. The 3-ns pharmacophore models from each simulation were overlaid and the pharmacophore elements recalculated by averaging all the original probes from the MUSIC simulations (Figure 4). As our 3-ns models are already highly similar, it is not surprising that we see the same 8 sites in our consensus of consensus model. The most variation in element sites is again found in the northwest and southeast sites, as is reflected in the larger radii at those positions compared to the northeast and southwest elements. The consensus of consensus model has larger radii in general and shows the symmetry one would expect given a homodimeric protein structure. Coordinates, radii, and raw screening data for the consensus of consensus model are provided as Supporting Information.

Again, this consensus of consensus model shows excellent performance in discriminating known HIV-1p inhibitors from false positives (Figure 5). The optimum consensus of consensus model identifies 88% known inhibitors with a false positive rate of 8% (1 × rmsd, 6/8). This model uses different specifications than the optimal models from the individual MD simulations. With the consensus of consensus model, a less stringent query is combined with the smallest radii size to give the best

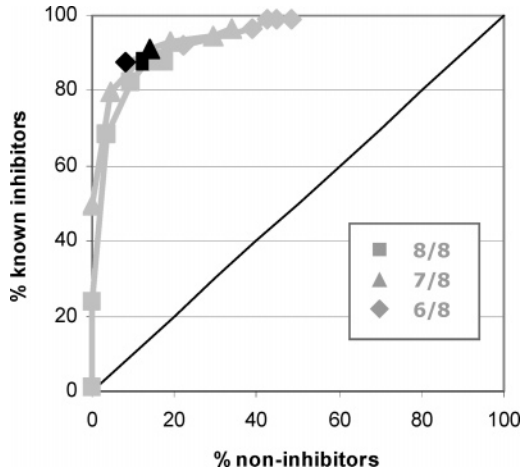

Figure 5. ROC curves for the consensus of consensus pharmacophore model. The black points indicate the three models with the best performance. They are approximately the same distance to the upper-left corner where an ideal model would identify 100% of the known inhibitors and 0% noninhibitors.

performance, although there is a cluster of various models which all demonstrate good selectivity (Figure 5). The best 7/8 model has radii of $1^2/_3 \times$ rmsd, identifying 91% of inhibitors and only 14% of noninhibitors. The best 8/8 model had $2^1/_3 \times$ rmsd radii, identifying 80% of inhibitors and 13% noninhibitors. Again, the pattern is seen that larger radii are needed when more elements are required. These three optimal, consensus of consensus models (requiring 8/8, 7/8, and 6/8 elements and shown in black in Figure 5) have the same excellent performance as seen for the individual models from 1, 2, and 3 ns of MD for 1HHP, 3PHV, and 3HVP.

Conclusions

We have extended the MPS method by studying three unbound HIV-1p structures and demonstrating the consistency of the method. Using three similar but unique starting structures and three independent MD trajectories of the unbound HIV-1p, we obtained nearly identical pharmacophore models each with excellent performance in discriminating known inhibitors from chemically similar noninhibitors. Thus, the MPS method is not overly dependent on a specific starting conformation or particular MD trajectory. On the basis of this work, we would recommend MPS pharmacophore queries requiring $n - 1$ of n of the pharmacophore elements with radii of $2.3 \times$ rmsd. We have further demonstrated that models that have an additional site or are missing a site continue to show good performance. This is encouraging as we automate our pharmacophore methods with the goal of providing an easy-to-use MPS pharmacophore method to the greater scientific community.

We chose unbound structures for this study to further probe our notable success with structure-based drug discovery from apo structures. However, it is also important to test the MPS method using bound structures of HIV-1p. At this time, there are over 170 high-resolution crystal structures of bound HIV-1p, and we are wrestling with the sizable task of combining these structures in statistically meaningful ways. We are also comparing the use of multiple crystal structures to ensembles of conformations from NMR studies.

Acknowledgment. This work has been supported by the National Institutes of Health (Grant GM65372) and the Beckman Young Investigator Program. We thank Allen Bailey for maintaining the computers used in this work. The American Chemical Society Division of Medicinal Chemistry and Bristol-Myers Squibb are gratefully acknowledged for a 2004−2005 predoctoral fellowship to K.L.M. K.L.M. appreciates receiving fellowships from Edward S. Blake, Fred W. Lyons, American Foundation for Pharmaceutical Education, and the University of Michigan Regents, as well as support from the Pharmacological Sciences Training Program (NIH Grant GM07767). M.G.L. is grateful for support from the Michigan Molecular Biophysics Training Grant (NIH Grant GM08270).

Supporting Information Available: A description of the metric used to compare models and the cross-validation of the data sets, static pharmacophore models for the three structures, the coordinates and rmsd of the pharmacophore elements for all models discussed, along with the raw data used to create the ROC plots, and the inhibitor and noninhibitor data sets used in this work. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Lyne, P. D. Structure-Based Virtual Screening: An Overview. *Drug Discov. Today* **2002**, *7*, 1047−1055.
(2) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813−1818.
(3) Teodoro, M. L.; Kavraki, L. E. Conformational Flexibility Models for the Receptor in Structure Based Drug Design. *Curr. Pharm. Des.* **2003**, *9*, 1635−1648.
(4) Teague, S. J. Implications of Protein Flexibility for Drug Discovery. *Nat. Rev. Drug Discov.* **2003**, *2*, 527−541.
(5) Wong, C. F.; McCammon, J. A. Protein Flexibility and Computer-Aided Drug Design. *Annu. Rev. Pharmacol.* **2003**, *43*, 31−45.
(6) Carlson, H. A. Protein Flexibility and Drug Design: How to Hit a Moving Target. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447−452.
(7) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. Developing a Dynamic Pharmacophore Model for HIV-1 Integrase. *J. Med. Chem.* **2000**, *43*, 2100−2114.
(8) Carlson, H. A.; Masukawa, K. M.; McCammon, J. A. Method for Including the Dynamic Fluctuations of a Protein in Computer-Aided Drug Design. *J. Phys. Chem. A* **1999**, *103*, 10213−10219.
(9) Meagher, K. L.; Carlson, H. A. Incorporating Protein Flexibility in Structure-Based Drug Discovery: Using HIV-1 Protease as a Test Case. *J. Am. Chem. Soc.* **2004**, *126*, 13276−13281.
(10) Mustata, G. I.; Briggs, J. M. A Structure-Based Design Approach for the Identification of Novel Inhibitors: Application to an Alanine Racemase. *J. Comput. Aided Mol. Des.* **2002**, *16*, 935−953.
(11) Deng, J. X.; Lee, K. W.; Sanchez, T.; Cui, M.; Neamati, N.; Briggs, J. M. Dynamic Receptor-Based Pharmacophore Model Development and Its Application in Designing Novel HIV-1 Integrase Inhibitors. *J. Med. Chem.* **2005**, *48*, 1496−1505.
(12) Schechner, M.; Sirockin, F.; Stote, R. H.; Dejaegere, A. P. Functionality Maps of the ATP Binding Site of DNA Gyrase B: Generation of a Consensus Model of Ligand Binding. *J. Med. Chem.* **2004**, *47*, 4373−4390.
(13) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(14) Lerner, M. G.; Carlson, H. A. Unpublished data.
(15) Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J. The 3-Dimensional Structure of the Aspartyl Protease from the HIV-1 Isolate BRU. *Biochimie* **1991**, *73*, 1391−1396.
(16) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. H. Conserved Folding in Retroviral Proteases—Crystal-Structure of a Synthetic HIV-1 Protease. *Science* **1989**, *245*, 616−621.
(17) Lapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Merson, J. R.; Whittle, P. J.; Danley, D. E.; Geoghegan, K. F.; Hawrylik, S. J.; Lee, S. E.; Scheld, K. G.; Hobart, P. M. X-ray-Analysis of HIV-1 Proteinase at 2.7 Å Resolution Confirms Structural Homology among Retroviral Enzymes. *Nature* **1989**, *342*, 299−302.
(18) Meagher, K. L.; Carlson, H. A. Solvation Influences Flap Collapse in HIV-1 Protease. *Proteins* **2005**, *58*, 119−125.
(19) Case, D. A.; Pearlman, D. A. C., J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER6*; University of California: San Francisco, CA, 1996.

(20) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc*. **1995**, *117*, 5179−5197.
(21) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys*. **1983**, *79*, 926−935.
(22) Jorgensen, W. L. *BOSS*, 4.2; Yale University: New Haven, CT, 2000.
(23) *Comprehensive Medicinal Chemistry Database*; MDL Information Systems, Inc.: San Leandro, CA, 2003.
(24) *Molecular Operating Environment, 2003.02*; Chemical Computing Group Inc.: Montreal, Canada, 2003.
(25) Halgren, T. A. Merck Molecular Force Field I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem*. **1996**, *17*, 490−519.
(26) DeLano, W. L. *The Pymol Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002.
(27) Wlodawer, A.; Gustchina, A. Structural and Biochemical Studies of Retroviral Proteases. *Biochim. Biophysica. Acta* **2000**, *1477*, 16−34.

JM050755M