J|A|C|S
ARTICLES
Published on Web 09/25/2004

Incorporating Protein Flexibility in Structure-Based Drug Discovery: Using HIV-1 Protease as a Test Case

Kristin L. Meagher and Heather A. Carlson*

Contribution from the Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, 428 Church Street, Ann Arbor, Michigan 48109-1065

Received May 24, 2004; E-mail: carlsonh@umich.edu.

Abstract: We have developed a receptor-based pharmacophore method which utilizes a collection of protein structures to account for inherent protein flexibility in structure-based drug design. Several procedures were systematically evaluated to derive the most general protocol for using multiple protein structures. Most notably, incorporating more protein flexibility improved the performance of the method. The pharmacophore models successfully discriminate known inhibitors from drug-like non-inhibitors. Furthermore, the models correctly identify the bound conformations of some ligands. We used unliganded HIV-1 protease to develop and validate this method. Drug design is always initiated with a protein−ligand structure, and such success with unbound protein structures is remarkable — particularly in the case of HIV-1 protease, which has a large conformational change upon binding. This technique holds the promise of successful computer-based drug design before bound crystal structures are even discovered, which can mean a jump-start of 1−3 years in tackling some medically relevant systems with computational methods.

Introduction

Structure-based drug design (SBDD) is classically based on an X-ray structure of a protein target complexed with a known agonist or antagonist. This complex is analyzed and novel compounds are designed to maximize interactions with the protein — thereby increasing selectivity and potency. The method has been used extensively in the pharmaceutical industry to develop successful new therapeutics.[1−3]

Crystal structures provide only a single conformation of the protein. Although this structure is a statistical average of many similar conformations present in the crystal lattice, the information on protein dynamics is limited. Conformations are also influenced by the crystallization conditions: crystal packing effects, pH, and temperature can alter the conformation and provide misleading information for SBDD. Furthermore, ligand binding can induce a conformational change in both the protein and ligand. Though ligand flexibility is often considered in SBDD, including protein flexibility is in its infancy. Therefore, new techniques are needed to push the frontiers of computer-aided drug discovery by incorporating ensembles of protein conformations to more accurately simulate the inherent motion of the system and potential induced fit.

Several groups have begun to incorporate protein flexibility into SBDD. Research in this area has been summarized in several recent reviews.[4−8] The first approach to employ an ensemble of protein structures for drug design created an "average" structure derived from several crystal structures or NMR structures.[9] Other techniques use a graph theory approach to generate multiple ligand conformations which are then rapidly screened against an active-site template, eliminating most compounds and retaining only a small number for docking.[10] The FLEXX docking software has also recently been extended to include protein flexibility by using a combinatorial collection of protein conformations (FLEXE).[11] Multiple protein structures have also been used to determine the "best" complex of ligand and enzyme by efficient docking and accurate scoring among an ensemble of conformational states.[12]

In 1999, Carlson et al. described a method to model the protein flexibility of human immunodeficiency virus (HIV-1) integrase and to develop a receptor-based pharmacophore model based on multiple protein structures (MPS).[13,14] Using MPS from molecular dynamics (MD) simulations, the pharmacophore model successfully predicted novel HIV-1 integrase inhibitors from the Available Chemicals Directory that were confirmed by biological testing.

(1) Babine, R. E.; Bender, S. L. *Chem. Rev.* **1997**, *97*, 1359−1472.
(2) Hardy, L. W.; Malikayil, A. *Curr. Drug Discov.* **2003**, *December*, 15−20.
(3) Jorgensen, W. L. *Science* **2004**, *303*, 1813−1818.
(4) Wong, C. F.; McCammon, J. A. *Annu. Rev. Pharmacol. Toxicol.* **2003**, *43*, 31−45.
(5) Teague, S. *Nat. Rev.* **2003**, *2*, 527−541.
(6) Carlson, H. A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447−452.
(7) Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Freer, S. T.; Rose, P. W. *Curr. Opin. Struct. Biol.* **2002**, *12*, 197−203.
(8) Carlson, H. A.; McCammon, J. A. *Mol. Pharmacol.* **2000**, *57*, 213−218.
(9) Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. *J. Mol. Biol.* **1997**, *266*, 424−440.
(10) Schnecke, V.; Kuhn, L. A. *Perspect. Drug Discovery Des.* **2000**, *20*, 171−190.
(11) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. *J. Mol. Biol.* **2001**, *308*, 377−395.
(12) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. *Biopolymers* **2003**, *68*, 47−62.
(13) Carlson, H. A.; Masukawa, K. M.; McCammon, J. A. *J. Phys. Chem. A* **1999**, *103*, 10213−10219.
(14) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. J.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. *J. Med. Chem.* **2000**, *43*, 2100−2114.

10.1021/ja0469378 CCC: $27.50

To generalize the method, we have applied it to the unbound structure of HIV-1 protease (HIVp). HIVp is an important drug target for the treatment of AIDS; seven drugs are currently marketed as HIVp inhibitors.[1] It is a system where protein flexibility is important for ligand binding.[15,16] Furthermore, it has been studied extensively both experimentally and computationally, and the wealth of data makes this an ideal case to test and refine the MPS method.

To develop a robust method for drug discovery, we have investigated several variables of the method, including MD simulation length, pharmacophore element size, number of required elements, and alignment mechanism. Using MPS pharmacophore models, we have succeeded in discriminating known ligands from drug-like non-inhibitors, showing better performance with the incorporation of more protein flexibility. The models were developed from uncomplexed HIVp, and the ability to do accurate SBDD with an unbound structure is a significant breakthrough. The models were not only able to identify the correct compounds but also to predict the proper bound conformation of known ligands in a system with significant structural reorganization upon ligand binding.

Computational Methods

Protein Preparation and MD. The apo HIVp structure (1HHP)[17] was solvated in a 62 × 62 × 82 Å^3 box of TIP3P[18] water using the AMBER6[19] suite of programs. Using the sander_classic module, the system was pre-equilibrated for 200 ps with the protein held fixed to allow the water to optimally orient around the protein. A second equilibration of 200 ps was performed with the protein restraints removed, and 3 ns of production phase MD were collected. The conformational fluctuations observed in the simulation are stable and consistent with NMR studies of unliganded HIVp.[20,21] A detailed discussion of the MD simulation will be published elsewhere.[22] Snapshots were saved from the MD simulation after equilibration and every 100 ps along the trajectory.

MUSIC Simulations. Each protein snapshot was used in a multiunit search for interacting conformers (MUSIC) simulation with the BOSS program,[23] using the OPLS force field.[24] The protein was held rigid during the MUSIC simulations (the use of many protein conformations took the place of explicit flexibility at this stage). The active site of each snapshot was flooded with hundreds of small-molecule probes which were minimized using low-temperature Monte Carlo sampling. All probe−probe interactions were ignored, so each probe experienced only the forces associated with the protein. This allowed the probe molecules to cluster to the surface of the protein and reveal positions of favorable interactions for a particular chemical functionality. A detailed description of the MUSIC methodology has been published elsewhere.[14]

For the HIVp study, three small-molecule probes were chosen to map the active site. Methanol probes revealed both hydrogen-bond-donating and -accepting sites, benzene elucidated favorable aromatic interactions, and ethane further clarified hydrophobic from aromatic interactions. Discrete clusters were observed for all three probes used, and an example of the results from a MUSIC simulation with benzene is shown in Figure 1A. As the figure shows, the clusters were obvious and did not require an RMSD cutoff to define them.

Clustering of Pharmacophore Elements. For each snapshot, probes with any atom within 10 Å of the two catalytic aspartate residues (25 and 25′) were examined. Each cluster is represented by its "parent", the probe molecule with the most favorable interaction energy with the protein as calculated in the MUSIC procedure.

Protein conformations taken along the MD trajectory were aligned, and the parent probes were clustered. These "clusters of clusters" identify regions of consensus where most conformations of the protein will be complemented by the probe functionality. The consensus clusters were also obvious and did not require an RMSD cutoff to define them. We should note that if we attempted to use a large number of snapshots for the MPS model, the consensus clusters became too diffuse to identify easily (data not shown). Each consensus region contained probes from different time points in the trajectory, and at least 40% of the snapshots were required to contribute to a given consensus cluster. Requiring representation of probes from the beginning, middle, and end of the trajectory ensured that the consensus features were representative of the entire conformational space sampled by the MD trajectory. An example of the consensus clusters for benzene is shown in Figure 1B.

To represent the consensus clusters as pharmacophore elements, the average position and the RMSD of the clusters were determined. For methanol, the pharmacophore center was defined as the average position of the O atoms in the cluster. For benzene, the centroid determined the average position, and for ethane the center point between the two carbons was chosen. The radii of the pharmacophore elements were based on the RMSD of the consensus clusters. It should be noted that we also created models using all probes from the individual snapshots — not just the parents — to define the pharmacophore elements. The models were essentially the same; the difference in the centers and RMSD were ≤0.1 Å for all elements (data not shown). We chose to work with the parents because it simplified the problem without losing precision.

Two excluded volumes were used to define the bottom of the binding site. The centers were defined as the average position of the γ-carbon of each catalytic aspartate. The radii of the excluded volumes were not based on the RMSD of the atoms (the RMSD was too small to really represent the bottom of the binding site). The radii were set to 1.5 Å, roughly the Cγ−Oδ bond length in the catalytic Asp side chains, so that it would represent the entire carboxylate group. Any pharmacophore element that was greater than 10 Å from the excluded volume centers was eliminated. Overlapping ethane and benzene clusters were combined into a single hydrophobic cluster by recalculating the element center and RMSD, including both the centroid of the benzene ring and the midpoint of the ethane C−C bond. Lone benzene clusters were left as aromatic elements, and extraneous ethane clusters were removed. Methanol clusters were defined as hydrogen-bond donors, acceptors, or doneptors (possessing both hydrogen-bond-donating and -accepting character). Overlapping pharmacophore elements were simplified by retaining the aromatic or aromatic/hydrophobic sites and eliminating hydrogen-bonding sites. A discussion of this issue is presented as Supporting Information. A representative pharmacophore model is shown in Figure 1C.

Figure 1D shows the superposition of co-crystallized ligands onto the 1HHP crystal structure. Using the MPS method, we have de novo created pharmacophores that are consistent with placement of aromatic, hydrophobic, and hydrogen-bonding features of the bound ligands from experimental structures.

(15) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. *J. Med. Chem.* **2004**, *47*, 45−55.
(16) Wlodawer, A.; Gustchina, A. *Biochim. Biophys. Acta* **2000**, *1477*, 16−34.
(17) Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J. *Biochimie* **1991**, *73*, 1391−1396.
(18) Jorgensen, W. L.; Chandrasekhar, J. D.; Madura, R. W.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(19) Case, D. A.; Pearlman, D. A.; Caldwell, 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. *AMBER 6*; University of California San Francisco: San Francisco, CA, 1996.
(20) Ishima, R.; Freedberg, D. I.; Wang, Y.-X.; Louis, J. M.; Torchia, D. A. *Structure* **1999**, *7*, 1047−1055.
(21) Freedberg, D. I.; Ishima, R.; Jacob, J.; Wang, Y.-X.; Kustanovich, I.; Louis, J. M.; Torchia, D. A. *Protein Sci.* **2002**, *11*, 221−232.
(22) Meagher, K. L.; Carlson, H. A. *Proteins: Struct., Funct. Bioinf.*, in press.
(23) Jorgensen, W. L. *BOSS*, Version 4.2; Yale University: New Haven, CT, 2000.
(24) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

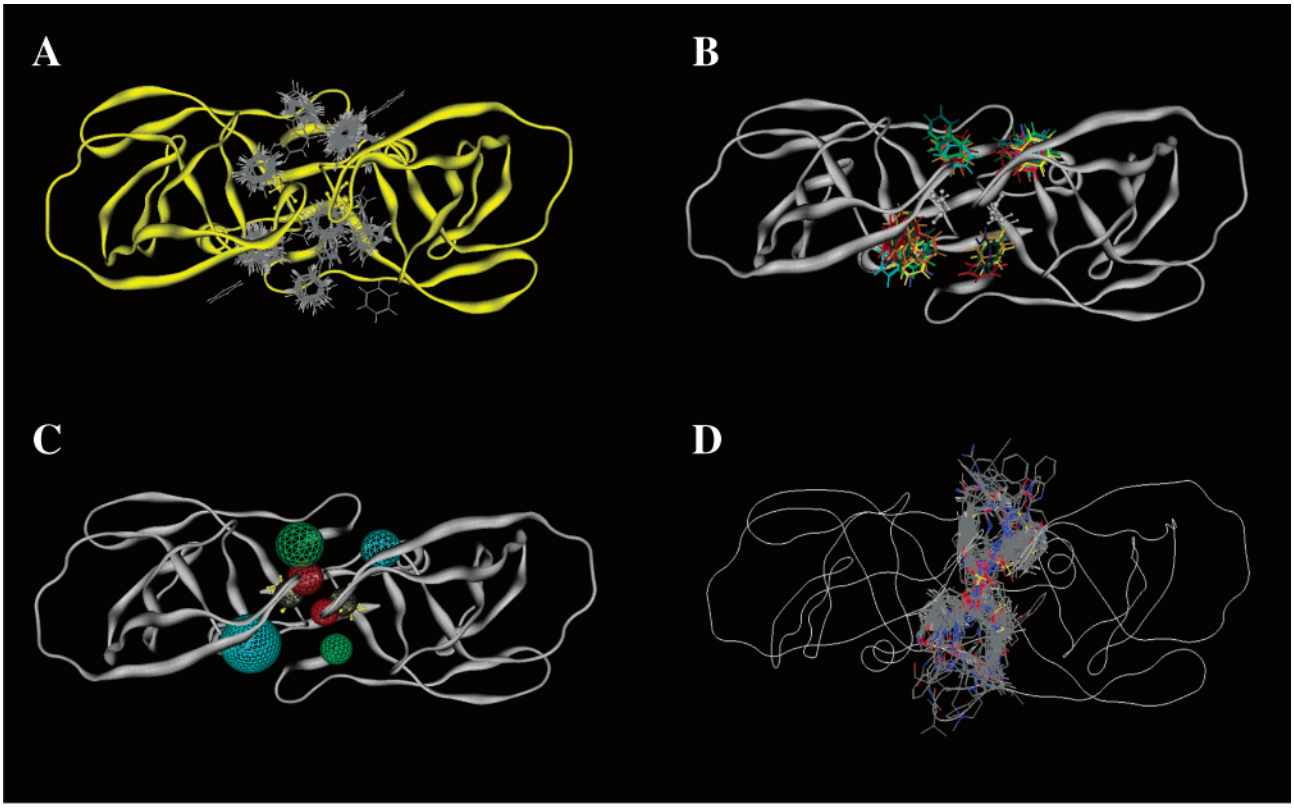

Figure 1. (A) Raw output of minimizing benzene probes to a single snapshot from the MD, demonstrating the observed clustering. (B) Parent probes from many MPS, overlaid to reveal consensus clusters. Note that some clusters in panel A are not retained when MPS are compared. Probes are colored blue to yellow to red according to their point in the MD simulation. (C) Representative pharmacophore model (1 ns, Cα alignment, radii of 2 × RMSD). Element spheres are color-coded according to chemical functionality: red, hydrogen-bond-donating; green, aromatic; cyan, aromatic or hydrophobic. (D) Superposition of ~50 co-crystallized HIVp inhibitors. General aromatic, hydrophobic, and hydrogen-bonding features agree well with the model in panel C.

Creation of the Ligand Databases. To evaluate the pharmacophore models, two ligand databases were created. The known-ligands database was created by extracting the bound conformation of 80 unique ligands from co-crystals present in the PDB. To augment this database, structures of known inhibitors from a review of HIV-1 protease inhibition[1] were built in MOE[25] and subjected to an 85% similarity criteria using fingerprint analysis. Seven ligands were subsequently removed to eliminate remaining structural redundancies in the database, yielding a database of 89 molecules. Multiple conformations of each ligand were generated by a stochastic conformational search in MOE using the MMFF force field.[26] The stochastic search was run with an energy cutoff of 25 kcal/mol and heavy-atom RMSD criterion of 2 Å. A maximum of 300 lowest-energy confomers were retained.

The drug-like non-inhibitor database was created from the Comprehensive Medicinal Chemistry Index (CMC).[27,28] Using the MDL DiscoveryGate[29] software, the CMC (~9000 compounds) was filtered to obtain an appropriate database of small-molecule ligands. Only molecules without listed HIVp activity, with drug-like molecular weight (1200 > MW > 300), at least one aromatic group, and at least one hydrogen-bond donor were considered. A spatial filter was used to ensure that the ligands were appropriately sized to interact with the large HIVp active site. Further refinement by fingerprint clustering, using a 65% similarity cutoff, resulted in 88 entries. These entries were examined by hand, and three structures were eliminated due to the presence of metals or fragmentation of the structure, yielding a set of 85 compounds. Confomers were generated using the same stochastic search method described above. The structures of both the known inhibitors and the drug-like non-inhibitors are provided as Supporting Information.

Criteria for Model Evaluation. Each pharmacophore model was screened against the two databases of potential ligands, known inhibitors (89 entries) and drug-like non-inhibitors (85 entries). MOE[25] was used to screen the pharmacophore models against the two databases. A ligand was counted as a hit if at least one pre-generated conformation could be rigidly aligned to the pharmacophore coordinates.

There is no single criterion for evaluating the predictive ability of these pharmacophore models. A delicate balance exists between achieving a high yield, identifying the maximum number of active compounds, and minimizing false positives. We have chosen to present the screening data as receiver operator characteristic (ROC) curves, with the percentage of known inhibitors identified by the models plotted against the percentage of false positives identified. Using this metric, a perfect model would predict 100% of the known inhibitors and zero non-inhibitors and would lie in the upper-left corner of the ROC plot (point 0,100). The raw data from the ROC plots are available as Supporting Information.

Results and Discussion

Creation of the Pharmacophore Models. Our goal in this study was to examine several factors used in creating the MPS pharmacophore models, to determine the optimum protocol for application to any protein system. When determining the consensus clusters, we chose to align the snapshots either by whole protein Cα position or by the catalytic aspartates (all atoms from the two residues). MPS models were created from snapshots from 1, 2, or 3 ns of the MD simulation. For each

(25) *Molecular Operating Environment*; Chemical Computing Group Inc.: Montreal, Canada, 2001.
(26) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490−519.
(27) *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press: Oxford, 1990; Vol. 1−6.
(28) *Comprehensive Medicinal Chemistry Database*; MDL Information Systems, Inc.: San Leandro, CA, 2003.
(29) *Discovery Gate*; MDL Information Systems, Inc.: San Leandro, CA, 2003.

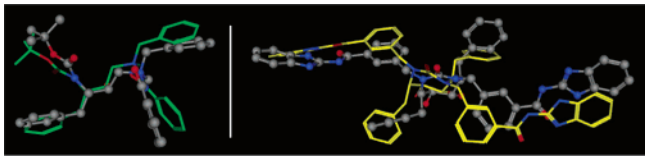

Figure 2. Conformers of two inhibitors that hit the pharmacophore model, overlaid with their actual crystal structure positions. (Proteins and pharmacophore models are not shown for clarity.) The crystal conformations are shown in gray ball-and-stick and the conformers that hit the MPS models are colored; AQ148 is in green (left, 3AID) and SD146 is in yellow (right, 1QBT). Ligand heteroatoms are colored to aid comparison. These results are from screening against our 2-ns, Cα-aligned, 1.67 × RMSD MPS pharmacophore model.

model, the size of the pharmacophore elements was varied by multiplying the consensus cluster RMSD by 1, 1.33, 1.67, 2, 2.33, 2.67, or 3. Also, the number of pharmacophore elements required for a "hit" was varied.

Below we describe in detail how the optimal MPS protocol was developed. The resulting models were highly successful in discriminating known HIVp inhibitors from drug-like non-inhibitors, correctly selecting 85−90% of the known ligands with a false positive rate of only 11−19%. Models derived from longer simulations (2 and 3 ns) exhibited better performance, as additional protein conformations are sampled. This is strong support of the methodology because incorporating more flexibility improves the MPS models.

The pharmacophore models derived from the unbound HIVp structure also accurately predict the bound conformation of some known ligands (Figure 2). To our knowledge, this is the first example of receptor-based pharmacophore models derived from unbound protein structure accurately predicting bound-ligand conformations.

Effect of Altering Model Specificity. The pharmacophore models created using the MUSIC methodology consisted of six elements representing favorable interaction areas for a given chemical functionality (Figure 1C). To determine how stringent the pharmacophore query needed to be, the number of matched pharmacophore elements required for a "hit" was varied. Each six-site pharmacophore model was screened, requiring all six pharmacophore elements to be present (6/6), any five out of six (5/6), and any four out of six (4/6). Thus, the 5/6 model actually screens six individual models, as each pharmacophore site is systematically omitted. Figure 3 clearly shows that requiring only 4/6 elements enables far too many non-inhibitors to fit the model, while some of the restrictive 6/6 models can miss many true inhibitors.

Effect of Varying the Pharmacophore Element Radii. The clustering method averages coordinates from the probes in the consensus cluster to provide the center and RMSD of the parents in the cluster. To determine the optimal radii size, we multiplied the cluster RMSD by values of 1, 1.33, 1.67, 2, 2.33, 2.67, and 3, creating seven models for each query (6/6, 5/6, or 4/6). Small radii result in highly specific pharmacophore models which identify high-quality hits. However, if the radii are too small, the models will be too specific to hit all the diverse known inhibitors. At the other extreme, using large radii creates large pharmacophore elements into which almost any molecule can fit, thereby losing all selectivity and specificity. As expected, we found that the optimal radii were dependent on the query stringency. For HIVp, optimal selectivity was achieved with

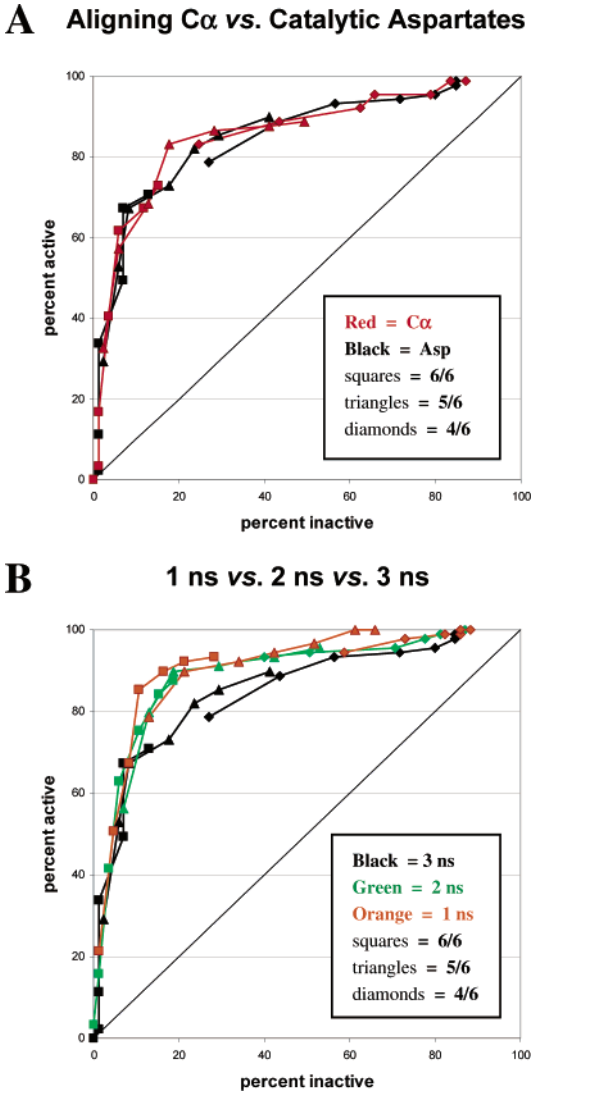

Figure 3. Screening data presented as ROC curves. Requiring 6/6 pharmacophore elements are squares, 5/6 triangles, and 4/6 diamonds. Points in each series are increasing RMSD values from 1× to 3×. For reference, a line of slope 1 indicates performance where there is no preference for known active compounds over inactives. (A) The 1-ns Cα (black) and Asp (red) alignment methods show nearly the same performance. (B) The 1-ns (black), 2-ns (green), and 3-ns (orange) pharmacophore models (Cα alignment). The 2-ns and 3-ns models show significant improvement over the 1-ns model.

two models: the most stringent query (6/6) coupled with larger radii (2, 2.67), and the 5/6 query with smaller radii (1.33, 1.67, Figure 3).

In the industrial setting, it is common to have a large database with few true positives. In these cases, one may wish to limit the false positives as much as possible. Though some actives will be sacrificed, the expense of testing of many false leads will be kept to a minimum. For this goal, the (6/6), 2 × RSMD, 1-ns model would be a good choice to reduce the number of false hits. Though only 34% of the actives are hit, only 1% of the inactives are identified. The smallest radii (6/6) models also show excellent enrichment for the 2-ns and 3-ns data (see data in Supporting Information).

Effect of Protein Alignment. The identification of consensus clusters is dependent on the method by which the protein snapshots are overlaid. The inherent flexibility of the flap domain of HIVp could influence protein alignment. Aligning

by Cα gives a global fit of the MPS, but the highly flexible flap region may skew the alignment inappropriately. In contrast, aligning by the catalytic residues isolates the active-site flexibility from the mobile flap region but may not be an appropriate frame of reference to align a large open binding site. We hypothesized that the different alignment methods would result in different pharmacophore models; however, minimal difference in model performance was observed (Figure 3A). Aligning by Cα is a more general method and is easily extended to other protein systems. We therefore chose to use alignment by Cα for our subsequent studies of simulation length.

Effect of Simulation Length. We predicted that, by using a longer simulation time, more conformational space of the enzyme would be explored, resulting in better performance of the resulting pharmacophore models. To support our claim that more conformational space is sampled, we examined the RMSD of the structures over the course of the MD. The maximal RMSD was 1.4 Å in the first nanosecond, 1.7 Å in the second nanosecond, and 1.95 Å in the third nanosecond (see the bold, black line in Figure 4 of ref 20).

Models were created using snapshots from 1, 2, or 3 ns of the simulation. The 1-ns models used the equilibrated structure and 10 additional snapshots taken every 100 ps along the trajectory. The 2-ns and 3-ns models used the equilibrated structure and snapshots taken every 200 ps (11 and 16 snapshots, respectively). The 2-ns and 3-ns models do indeed perform better than the 1-ns model (Figure 3B). The optimum 2-ns model (5/6, 1.67 × RMSD) predicts 90% of the known inhibitors while identifying only 19% of the non-inhibitors as false positives. Similarly, the best 3-ns model (6/6, 2 × RMSD) predicts 85% of the known inhibitors and 11% of the non-inhibitors. In contrast, the optimum 1-ns model (5/6, 2.33 × RMSD) identifies 84% of the known inhibitors and has a much higher false positive rate of 24%. The use of 2 or 3 ns of data is a significant improvement over only 1 ns. We were pleased to see that incorporating more protein flexibility improves the models.

The most noticeable difference between the 2-ns and 3-ns models is that the most stringent model (6/6) exhibits the best performance for 3 ns, but the (5/6) criterion is the better choice for 2 ns of data (Figure 3B). The 3-ns model incorporates 16 snapshots, compared to 11 for the 2-ns model. The additional snapshots result in greater spread in the consensus clusters and therefore larger RMSDs for the pharmacophore elements. This may explain why (6/6) is more generous with 3 ns of data.

Prediction of Bound Ligand Conformations. Docking to an apo protein structure is challenging, particularly for proteins like HIVp that have a large conformational change upon binding.[15] The flap domain of HIVp is known to be highly flexible and closes 5−7 Å upon ligand binding.[16] It is therefore remarkable that the simple six-point pharmacophore models we have created from the unbound 1HHP crystal structure of HIVp identify the correct binding mode of known inhibitors. As shown in Figure 2, some ligand conformations predicted by the pharmacophore model overlay well with the known conformations extracted from the co-crystals. In the case where co-crystals were not available for known inhibitors, we see good agreement comparing predicted bound conformations to crystal conformations of structurally related inhibitors.

Investigation of False Positives. To thoroughly evaluate our pharmacophore models, we examined the false positives identified from the drug-like non-inhibitor database. For the 2-ns model, 16 compounds (19%) of the non-inhibitors were identified as hits, while for the 3-ns model, 9 compounds (11%) were identified. The structures of both sets of false positives are available as Supporting Information. In both cases, two of the false positives are renin inhibitors — inhibitors of a homologous aspartic protease. Others are peptide analogues that can inhibit peptide cleavage. It is understandable that these molecules hit our pharmacophore model, given the proteolytic activity of HIVp, and it is quite likely that several of our "false positives" are indeed inhibitors.

It should also be noted that we defined the non-inhibitor database to provide a particularly rigorous test set. The CMC database was "pre-screened" to identify compounds that were the correct size and composition — the compounds most likely to fit the MPS models. This eliminated many drug-like compounds that could not possibly fit the HIVp pharmacophores. Had we included those thousands of compounds in the test set, our false-positive rate would have been incredibly low, but for superficial reasons.

Comparison to a Static Pharmacophore Model. The advantages of using MPS in pharmacophore development were most apparent when we created a model from a single HIVp crystal structure. The same three probes were run in MUSIC with the 1HHP crystal structure (data not shown). The resulting static model had far too many sites with very small RMSD, and it was not appropriate to use in screening. There was no clear way to identify the most important features of the model and thereby reduce the complexity. Using MPS, features conserved over the entire simulation were retained, distilling the complicated individual pharmacophore maps into a consensus map incorporating the most important and consistent complements to the inherent protein flexibility.

Conclusions

By incorporating MPS, we have developed pharmacophore models which successfully discriminate known HIVp inhibitors from drug-like non-inhibitors. Starting from an unbound structure of HIVp, our pharmacophore models are able to predict the actual ligand conformations observed in co-crystals. We have also investigated several variables of the MPS method and proposed optimized parameters to enable this technique to be applied to other medicinally important systems. Increasing the simulation time allows for the sampling of more conformational space and results in a better model, reinforcing the importance of protein flexibility in SBDD.

The pharmacophore models focus on the consensus regions, where all protein structures have similar requirements, but they do not place restrictions on the flexible regions of the binding site. Creating pharmacophore models in this way provides a means to identify hits with very different scaffolds and chemical features — to explore new chemistry space without bias toward reproducing the same size and functional groups of other co-crystallized ligands. This method is ideal for rapidly screening large databases of compounds, broadening and enriching the pool of compounds subsequently tested in biological assays.

We will continue developing the MPS pharmacophore method. To determine the dependence of the methodology on the starting structure used in the MD simulations, we will apply this method to the other two apo structures in the PDB (3HVP

and 3PHV). We will also examine the optimal number of snapshots to use as well as other computational methods to generate MPS. Protocols for using MPS from experimental structures will also be developed to make this method more broadly applicable.

Acknowledgment. This work has been supported by the National Institutes of Health (GM65372) and the Beckman Young Investigator Program. We thank Michael Lerner for the python scripts used in the pharmacophore clustering and Richard Smith for help generating low-energy conformers for the databases. K.L.M. would like to thank the ACS Division of Medicinal Chemistry and Bristol-Myers Squibb for a predoctoral fellowship. K.L.M. is grateful for fellowships from Edward S. Blake, Fred W. Lyons, AFPE, and the University of Michigan Regents, as well as support from the Pharmacological Sciences Training Program, GM07767 NIGMS.

Supporting Information Available: Simplification of overlapping pharmacophore elements, structures of molecules in the database, and structures of false-positive hits. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0469378