Published in final edited form as:

J Comput Aided Mol Des. 2017 November ; 31(11): 979–993. doi:10.1007/s10822-017-0077-7.

Comparing pharmacophore models derived from crystallography and NMR ensembles

Phani Ghanakota and Heather A. Carlson*

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

Abstract

NMR and X-ray crystallography are the two most widely used methods for determining protein structures. Our previous study examining NMR vs X-Ray sources of protein conformations showed improved performance with NMR structures when used in our Multiple Protein Structures (MPS) method for receptor-based pharmacophores [1]. However, that work was based on a single test case, HIV-1 protease, because of the rich data available for that system. New data for more systems are available now, which calls for further examination of the effect of different sources of protein conformations. The MPS technique was applied to Growth factor receptor bound protein 2 (Grb2), Src SH2 homology domain (Src-SH2), FK506-binding protein 1A (FKBP12), and Peroxisome proliferator-activated receptor- γ (PPAR- γ). Pharmacophore models from both crystal and NMR ensembles were able to discriminate between high-affinity, low-affinity, and decoy molecules. As we found in our original study, NMR models showed optimal performance when all elements were used. The crystal models had more pharmacophore elements compared to their NMR counterparts. The crystal-based models exhibited optimum performance only when pharmacophore elements were dropped. This supports our assertion that the higher flexibility in NMR ensembles helps focus the models on the most essential interactions with the protein. Our studies suggest that the “extra” pharmacophore elements seen at the periphery in X-ray models arise as a result of decreased protein flexibility and make very little contribution to model performance.

Introduction

Pharmacophore models are a spatial description of the chemical features required for a molecule to have optimal binding interactions with a given pocket on a biological target [2–5]. Structure-based pharmacophore models can be created from conformations of that protein pocket with or without bound ligands. Pharmacophore points can be defined through geometric rules based on the orientations of the residues in the binding sites or through information about ligand contacts with those residues. The contact information can come from bound ligands [6–10] or from the positions of fragment molecules in the binding sites

*To whom correspondence should be addressed. Tel: +1 734 615 6841; Fax: +1 734 763 2022; carlsonh@umich.edu.

Supplementary Information

Supplementary information provides additional information on pharmacophore model coordinates, used in this study, ROC plot data, and a list of PDB structures used for the NMR and crystal structures used to build the pharmacophore models.

[1,5,11–13]. A growing trend is to use more than one protein structure when creating pharmacophore models to represent the conformational flexibility inherent to each system. The structures can come from crystallography [1,7,8,10,14], NMR [1], or molecular dynamics (MD) simulations [6,9,11–16]. The resulting pharmacophore models from each structure can be combined into a single, overarching model [7,8,11,12,14,16] or used as a set. When used as a set, scientists can simply combine all hits [15], look for common hits across the pharmacophore models [6], or use machine-learning algorithms to develop a combined model for ligand binding [9,10,13].

In this study, we compare ensembles of protein conformations from crystal and NMR structures, which were readily available. Our Multiple Protein Structures (MPS) method for creating structure-based pharmacophore models (originally called “dynamic pharmacophore” models) is an experimentally verified, computational technique that leverages ensembles of protein conformations. The use of many protein conformations reveals areas of the binding site that have consistent criteria for complementarity and cause the least entropic penalty [17,18]. Each conformation of the protein binding site is mapped to determine the essential pharmacophore elements required to complement the pocket. MPS then overlays all the structures of the ensemble to identify pharmacophore sites that are common to more than 50% of the structures. This consensus of pharmacophore sites describes the essential elements that a ligand must contain to bind the target.

One of our previous studies compared the performance of pharmacophore models of HIV-1 protease derived from a collection of crystal structures and an NMR ensemble, using the MPS technique.[1] For that system, the pharmacophore models from the NMR ensemble encoded a more accurate representation of the essential features of the active site while maintaining selectivity for inhibitors over decoy molecules. This was a direct consequence of the greater flexibility observed in the NMR ensemble over the collection of crystal structures. HIV-1 protease is more flexible than most protein targets, and it is important to determine how universal this finding may be. In this study, we extended the MPS technique to several new protein targets. Again, we find that incorporating the greater protein flexibility of an NMR ensemble in the MPS method translates into an improvement in the quality and performance of pharmacophore models over those created with collections of crystal structures.

There are very few systems with both NMR and crystal structures available in the Protein Data Bank [19]. Even fewer are of biomedical interest so that databases of known inhibitors can be generated to test method performance. Growth factor receptor bound protein 2 (Grb2), Src SH2 homology domain (Src-SH2), FK506-binding protein 1A (FKBP12), and Peroxisome proliferator-activated receptor γ (PPAR- γ) protein targets met all the required criteria. We used ligand-bound crystal structures and NMR models whenever possible in order to ensure a fair comparison. However, due to the lack of such structures for FKBP12 and PPAR- γ , apo NMR structures were used for these particular proteins. Here, we demonstrate that the lesser protein conformational sampling seen in the crystal-structure ensemble leads to the identification of non-essential pharmacophore elements. In order to further probe the location and origin of these extraneous pharmacophore elements, the crystal pharmacophore models were systematically truncated. This resulted in retention of

crystal pharmacophore model performance in most of the target cases studied. In the following sections, we discuss reasons for the retention in performance upon truncation of crystal pharmacophore models and present results across all the protein targets examining the ability of NMR and crystal pharmacophore models to identify inhibitors/agonists over decoy molecules.

Lastly, we note that the only probes used to create the MPS pharmacophore models are benzene, ethane, and methanol. These simple molecules provide the most basic features required for any pharmacophore search: aromatic, hydrophobic, and hydrogen-bonding. All search methods accommodate these features, and these features can be mapped on all protein surfaces. Below, we show that this simple approach to “universal utility” has excellent discrimination. Though more complicated probes could be incorporated for a fragment-based design approach, they are not required.

Methods

Protein Preparation.

In order to ensure a uniform setup across NMR and crystal structures, all protein-ligand complexes were stripped of hydrogen atoms which were added later using Molprobit [20]. Additionally, structures were manually assigned histidine protonation state and histidine/asparagine/glutamine flips. These protein-ligand complexes were then visually inspected and corrected where necessary for errors in hydrogen placement and ligand bond orders, followed by partial charge assignment based on MMFF94 force field [21] for the ligand and AMBER ff99 force field for the rest of the protein as implemented in Molecular Operating Environment (MOE 2010.10) [22]. The hydrogen atoms were then minimized while the heavy atoms were restrained.

Probe Flooding, Minimization, and Clustering.

The protein structures extracted from the minimized protein-ligand complexes were then flooded separately with 500 molecules of benzene, ethane, and methanol probes with a flooding radius of 10 Å from the center of the active site using PyMOL (Figure 1A) [23]. The atom used to define the center of the active site for flooding in each protein target is provided in the supplemental information. The probe molecules were then subjected to a low temperature, gas-phase minimization using the Multi Unit Search for Interacting Conformers (MUSIC) routine in BOSS [24]. This type of minimization keeps the protein fixed and does not allow the probe molecules to interact with one another. The minimized probes in each individual protein structure were then grouped by a Jarvis-Patrick clustering method, and the group was represented by the probe with the best interaction energy (the “parent” of the cluster). (Figure 1C) [25]. This process was repeated for all the structures in the crystal and NMR ensembles separately. Structures from each ensemble were then aligned using our weighted RMSD method (wRMSD) (Figure 1D) [26]. The parent probes for each structure in the ensemble were then compared across the ensemble to identify consensus clusters that represent conserved interactions in more than 50% of all protein conformations/models (Figure 1E). Pharmacophore elements were centered at the average coordinates for the corresponding parent probes. The RMS deviation of the parent probes in each consensus

cluster formed the radii of the corresponding pharmacophore elements. Using this approach, Donor, Acceptor, and Doneptor (both donor and acceptor) pharmacophore elements were derived from Methanol clusters. Aromatic pharmacophore elements were derived from Benzene clusters, and hydrophobic pharmacophore elements were created from overlapping Benzene and Ethane clusters. Any pharmacophore element lying outside the 10 Å flooding sphere was not included in the final pharmacophore model.

Creation of ligand and decoy databases.

Databases of inhibitors for the four protein targets were gleaned from the ChEMBL Database [27]. $IC_{50} \leq 50$ nM was the cutoff between high-affinity and low-affinity inhibitors. As a sufficient number of high-affinity inhibitors with $IC_{50} \leq 50$ nM could not be found for Src SH2, the cutoff for high-affinity inhibitors was relaxed to $IC_{50} \leq 500$ nM (only for Src SH2). In order to ensure that the compounds in the inhibitor databases were structurally diverse, they were grouped by 85% Tanimoto similarity, calculated using the MACCS fingerprint [28] in MOE 2010.10 [22]. The inhibitors with the highest affinity in each set of similar structures were retained. A previously reported decoy database of 2324 drug like molecules obtained from the Comprehensive Medicinal Chemistry Index [29] was used as the decoy set [30]. To rule out the possibility of finding inhibitors in this decoy set that bind to protein targets under consideration, a similarity search was performed for each of the 2324 decoy molecules against the ChEMBL-derived database using a path based similarity fingerprint implemented in OEGraphSim toolkit [31]. Using a Tanimoto coefficient of 85% as a measure of similarity, if a ChEMBL molecule similar to the decoy molecules existed, its activity profile was searched in the ChEMBL database and any reported activity against the pfam family of the target under consideration resulted in the removal of the compound from the decoy set. This allowed us to establish a decoy set of molecules for each protein that is unlikely to contain any true positives. OMEGA [32] was then used to generate multiple conformations for the inhibitors and decoys, limiting the number of conformations to 300 while imposing a heavy-atoms RMSD constraint of 2 Å for rejection of similar conformers along with a 25 kcal/mol energy cutoff. The number of molecules in the high-affinity, low-affinity, and decoy sets used for each protein target are provided in Table 1.

Evaluation of Pharmacophore Models.

Performance of the pharmacophore models were evaluated using Receiver Operator Characteristic (ROC) plots. For pharmacophore models, each ligand is evaluated with simple fit/no fit criteria, rather than a score. With scores, ROC plots are generated by plotting more and more relaxed scores (Figure 2). In our use of MPS, relaxation is accomplished by two means: 1) systematically increasing the radii of the pharmacophore elements and 2) varying the number of pharmacophore elements that need to be matched to identify a molecule as a hit. Each line on the ROC plot corresponds to results obtained from a model as the radii of its elements are systematically increased from $1 \times$ to $3 \times$ RMSD (see Figure 2). Each line shows hits for ligands required to match all N features of the pharmacophore model, N-1 matches, N-2 matches, etc. To illustrate this point, the label “5/6 sites, $2.66 \times$ RMSD” indicates hits fit five of a six-feature pharmacophore model with radii set to 2.66 times the RMSD. Using this approach, each pharmacophore model under varying degrees of

relaxation was then screened against inhibitor and decoy databases using MOE 2010.10. As the database was split into high- and low-affinity data sets, this allowed us to observe if pharmacophore models exhibited selective preference for high-affinity over low-affinity inhibitors. The number of hits in the inhibitor and decoy sets were then plotted in a ROC plot. Ideally, one would expect the best results to lie in the upper left quadrant of the ROC plot, indicating that the pharmacophore models exhibit greater selectivity for inhibitors/agonists over decoys. The distance of all the data points

(distance = $\sqrt{(\%falsepositives - 0)^2 + (\%truepositives - 100)^2}$) on the ROC plot were calculated from (0,100) which represented the ideal case scenario where all inhibitors are identified and no decoys are detected by the pharmacophore model, the pharmacophore model with the least distance from (0,100) was considered to have the best performance in screening the ligand sets.

Results and Discussion

Protein conformational sampling in crystal structures and NMR ensembles.

As expected, crystal structures had limited conformational variation in their C_α backbone as compared to the NMR structures. The C_α RMSD of the crystal structures and NMR ensembles for each protein was wRMSD aligned to the highest resolution crystal structure and the first model from the NMR ensemble, respectively. The results are summarized in Table 2 and illustrate the greater flexibility in the NMR ensembles. The greater flexibility in the NMR structures is also reflected in the heavy atom RMSD of the binding-site residues in the NMR ensemble as seen in Table 3. The NMR ensembles appear to be derived from high quality data, and we believe the conformational sampling exhibited in the NMR ensemble represents true sampling of these proteins in solution. While there were no unusually large conformational changes between the NMR and crystal structures of Grb2 SH2, Src SH2, and FKBP12, the differences in the conformations of PPAR-γ NMR and crystal structures represented a more important conformational change from the perspective of identifying MPS pharmacophore models. The PPAR-γ NMR structures represented an inactive conformation of the protein, while the crystal structures represented an active conformation of the protein.

Comparison of crystal and NMR pharmacophore models.

In general, the pharmacophore models had certain characteristic traits that were common across all protein targets investigated in this study. Pharmacophore models from NMR structures had fewer pharmacophore elements which were greater in size, as compared to their crystal-pharmacophore counterparts.

Comparison of the Src SH2 pharmacophore models

Src SH2 is an important component in the auto regulation of its kinase domain; upon phosphorylation the C-terminus of the protein binds to the SH2 domain and results in the distortion of the kinase active site [33]. The SH2 domain of Src binds with high affinity to phosphorylated peptides and recognizes the peptide sequence pYEEI with high affinity [34]. The phosphotyrosine moiety of peptide ligands bind to the pY site, and the pY+1, pY+2, pY+3 sub-sites that determine specificity lie C terminal to this phosphotyrosine binding site

(Figure 3A, Figure 3B). Pharmacophore models from both crystal and NMR structures reproduced key essential features in the different sub-sites required for binding substrates. As shown in Figure 3B, the NMR pharmacophore model for Src SH2 had six pharmacophore elements compared to the ten pharmacophore elements in the crystal model (Figure 3A). The pY site in the crystal pharmacophore model had two extra elements, a donor and a doneptor that were absent in the NMR model. Additionally, an extra hydrophobic element in the pY+2 and an aromatic element in the pY+3 pockets were located in the crystal pharmacophore model. The elements in the NMR pharmacophore model represented a subset of those seen in the crystal pharmacophore model. Interestingly, while the exact location of the phosphotyrosine moiety was not mapped in the NMR and crystal pharmacophore models, a doneptor element in close proximity was identified whose location appeared shifted between the two pharmacophore models. The lone hydrogen-bonding element in the pY+3 pocket changed from a hydrogen-bond donor in the crystal pharmacophore model to a hydrogen-bond acceptor in the NMR pharmacophore model. This resulted from differing positions of a tyrosine residue lining the pY+3 pocket. Overlaying the crystal-structure ligands with the pharmacophore models as shown in Figure 3C emphasizes the observation that most elements from the NMR model overlap with the ligands in contrast to the crystal model where many elements failed to do so. This is particularly surprising because the ligands in Figure 3D are from the crystal structures, not the NMR model.

Comparison of the Grb2 SH2 pharmacophore models

Grb2 is an adaptor protein and consists of two SH3 domains and one SH2 domain. The SH2 domain of Grb2 binds to phosphorylated peptides of the general sequence pYXNX and adopts a similar fold seen in other SH2 domains such as Src SH2. Sub-sites in the protein active site that accommodate ligands follow a similar nomenclature as noted above for Src SH2 and are named pY, pY+1, pY+2 (shown in Figure 4A and Figure 4B). A key difference between the Grb2 and Src SH2 domains is Trp 121 in Grb2 which is part of the specificity determining EF loop (according to the naming convention described in Eck et al. [35]) that blocks the large pY+3 pocket seen in Src SH2. As a result, phosphotyrosine peptides that bind to Grb2 SH2 domain adopt a beta turn instead of binding in an extended conformation occupying the pY+3 sub-site in Src SH2 [36]. MPS pharmacophore models in this study were obtained for the SH2 domain of Grb2. Pharmacophore models reproduced key features of the active site which included the phosphotyrosine binding location in the pY subsite and essential interactions seen across all ligands in the pY+1 pocket (shown in Figure 4A and Figure 4B). A hydrophobic element that overlapped the benzene ring of the phosphotyrosine residue was seen in the pY subsite of both pharmacophore models (Figure 4C and Figure 4D). The pY subsite of the crystal pharmacophore model displayed an additional acceptor element not found in the NMR model. As seen in Figure 4C, this acceptor element overlaps with the carbonyl group of the amide bond linking the phosphotyrosine residue to the residue preceding it. Three pharmacophore elements in the pY+1 pocket appear in similar locations in the crystal and NMR models. The only difference between the two pharmacophore models in the pY+1 pocket was an additional doneptor element in the NMR model. The pY+2 sub site followed a similar trend and had more elements in the crystal pharmacophore model. A doneptor and hydrophobic element consistent with a key

interaction is seen in both NMR and Crystal pharmacophore model (see Figure 4C and Figure 4D).

Comparison of the FKBP12 pharmacophore models

FKBP12 is a peptidyl prolyl cis/trans-isomerase that catalyzes the isomerization of proline amide bonds in proteins and peptides, and it is known to act as an immunosuppressant in complex with FK506 or Rapamycin [37,38]. The proline ring of a substrate sits in the center of the active site, which is a hydrophobic pocket lined with aromatic and hydrophobic residues with a tryptophan residue forming the base of the pocket [39,40]. Pharmacophore models from both crystal and NMR structures identify this key hydrophobic pocket in the center of the active site as illustrated in Figure 5A and Figure 5B. While the crystal pharmacophore model identifies several elements at the periphery of the active site, the absence of most of these elements from the NMR pharmacophore model is quite apparent and can be attributed to the lack of consensus clusters in the more flexible regions of the NMR ensemble of FKBP12. An overlay of the pharmacophore models with the ligands bound to FKBP12 in the crystal structures (as seen in Figure 5C and Figure 5D) provides a more detailed understanding of the location of the elements. The only hydrogen-bonding element in the NMR pharmacophore model is a doneptor element that closely overlaps the carboxylic acid region of the proline residue in the ligands, making hydrogen-bonding interactions with the backbone of the protein (Figure 5D). It is interesting to note that some of the elements that are exclusive to the crystal pharmacophore model do not overlap with any ligands from the FKBP12 protein ligand complexes. This does not appear to be the case for the NMR model where most ligands from the crystal structures overlap with all pharmacophore elements of the NMR model.

Comparison of the PPAR- γ pharmacophore models

PPAR- γ is a ligand-activated transcriptional factor. It primarily consists of a ligand-binding domain and a DNA-binding domain [41]. PPAR- γ agonists bind to the ligand-binding domain and stabilize helix 12 located at the C-terminus, resulting in a conformational change to a closed form of helix 12. PPAR- γ agonists stabilize helix 12 through a network of hydrogen bonds involving Tyr 473 in helix 12 and several polar residues in the vicinity (His 449, His 323, and Ser 289) typically through a carboxylic acid or thiazolidinedione moiety [42]. MPS pharmacophore models in this study were obtained by flooding at the center of the active site in the ligand-binding domain. The crystal pharmacophore model displayed six elements which included a doneptor element that mapped the functional moiety of PPAR- γ agonists that stabilizes helix 12 (see Figure 6A). Four of the six elements in the crystal pharmacophore model were seen to overlay well with the crystal structure ligands. As a ligand-bound NMR ensemble for PPAR- γ was not available, we had to use an apo NMR ensemble for creating the pharmacophore model. The NMR pharmacophore model had three elements and the doneptor element mimicking the key functional moiety of PPAR- γ agonists was absent (see Figure 6). Helix 12 in the NMR ensemble (nine apo structures) sampled a wide variety of open conformations (see Figure 6) that did not resemble the well-ordered, hydrogen-bonding network seen in the agonist bound PPAR- γ crystal structures. Consequently, the flexibility and the absence of a doneptor element that mapped the PPAR- γ agonist functional moiety was expected. A hydrophobic element that

overlapped with the agonists near helix 12 in the crystal pharmacophore model appeared shifted in the NMR model and more closely mapped the location of the Tyr 473 seen in the crystal structures. It is important to note that while the NMR pharmacophore model mapped important locations of the protein, these locations were less important for ligand binding and of more relevance for the conformational change going from the inactive form of the protein to the ligand-bound, activated form. The only element in common between the NMR and crystal pharmacophore model was a hydrophobic element located at the entrance to the active site of PPAR- γ . Interestingly, the aromatic element in the center of the active site of the crystal pharmacophore model was replaced by a donor element in the NMR pharmacophore model, presumably due to the bent nature of the helix in this region that exposes a cysteine residue (Cys 285) backbone amide in the NMR ensemble.

Evaluation of MPS pharmacophore models

Pharmacophore models from NMR ensembles and crystal structures were evaluated by their ability to identify inhibitors/agonists over decoy molecules. In addition, the databases were split into high- and low-affinity data sets to provide a further measure of the effectiveness of the models. Certain unifying trends were observed across all models in general. This included the ability of the pharmacophore models to identify high-affinity inhibitors over low-affinity inhibitors while maintaining selectivity over decoy molecules. Furthermore, NMR models consistently had fewer pharmacophore elements. The “extra” elements in the X-ray models typically did not improve performance. Indeed, most elements on the periphery of the X-ray models were unnecessary. This gives the NMR models a distinct advantage, particularly for prospective application of the MPS method.

The increased flexibility in the NMR ensemble, coupled with the requirement of defining an element only when similar interactions are made across more than 50% of the protein conformations, decreases the likelihood of requiring elements in locations where proteins exhibit greater flexibility. Our goal with MPS is to identify the most essential interactions, so these more variable regions are inherently down played. Regions like these may complement some ligands, but they are not essential to all ligand binding. Furthermore, locations such as these may be associated with entropic costs. With the exception of PPAR- γ , NMR models for all protein targets exhibited optimum performance when all pharmacophore elements were required to identify a hit. This provided support to the argument that conformational flexibility in the NMR ensemble enables one to identify hot spots that closely map functional groups retained across diverse inhibitors and represent the most essential features of the active site. The poor performance of PPAR- γ NMR pharmacophore model can be attributed to the fact that the apo NMR ensemble samples an inactive form of the protein and the resulting pharmacophore model is devoid of the ability to identify agonists that bind to the active form of the protein. Nevertheless, hot spots mapped by the PPAR- γ NMR pharmacophore model represent locations of protein side chains in the active form of PPAR- γ as seen in the crystal structures and hence resemble sites relevant for conformational transition from the inactive to the active form of the protein.

Conversely, crystal pharmacophore models were too limited in protein flexibility, which resulted in the identification of more non-essential elements. The crystal model's

performance was poor when all the pharmacophore elements were used to identify hits. In contrast to NMR pharmacophore models, elements had to be dropped in order to obtain optimum performance for crystal models, which suggested that the extraneous elements identified in the crystal pharmacophore model represented non-essential sites, meaning that inhibitors/agonists may or may not have the chemical features. The addition of these elements hindered performance. Dropping elements from the crystal pharmacophore model allowed a more diverse set of inhibitors to be identified; however, it also resulted in a concomitant increase in the number of decoy molecules identified.

Across all protein targets, there was no trend on the number of elements required to be dropped from the crystal pharmacophore model in order to achieve optimum performance. The lack of such a trend highlights the uncertain nature of the task of optimizing crystal pharmacophore models for protein targets where sufficient data to evaluate model performance does not exist. It is clear that pharmacophore models derived from NMR ensembles do not suffer from such drawbacks since they can be screened by using all of their pharmacophore elements to identify a diverse set of inhibitors, completely obviating the need to drop elements to achieve optimum performance.

Our objective was to approach the problem in an unbiased manner excluding any prior knowledge of the binding mode of ligands and the function of the protein target while preparing MPS pharmacophore models. Hence, we flooded the protein structures with probe molecules at the center of the binding cavity using a 10Å flooding radius across all protein targets the same as one might center a “docking box” on the middle of a binding site. While flooding in this study was performed with 500 probe molecules, we also evaluated the effect of flooding with 2000 probe molecules and found that several non-essential sites were identified in both NMR and crystal structures, limiting the performance of crystal and NMR pharmacophore models (data not shown).

ROC plots for each protein target, characteristics that deviate from general trends discussed above, and reasons for such anomalies are elaborated in further detail for each protein target below.

Performance of the pharmacophore models of the Src SH2

The optimal NMR pharmacophore model (6/6 sites, $2.66 \times \text{RMSD}$, Figure 7) identified 93.7% of high-affinity inhibitors and 10.5% decoys. The best performing crystal pharmacophore model (7/10, $2.66 \times \text{RMSD}$, Figure 7) identified a similar number of inhibitors (93.7%) at the expense of 11.9% decoys. Pharmacophore models from both NMR and crystal structures were able to distinguish high-affinity inhibitors from low-affinity inhibitors. While optimal pharmacophore models from both models identified a similar number of inhibitors, it is important to note that the NMR pharmacophore model achieved this task using all the elements in the pharmacophore model, unlike the crystal pharmacophore model where three elements had to be dropped in order to achieve a similar result.

Performance of the pharmacophore models for Grb2

Pharmacophore models from both crystal and NMR Grb2 models were successful in differentiating inhibitors over decoys and high-affinity over low-affinity inhibitors. The NMR pharmacophore model of Grb2 displayed optimum performance when all elements of the pharmacophore model were included (8/8 sites, $3 \times$ RMSD, Figure 8). This pharmacophore model identified 98.3% of high-affinity inhibitors and 5.6 % of decoys. In contrast, the crystal pharmacophore model exhibited optimum performance identifying 98.3% of high-affinity inhibitors and 7.5% of decoys when screened with a model (8/9, $3 \times$ RMSD, Figure 8) where one element was dropped. The crystal pharmacophore models displayed similar characteristics observed for Src SH2 where progressively dropping pharmacophore elements improved the ability of the models to identify a diverse set of inhibitors.

Performance of the pharmacophore models of FKBP12

The FKBP12 NMR pharmacophore model was rather simple and had only four pharmacophore elements. Its optimum model (4/4, $1 \times$ RMSD, Figure 9) identified 73% of the high-affinity inhibitors and 10.3 % of decoys. In contrast, the crystal pharmacophore model identified 14 pharmacophore elements. Elements of the NMR model represented a subset of the crystal pharmacophore model. The inability to identify either inhibitors or decoys using the full crystal pharmacophore model points to the fact that several of these sites are not required for inhibitors to exhibit activity, and they severely limit performance of the crystal pharmacophore model. The crystal pharmacophore model showed decent performance only when half of the pharmacophore elements were dropped, and the optimum model (7/14, $2.66 \times$ RMSD, Figure 9) identified 80.7% of high-affinity inhibitors at the expense of 28.4% decoys.

Performance of the pharmacophore models of the PPAR- γ

PPAR- γ presented a unique case where the crystal pharmacophore model displayed better performance than the NMR pharmacophore model. These differences in performances reflect the different conformations upon which crystal and NMR pharmacophore models were built. The crystal model was derived from protein conformations of the active form where helix 12 sampled a “bound conformation” in all of the ligand-bound crystal structures. In contrast, the NMR pharmacophore model was based on apo, inactive conformations in the NMR ensemble where helix 12 was found in the open conformation.

The optimum crystal pharmacophore model (5/6, $2.33 \times$ RMSD, Figure 10) identified 90.7% of high-affinity agonists and 21.6 % decoys. When all six pharmacophore elements of the crystal model were used to screen for agonists, very few were identified as hits. Moreover, under such a constraint, the crystal pharmacophore displayed a selective preference for low-affinity agonists over high-affinity agonists. Similar results were found for our previous MPS models for the protein dihydrofolate reductase derived from crystal structures [30,43].

The NMR pharmacophore model performed poorly in identifying both high-affinity and low-affinity agonists, and again the flipped specificity was seen. The optimum NMR pharmacophore model (3/3, $1.33 \times$ RMSD, Figure 10) identified 69.7% low-affinity

inhibitors and 37.5% decoy molecules. This emphasizes the importance of pharmacophore elements absent in the NMR pharmacophore model that interact with the tyrosine residue (Tyr 473) in helix 12 and their key role in defining a high-affinity PPAR- γ agonists [42].

Locating and characterizing extraneous elements in crystal pharmacophore models.

Extraneous elements limited the performance of crystal pharmacophore models, and the models required these extra elements to be dropped (at random) in order to achieve optimal performance. Given that crystal structures represent the predominant method of structure determination, it is important to develop methods that circumvent these issues. While dropping elements at random to achieve optimum performance presents one such alternative, it is unclear how to do this in a prospective case since this seems to vary by protein target. In order to investigate the location of these extraneous elements and their impact on model performance, we removed pharmacophore elements beyond a defined cutoff radius from the center of the active site. Each resulting truncated model was screened against inhibitor and decoy molecules as was done earlier.

Removing elements at the periphery of the active site resulted in pharmacophore models that exhibited optimum performance when screened using all of their pharmacophore elements. PPAR- γ (Figure 11) exhibited such behavior when screened after using a cutoff of 7 Å and 8 Å, respectively, to truncate pharmacophore models. Interestingly, minor imperfections in the pharmacophore model such as the identification of low-affinity inhibitors over high-affinity inhibitors seen with the original PPAR- γ crystal pharmacophore model disappeared using a cutoff-based truncation (compare lower, left-hand corner of the ROC plots in Figure 10 and Figure 11). This is particularly important when the active form was only available in crystal structures.

For Grb2, although the exact data points are slightly different in the ROC plots of the full X-ray model (Figure 8) vs the truncated models (Figure 12), they all show excellent performance with optimal models close to the upper, left-hand corner. Similar results were seen for FKBP12, where removing pharmacophore elements beyond 8 or 9 Å of the center of the active site resulted in models that required less elements to be dropped in order to achieve optimal performance (see Figure 13). In fact, the 9 Å truncated model gives the exact same data points as the full models. It appears that crystal packing effects cause side chains at the periphery of this active site to be artificially constrained, creating false consensus elements in the pharmacophore models. This is clearly seen in Figure 14. Coloring the backbone based on C α RMSD shows that most of the pharmacophore elements unique to the crystal pharmacophore model drop out in the NMR pharmacophore model due to higher protein flexibility.

Interestingly, when cutoffs were applied to the model of Src SH2, degradation in performance was observed (see Figure 15). This was due to the removal of pharmacophore elements located at the periphery in the pY+3 pocket. The pY+3 pocket is known to determine specificity and contributes significantly to the binding affinity of inhibitors. Inhibitors with larger hydrophobic residues in the pY+3 pocket are known to bind with higher affinity to Src SH2 [44]. Clearly, all the elements of the initial X-ray model are relevant for this system.

Conclusions

MPS pharmacophore models displayed selective preference for high-affinity inhibitors over low-affinity inhibitors. NMR pharmacophore models exhibited optimum performance when screened using all of their pharmacophore elements, an observation that lends support to the argument that greater flexibility in the NMR ensembles aids in identifying only essential pharmacophore elements. In contrast, crystal pharmacophore models identified a greater number of pharmacophore elements, some of which had to be dropped in order to improve performance. However, the number of elements to drop varied across the protein targets, which may make them more difficult to use in prospective structure-based drug discovery.

In order to understand the location and impact of these extraneous pharmacophore elements on model performance, we have truncated pharmacophore models using different cutoffs from the center of the active site. This improved the PPAR- γ model significantly. The Grb2 and FKBP12 models retained their optimum performance, clearly showing the distal elements were inconsequential. However, all the elements in the Src SH2 model were essential. These results confirm that the extraneous pharmacophore elements are primarily located at the periphery and frequently do not contribute any value to model performance. This study highlights the relationship of protein flexibility with MPS pharmacophore model performance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We greatly appreciate the generous donation of the MOE software from Chemical Computing Group and OMEGA and OEGraphSim from OpenEye Scientific Software, Inc. This work has been supported by the National Institutes of Health (GM65372).

References

1. Damm KL, Carlson HA. (2007) Exploring Experimental Sources of Multiple Protein Conformations in Structure-Based Drug Design. *J Am Chem Soc* 129: 8225–8235 [PubMed: 17555316]
2. Sanders MPA, McGuire R, Roumen L, de Esch IJP, de Vlieg J, Klomp JPG, de Graaf C. (2012) From the Protein's Perspective: the Benefits and Challenges of Protein Structure-Based Pharmacophore Modeling. *Med Chem Commun* 3: 28–38
3. Qing X, Lee XY, De Raeymaeker J, Tame JRH, Zhang KYJ, De Maeyer M, Voet ARD. Pharmacophore Modeling: Advances, Limitations, and Current Utility in Drug Discovery. (2014) *J Recept Lig Channel Res* 7: 81–92
4. Meslamani J, Rognan D. (2015) Protein-Ligand Pharmacophores: Concept, Design and Applications. *CICSJ Bulletin* 33: 27–32
5. Koes DR. (2016) Pharmacophore Modeling: Methods and Applications. In: Zhang W (eds) *Computer-Aided Drug Discovery. Methods in Pharmacology and Toxicology* Humana Press, New York, NY pp. 167–188
6. Wieder M, Garon A, Perricone U, Boresch S, Seidel T, Almerico AM, Langer T.. (2017) Common Hits Approach: Combining Pharmacophore Modeling and Molecular Dynamics Simulations. *J Chem Inf Model* 57: 365–385 [PubMed: 28072524]
7. Zou J, Xie H-Z, Yang S-Y, Chen J-J, Ren J-X, Wei Y-Q. (2008) Towards More Accurate Pharmacophore Modeling: Multicomplex-Based Comprehensive Pharmacophore Map and Most-

- Frequent-Feature Pharmacophore Model of CDK2. *J Mol Graph Model* 27: 430–438 [PubMed: 18786843]
8. Wu F, Xu T, He G, Ouyang L, Han B, Peng C, Song X, Xiang M. (2012) Discovery of Novel Focal Adhesion Kinase Inhibitors Using a Hybrid Protocol of Virtual Screening Approach Based on Multicomplex-Based Pharmacophore and Molecular Docking. *Int J Mol Sci* 13: 15668–15678 [PubMed: 23443087]
 9. Rakers C, Schumacher F, Meinel W, Glatt H, Kleuser B, Wolber G. (2016) In Silico Prediction of Human Sulfotransferase 1E1 Activity Guided by Pharmacophores from Molecular Dynamics Simulations. *J Biol Chem* 291: 58–71 [PubMed: 26542807]
 10. Tian S, Sun H, Li Y, Pan P, Li D, Hou T. (2013) Development and Evaluation of an Integrated Virtual Screening Strategy by Combining Molecular Docking and Pharmacophore Searching Based on Multiple Protein Structures. *J Chem Inf Model* 53: 2743–2756 [PubMed: 24010823]
 11. Wenbo Y, Lakkaraju SK, Raman EP, MacKerell AD Jr. (2014) Site-Identification by Ligand Competitive Saturation (SILCS) Assisted Pharmacophore Modeling. *J Comput Aided Mol Des* 28:491–507 [PubMed: 24610239]
 12. Wenbo Y, Lakkaraju SK, Raman EP, Fang L, MacKerell AD Jr. (2015) Pharmacophore Modeling Using Site-Identification by Ligand Competitive Saturation (SILCS) with Multiple Probe Molecules. *J Chem Inf Model* 55:407–420 [PubMed: 25622696]
 13. Spyraakis F, Benedetti P, Decherchi S, Rocchia W, Cavalli A, Alcaro S, Ortuso F, Baroni M, Cruciani G. (2015) A Pipeline to Enhance Ligand Virtual Screening: Integrating Molecular Dynamics and Fingerprints for Ligand and Proteins. *J Chem Inf Model* 55: 2256–2274 [PubMed: 26355717]
 14. Bowman AL, Makriyannis A. (2011) Approximating Protein Flexibility through Dynamic Pharmacophore Models: Application to Fatty Acid Amide Hydrolase (FAAH). *J Chem Inf Model* 51: 3247–3253 [PubMed: 22098169]
 15. Choudhury C, Priyakumar UD, Sastry GN. (2015) Dynamics Based Pharmacophore Models for Screening Potential Inhibitors of Mycobacterial Cyclopropane Synthase. *J Chem Inf Model* 55: 848–860 [PubMed: 25751016]
 16. Baptista SJ, Silva MMC, Moroni E, Meli M, Colombo G, Dinis TCP, Salvador JAR. (2017) Novel PARP-1 Inhibitor Scaffolds Disclosed by a Dynamic Structure-Based Pharmacophore Approach. *PLoS ONE* 12: e0170846 [PubMed: 28122037]
 17. Carlson HA, Masukawa KM, McCammon JA. (1999) Method for Including the Dynamic Fluctuations of a Protein in Computer-Aided Drug Design. *J Phys Chem A* 103: 10213–10219
 18. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, Briggs JM, McCammon JA. (2000) Developing a Dynamic Pharmacophore Model for HIV-1 Integrase. *J Med Chem* 43: 2100–2114 [PubMed: 10841789]
 19. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235–242 [PubMed: 10592235]
 20. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. (2010) MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr D Biol Crystallogr* 66: 12–21 [PubMed: 20057044]
 21. Halgren TA. (1996) Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J Comput Chem* 17: 490–519
 22. Molecular Operating Environment (2010) Chemical Computing Group Inc: Montreal, Canada
 23. The PyMOL Molecular Graphics System (2010) Schrödinger LLC: New York, New York
 24. Jorgensen WL. (2000) BOSS; Yale University: New Haven, CT
 25. Lerner MG, Meagher KL, Carlson HA. (2008) Automated Clustering of Probe Molecules from Solvent Mapping of Protein Surfaces: New Algorithms Applied to Hot-Spot Mapping and Structure-Based Drug Design. *J Comput-Aided Mol Des* 22: 727–736 [PubMed: 18679808]
 26. Damm KL, Carlson HA. (2006) Gaussian-Weighted RMSD Superposition of Proteins: A Structural Comparison for Flexible Proteins and Predicted Protein Structures. *Biophys J* 90: 4558–4573 [PubMed: 16565070]

27. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 40: D1100–D1107 [PubMed: 21948594]
28. Durant JL, Leland BA, Henry DR, Nourse JG. (2002) Reoptimization of MDL Keys for Use in Drug Discovery. *J Chem Inf Model* 42: 1273–1280
29. Comprehensive Medicinal Chemistry Database (2003) MDL Information Systems: San Leandro, CA
30. Bowman AL, Lerner MG, Carlson HA. (2007) Protein Flexibility and Species Specificity in Structure-Based Drug Discovery: Dihydrofolate Reductase as a Test System. *J Am Chem Soc* 129: 3634–3640 [PubMed: 17335207]
31. OEGraphSim (2010) OpenEye Scientific Software Inc: Santa Fe, NM
32. OMEGA (2010) OpenEye Scientific Software Inc: Santa Fe, NM
33. Xu W, Harrison SC, Eck MJ. (1997) Three-Dimensional Structure of the Tyrosine Kinase c-Src. *Nature* 385: 595–602 [PubMed: 9024657]
34. Zhou S, Shoelson SE, Chaudhuri M, Gish G, Pawson T, Haser WG, King F, Roberts T, Ratnofsky S, Lechleider RJ, Neel BG, Birge RB, Fajardo JE, Chou MM, Hanafusa H, Schaffhausen B, Cantley LC. (1993) SH2 Domains Recognize Specific Phosphopeptide Sequences. *Cell* 72: 767–778 [PubMed: 7680959]
35. Eck MJ, Shoelson SE, Harrison SC. (1993) Recognition of a High-Affinity Phosphotyrosyl Peptide by the Src Homology-2 Domain of p56lck. *Nature* 362: 87–91 [PubMed: 7680435]
36. Rahuel J, Gay B, Erdmann D, Strauss A, Garcia-Echeverria C, Furet P, Caravatti G, Fretz H, Schoepfer J, Grutter MG. (1996) Structural Basis for Specificity of GRB2-SH2 Revealed by a Novel Ligand Binding Mode. *Nat Struct Mol Biol* 3: 586–589
37. Brown EJ, Albers MW, Shin TB, Ichikawa K, Keith CT, Lane WS, Schreiber SL. (1994) A Mammalian Protein Targeted by G1-Arresting Rapamycin-Receptor Complex. *Nature* 369: 756–758 [PubMed: 8008069]
38. Liu J, Farmer JD Jr, Lane WS, Friedman J, Weissman I, Schreiber SL. (1991) Calcineurin Is a Common Target of Cyclophilin-Cyclosporin A and FKBP-FK506 Complexes. *Cell* 66: 807–815 [PubMed: 1715244]
39. Hamilton GS, Steiner JP. (1998) Immunophilins: Beyond Immunosuppression. *J Med Chem* 41: 5119–5143 [PubMed: 9857082]
40. Schreiber SL. (1991) Chemistry and Biology of the Immunophilins and Their Immunosuppressive Ligands. *Science* 251: 283–287 [PubMed: 1702904]
41. Tontonoz P, Spiegelman BM. (2008) Fat and Beyond: The Diverse Biology of PPAR γ . *Annu Rev Biochem* 77: 289–312 [PubMed: 18518822]
42. Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH, Kurokawa R, Rosenfeld MG, Willson TM, Glass CK, Milburn MV. (1998) Ligand Binding and Co-Activator Assembly of the Peroxisome Proliferator-Activated Receptor- γ . *Nature* 395: 137–143 [PubMed: 9744270]
43. Lerner MG, Bowman AL, Carlson HA. (2007) Incorporating Dynamics in E. Coli Dihydrofolate Reductase Enhances Structure-Based Drug Discovery. *J Chem Inf Model* 47: 2358–2365 [PubMed: 17877338]
44. Bradshaw JM, Grucza RA, Ladbury JE, Waksman G. (1998) Probing the “Two-Pronged Plug Two-Holed Socket” Model for the Mechanism of Binding of the Src SH2 Domain to Phosphotyrosyl Peptides: A Thermodynamic Study. *Biochemistry* 37: 9083–9090 [PubMed: 9636054]

Figure 1.

A detailed description of the MPS method is presented using benzene probes as an example.

A) The protein active site is flooded with 500 probe molecules. B) These molecules are minimized, independent of each other using the MUSIC routine in BOSS. C) The probe molecules are clustered and represented by “parent probes” which are the single benzene molecule with the best interaction energy in each cluster in B. D) Steps A-C are carried out on all structures in the ensemble, and they are overlaid using the wRMSD method. E) Clusters of “parent probes” are identified manually. F) Consensus clusters are identified when at least 50% of the conformations contain a parent probe in the same location. All probes farther than 10Å of the center of the active site are ignored. The center of each pharmacophore element is derived from the center of mass of the parents in the consensus cluster, and the radius of the element is set by the RMSD of the parents.

Figure 2.

Receiver Operator Characteristic plots are shown for two cases, the first one corresponds to an exaggerated case of the more traditional use of ROC plots for continuous scores such as those obtained from docking results. The second plot illustrates the ROC plots generated by screening MPS pharmacophore models. Each discrete point on this line corresponds to the pharmacophore screening results obtained by gradually increasing the radii of the pharmacophore elements from 1× to 3× RMSD. Thus a label “5/6 2.33×” denotes the screening results from a MPS pharmacophore model whose radii have been multiplied by 2.33 and requires 5 of its 6 pharmacophore elements to be matched for a hit to be identified.

Figure 3.

MPS pharmacophores are shown with $1 \times$ RMSD radii, which indicates tighter or looser position constraints. Pharmacophore models for all the protein targets are color coded to represent different interactions: Red – Donor, Blue – Acceptor, Purple – Doneptor, Green – Aromatic, and Cyan – Hydrophobic. A) The MPS pharmacophore model for Src SH2 derived from X-ray structures. B) The MPS pharmacophore model for Src SH2 derived from the NMR ensemble. C) The ligands from X-ray structures overlaid on top of the Src SH2 X-ray model. D) The ligands from X-ray structures overlaid on top of the Src SH2 NMR model.

Figure 4.

Coloring and radii of the pharmacophore elements are the same as in Figure 3. A) The MPS pharmacophore model for GRB2 SH2 derived from X-ray structures. B) The MPS pharmacophore model for GRB2 SH2 derived from the NMR ensemble. C) The ligands from X-ray structures overlaid on top of the GRB2 SH2 MPS X-ray pharmacophore model. D) The ligands from X-ray structures overlaid on top of the GRB2 SH2 MPS NMR pharmacophore model. In Figure 4A and Figure 4B, tryptophan 121 is rendered as sticks between the pY+1 and pY+2 surfaces.

Figure 5.

Coloring and radii of the pharmacophore elements are the same as in Figure 3. A) The MPS pharmacophore model for FKBP12 derived from X-ray structures. B) The MPS pharmacophore model for FKBP12 derived from the NMR ensemble. C) The ligands from X-ray structures overlaid on top of the FKBP12 X-ray model. D) The ligands from X-ray structures overlaid on top of the FKBP12 NMR model.

Figure 6.

Coloring and radii of the pharmacophore elements are the same as in Figure 3. MPS pharmacophore models derived from PPAR- γ X-ray and NMR ensembles are shown along with a representative protein conformation. The tyrosine residue 473 which is part of Helix 12 is shown in pink. In the X-ray ensemble, there is limited sampling of the tyrosine residue 473, which corresponds to the active form of the protein. In the NMR ensemble; this residue samples the inactive conformation of the protein. A) The MPS pharmacophore model for PPAR- γ derived from X-ray structures. B) The MPS pharmacophore model for PPAR- γ derived from the NMR ensemble. C) The ligands from X-ray structures overlaid on top of the PPAR- γ MPS X-ray pharmacophore model. D) The ligands from X-ray structures overlaid on top of the PPAR- γ MPS NMR pharmacophore model. In Figure 6A and Figure 6B, the location of the binding site is shown by rendering rosiglitazone as a stick model obtained from the PDB ID: 1ZGY (colored brown).

Figure 7.

ROC plots of crystal and NMR pharmacophore models of Src SH2 are shown along with a label for the model that displays the best performance. The best performing crystal pharmacophore model was at $2.66 \times$ RMSD using seven out of ten pharmacophore elements. The best performing NMR pharmacophore model was achieved at $2.66 \times$ RMSD using all six pharmacophore elements. The average area under the curve (AUC) for the best performing pharmacophore models are shown in parenthesis.

Figure 8.

ROC plots of crystal and NMR pharmacophore models of Grb2 SH2 are shown along with a label for the model that displays the best performance. The best performing crystal pharmacophore model was at $3.00 \times$ RMSD using eight out of nine pharmacophore elements. The best performing NMR pharmacophore model was achieved at $3.00 \times$ RMSD using all six pharmacophore elements. The average AUC for the best performing pharmacophore models are shown in parenthesis.

Figure 9.

ROC plots of crystal and NMR pharmacophore models of FKBP12 are shown along with a label for the model that displays the best performance. The best performing crystal pharmacophore model was at $2.66 \times$ RMSD using seven out of fourteen pharmacophore elements. The best performing NMR pharmacophore model was achieved at $1.00 \times$ RMSD using all four pharmacophore elements. The average AUC for the best performing pharmacophore models are shown in parenthesis.

Figure 10.

ROC plots of crystal and NMR pharmacophore models of PPAR- γ are shown along with a label for the model that displays the best performance. The best performing crystal model was at $2.33 \times$ RMSD using five out of six pharmacophore elements. The best performing NMR model was achieved at $1.33 \times$ RMSD using all three pharmacophore elements. The NMR pharmacophore model was built from an NMR ensemble that samples the inactive conformation, so this model was expected to perform poorly. The average AUC for the best performing pharmacophore models are shown in parenthesis.

Figure 11.

ROC plots are shown for the PPAR- γ crystal pharmacophore model with cutoffs of 8 Å and 7 Å from the center of the binding site. The best performing crystal pharmacophore model at a cutoff of 8 Å was $1.00 \times \text{RMSD}$ using four out of five pharmacophore elements and $1.33 \times \text{RMSD}$ using four out of four pharmacophore elements for a cutoff of 7 Å. In the 8 Å ROC plot, the AUC are averages of the five of five and four of five curves.

Figure 12.

ROC plots of Grb2 SH2 crystal model truncated to 8Å and 7Å are shown. The best performing model at a cutoff of 8Å was $3.00 \times \text{RMSD}$ using eight out of eight pharmacophore elements and $2.33 \times \text{RMSD}$ using six out of six pharmacophore elements for a cutoff of 7Å. The AUC are averages for the three curves in the 8Å ROC plot and two curves in the 7Å ROC plot.

Figure 13.

ROC plots of FKBP12 crystal pharmacophore model truncated to 9Å and 8Å are shown along with a label for the model that displays the best performance. The best performing crystal pharmacophore model at a cutoff of 9Å was $2.66 \times \text{RMSD}$ using seven out of thirteen pharmacophore elements and $3.00 \times \text{RMSD}$ using seven out of ten pharmacophore elements for a cutoff of 8Å.

Figure 14.

A) FKBP12 crystal pharmacophore model and B) FKBP12 NMR pharmacophore model. The models are overlaid on the protein which is color coded by C_α RMSD after wRMSD alignment. The color scale ranges from Blue (0.1 Å RMSD) to Red (3.8 Å RMSD). The increased flexibility of the NMR ensemble reduces the consensus across the probes used in constructing the model, which removes several elements present in the crystal pharmacophore model.

Figure 15.

ROC plots of Src SH2 crystal pharmacophore model truncated to 9Å and 8Å are shown. The best performing crystal pharmacophore model at a cutoff of 9Å was $2.66 \times \text{RMSD}$ using six out of eight pharmacophore elements and $2.00 \times \text{RMSD}$ using five out of six pharmacophore elements for a cutoff of 8Å.

Table 1.

The number of high-affinity and low-affinity inhibitors/agonists used to validate MPS pharmacophore models are reported below for every protein target.

Protein	# High-Affinity	# Low-Affinity	# Decoys
Grb2	61	97	2303
Src-SH2	16	145	2303
FKBP12	78	96	2324
PPAR-γ	54	119	2149

Table 2.

The range of pair-wise C_α RMSDs for all the crystal and NMR structures reflects the greater flexibility in the NMR ensemble.

Protein	Crystal Structures C _α RMSD	NMR Ensemble C _α RMSD
Grb2	0.22 – 0.68 Å	0.74 – 2.32 Å
Src SH2	0.16 – 0.91 Å	0.82 – 1.68 Å
FKBP12	0.51 – 0.98 Å	0.95 – 1.43 Å
PPAR-γ	0.69 – 2.43 Å	1.36 – 2.8 Å

Table 3.

The range of heavy-atom RMSDs in the binding site for the crystal structures and NMR ensembles provides support for the greater flexibility of the NMR ensemble.

Protein	Crystal Structures Binding Site RMSD	NMR Ensemble Binding Site RMSD
Grb2	0.33 – 1.71 Å	1.02 – 2.08 Å
Src SH2	0.29 – 1.18 Å	1.00 – 1.79 Å
FKBP12	0.64 – 2.57 Å	1.58 – 3.95 Å
PPAR-γ	1.37 – 2.03 Å	1.49 – 3.21 Å