JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

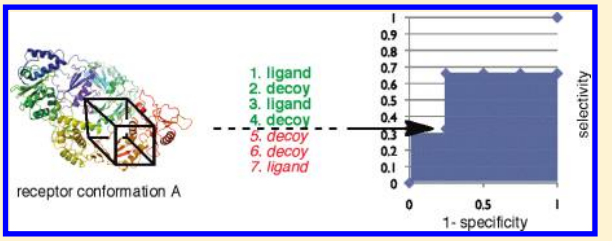
Predictive Power of Molecular Dynamics Receptor Structures in Virtual Screening

Sara E. Nichols,[*,†,‡] Riccardo Baron,[*,†,§] Anthony Ivetac,[†,‡] and J. Andrew McCammon[†,‡,§,⊥]

[†]Department of Chemistry and Biochemistry, [‡]Department of Pharmacology, [§]Center for Theoretical Biological Physics, and [⊥]Howard Hughes Medical Institute[4], University of California San Diego, La Jolla, California 92093-0365, United States

Supporting Information

ABSTRACT: Molecular dynamics (MD) simulation is a well-established method for understanding protein dynamics. Conformations from unrestrained MD simulations have yet to be assessed for blind virtual screening (VS) by docking. This study presents a critical analysis of the predictive power of MD snapshots to this regard, evaluating two well-characterized systems of varying flexibility in ligand-bound and unbound configurations. Results from such VS predictions are discussed with respect to experimentally determined structures. In all cases, MD simulations provide snapshots that improve VS predictive power over known crystal structures, possibly due to sampling more relevant receptor conformations. Additionally, MD can move conformations previously not amenable to docking into the predictive range.



INTRODUCTION

Molecular docking algorithms attempt to determine the binding modes of small organic molecules relative to a biomolecular receptor and to evaluate a score representing their relative binding propensity. In an effort to find novel binders for "hit" identification in structure-based drug discovery, virtual screening (VS) entails utilizing a docking algorithm to rank large libraries of compounds. Receptor coordinates are most commonly provided by X-ray crystallography experiments as well as homology modeling or computer simulation. The nature of the receptor model employed affects the predictive performance of dock-based approaches, as different conformations can produce alternative rankings of possibly active and inactive compounds, and only approximates the dynamic process occurring. Although a number of strategies to incorporating protein flexibility have been developed in this context (see e.g. refs 1−4 and references therein), defining protocols to select receptor structures for blind VS predictions is difficult.[5−10]

Modeling the natural dynamics of a protein for ligand-binding events can benefit from methods that use multiple target configurations, so-called ensemble approaches, but not without limitations and trade-offs between sufficient model reliability and computational costs.[11,12] Previous studies focused on crystallographic and homology models to examine single receptor effects on VS ranking performance[13−18] as well as on the advantages of using multiple structures.[11,13,17,18] Yet, to our knowledge, no critical assessment of VS predictive power using individual receptor conformations from molecular dynamics (MD) simulations has been reported to date. This raises two general questions: Are snapshots from MD simulations predictive, and how do they compare to X-ray structures in affecting VS predictive power? How do structures from the different types of MD ensembles affect VS predictions?

VS of MD snapshots have been successfully used for pose prediction and compound library ranking.[3,19−22] In some cases, clustering algorithms can alleviate computational costs by reducing the MD ensemble with no significant loss of information for VS approaches.[3,23] However, depending on molecular flexibility and binding properties, favorable protein−ligand complexes can form at varying frequencies along typical MD sampling time scales. For example, rare protein configurations have been shown to determine ligand binding in FKBP.[21] In other cases the dominant, frequent protein configurations are those promoting the best binding conditions for a variety of ligands.[3,23] In the present study VS predictive power, using MD snapshots and X-ray structures for two model systems, was explored.

The first model system selected was HIV-1 reverse transcriptase (RT; Figure 1a). RT catalyzes the transcription of the single-stranded RNA viral genome into a double-stranded DNA form and is essential for HIV replication. As a major drug target, RT is the subject of substantial structural biology efforts, resulting in more than a hundred related crystal structures to date. Together with computational studies, the heterogeneous properties of RT structures suggest considerable plasticity, which has been interpreted in the context of its function as both a DNA polymerase and ribonuclease. Current FDA-approved anti-RT drugs bind to one of two identified sites: the polymerase active site or a nearby hydrophobic allosteric site targeted by non-nucleoside reverse transcriptase inhibitors (NNRTIs).[24] The NNRTI binding pocket (NNIBP; Figure 1c) was the focus of the current work reported here, as it is of significant pharmaceutical interest and was suggested to be remarkably flexible, fluctuating between a

Received: March 9, 2011
Published: May 02, 2011

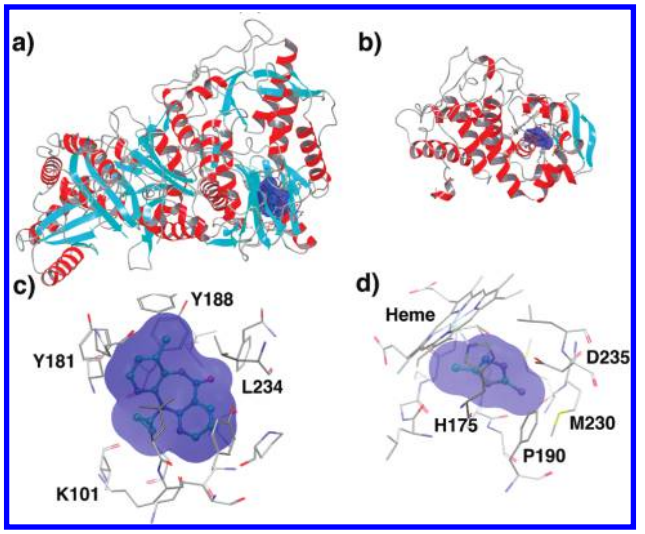


Figure 1. Protein receptors considered in this study: (a) RT and (b) W191G overall representations on the same scale. Secondary structure elements and the location of the binding sites are highlighted (red: helices; cyan: sheets; and gray: loops and turns). Insight views for: (c) the RT NNRTI binding pocket (NNIBP) with nevirapine bound and (d) the W191G cation-binding pocket with 2a5mt bound. Ligands (balls and sticks) and pocket volumes (blue surfaces) are also shown.

"collapsed" inhibitor-free state and an "open" inhibitor-bound state (see, e.g., refs 24 and 25 and references therein). Moreover, the NNIBP has been shown to bind to a broad range of NNRTIs, which bear structurally diverse scaffolds and was considered representative of allosteric binding sites.[24]

The second model system was the less flexible W191G artificial cavity mutant of cytochrome c peroxidase (W191G; Figure 1b) for which ligand binding was characterized by X-ray crystallography[26−29] and isothermal titration calorimetry experiments[28] as well as by computational techniques.[23,26,29,30] Taken as a whole, numerous X-ray structures[28,29,31] and MD simulations[3,23] showed that no major backbone conformational rearrangement occurs upon binding of a large number of small cation compounds. An exception to this general observation is made for three bulky ligands, which upon binding were found to induce the partial opening of the 190−195 gating loop.[27,28] In all cases, the electrostatic interaction between the cation ligands with Asp235, at the bottom of the binding pocket (Figure 1d), is crucial for binding.[31] In line with these distinctive properties, W191G ligand binding is a prototype of lowest-possible change of receptor entropy.[31] RT and W191G binding sites are ideal for evaluating the dependence of VS predictive power on diverse input receptor structures.

The subsequent sections describe analysis of blind VS to different (unbound and bound) MD and crystallographic structures of the RT and W191G systems. Predictive power was characterized using receiver operating characteristic (ROC) curves to determine the probability of ranking actives over inactives.[32,33] In this regard, the goal of this study was to determine if individual MD snapshots were preferable for ranking actives over inactive compared to individual crystallographic structures. Additionally, strategies for selection of MD conformations a priori were examined. The current article presents a procedure for nonarbitrary assessment and comparison of VS predictive power, of general interest for computational drug discovery.

MATERIALS AND METHODS

Molecular Models. A total of 15 representative RT crystal structures were selected for initial analysis. These crystallographic conformations represent a combination of 10 diverse NNRTI-bound and five NNRTI-free states. In agreement with previous studies,[25,34] the NNIBP was defined as including Pro 95, Leo 100, Lys 101, Lys 103, Val 106, Val 108, Val 179, Tyr 181, Tyr 188, Gly 190, Phe 227, Trp 229, Leu 234, His 235, Pro 236, and Tyr 318 and is displayed in Figure 1c.

For W191G, a total of 49 X-ray crystallographic structures were deposited in the Protein Data Bank (PDB) and considered for our study (45 of which were cocrystallized with a ligand; for a complete list of the corresponding PDB entries see Supporting Information, Table SI).[26−29] Out of the 45 binders reported, modest binding site rearrangements are limited to amino acid side chains in 42 cases, thus excluding the peptide backbone. In agreement with previous studies,[3,23,30] W191G binding pocket includes His 175, Leu 177, Lys 179, Thr 180, Pro 190, Asn 195, Phe 202, Met 230, Met 231, Leu 232, and Asp 235 displayed in Figure 1d. Pro 190 and Asn 195 residues are the hinge points of W191G cavity gating loop. Crystallographic structures are listed in the Supporting Information, Table SI.

MD Simulations. MD simulations of four HIV-1 RT systems (983 residues) in explicit water at 300 K were performed using the GROMACS simulation software,[35,36] the GROMOS 53A6 force field,[37,38] and a compatible simple point charge (SPC) water model.[39] The four different systems included two bound systems, 1VRT[40] structure with α-APA bound to NNIBP (α-APA bound) and 1VRT structure with UC-781 bound to NNIBP (UC-781 bound) as well as two unbound systems, 1VRT structure with the NNRTI extracted from NNIBP (unbound open) and 1DLO structure[41] (unbound closed). For each system, four independent 30 ns trajectories were generated, with the final 12.5 ns of each used in our analysis. MD snapshots were extracted every 20 ps from this cumulative total of 50 ns of simulation time, yielding 2500 snapshots for each system and a total of 10 000 RT snapshots. For computational details refer to ref 25.

Three MD trajectories of the W191G cavity mutant of cytochrome c peroxidase from *Escherichia coli* (W191G; 290 residues) in explicit water at 300 K were generated using the GROMOS05 software,[42] the GROMOS 45A4 force field,[43] and a compatible SPC water model.[44] The three systems included an initial closed-gate W191G-2a5mt complex from PDB: 1AEN (2a5mt bound), an initial open-gate W191G from PDB: 1RYC after extracting ligand (unbound open), and an initial closed-gate W191G from PDB: 1AA4 (unbound closed).[27] MD snapshots were extracted every 20 ps, for a total of 2500 snapshots from 50 ns of cumulative trajectory for each complex, totaling 7500 W191G receptor structures. MD simulations of both RT and W191G systems have been previously described and analyzed. For computational details refer to refs 23 and 30.

Known Ligand and Decoy Compounds. In the case of RT, 20 diverse known inhibitors were combined with the National Cancer Institute Diversity Set II (http://dtp.cancer.gov), assumed to be decoys for the RT ligand set. The W191G ligand set comprised 53 known actives and 21 known inactives.[26−29] In all cases, bond orders, stereochemistry, hydrogen atoms, and protonation states were assigned by Ligprep.[45] Properties of the ligand sets are summarized in Table 1. These compound sets were chosen to represent prototypes of challenging sets for blind VS.

Table 1. Mean and Standard Deviation of Physicochemical Properties of Ligand Sets

	no. of compounds	molecular weight (g/mol)	no. of rotatable bonds	no. of hydrogen-bond acceptors	no. of hydrogen-bond donors
RT actives[a]	20	371.9 ± 75.3	5.2 ± 2.4	5.6 ± 2.1	1.8 ± 0.8
RT decoys[a]	1323	280.7 ± 80.6	3.4 ± 2.2	4.4 ± 2.1	1.4 ± 1.3
W191G actives[a]	53	106.8 ± 13.9	1.0 ± 0.9	1.7 ± 1.0	1.5 ± 1.2
W191G decoys[a]	23	107.0 ± 30.4	1.1 ± 1.1	2.3 ± 1.2	1.9 ± 1.3

[a] See Supporting Information for a complete list of X-ray structures, actives, and decoys considered.

Molecular Docking. For all receptor structures, crystallographic and trajectory water molecules, ions and ligand compounds were removed. Proteins were prepared using Schrodinger software, Maestro 9.0 and Glide 5.5.[46,47] This blind setup reflects the typical scenario for VS of newly discovered protein targets, in which no additional chemical information can be used a priori for docking. Receptors were processed using the default protein preparation workflow, which employs a restrained, partial energy minimization. The Glide SP algorithm was employed using a grid box volume of 10 × 10 × 10 Å. All structures were fitted using the C_α atoms of the binding pocket, as defined above, and grid center coordinates were based on this alignment. Only the best, or lowest energy, pose for each docking run was retained. The use of Glide as a docking software might influence the quantitative results of this study, however, the more general qualitative conclusions are likely transferable to different algorithms and scoring functions. A total of 2500 MD snapshots for each of the 7 systems were processed as well as 15 RT and 49 W191G X-ray structures, totaling 17564 different receptor docking runs performed and analyzed in this study.

Predictive Power Analysis. The receptor conformation affects the ranking of actives and inactives. In practice, after the algorithm ranks the compounds, the top X will be pursued further, by experimental assay or as a starting point for rescoring calculations, the rest are discarded. The top X are thus deemed positives, and the rest negatives. If the activity of the compounds ranked is previously known, then such classifications are either true or false. Using the four categories true positive (TP), false positive (FP), true negative (TN), and false negative (FN), the true positive rate (i.e., the selectivity):

$$\mathrm{TP}/(\mathrm{TP}+\mathrm{FN}) \quad (1)$$

and the false positive rate (i.e., 1-specificity):

$$1-(\mathrm{TN}/(\mathrm{FP}+\mathrm{TN})) \quad (2)$$

are determined based on the X threshold.

The ROC plots eq 1 with respect to eq 2 as X is perturbed, producing a curve that is threshold independent.[32,33] The integral of a ROC plot, or the area under the curve (AUC), gives the probability that a classifier, in this case the docking algorithm, will rank a randomly chosen active over a randomly chosen inactive. AUC is a useful metric to cross-compare different receptors relative predictive performances. Figure 2 summarizes the practical aspects of comparing VS predictive power for two conformations of a receptor.

The mean AUC for different groups of structures, bound and unbound, was compared for quantitative assessment. The best, worst, and mean AUC for 10 and 45 bound X-ray structures as well as 5 and 4 unbound X-ray structures for RT and W191G, respectively, were calculated for comparison. All of these X-ray groups were compared to averages based on 2500 snapshots for each of the 7 systems.

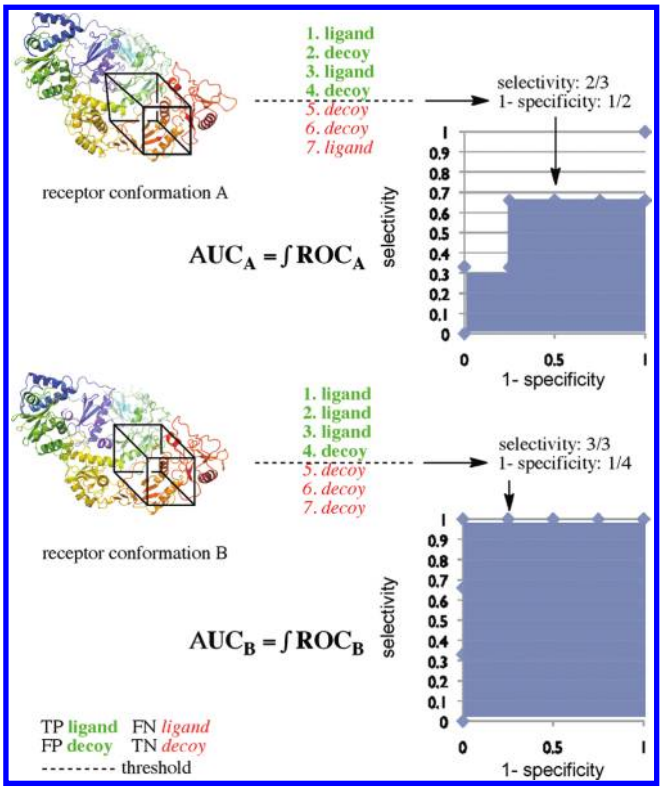


Figure 2. Schematic representation of the practical approach employed for comparison of VS predictive power. Different A and B receptor conformations used in docking affect compound ranking, thus prediction selectivity and specificity. The integral of a ROC curves (also named AUC) is the metric employed for nonarbitrary comparison of VS predictive power.

RESULTS

Predictive Power of MD Structures. Docking the compound sets to 2500 MD trajectory snapshots results in a distribution of AUC values for each of the seven systems. These distributions arise from the diversity of the conformational space sampled by the receptor and allow for quantification of bound versus unbound VS predictive power. In addition, different protein systems AUC distributions can be compared on a qualitative basis, keeping in mind that both receptor configuration sampling and ligand diversities determine such representations. AUC values of 0.5, corresponding to a random ranking of the compounds, or smaller signify worse than random predictive power for a give ligand set.

Figure 3 shows AUC distributions based on four RT and three W191G trajectories and compared with the range of and average

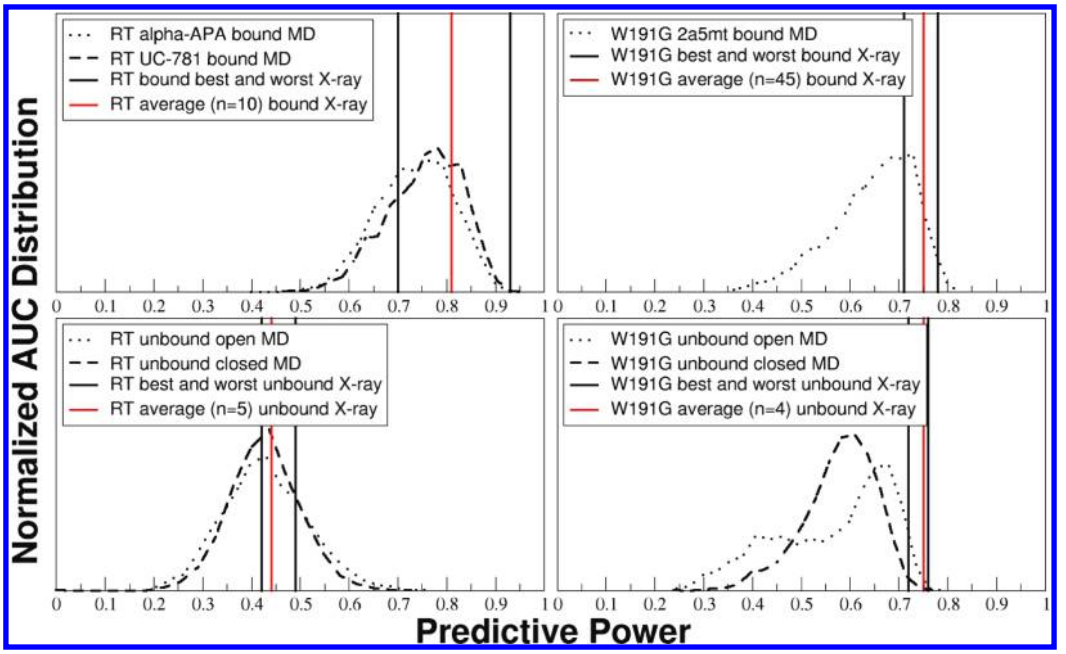


Figure 3. Normalized distributions of AUC values for blind VS of RT (left panels) and W191G (right panels) systems based on seven molecular dynamics ensembles. Top left: RT simulations, α-APA bound and UC-781 bound. Bottom left: RT unbound open and unbound closed. Top right: W191G simulation with 2a5mt bound. Bottom right: W191G unbound open and unbound closed. AUC value ranges (delimited by vertical black lines corresponding to worst and best values) and the AUC averages (vertical red lines, *n* is the number of crystallographic structures available) from corresponding X-ray structures are also displayed as reference. Area under the distribution to the right of the vertical red lines indicates MD structures with better predictive power than average X-ray structures from corresponding ensembles. See Figure 4 for additional details.

Table 2. Maximum, Minimum, and Average AUC Values from X-ray and MD Structure Ensembles[a]

structures	X-ray worst AUC	X-ray best AUC	MD worst AUC	MD best AUC	X-ray <AUC>	MD <AUC>
RT bound[b]	0.70	0.93	0.40	0.96	0.81	0.76
RT unbound[c]	0.42	0.49	0.21	0.77	0.44	0.44
W191G bound[d]	0.71	0.78	0.34	0.83	0.75	0.66
W191G unbound[e]	0.72	0.76	0.24	0.79	0.75	0.58

[a] Refer to the Methods Section for details on the different ensembles considered and to Figure 3 for corresponding distributions. [b] MD RT bound trajectories consist of α-APA bound and UC-781 bound. [c] MD RT unbound trajectories. [d] MD W191G 2a5mt-bound trajectory. [e] MD W191G unbound of gating loop open and closed.

corresponding AUC values from individual X-ray bound and unbound structures. All trajectory distributions are largely unimodal with the exception of the W191G unbound, started from the PDB structure 1RYC. The presence of a second peak can be explained considering the comparatively greater flexibility of the open 190−195 gating loop.[30]

For the more flexible RT system, a pronounced shift toward lower AUC values and poorer predictive power can be observed by comparing the bound systems peaks of 0.76 and 0.78 AUC for α-APA and UC-781 bound systems, respectively, and the unbound ensemble peak of 0.43 and 0.44 AUC for open and closed unbound simulations, respectively. This is consistent with previous studies suggesting improved VS predictive power of complex structures compared with unbound receptors.[15] In the case of W191G, distributions for bound and unbound systems are more similar, consistent with the fact that receptor relaxation upon ligand binding is less pronounced for more rigid binding sites.

MD Snapshots Versus X-ray Structures. The mean, maximum, and minimum AUC values from the bound and unbound ensembles are reported in Table 2. Interestingly, in all cases, the best MD AUC value outperforms the best X-ray available. This observation suggests that MD simulations can be used to improve VS results. However, a significant percentage of the MD AUCs was less predictive than the worst X-ray structure available. The mean values indicate that on average the predictive power of MD snapshots is poorer than that of the X-ray structures. An exception to this observation is made for RT unbound, for which the mean values are comparable.

The range of predictive power of X-ray structures can also be seen in Figure 3 with black vertical lines at maximum and minimum AUC values found in this study. The bound X-ray ranges for both systems are wider than unbound X-ray AUC ranges, although RT has a considerable larger range, consistent with greater flexibility in this case. Interestingly, similar distribution ranges of ca. 0.5 AUC values for MD snapshots are found for both systems.

The predictive power of MD versus X-ray structures is represented in Figure 4 in terms of the relative number of MD

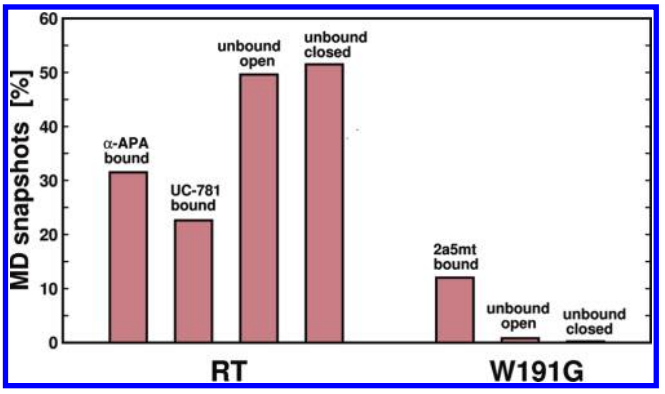


Figure 4. Relative number of molecular dynamics snapshots with higher predictive power (AUC) than the mean AUC from all available corresponding X-ray structures. See also Supporting Information, Table SII, for detailed data.

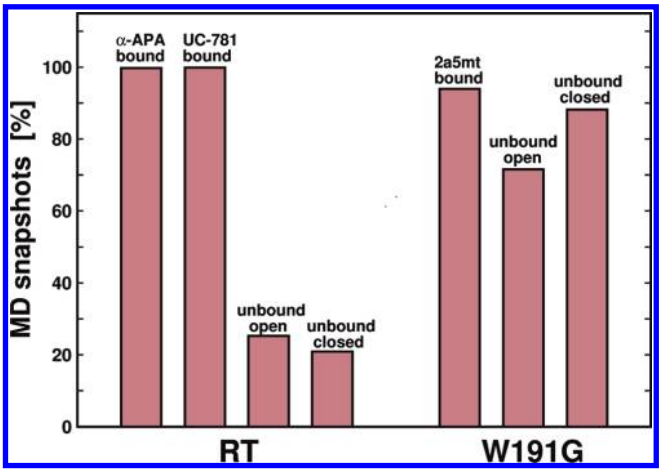


Figure 5. Relative number of molecular dynamics snapshots with higher predictive power (AUC) than random ranking. See also Supporting Information, Table SII, for detailed data.

snapshots with AUC values larger than the average value from corresponding X-ray models. This is a measure of the advantage of using MD structures versus using a randomly selected static X-ray model. For RT ligand-bound systems at least 23% of MD snapshots lead to improved VS prediction, this value being somewhat higher for the α-APA versus UC-781 system (31%). For RT unbound systems, at least 50% of MD snapshots outperform static X-ray models with similar values for unbound open (51%) and unbound closed (50%) simulations.

For the rigid W191G, the majority of the MD snapshots were less predictive than the average X-ray AUC, with 12% for the bound simulation better than the average X-ray structure AUC, and less than 1% for both unbound simulations compared to the average AUC of bound X-ray structures. It is important to stress that AUC values are randomly distributed along the seven trajectories (not shown). Therefore, there is no correlation of AUC value with the evolution of the MD run from the same initial X-ray coordinates.

MD Snapshots Versus Random. As described previously, the AUC value of 0.5 corresponds to the predictive power of randomly ranking compounds. Figure 5 summarizes the VS predictive power of MD snapshots in terms of the number of MD snapshots with AUC values greater than 0.5 or random. RT percentages of predictive conformations were 99.8% α-APA and 99.9% for UC-781 ligands versus 25.3% and 20.9% for unbound open and closed, respectively. W191G percentages of predictive conformations were 94.0% for the bound system versus 71.6% and 88.2% for unbound open and closed, respectively. For both systems, at least 94% of bound MD snapshots were more predictive than random. The unbound simulation predictive structures were less in number, particularly for the unbound RT structures.

Remarkably, in both unbound RT simulations, approximately 50% of the structures processed were more predictive than the average of the unbound X-ray AUCs. More significantly, 21% of the closed unbound and 25% of the open unbound simulations were more predictive than random, where as all of the X-ray unbound structures were worse than random. This case indicates that MD can move a conformation seemingly not amenable to docking, into the predictive range.

McGovern et al. found that the majority of X-ray structures studied identified known binders over nonbinders better than random selection of rank.[15] The X-ray structures for the two systems studied here were all more predictive than random, having a AUC of greater than 0.5, with the exception of unbound RT structures which were all worse than random. The reason for the poor performance of unbound RT X-rays is likely due to the occluded binding site, where key side chains must move out of the way in order for a ligand to fit.

Influence of Molecular Flexibility and Properties of the Compound Sets. Overall system flexibility affects VS predictive power. Figure 6 reports a comparative analysis of backbone fluctuations for RT and W191G systems. Backbone flexibility is clearly greater for the larger RT receptor overall. However, the all-atom binding site fluctuations are comparable for both systems (Supporting Information, Figure S1), and a pronounced decrease in side chain flexibility upon ligand binding can be observed, when comparing bound and unbound trajectories.

Looking at fluctuations of all binding site atoms and the α-carbon subset represents motion with respect to the entire binding site and the backbone chain that makes up the area. Figure 7 plots the average binding site fluctuations for each of the trajectories, comparing it with the average AUC of the ensemble. A correlation was found between the average predictive power <AUC> and the averaged binding site flexibility <RMSD>. Averaged fluctuations and predictive power values (Supporting Information, Table SIII) show generally that the more flexible the binding site, the less predictive the MD ensemble. This is in agreement with previous findings that bound or holo structures, traditionally more stable and less flexible, are more amenable to predictive docking experiments than unbound or apo structures.[15]

While the two systems display variation in flexibility, the different chemical nature of the compounds docked as well as of the active sites should also be noted. Properties of the ligand sets are summarized in Table 1. Compounds in the RT set represent a more diverse range of physicochemical properties than for W191G. This structural diversity results in a more numerous conformations per ligand to evaluate. Additionally, W191G ligand binding is dominated by a specific electrostatic interaction, with all known binders carrying a positive charge. While ligand ranking might be equally challenging for both systems, the more pronounced difference between ligand and decoys for W191G may in part explain a higher overall VS predictive power in this case.

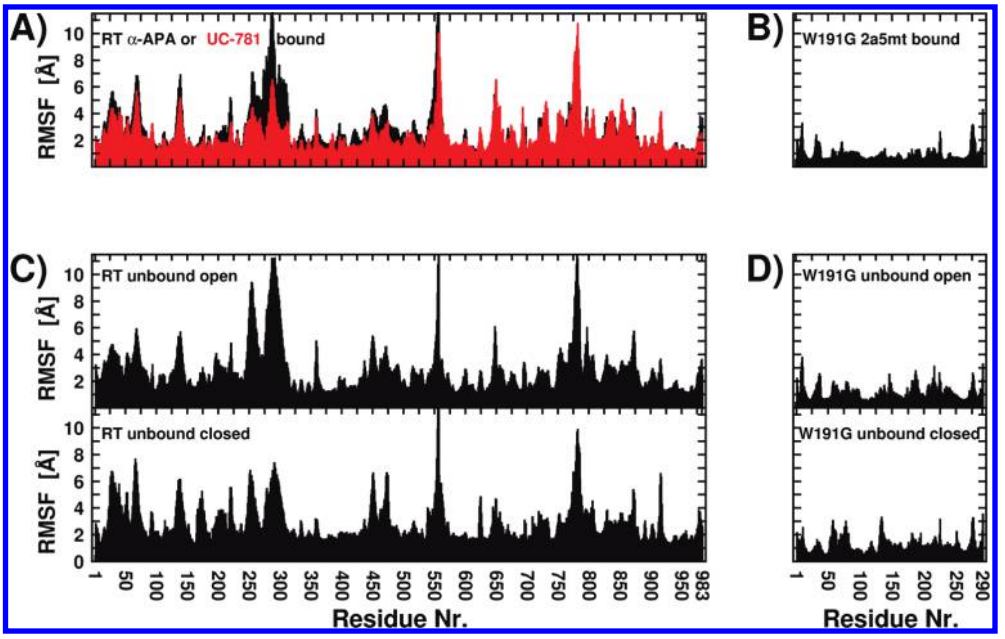


Figure 6. Root-mean-square fluctuations (RMSF) of the backbone C_α atom positions from the seven molecular dynamics simulations. RT (left) and W191G (right), while (A and B) show bound simulation fluctuations and (C and D) show unbound simulation fluctuations. RMSF values were calculated using sampling of 50 ns for each of the seven cases. For analysis of these simulations see ref 25 (RT) and refs 23, 30 (W191G).

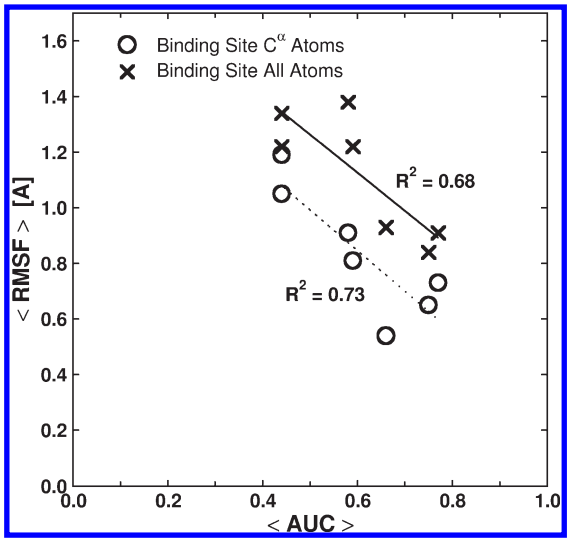


Figure 7. Relationship between ensemble averaged predictive power, <AUC>, and ensemble averaged binding site flexibility, <RMSF>. Values were calculated using sampling of 50 ns for each of the seven cases. See Supporting Information, Table SIII, for data values.

DISCUSSION

Analyses of root-mean-square deviation (RMSD) from starting structure, binding site volume, radius of gyration, cognate ligand size, mean score of top 1% of docked compounds, RMSD binding site clustering, and flexibility descriptors from principal component analysis were not able to distinguish better performing snapshots from weakly performing snapshots. Overall, it was not possible to identify any evident relationship between characterization of MD snapshots and VS predictive performance. This is in agreement with other studies for different protein receptor and compound sets.[11,13]

The results suggest that MD conformations can improve VS results compared to the exclusive use of X-ray structures, yet a general system independent protocol to this goal is still challenging. The present data strongly confirms previous observations that the VS "quality" of a protein receptor configuration is strictly dependent on the chemical properties of the system and the ligand sets considered. The only general consideration demonstrated by this study is that no simple property can be used to separate the best receptor structures from the ensemble, and as such, none should be discarded a priori. While no method for selecting MD structures before hand was determined, exploratory screening using the AUC measurement on an MD ensemble will be used to identify the best MD conformations prior to a more extensive VS computation, in a hierarchical fashion in future work. Additionally, while individual receptor structures were assessed here, evaluation of efficient methods for combining information from multiple ensembles is on going.

One practical concern for users seeking enrichment is that generating more conformations using MD simulations may be computationally time intensive. As technology advances, time scales achieved in even one days worth of simulation time are rapidly increasing, and it is plausible that in the future generating and choosing a receptor structure may be as practical as choosing and generating the number of compounds in a VS database; this was the main motivation for evaluating the different ensembles, and indeed it was shown here that MD can provide snapshots that are different from crystal structures. In most cases, the extent of MD sampling will determine the diversity of conformations, relying on how well the sampling has explored phase space. Enhanced sampling methods, such as accelerated molecular dynamics, allow for more rapid exploration of conformational space by reducing energy landscape barriers and subsequently effective simulation time scale.[48]

Finally, one could argue that the user becomes more intimately familiar with the target through running a dynamics simulation; standard practice for docking protocol often involves

manual review and processing as well as familiarity with the system. This was highlighted by Head's results from SAMPL-1 cross-docking challenge, which showed that successful use of docking programs requires human expertise and experience.[49] Furthermore, some publications of MD VS studies have appeared subsequent to an MD study that aimed to biophysically characterize a target; in such cases simulation data were available for VS, and the computation time is dependent just on the analysis or selection of structures, which is done in the case of multiple X-ray structures as well.

CONCLUSION

This study presented an explorative analysis of blind virtual screening predictive power for ligand bound and unbound molecular dynamics snapshots of two systems with diverse molecular flexibility and binding site properties. Predictive power was quantified based on virtual screening using receiver operating characteristic curves to determine the probability of ranking actives over inactives. Results from predictive power analysis were compared with crystallographic structure results as well as with respect to random compound rank.

The data demonstrate that molecular dynamics snapshots can improve blind virtual screening predictive power. Most interestingly, the best molecular dynamics snapshot was systematically more predictive than the best X-ray structure in all cases examined. However, particularly for the less flexible of the two systems considered, a significant number of molecular dynamics snapshots led to less predictive virtual screening results than the least predictive X-ray structure available. Therefore, the advantage of molecular dynamics simulations for virtual screening seems to rely on the enrichment that a reduced number of receptor structures supply, rather than the whole configurational ensemble generated. While no method for selecting the best molecular dynamics structures a priori was found, exploratory screening will be used to identify the best conformations in a hierarchical approach. For the more flexible RT system, enhanced sampling methods are being employed to improve virtual screening predictive power to an even greater extent.

ASSOCIATED CONTENT

Supporting Information. Detailed lists of all relevant chemical structures is included, as well as figure data points. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: senichols@ucsd.edu; rbaron@mccammon.ucsd.edu.

ACKNOWLEDGMENT

We thank the members of the McCammon group for useful discussions. This work was supported by the National Institutes of Health, the National Science Foundation, the Howard Hughes Medical Institute, the National Biomedical Computation Resource, and the NSF Supercomputer Centers. Computational resources were supported, in part, by the National Science Foundation grant PHY-0822283, the Center for Theoretical Biological Physics.

REFERENCES

(1) Bowman, A.; Lerner, M.; Carlson, H. Protein flexibility and species specificity in structure-based drug discovery: Dihydrofolate reductase as a test system. *J. Am. Chem. Soc.* **2007**, *129*, 3634–3640.

(2) Carlson, H. Protein flexibility and drug design: how to hit a moving target. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447–452.

(3) Amaro, R. E.; Baron, R.; McCammon, J. A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 693–705.

(4) Wong, C. F. Flexible ligand-flexible protein docking in protein kinase systems. *Biochim. Biophys. Acta* **2008**, *1784*, 244–251.

(5) Schneider, G. Virtual screening: an endless staircase? *Nat. Rev. Drug Discovery* **2010**, *9*, 273–276.

(6) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.

(7) Guvench, O.; MacKerell, A. D. Computational evaluation of protein-small molecule binding. *Curr. Opin. Struct. Biol.* **2009**, *19*, 56–61.

(8) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813–1818.

(9) Armen, R. S.; Chen, J. H.; Brooks, C. L. An evaluation of explicit receptor flexibility in molecular docking using molecular dynamics and torsion angle molecular dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2909–2923.

(10) Bolstad, E. S. D.; Anderson, A. C. In pursuit of virtual lead optimization: The role of the receptor structure and ensembles in accurate docking. *Proteins* **2008**, *73*, 566–580.

(11) Rao, S.; Sanschagrin, P. C.; Greenwood, J. R.; Repasky, M. P.; Sherman, W.; Farid, R. Improving database enrichment through ensemble docking. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 621–627.

(12) Huang, S.; Zou, X. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins* **2007**, *66*, 399–421.

(13) Rueda, M.; Bottegoni, G.; Abagyan, R. Recipes for the selection of experimental protein conformations for virtual screening. *J. Chem. Inf. Model.* **2010**, *50*, 186–193.

(14) Rockey, W. M.; Elcock, A. H. Structure selection for protein kinase docking and virtual screening: Homology models or crystal structures? *Curr. Protein Pept. Sci.* **2006**, *7*, 437–457.

(15) McGovern, S. L.; Shoichet, B. K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*, 2895–2907.

(16) Kairys, V.; Fernandes, M.; Gilson, M. Screening drug-like compounds by docking to homology models: A systematic study. *J. Chem. Inf. Model.* **2006**, *46*, 365–379.

(17) Craig, I.; Essex, J.; Spiegel, K. Ensemble docking into multiple cystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 511–524.

(18) Barril, X.; Morley, S. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J. Med. Chem.* **2005**, *48*, 4432–4443.

(19) Ivetac, A.; McCammon, J. A. Mapping the druggable allosteric space of G-protein coupled receptors: a fragment-based molecular dynamics approach. *Chem. Biol. Drug Des.* **2010**, *76*, 201–217.

(20) Gerek, Z. N.; Ozkan, S. B. A flexible docking scheme to explore the binding selectivity of PDZ domains. *Protein Sci.* **2010**, *19*, 914–928.

(21) Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632–5633.

(22) Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers* **2003**, *68*, 47–62.

(23) Baron, R.; McCammon, J. A. Dynamics, hydration, and motional averaging of a loop-gated artificial protein cavity: the W191G mutant of cytochrome c peroxidase in water as revealed by molecular dynamics simulations. *Biochemistry* **2007**, *46*, 10629–10642.

(24) De Clercq, E. Non-nucleoside reverse transcriptase inhibitors (NNRTIs): past, present, and future. *Chem. Biodiversity* **2004**, *1*, 44–64.
(25) Ivetac, A.; McCammon, J. A. Elucidating the inhibition mechanism of HIV-1 non-nucleoside reverse transcriptase inhibitors through multicopy molecular dynamics simulations. *J. Mol. Biol.* **2009**, *388*, 644–658.
(26) Brenk, R.; Vetter, S. W.; Boyce, S. E.; Goodin, D. B.; Shoichet, B. K. Probing molecular docking in a charged model binding site. *J. Mol. Biol.* **2006**, *357*, 1449–1470.
(27) Fitzgerald, M. M.; Musah, R. A.; McRee, D. E.; Goodin, D. B. A ligand-gated, hinged loop rearrangement opens a channel to a buried artificial protein cavity. *Nat. Struct. Biol.* **1996**, *3*, 626–631.
(28) Musah, R. A.; Jensen, G. M.; Bunte, S. W.; Rosenfeld, R. J.; Goodin, D. B. Artificial protein cavities as specific ligand-binding templates: Characterization of an engineered heterocyclic cation-binding site that preserves the evolved specificity of the parent protein. *J. Mol. Biol.* **2002**, *315*, 845–857.
(29) Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. Rescoring docking hit lists for model cavity sites: Predictions and experimental testing. *J. Mol. Biol.* **2008**, *377*, 914–934.
(30) Baron, R.; McCammon, J. A. (Thermo)dynamic role of receptor flexibility, entropy, and motional correlation in protein-ligand binding. *ChemPhysChem* **2008**, *9*, 983–988.
(31) Fitzgerald, M. M.; Trester, M. L.; Jensen, G. M.; McRee, D. E.; Goodin, D. B. The role of aspartate-235 in the binding of cations to an artificial cavity at the radical site of cytochrome c peroxidase. *Protein Sci.* **1995**, *4*, 1844–1850.
(32) Fawcett, T. An introduction to ROC analysis. *Pattern Recogn. Lett.* **2006**, *27*, 861–874.
(33) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* **2005**, *48*, 2534–2547.
(34) Nichols, S.; Domaoal, R.; Thakur, V.; Tirado-Rives, J.; Anderson, K.; Jorgensen, W. Discovery of wild-type and Y181C mutant non-nucleoside HIV-1 reverse transcriptase inhibitors using virtual screening with multiple protein structures. *J. Chem. Inf. Model.* **2009**, *49*, 1272–1279.
(35) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. Gromacs - a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
(36) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306–317.
(37) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
(38) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
(39) Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. A consistent empirical potential for water-protein interactions. *Biopolymers* **1984**, *23*, 1513–1518.
(40) Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.;High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat. Struct. Biol.* **1995**, *2*, 293–302.
(41) Arnold, E.; Das, K.; Ding, J.; Yadav, P. N.; Hsiou, Y.; Boyer, P. L.; Hughes, S. H. Targeting HIV reverse transcriptase for anti-AIDS drug design: structural and biological considerations for chemotherapeutic strategies. *Drug Des. Discovery* **1996**, *13*, 29–47.
(42) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719–1751.
(43) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation. The GROMOS96 Manual and User Guide*; University of Groningen: Groningen, The Netherlands, 1996.
(44) Berendsen, H. J. C. Interaction models for water in relation to protein hydration. *Intermol. Forces* **1981**, 331−342.
(45) *LigPrep*, version 1.6; Schrodinger LLC: New York, 2009.
(46) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
(47) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.
(48) Hamelberg, D.; Mongan, J.; McCammon, J. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929.
(49) Head, M. s., Docking: a doomsday report. In *Drug Design: Structure- and Ligand-Based Approaches*, Merz, K. M., Ringe, D., Reynolds, C. H., Eds. Cambridge University Press: Cambridge, U.K., 2010; pp 98−119.