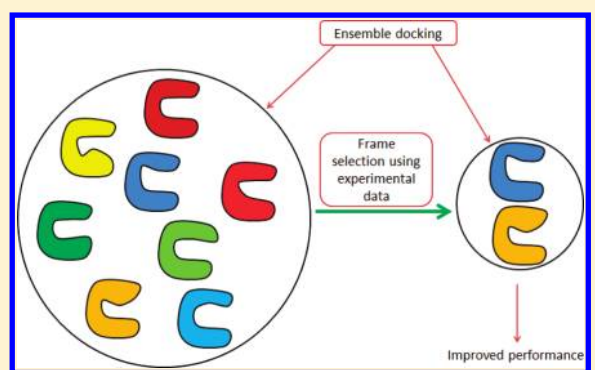Article

Utilizing Experimental Data for Reducing Ensemble Size in Flexible-Protein Docking

Mengang Xu and Markus A. Lill*

Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Purdue University, 575 Stadium Mall Drive, West Lafayette, Indiana 47907, United States

Ⓢ *Supporting Information*

ABSTRACT: Efficient and sufficient incorporation of protein flexibility into docking is still a challenging task. Docking to an ensemble of protein structures has proven its utility for docking, but using a large ensemble of structures can reduce the efficiency of docking and can increase the number of false positives in virtual screening. In this paper, we describe the application of our new methodology, Limoc, to generate an ensemble of holo-like protein structures in combination with the relaxed complex scheme (RCS), to virtual screening. We describe different schemes to reduce the ensemble of protein structures to increase efficiency and enrichment quality. Utilizing experimental knowledge about actives for a target protein allows the reduction of ensemble members to a minimum of three protein structures, increasing enrichment quality and efficiency simultaneously.

INTRODUCTION

Protein flexibility plays a key role in molecular recognition between a ligand and its target receptor.[1] Levels of protein flexibility range from small-scale changes, such as side-chain rotamer changes, to large-scale changes, such as hinge-bending motions and conformational changes in loop regions.[2–5] Recent approaches for improving docking methods have focused on incorporating both small and large scale protein flexibility in search and scoring procedures.[4,6–12] One widely used methodology aims to include backbone and side-chain flexibility into docking by using an ensemble of protein structures (EPS) as docking templates.[7] The templates are generated using methods such as molecular dynamics (MD), Monte Carlo (MC), elastic network model (ENM)[13,14] simulations, or normal-mode analysis (NMA).

Ensemble docking studies[15] have found that using an EPS with only a few protein conformations can increase the enrichment in virtual screening, but using a very large EPS often hurts docking performance. The drop in virtual screening performance is due to the inherent inaccuracy of the simplistic scoring methods used in docking. With an increasing number of protein conformations, the potential to generate non-native protein–ligand conformations also increases. Thus, using inaccurate scoring methods in virtual screening may result in an increased potential to generate high ranked decoys and thereby reduce the enrichment quality. Thus, in ensemble docking, it is important to consider how many and which protein conformations should be used.

Different strategies have been developed to identify a subset of protein structures that provides a compromise between the amounts of protein flexibility incorporated into docking and the number of structures considered. Amaro et al.[16] used methods such as hierarchical clustering based on pairwise root-mean square-deviations (RMSD) between different EPS members to reduce the initial EPS generated by MD simulations to a structurally diverse subset of protein structures. Armen et al.[17] investigated different strategies to limit the modeled protein flexibility to the fewest degrees of freedom necessary to adequately represent the experimentally observed flexibility (i.e., binding-site side chains and two flexible loops in the studied test case). Bolstad et al.[18] developed a selection strategy based on the conservation of the relative orientations of a core of amino acids critical for ligand binding. Atoms of binding site residues that have a conserved relative position in experimentally determined complex structures of the same protein with different bound ligands were manually identified. Distances between the identified atoms were computed for all MD snapshots and compared with the conserved distances in the various X-ray structures. Only MD snapshots that approximately preserve the interatomic distances between the selected protein atoms were chosen as members of the EPS used for docking. This procedure allowed the reduction of the size of the initial EPS, increasing the efficiency and ranking accuracy of ensemble docking.

The relaxed complex scheme (RCS)[19,20] can be viewed as an extension of ensemble docking where MD simulations of the apo form of the protein are used to generate an EPS. Sequential docking to the EPS is performed, and for each ligand, similar binding poses are clustered across all EPS protein templates.

The scores of all members of a cluster are averaged to approximate the thermodynamic states of ligand binding. Compared to using individual scores, the average produces a more accurate estimation of the free energy of binding.[19,20]

We recently performed docking studies utilizing RCS to an EPS generated by MD simulations on protein systems that involve small-scale conformational changes between the apo and holo structures (RMSD < 1 Å).[21] Although the apo form of the protein is typically used to generate the EPS for the RCS scheme, we observed that docking to an EPS generated by short MD simulations with bound ligands outperformed docking to the EPS generated by long MD simulations on the apo form of the protein. We concluded that either there are too many alternative protein conformations generated by the long MD simulations, making the identification of the holo-like structures relevant for docking difficult, or ligand binding was required to induce the holo conformation of the protein.

These results led us to the development of the ligand-model concept (Limoc), a method that is capable of sampling protein conformations that are relevant for binding structurally diverse ligands.[21] In this method, MD simulations are performed with a dynamically changing set of restrained functional groups in the binding site of the protein, essentially representing a large hypothetical ensemble of different chemical species binding to the same target protein. Starting from an apo structure, the ligand-model approach is used to derive an EPS used for docking. The combination of Limoc and RCS has been successfully applied to the prediction of binding modes and affinities of structurally diverse compounds. However, ensemble docking using large EPS and RCS analysis requires significantly more computational resources compared to virtual screening performed on a single static protein structure.

In this manuscript, we want to address the question of whether the reduction of ensemble size can improve efficiency and enrichment quality of virtual screening using the Limoc-RCS scheme. We investigated three different strategies to select a subset of protein structures from the original EPS. First, we clustered similar frames based on their pairwise RMSD. Second, we developed a training process in which protein structures are selected on the basis of their performance to reproduce the experimentally known binding mode of a small set of diverse ligands. Third, frames are selected on the basis of their performance to distinguish actives and decoys using a small training set of ligands.

■ MATERIAL AND METHODS

Data Sets and Target Proteins. We selected five protein systems as targets for the virtual screening studies: thrombin (PDB code: 2UUF), estrogen receptor α (ERa, 1GWR), cyclin-dependent kinase 2 (CDK2, 1HCL), acetylcholine esterase (AchE, 1EA5), and *Pneumocystis carinii* dihydrofolate reductase (DHFR, 1DAJ). The protein structures chosen represent the apo (ligand-free) form for thrombin, CDK2, and AChE. As there is no apo structure for ERa and *Pneumocystis carinii* DHFR in the PDB available, we used a holo structure for our screening studies. The side-chain conformations, tautomers, and protonation states of ASN, GLN, and HIS were adjusted using the Reduce program.[22] Protein structures were subsequently minimized using Amber 10.[23] The dictionary of useful decoys (DUD)[24] was used to perform the virtual screening studies. For each active in DUD, a set of decoys with similar physical properties but dissimilar topology is added to the data set of compounds used for screening. To reduce computational time, a subset of DUD DHFR ligands was used in this study by randomly selecting one-fourth of the total number of actives and decoys. For AChE, the data set used in our study contains 107 actives and 3892 decoys; for CDK2, 72 and 2072; for ERa, 67 and 2570; for DHFR, 102 and 2100; and for thrombin, 72 and 2456.

Docking. All docking simulations were performed using AutoDock Vina.[25] Protein and ligand preparation for docking was performed using our in-house PyMOL plugin that automatically calls the programs prepare_receptor4.py and prepare_ligand4.py that are part of AutoDockTools.[26] To define the search volume for AutoDock Vina, the apo form of a protein is aligned to several holo forms of the same protein, and then a cubic box of 25 Å × 25 Å × 25 Å is placed around the cocrystallized ligands of the associated holo forms. Standard docking settings were used, and the 10 energetically most favorable binding poses were outputted.

Limoc. MD simulations with Limoc[21] were used to generate an EPS for ensemble docking and RCS analysis. In this study, we started from the apo protein structure (except for ERa and DHFR). First, the binding site is defined in Limoc by specifying residues within a given radius (default is 5 Å) of any atom of the chosen ligand. Then, the solvent accessible surface (SAS) of the binding site is computed. Physical−chemical probes representing the ligand model are evenly distributed onto the SAS with properties that are complementary to the nearest binding pocket residue. In the current version of Limoc, we consider three types of probes: hydrogen-bond donor, acceptor, and hydrophobic group. To sample the interactions of protein residues with different hydrogen-bonding groups from different ligands, we employed the local enhanced sampling (LES) method.[27] To simulate the protein adapting to ligands differing in size and shape, we use locally steered MD simulations to enlarge or decrease the size of the ligand model. The size and direction of movement of the binding pocket residues is determined by the first principle component derived from analyzing the covariance matrix of a 50 ps ligand free simulation trajectory. The positions of the ligand model probes are harmonically restrained with a force constant 4 times greater in the direction parallel to the SAS than that perpendicular to the SAS. The steered MD is performed in both directions (locally expanding and reducing the size of the ligand model). The total simulation length (including all induced fit scenarios) was 20 ps using a time step of 1 fs. The Amber03 force field was used, and solvation was modeled by a water cap of 25 Å radius. For additional details on Limoc and its default settings, we refer to our recent publication on this method.[21]

Relaxed Complex Scheme. The MD trajectories using Limoc were clustered using quality threshold (QT) clustering.[28] The RMSD criterion was adjusted automatically to generate 180−250 distinct protein conformations, defining the initial EPS. On the basis of this EPS, different frame selection methods were used to derive subensembles with 1, 2, 3, 5, 10, 20, and 50 protein structures. AutoDock Vina was used to perform docking to the full initial ensemble and each subensemble. The top-10 ranked binding poses from each individual docking simulation were outputted and considered for subsequent clustering. The binding poses were clustered using QT clustering (1 Å cluster radius). For the full ensemble (180−250 frames) and the 50 and 20 subensembles, clusters with less than 15, 5, and 3 members were discarded, respectively. The predicted score for a binding-mode cluster i, $\langle \Delta G_i \rangle$, was
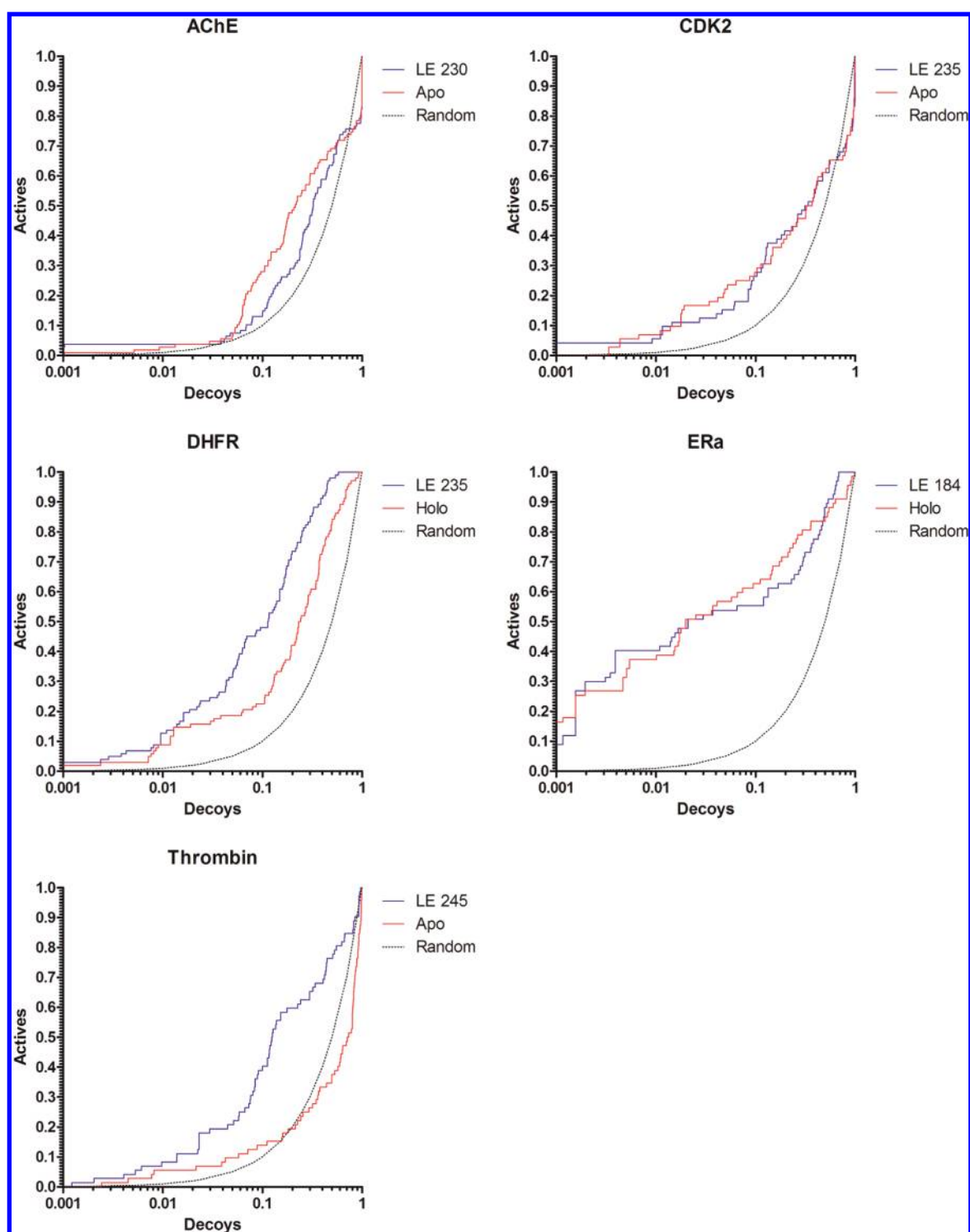
Figure 1. Docking enrichment plots for five protein systems using DUD. The fraction of actives ranked by docking score (*y* axis) is plotted against the fraction of ranked decoys (*x* axis; logarithmic scale). Both results derived from docking to the Limoc ensemble (LE; blue line) and a static apo/holo (red) structure are shown. Theoretical results from random selection (black dots) are shown for comparison.

calculated on the basis of the scores of all binding modes *j* of a cluster *i*, ΔG_j, using

$$\langle \Delta G_i \rangle = \sum_{j \in \text{cluster} i} p_{i,j} \Delta G_j$$

$$\text{with } p_{i,j} = \frac{\exp(-\Delta G_j / k_B T)}{\sum_{k \in \text{cluster } i} \exp(-\Delta G_k / k_B T)} \quad (1)$$

where $p_{i,j}$ is the probability of identifying a pose *j* in cluster *i*, and ΔG_j is the score of binding pose *j*. As an alternative scheme to rank the docked ligands, we used the overall lowest score of any binding pose of a ligand.

Frame Selection Using RMSD. To test the impact of reducing the size of the EPS on virtual screening results using the Limoc-RCS scheme, we used three different methods to select frames from the full ensemble of 180–250 protein structures.

Table 1. Indicators of Enrichment Quality for Five Systems Utilizing the Full EPS Generated Using Limoc (LE = Limoc Ensemble) and a Static Apo or Holo Protein Structure: EF1 (percentage of ranked actives at 1% of the ranked decoys), EF10 (at 10% of the ranked decoys), EF20 (at 20% of the ranked decoys), AUC (area under the curve), and log-AUC (area under the curve using logarithmic scale for the x axis)

	AChE		CDK2		DHFR		ERa		thrombin	
	apo	LE	apo	LE	apo	LE	holo	LE	apo	LE
EF1	2.80	3.70	6.90	5.60	8.82	12.8	37.3	40.3	5.60	8.30
EF10	2.80	1.40	2.78	2.64	2.25	4.80	6.27	5.52	1.39	3.89
EF20	2.43	1.45	1.95	2.10	2.11	3.68	3.58	3.16	0.97	2.99
AUC	0.61	0.58	0.56	0.56	0.71	0.85	0.82	0.82	0.40	0.71
logAUC	0.22	0.19	0.23	0.23	0.27	0.38	0.53	0.52	0.14	0.30

In the first scheme, QT clustering is performed on the full EPS using the g_cluster program, which is part of the Gromacs 4.5 software suite. The heavy atom RMSD was adjusted to generate 1, 2, 3, 5, 10, 20, and 50 clusters. The centroid of each cluster was used as a representative structure in the resulting subensemble.

Frame Selection by Training on Experimentally Known Binding Modes. In the second scheme for reducing the size of the original EPS, several holo forms of each protein system were retrieved from the PDB, and up to three structurally diverse ligands were chosen on the basis of visual comparison for subsequent training purposes. The selected ligands were docked to the original EPS using AutoDock Vina, and 10 poses were retrieved from docking to each member of the EPS. RMSD between the predicted docking poses and the experimentally known binding mode was calculated, and the docking pose with the smallest RMSD was identified for each EPS member. The sum of lowest RMSD values over all utilized ligands (one to three ligands in our study) was computed and used to rank protein structures in the original EPS. Subensembles were generated from the top ranked structures. As an example, assume that three ligands A, B, and C, are docked to two different protein structures, yielding the smallest RMSD values among the top-10 ranked poses for the two protein structures of 1.4, 2.4, and 2.7 Å and 1.8, 1.7, and 2.3 Å, respectively. Taking the sum of those RMSD values (6.5 Å vs 5.8 Å) ranks the second protein structure before the first one.

Frame Selection by Training on Enrichment of Small Subsets. In the third scheme for reducing the size of the full EPS, structures are selected on the basis of the degree of separation achieved between actives and decoys in a virtual screen. Experimentally known actives are mixed with a set of decoy ligands to build a training set; 10 or 20 actives and a corresponding number of decoys were randomly selected from the DUD data set to construct the training subsets for each protein system. The active/decoy ratio of subsets was kept the same as the ratio of the complete DUD set for each protein system. Mini-enrichment of each ligand training subset was performed using ensemble docking and RCS. The lowest docking score of all binding pose clusters, k, for a ligand, l, was identified for each protein structure, p:

$$\langle \Delta G_{p,l} \rangle = \min_k \{ \langle \Delta G_{p,l,k} \rangle \} \qquad (2)$$

Third, the average score over all actives and decoys is computed for each protein conformation p:

$$\overline{S_{actives,p}} = \frac{1}{N_{actives}} \sum_{l=1}^{N_{actives}} \langle \Delta G_{p,l} \rangle \qquad (3a)$$

$$\overline{S_{decoys,p}} = \frac{1}{N_{decoys}} \sum_{l=1}^{N_{decoys}} \langle \Delta G_{p,l} \rangle \qquad (3b)$$

Finally, the difference between $\overline{S_{actives,p}}$ and $\overline{S_{decoys,p}}$ was calculated. The different members of the EPS are ranked according to those differences, and the structures with largest negative difference are selected for inclusion in a subensemble.

Measures of Virtual Screening Success. The following standard descriptors for enrichment are utilized to measure the success of our virtual screening studies: the enrichment factors of actives at 1% (EF1), 10% (EF10), and 20% (EF20) of ranked decoys and the area-under-the-curve (AUC) of the ROC curve displaying the fraction of ranked actives at a given fraction of ranked decoys. A slight modification of the AUC value is computed to emphasize the importance of early enrichment in data sets that typically contain a majority of decoys: The log AUC value is the area-under-the-curve using a logarithmic scale of the x axis (fraction of ranked decoys).

■ RESULTS AND DISCUSSION

Enrichment Using Full Ensemble. Introducing protein flexibility by ensemble docking can enhance correct prediction of native binding poses and quantification of ligand binding affinities[7,12] but can also result in an increased number of false positives in virtual screening.[15] The latter can result in a poor enrichment performance compared to using a single static structure for virtual screening. RCS attempts to mitigate this drawback by clustering similar poses and averaging their scores.

Testing RCS with our Limoc-generated EPS on five protein systems (Figure 1 and Table 1) revealed that, compared to virtual screening using a single protein structure, using the full EPS resulted in improved enrichment for two protein systems (thrombin and DHFR) and gave comparable results for the other three systems (AChE, ERa, CDK2). The AUC is similar for AChE (EPS, 0.58; single structure, 0.61), CDK2 (0.56, 0.56), and ERa (0.82, 0.82) but significantly larger if utilizing the full EPS for DHFR (0.85, 0.71) and thrombin (0.71, 0.40).

We also compared the enrichment plots using the EPS with and without clustering similar poses and averaging their scores (see the Supporting Information, Figure S1). Interestingly, only a slight improvement in enrichment for DHFR (0.85 compared to 0.82 for using the lowest overall score) and thrombin (0.71 compared to 0.69) is observed after clustering and averaging the scores of similar poses. No differences could be identified for the other three protein systems. In summary, averaging seems to have only a small effect on the enrichment quality for the five protein systems studied.

The improved enrichment for thrombin is consistent with the increased binding pose prediction quality previously observed using Limoc-RCS.[21] However, the same trend was not observed for AChE, where increased binding pose prediction quality[21] did not translate into improved enrichment.
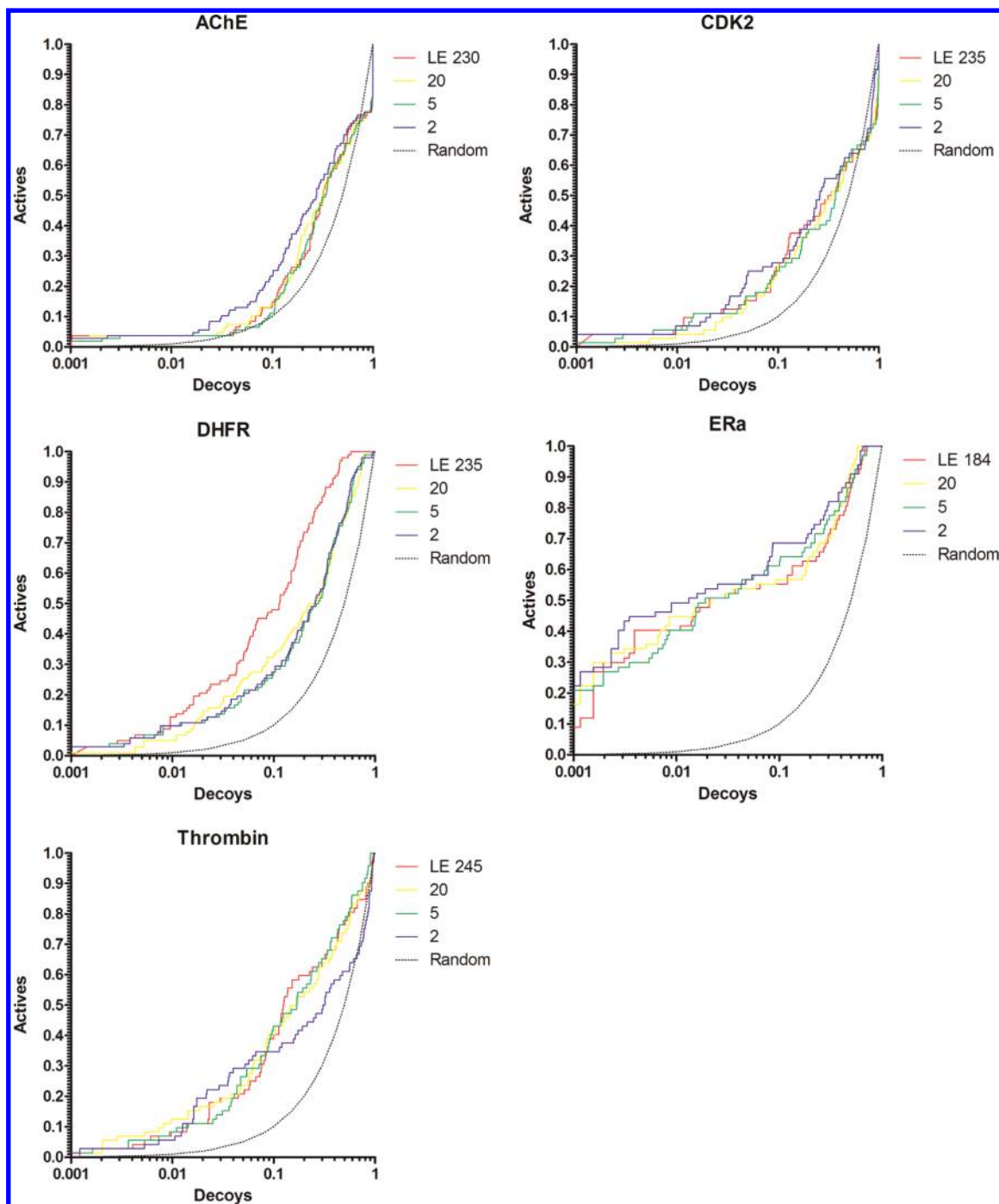
Figure 2. Docking enrichment plots for five protein systems using DUD. The fraction of actives ranked by docking score (y axis) is plotted against the fraction of ranked decoys (x axis; logarithmic scale). Results docking to the full Limoc EPS (LE; red) and most diverse subensembles with 20 (yellow), 5 (green), and 2 (purple) members are shown.

Importantly, no significant reduction in enrichment performance was observed for any of the five protein systems.

Frame Selection Using RMSD. Although the inclusion of protein flexibility in the Limoc-RCS scheme shows some benefit for enrichment in virtual screening, it is computationally demanding, requiring several hours per ligand using AutoDock Vina on a dual quad-core machine. Thus, it is impractical to use a large EPS for virtual screening of tens to hundreds of thousands of ligands. Reducing the number of protein structures in the ensemble is essential for efficient ensemble docking. The most naive way is to generate subensembles by clustering the

full EPS on the basis of their pairwise RMSD values and use the resulting cluster centers as the subensemble for virtual screening.

Figure 2 displays the enrichment using the most diverse ensemble members by QT clustering. With the exception of DHFR and for very small number of frames in the thrombin system, there is no significant drop in enrichment in comparison with the full EPS (Figure 3, yellow lines). This is consistent with the results in the previous section, where no significant difference in enrichment between static docking and docking using Limoc-RCS was observed for AChE, ERa, and CDK2.
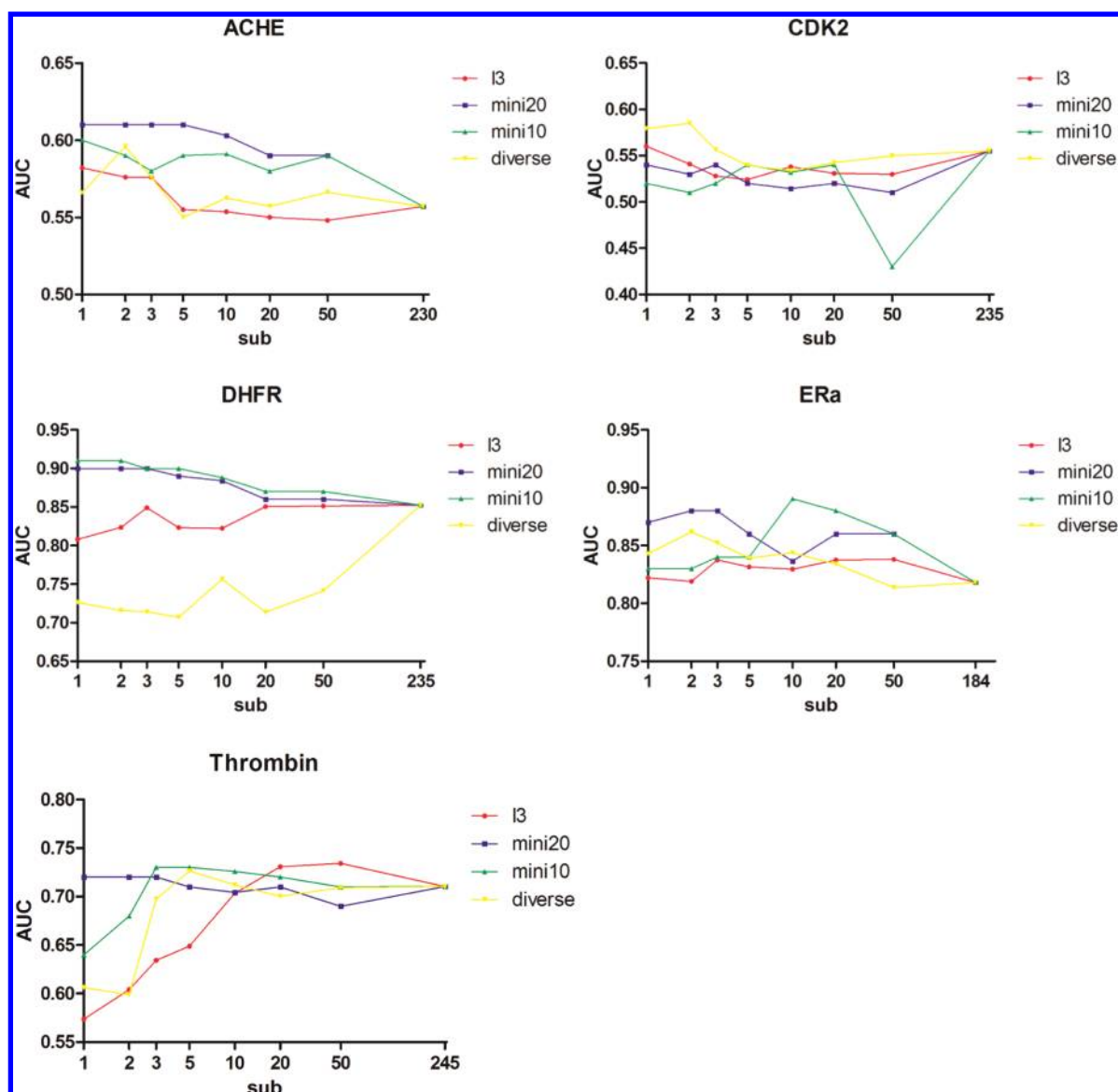
Figure 3. AUC as a function of the number of members in the subensembles of protein structures. The results for three different selection schemes are presented: selection by diversifying the EPS using pairwise RMSD values (yellow line), by training using binding-pose reproduction of three different cocrystallized ligands (red), and by optimizing the difference in scores between a small set of known actives and decoys (using 20 actives, blue; 10 actives, green).

Frame Selection by Training on Experimentally Known Binding Modes. Although clustering for EPS generation is promising, in some protein systems, this method resulted in reduced enrichment quality. Motivated by the studies of Anderson and Bolstad,[18] who used visual inspection of experimentally known holo structures of DHFR to guide the ensemble selection process, we hypothesized that automatic selection using a small set of experimental data may increase the accuracy of Limoc-RCS screening. In a first scheme, we used a training set composed of 1−3 holo structures for structure selection. The training set ligands were docked to all members of the EPS, and the members that are in best agreement with the experimentally determined binding mode were selected for subensembles. Figures 3 and 4 display the results for using a training set of three ligands. Results using one or two ligands are shown in the Supporting Information, Figure S2. Similar to the enrichment results for using the diversified subensembles (see previous section), no significant change in enrichment was observed for AChE, CDK2, and ERa. For DHFR, the decrease in enrichment quality was reduced compared to the diversified subsets but remained the same for thrombin. In summary, no significant improvement in enrichment was observed between the RMSD based selection scheme and the selection scheme based on binding-pose training.

Frame Selection by Training on Enrichment of Small Subsets. Although no significant improvement in enrichment was observed for docking to selected members of the EPS using training of known binding modes, we question whether utilizing training data that is strongly related to the real experiment, i.e., virtual screening, will increase enrichment of the full DUD data set. Sometimes medicinal chemists gather experimental knowledge about several actives before performing a full (virtual) high screening study. However, this useful information is seldom included in the virtual screening procedure. We tested whether selecting structures from the EPS on the basis of a mini-enrichment study using only a few actives might improve
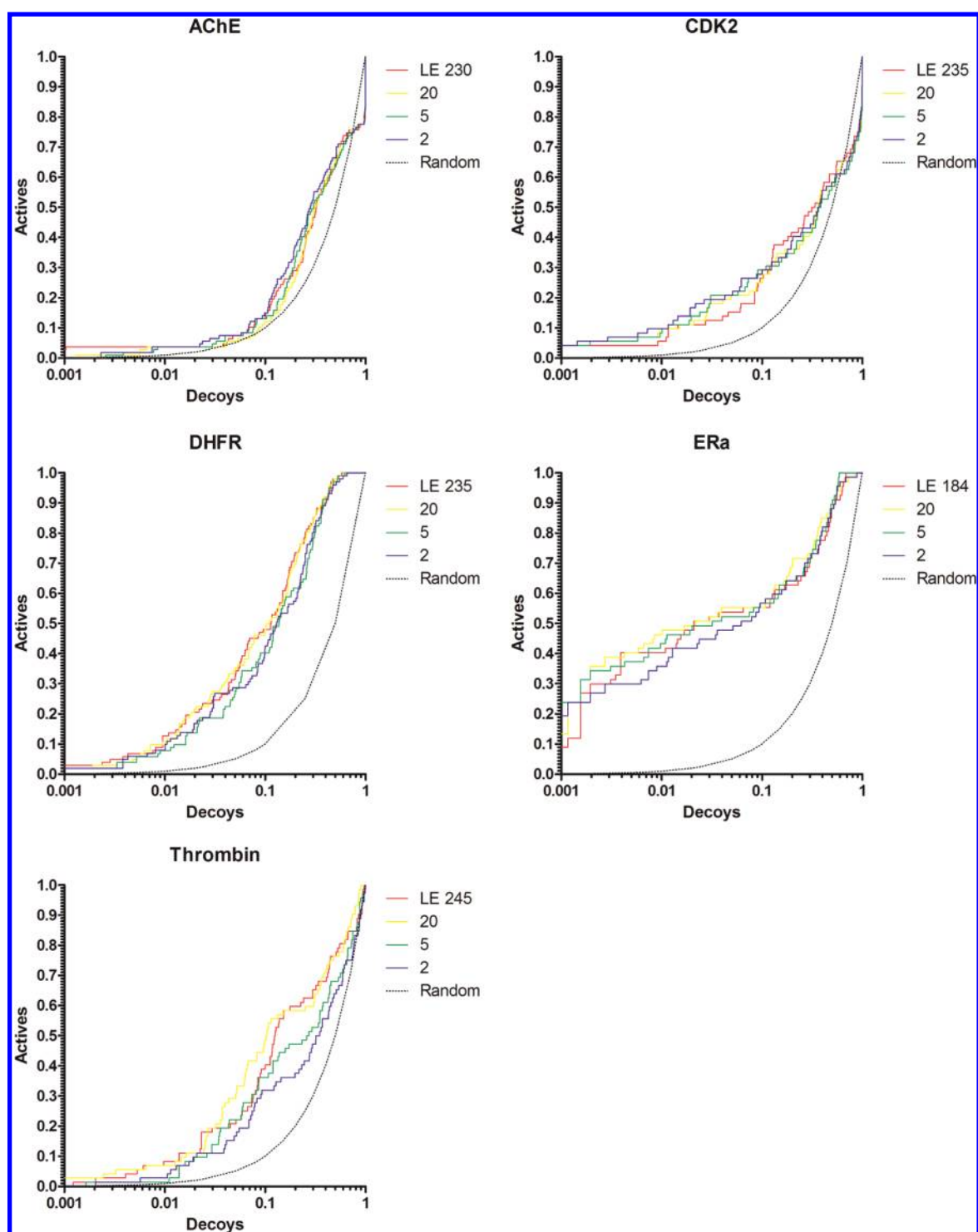
Figure 4. Docking enrichment plots for five protein systems using DUD. The fraction of actives ranked by docking score (*y* axis) is plotted against the fraction of ranked decoys (*x* axis; logarithmic scale). Results docking to the full Limoc EPS (LE; red line) and subensembles with 20 (yellow), 5 (green), and 2 (purple) members are shown. Members of the subensembles were selected on the basis of their potential to predict known binding poses correctly.

the overall virtual screening quality utilizing subensembles. A total of 10 or 20 actives with the same ratio of actives to decoys, both randomly picked from the DUD data set, were used in an initial enrichment experiment. Members of the EPS were selected on the basis of their ability to separate actives from decoys using scoring. These subensembles were used for virtual screening of the full ligand data set not including the actives and decoys utilized in the training process. Figures 3 and 5 and

Supporting Information Figure S3 show a general positive trend in using those subensembles, increasing or at least not reducing enrichment compared to the virtual screening results on the full EPS. As expected intuitively, utilizing a larger set of compounds for frame selection (20 vs 10 actives) typically resulted in a more stable increase of enrichment. For AChE, the AUC value increases with decreasing size of the subensemble, with the larger training set producing a stronger trend. For DHFR, the
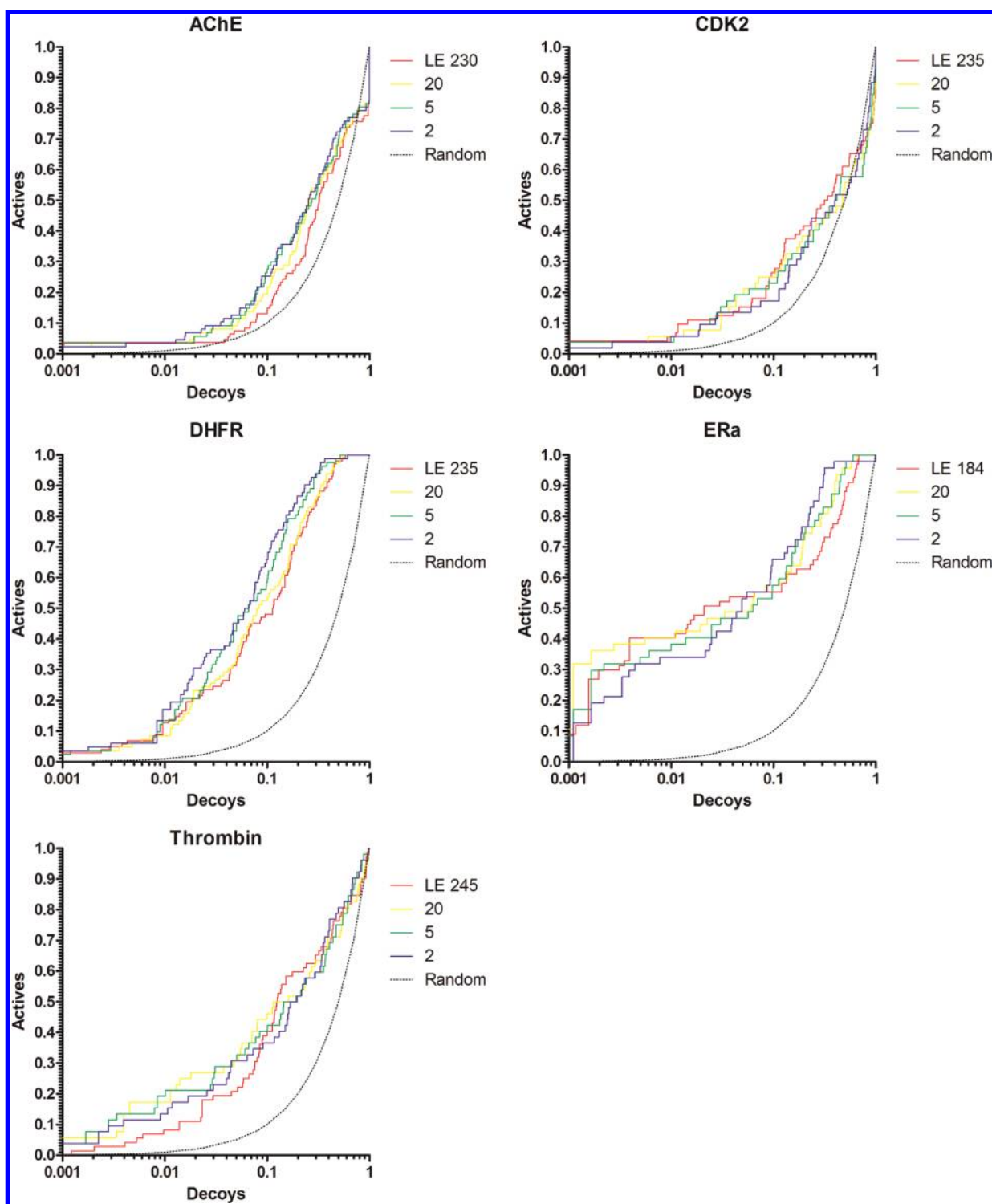
Figure 5. Docking enrichment plots for five protein systems using DUD. The fraction of actives ranked by docking score (y axis) is plotted against the fraction of ranked decoys (x axis; logarithmic scale). Results docking to the full Limoc EPS (LE; red line) and subensembles with 20 (yellow), 5 (green), and 2 (purple) members are shown. Members of the subensembles were selected on the basis of their potential to actives from decoys of a small subset of ligands from the DUD data set. The results for using 20 actives are displayed.

same positive trend of reducing ensemble size was observed for both training set sizes. For CDK2 and ERa, no significant decrease or increase in AUC with reducing ensemble size was observed. For thrombin, using a training set containing 20 actives results in a stable AUC vs ensemble-size curve. Although the AUC values of a training set with 10 actives decrease for very small ensemble sizes (one or two frames), the curve

demonstrates stability for subensembles with 3–50 frames. This displays a significant improvement over the previously tested frame-selection schemes. In general, using 20 actives yields stable enrichment for a small number of selected protein structures, and with some limitations this observation can also be extended to the use of 10 actives. For several systems, the AUC values are even increasing with reducing number of frames.

Comparison of Three Frame-Selection Methods. In the previous sections, we have qualitatively compared the results of three methods for reducing the number of structures in the EPS. Figure 3 provides the first quantitative assessment of how the AUC changes as a function of subensemble size. To extend this analysis, additional quantitative descriptors for enrichment quality are discussed in this section: EF1, EF10, EF20, and log AUC. Graphs for each descriptor as a function of subensemble size were plotted, similar to Figure 3, and linear regression was derived for all quantitative descriptors of virtual screening performance. Slopes of the linear regression models are displayed in Table 2 and Supporting Information Table S4.

Table 2. Slopes of Linear Regression Models on Quantitative Indicators for Enrichment Quality Using Different Training Methods (diverse = diversification of EPS using pairwise RMSD, l3 = selection of EPS members based on binding pose predictions of three ligands, mini20/min10 = selection of EPS members based on set of 20/10 known actives and decoys)[a]

diverse	Thrombin	DHFR	ERa	CDK2	ACHE
EF1	0.014	0.018	0.007	0.004	0.000
EF10	0.012	0.080	-0.031	0.019	-0.009
EF20	0.053	0.103	-0.039	0.041	-0.032
AUC	0.041	0.047	-0.018	-0.011	-0.008
LogAUC	0.011	0.040	-0.018	-0.011	-0.008

l3	Thrombin	DHFR	ERa	CDK2	ACHE
EF1	0.023	0.020	0.025	-0.023	0.000
EF10	0.054	0.037	0.007	-0.030	-0.018
EF20	0.113	0.063	0.000	0.006	-0.051
AUC	0.066	0.016	0.001	0.000	-0.013
LogAUC	0.037	0.020	0.011	-0.013	-0.013

mini20	Thrombin	DHFR	ERa	CDK2	ACHE
EF1	-0.002	-0.030	0.046	0.003	0.001
EF10	0.018	-0.088	-0.031	0.041	-0.063
EF20	0.024	-0.060	-0.048	0.055	-0.053
AUC	-0.008	-0.022	-0.025	0.001	-0.021
LogAUC	-0.001	0.017	-0.007	0.011	-0.022

mini10	Thrombin	DHFR	ERa	CDK2	ACHE
EF1	-0.026	-0.021	-0.008	0.005	-0.005
EF10	0.008	-0.088	-0.034	0.002	-0.041
EF20	0.034	-0.060	-0.042	0.040	-0.048
AUC	0.018	-0.004	0.009	-0.002	-0.013
LogAUC	-0.009	-0.016	-0.015	0.006	-0.013

[a]Negative regression (blue color) means increasing enrichment quality with reducing number of EPS members; positive regression in red.

It should be noted that the graphs of descriptor versus subensemble size do not necessarily always follow a linear functional form, as can be easily observed in Figure 3. It is not our aim in this analysis to derive optimal regression models but rather to provide trends for how the descriptors change as a function of subensemble size for the three different frame-selection methods. Table 2 demonstrates that among the three

selection methods, training on known actives yields the strongest trend for increasing enrichment performance when reducing the number of structures in the subensemble (negative sign), whereas for the other two selection schemes (diversification and training on binding modes), a negative trend (reduced enrichment with reducing subensemble size) was observed for several protein systems.

Structural and Energetic Analysis. In this section, we aim to further discuss two observations made throughout our studies: First, why do DHFR and thrombin in particular show improved enrichment quality using the Limoc ensemble compared to docking to a single apo or holo structure? Second, using the scheme in which protein structures are chosen that are able to separate actives from decoys in prescreening enrichment studies on a small subset of ligands, why do DHFR and AChE show most consistent improvement among all protein systems using a small number of selected protein structures (cf. Table 2).

To address the second question, we computed the difference between $\overline{S}_{actives,p}$ and $\overline{S}_{decoys,p}$ for all protein structures p. The cumulative probability over all protein structures is shown in Figure 6. Negative values in the plot correspond to protein
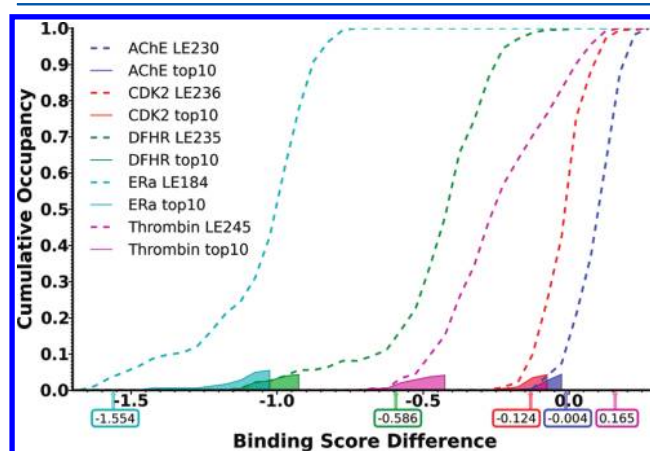

Figure 6. Cumulative distribution of protein structures as a function of the difference between $\overline{S}_{actives,p}$ and $\overline{S}_{decoys,p}$. Distributions of all structures of a Limoc ensemble are displayed as dashed lines; the corresponding subpopulations of 10 structures selected by training on subsets of actives are displayed as solid lines with transparent fillings. The difference between $\overline{S}_{actives,p}$ and $\overline{S}_{decoys,p}$ for docking to a single apo or holo structure is displayed as vertical arrows below the x axis including exact values.

structures for which, on average, actives are better ranked (more negative score) than decoys. DHFR (Figure 6, green dashed line) displays a large tail toward negative score differences, and using virtual screening of a subset of actives for protein structure selection, we were able to select protein structures from this tail (Figure 6, green shaded region). By selecting protein structures within the tail region, there is a larger separation distinguishing actives from decoys which ultimately leads to higher enrichment. ERa (cyan) also displays a long tail with negative score differences; however, it should be noted that all structures have a score difference smaller than −0.75. Thus, any selection of protein structures results in an overall good enrichment performance, and selecting few ERa protein structures does not improve enrichment quality with the same significance as for DHFR. The reason why AChE shows significant improvement, when selecting a few protein structures, is that only about 10% of all protein structures display a

slightly negative score difference (Figure 6, blue dashed line). Thus, the whole Limoc ensemble includes many structures which actually favor decoys over actives. Consequently, we observe poor enrichment for AChE using the whole Limoc ensemble (Figure 1), but selecting the few protein structures with a negative score difference (Figure 6, blue shaded region) can actually result in increased enrichment performance (Figure 5 and Table 2).

To address the question of why DHFR and thrombin display improved enrichment when docking to the Limoc ensemble compared to docking to a single static protein structure, we performed a detailed analysis characterizing the structural changes of the protein relevant to ligand binding by comparing distances between residues of the binding site. This analysis is similar to that used in our previous paper on Limoc RCS docking.[21] In detail, the backbone of each amino acid is represented by the C_α atom, and one atom represents the side-chain of each residue (except glycine): C_β for Ala, N_ε for Arg, C_γ for Asn, C_γ for Asp, S_γ for Cys, C_δ for Gln, C_δ for Glu, C_γ for His, $C_{\gamma 1}$ for Ile, C_γ for Leu, C_ε for Lys, S_δ for Met, C_γ for Phe, C_γ for Pro, O_γ for Ser, C_β for Thr, $C_{\delta 2}$ for Trp, C_ζ for Tyr, and C_β for Val. Distances d_{ij} are computed between all of those representing atoms of the residues that directly interact with a diverse set of ligands in their bioactive conformations. In the thrombin system, for example, residues His57, Tyr60A, Trp60D, Lys60F, Leu99, Ile174, Asp189, Ala190, Cys191, Glu192, Ser195, Val213, Trp215, Gly216, Glu217, Gly219, and Gly226 were considered to directly interact with the ligands of the three selected PDB structures used for training on known binding modes. We then identified the 10 protein structures for each system that displayed the largest negative difference between $\overline{S_{actives,p}}$ and $\overline{S_{decoys,p}}$. From these 10 structures, we computed the range of distances for d_{ij}; the minimum $d_{ij,min}$ and maximum $d_{ij,max}$ value for each distance is identified. Two variables describing the deviations from the optimal range of $d_{ij} \in \lfloor d_{ij,min}; d_{ij,max} \rfloor$ in the positive and negative directions were computed:

$$d_{ij}^{<} = \begin{cases} d_{ij} - d_{ij,min} & \text{if } d_{ij} < d_{ij,min} \\ 0 & \text{if } d_{ij} > d_{ij,min} \end{cases} \qquad (4a)$$

and

$$d_{ij}^{>} = \begin{cases} d_{ij} - d_{ij,max} & \text{if } d_{ij} > d_{ij,max} \\ 0 & \text{if } d_{ij} < d_{ij,max} \end{cases} \qquad (4b)$$

This separation into smaller and larger than optimal d_{ij} values is due to the assumption that values that are too small might not provide enough space for binding the ligand in this particular moiety of the binding site, whereas excessively large distances might not provide optimal interactions between the ligand and protein, resulting in a shift to more unfavorable score values.

Using all protein structures of the Limoc ensemble, we correlated the difference between $\overline{S_{actives,p}}$ and $\overline{S_{decoys,p}}$ for each structure with $d_{ij}^{<}$ and $d_{ij}^{>}$, to characterize the relative positions of amino acids in the binding site and their deviations from the best 10 protein structures. We performed a multilinear regression analysis for all protein systems using SAS software[29] correlating $d_{ij}^{<}$ and $d_{ij}^{>}$ with the difference between $\overline{S_{actives,p}}$ and $\overline{S_{decoys,p}}$:

$$\overline{S_{active,p}} - \overline{S_{decoys,p}}$$
$$= c_0 + \sum_{ij} c_{ij}^{<} \times d_{ij}^{<} + \sum_{ij} c_{ij}^{>} \times d_{ij}^{>} \qquad (5)$$

where c_0, $c_{ij}^{<}$, and $c_{ij}^{>}$ are the regression coefficients. The results for all five protein systems are presented as Supporting Information Figure S5. For thrombin, nine descriptors are identified in the multilinear regression model with an overall regression coefficient of $r^2 = 0.66$. The single descriptor $d^{<}(C_\delta,Glu192-C_\alpha,Trp215)$ explains about 30% of the total variance in the score-difference data. This particular distance varies between 12.3 Å and 14.7 Å for the 10 selected structures, with the largest average score difference between actives and decoys. The same distance in the apo thrombin structure is 1.5 Å smaller than $d_{ij,min}$. This is the largest difference to $d_{ij,min}$ or $d_{ij,max}$ of any descriptor for any of the five protein systems and may strongly contribute to the large difference in enrichment quality between docking to the Limoc ensemble and docking to the static apo structure.

In the context of the X-ray structure in two of the three selected protein structures of thrombin (1a4w, 1d3t; see Figure 7),
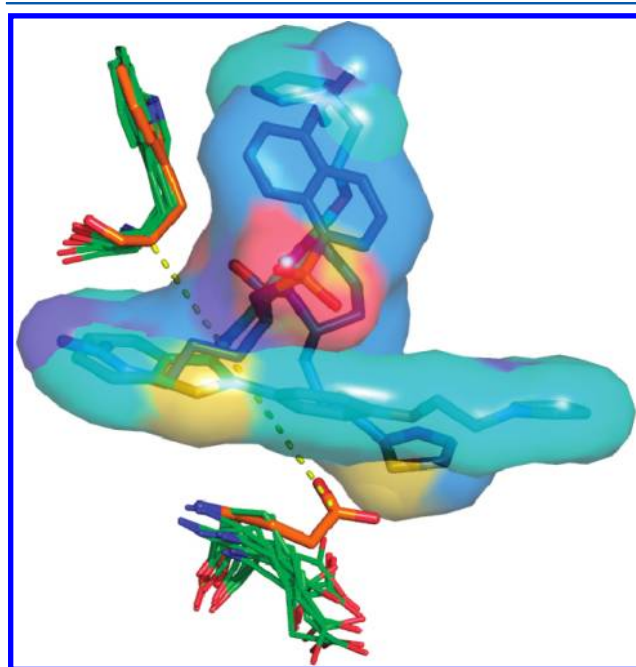

Figure 7. Comparison of the conformations of Glu192 and Trp215 in the apo structure of thrombin (carbon atoms colored in orange) and in the 10 structures chosen from the Limoc ensemble with the largest deviation between average scores of actives and decoys (carbon atoms in green). Two representative thrombin ligands in their native binding pose from the PDB databank (1a4w, 1d3t) are displayed with carbon atoms colored in light and dark blue, respectively. Glu192 in the apo structure sterically overlaps with the two ligands. The distance between C_δ of Glu192 and C_α of Trp215 in the apo form of thrombin is shown as a yellow dashed line.

the side-chain of Glu192 is oriented closer to the opposite site of the binding pocket (here represented by Trp215) and prevents certain classes of thrombin ligands from binding in their native conformation. This inward-pointing conformation of Glu192 is identified in the apo structure, whereas outward-pointing conformations are identified for the 10 protein structures with largest difference between $\overline{S_{actives,p}}$ and $\overline{S_{decoys,p}}$. Thus, docking to the apo structure of thrombin will result in poor enrichment performance (cf. Figure 1).

■ CONCLUSIONS

We have presented three different schemes to automatically select protein structures as alternative templates for ensemble

docking. The initial ensemble of protein structures was generated using our Limoc concept which is able to generate holo-like protein structures. In combination with AutoDock Vina and the RCS concept, we demonstrate that this methodology is capable of enriching data sets of compounds using virtual screening. Furthermore, we found that a balance can be achieved between the amount of protein flexibility incorporated into docking and the increased potential to generate false positives if too many template structures are considered in docking. If the protein structures optimal for virtual screening are selected by a feasible training process on experimental data, ensemble docking can be successfully performed on a very small number (3−5) of protein structures. The most successful selection method in our studies has been to choose protein structures that are able to separate actives from decoys in prescreening enrichment studies on a small subset of ligands. As a consequence of selecting a small number of protein structures for ensemble docking, the efficiency of the screening process can be significantly improved. In the most successful scheme, 10−20 active ligands have to be identified experimentally, but no exact affinity data or knowledge on binding modes of these actives is necessary.

It should be noted that the scoring function of AutoDock Vina, used in our study, was optimized for the correct prediction of binding poses and not for enrichment experiments. Thus, scoring functions that are optimized for virtual screening might increase the observed enrichment rates. We further speculate on whether a scoring function optimized in the context of the RCS scheme, considering protein flexibility and dynamics, has the potential to further improve enrichment quality. Research in this direction is ongoing in our lab.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Enrichment performance of docking to the Limoc ensemble using the average score and lowest score; enrichment performance for docking to subensembles obtained by frame selection by training using binding-pose reproduction of one, two, and three different cocrystallized ligands; docking enrichment plots for docking to subensembles obtained by frame selection by training on potential of protein structures to separate 10 actives from decoys; slopes of linear regression models on quantitative indicators for enrichment quality using selection of EPS members based on binding pose predictions of one and two ligands; and results of multilinear regression analysis correlating protein structure differences with average differences between scores for actives and decoys. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 765 496-9375. Fax: 765 494-1414. E-mail: mlill@
purdue.edu.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Teague, S. J. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discovery* **2003**, *2* (7), 527−541.

(2) Bossart-Whitaker, P.; Carson, M.; Babu, Y. S.; Smith, C. D.; Laver, W. G.; Air, G. M. Three-dimensional Structure of Influenza A N9 Neuraminidase and Its Complex with the Inhibitor 2-Deoxy 2,3-Dehydro-N-Acetyl Neuraminic Acid. *J. Mol. Biol.* **1993**, *232* (4), 1069−1083.

(3) Taylor, N. R.; Cleasby, A.; Singh, O.; Skarzynski, T.; Wonacott, A. J.; Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Bethell, R.; Colman, P.; Varghese, J. Dihydropyrancarboxamides Related to Zanamivir: A New Series of Inhibitors of Influenza Virus Sialidases. 2. Crystallographic and Molecular Modeling Study of Complexes of 4-Amino-4H-pyran-6-carboxamides and Sialidase from Influenza Virus Types A and B. *J. Med. Chem.* **1998**, *41* (6), 798−807.

(4) Carlson, H. A. Protein flexibility and drug design: how to hit a moving target. *Curr. Opin. Chem. Biol.* **2002**, *6* (4), 447−452.

(5) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337* (1), 209−225.

(6) Teodoro, M. L.; Kavraki, L. E. Conformational flexibility models for the receptor in structure based drug design. *Curr. Pharm. Des.* **2003**, *9* (20), 1635−1648.

(7) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 178−184.

(8) Beier, C.; Zacharias, M. Tackling the challenges posed by target flexibility in drug design. *Expert Opin. Drug Discovery* **2010**, *5* (4), 347−359.

(9) Rao, C.; Subramanian, J.; Sharma, S. D. Managing protein flexibility in docking and its applications. *Drug Discovery Today* **2009**, *14* (7−8), 394−400.

(10) Sotriffer, C. A. Accounting for Induced-Fit Effects in Docking: What is Possible and What is Not? *Curr. Top. Med. Chem.* **2011**, *11* (2), 179−191.

(11) Lin, J. H. Accommodating Protein Flexibility for Structure-Based Drug Design. *Curr. Top. Med. Chem.* **2011**, *11* (2), 171−178.

(12) Lill, M. A. Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. *Biochemistry* **2011**, *50* (28), 6157−6169.

(13) Bahar, I.; Lezon, T. R.; Yang, L. W.; Eyal, E. Global Dynamics of Proteins: Bridging Between Structure and Function. *Ann. Rev. Biophys.* **2010**, *39*, 23−42.

(14) Chennubhotla, C.; Rader, A. J.; Yang, L. W.; Bahar, I. Elastic network models for understanding biomolecular machinery: From enzymes to supramolecular assemblies. *Phys. Biol.* **2005**, *2* (4), S173−S180.

(15) Barril, X.; Morley, S. D. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J. Med. Chem.* **2005**, *48* (13), 4432−4443.

(16) 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* (9), 693−705.

(17) Armen, R. S.; Chen, J.; 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* (10), 2909−2923.

(18) Bolstad, E. S.; Anderson, A. C. In pursuit of virtual lead optimization: pruning ensembles of receptor structures for increased efficiency and accuracy during docking. *Proteins* **2009**, *75* (1), 62−74.

(19) 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* (20), 5632−5633.

(20) 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* (1), 47−62.

(21) Xu, M.; Lill, M. A. Significant enhancement of docking sensitivity using implicit ligand sampling. *J. Chem. Inf. Model.* **2011**, *51* (3), 693−706.

(22) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285* (4), 1735−1747.

(23) Case, D. A.; Cheatham, T. E. III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668−1688.

(24) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49* (23), 6789−6801.

(25) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455−461.

(26) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785−2791.

(27) Elber, R.; Karplus, M. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* **1990**, *112* (25), 9161−9175.

(28) Heyer, L. J.; Kruglyak, S.; Yooseph, S. Exploring expression data: identification and analysis of coexpressed genes. *Genome Res.* **1999**, *9* (11), 1106−1115.

(29) The multilinear regression data analysis for this paper was generated using SAS software, version 9.2 of the SAS system for Windows, copyright 2002−2008, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.