Incorporating Dynamics in *E. coli* Dihydrofolate Reductase Enhances Structure-Based Drug Discovery

Michael G. Lerner,[†,§] Anna L. Bowman,[‡,§] and Heather A. Carlson*[,†,‡]

Biophysics Research Division and Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1065

Received May 17, 2007

Escherichia coli dihydrofolate reductase (DHFR) is a long-standing target for enzyme studies. The influence of protein motion on its catalytic cycle is significant, and the conformation of the M20 loop is of particular interest. We present receptor-based pharmacophore models—an equivalent of solvent-mapping of binding hotspots—based on ensembles of protein conformations from molecular dynamics simulations of DHFR·NADPH in both the closed and open conformation of the M20 loop. The optimal models identify DHFR inhibitors over druglike noninhibitors; furthermore, high-affinity inhibitors of *E. coli* DHFR are preferentially identified over general DHFR inhibitors. As expected, models resulting from simulations with DHFR in the productive conformation with a closed M20 loop have better performance than those from the open-loop simulations. Model performance improves with increased dynamic sampling, indicating that including a greater degree of protein flexibility can enhance the quest for potent inhibitors. This was compared to the limited conformational sampling seen in crystal structures, which were suboptimal for this application.

INTRODUCTION

Dihydrofolate reductase (DHFR; EC. 1.5.1.3) catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) (and more slowly, folate to dihydrofolate) while concurrently oxidizing the nicotinamide adenine dinucleotide phosphate (NADPH) cofactor.[1] DHFR is a small enzyme that plays an essential role in various cellular processes including the biosynthesis of DNA.[2] It is the only source of THF, an essential precursor for purine and thymidylate biosynthesis, and hence, it is a long-standing target for anticancer drugs and antibacterial agents. Moreover, *Escherichia coli* DHFR (ecDHFR) is a canonical system for studying enzyme structure, dynamics, and catalysis. In fact, ecDHFR was recently the focus of a data-mining and docking competition[3] which highlighted possible improvements needed in the field.[4]

Protein flexibility and dynamics are of utmost importance in understanding the structure and mechanism of DHFR. The role of protein dynamics in DHFR catalysis has been well studied both computationally[5−7] and experimentally through NMR spectroscopy[8−10] and multiple X-ray crystallography structures.[11] The M20 loop of ecDHFR (residues 9−24) adopts three different conformations: closed, open, and occluded.[11] When the substrate and cofactor are both bound, the closed conformation is adopted, in which the M20 loop is packed against the nicotinamide ring of the cofactor.[11] The closed form presents the only conformation in which the substrate and cofactor are arranged favorably for reaction, and it is also present in the preceding state of the catalytic cycle, DHFR·NADPH.[6] The closed form is also the only conformation seen in DHFR from all other species, regardless of the crystal packing and ligands bound in the substrate and cofactor binding sites.[11] The occluded M20 loop is observed in the product complexes (DHFR·$NADP^+$·THF, DHFR·THF, and DHFR·NADPH·THF). It is an unproductive conformation where the central part of the M20 loop forms a helix, blocking access to the binding site for the nicotinamide moiety of the cofactor. This forces the nicotinamide of the cofactor out into solvent, making it unresolved in the crystal structures of DHFR·$NADP^+$·THF and DHFR·NADPH·THF.[11] The open loop is a conformational intermediate between the extremes of the closed and occluded loops. The significance of M20 loop dynamics in ligand binding and catalysis has been an area of great interest, driven by the persuasive evidence of conformational change in the loop during the catalytic cycle and its interaction with the substrate and cofactor.[11]

Furthermore, there is evidence that the degree and pattern of protein flexibility changes throughout the catalytic cycle.[11] All regions of DHFR dynamically contribute to substrate binding and chemical turnover.[12] The M20 loop conformations and subdomain rotations are both pivotal during catalysis, implying coupling to the reaction coordinate.[11,13] The M20 loop undergoes fluctuations on the ps−ns time scale, some of which are analogous to the rate-limiting step of product release.[14] During molecular dynamic (MD) simulations of the DHFR·NADPH·DHF complex, strongly correlated and anticorrelated motions were observed across distant regions of the protein.[5] This extensive coupling was only seen in the reactant form, not in either the DHFR·$NADP^+$·THF or DHFR·NADPH·THF complexes, implying that these motions may be necessary for catalysis.[5] NMR relaxation experiments indicate that binding events influence not only active-site dynamics but also distal motions.[9,10]

In light of this evidence, it is imperative that protein flexibility be included when considering inhibitor binding in ecDHFR. Here, we present a computational study of the influence of conformational behavior of ecDHFR on inhibitor identification. We employ multiple protein structures (MPS) to incorporate protein flexibility in structure-based drug discovery. Our technique identifies complementary interac-

* Corresponding author phone: (734)615-6841; fax: (734)763-2022; e-mail: carlsonh@umich.edu.
† Biophysics Research Division.
‡ Department of Medicinal Chemistry.
§ These two authors contributed equally to this work.

10.1021/ci700167n CCC: $37.00
Published on Web 09/18/2007

tions within the binding site of an ensemble of conformational states.[15,16] Conserved regions, where the same complementary interactions are consistently made in a majority of the MPS, define the most essential binding hotspots for different chemical functionalitites. The conformational states can be taken from MD simulation,[16−18] X-ray crystallographic structures,[15,19,20] or an NMR ensemble.[20] The method, initially applied to HIV integrase,[15,16] has been successfully extended to HIV-1 protease[17,18] and other species of DHFR.[19]

The dynamic motions of ecDHFR are crucial to ligand binding and subsequent catalysis. Additionally, incorporating protein flexibility expands the chemical and conformational space of the predicted ligands. The use of pharmacophore models, rather than specific, predetermined chemical scaffolds, also greatly increases the available chemical space for the hit list. The incorporation of greater flexibility through extending the length of the sampled MD simulation has shown an improvement of the identification of known HIV-1 protease inhibitors over druglike noninhibitors with our method.[17,18] Our recent work with DHFR showed that the inclusion of protein flexibility did not result in a loss of specificity across DHFR from different species.[19] In this study, we have used snapshots from an MD simulation to develop models from the unliganded complex DHFR·NADPH with the M20 loop in both the closed and open conformations. This use of the DHFR·NADPH complex is analogous to that of using *apo* structures; the substrate binding site has no predetermined conformation relating to a specific inhibitor. We also utilize experimental structures to develop a pharmacophore model, in a similar way to our previous work based on DHFR from other species.[19] The MPS method has already been shown to be successful when investigating *apo* HIV-1 protease structures,[17,18] and the ecDHFR pharmacophore models developed here are highly selective for known DHFR inhibitors over druglike noninhibitors. We also show that extending the length of the MD simulation to increase sampling of protein flexibility improves model performance, especially in the model resulting from the simulation of DHFR with the M20 loop in the open, unproductive orientation. This study is a further indication of how important the inclusion of conformational behavior can be in enhancing drug discovery.

METHODS

Protein Preparation. The starting models for the MD simulations were based on two crystal structures of the wild-type *E. coli* DHFR·NADPH complex, representing both the closed (1RX1) and open (1RA1) M20 loop.[11] The 1RA1 structure was altered to include the entire cofactor, building in the nicotinamide ribose group (Figure 1) by direct alignment with the coordinates in the 1RX1 structure. In complexes with the open loop, the nicotinamide ribose group is mobile and moves in and out of the cofactor binding site. We chose to represent it in the binding pocket, as electron density indicates that the pocket is occupied 75% of the time.[11] The occupancy is increased to 100% upon binding of the substrate and closure of the M20 loop.[11] This conformation is more suited to modeling inhibitor binding and hence more appropriate for this study.

Side-chain orientations and protonation states were checked with a custom script in PyMOL.[21] Hydrogen atoms were placed using the LEaP routine[22] and optimized with conjugate gradient energy minimization (convergence criterion: rms gradient <0.0001 kcal/mol Å). The force-field parameters

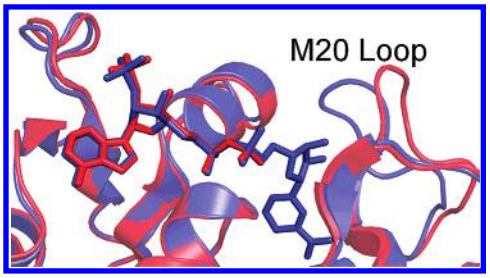

Figure 1. The crystal structures of 1RX1[11] in blue with the closed M20 loop conformation and 1RA1[11] in red in the open M20 loop conformation. NADPH is shown in stick representation; the nicotinamide ribose moiety of the cofactor is not resolved in the 1RA1 structure and was built in using the appropriate coordinates from the 1RX1 structure. The M20 loop is to the right of the nicotinamide.

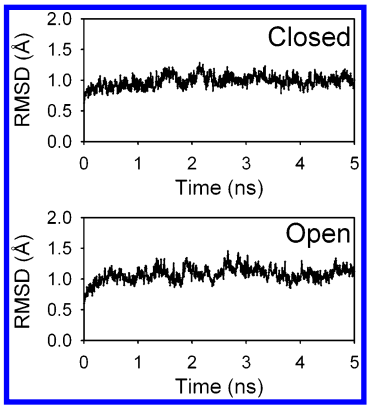

Figure 2. Plots reflecting the rmsd between the equilibration structure of (a) M20 closed-loop and (b) M20 open-loop ecDHFR and respective trajectory snapshots versus time.

for the cofactor were provided by Ryde et al.[23] The protein was then solvated in a truncated octahedral box of TIP3P[24] water extending 15 Å from the surface of the protein. To give the system a net neutral charge, 15 sodium counterions were added 10 Å from the protein surface. A preliminary minimization of the entire system was performed with 300 steps of a conjugate gradient method to remove initial bad contacts. The proteins heavy atoms were then held fixed. The water molecules were heated from 10 to 310 K over 50 ps, followed by equilibration (310 K) for 50 ps at constant volume and a subsequent 300 ps with constant pressure. The entire system then underwent heating from 10 to 310 K over 50 ps, followed by equilibration (310 K) for 50 ps at constant volume. Finally, 500 ps of all-atom equilibration at constant pressure was performed, and 5 ns of production phase MD were collected. Analysis of the MD (Figure 2) trajectory showed that the system takes an unusually long time to equilibrate, particularly the more mobile open conformation. This same slow equilibration was also found in other dynamics studies of ecDHFR.[5] For this reason, we discarded the first nanosecond of production dynamics, leaving 4 ns of sampling dynamics. This was done for both simulations to maintain consistency. Snapshots were saved from the dynamics trajectory every 100 ps. All simulations were carried out with AMBER.[25]

Comparison to Pharmacophore Models from X-ray Crystallographic Structures. In our previous work, we utilized X-ray crystallographic structures of DHFR from other species which had seven or more complexes in the PDB.[19] There are only crystal structures of wild-type ecDHFR bound to four unique ligands,[11,26] but it is important to compare models based on MD to those based on

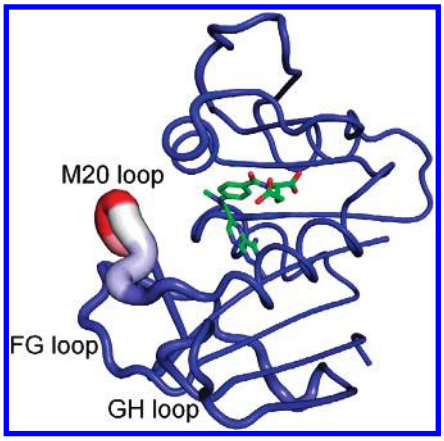

Figure 3. Average backbone structure for four X-ray crystallographic structures of ecDHFR. A red, thicker tube indicates greater rmsd across the ensemble, whereas a blue, narrow tube shows limited flexibility. Loop regions are labeled in the model. As expected, greater flexibility is seen in the loops than in the core of the protein. The bound conformation of methotrexate is superimposed on the model to orient the reader.

experimental structures. X-ray crystallographic structures of the wild-type ecDHFR with unique ligands bound and in which the cofactor was fully resolved were downloaded from the PDB.[27] If a structure with an identical ligand existed, the structure with the better resolution was taken. The four resultant structures (1RA2,[11] 1RC4,[11] 1RX3,[11] and 2ANQ[26]) had an average Cα rmsd of 0.56 Å. Figure 3 shows that the flexibility comes almost entirely in the M20 loop region; there is very little conformational variation in the rest of the structure.

Probe Flooding, Minimization, and Clustering. The procedure for the model setup was similar to that for our recent work on DHFR from other species.[19] The active site of each snapshot was flooded with 1000 small molecule probes using an 11 Å sphere centered at the midpoint of Phe31 and the nicotinamide ring of the cofactor, using an in-house program to randomly pack the probes in their initial locations. The size and shape of the binding site (in particular, the fact that it is relatively small and deep) required a denser placement, in order to fully explore relevant interactions, as compared to our earlier studies on HIV protease. This was achieved by doubling the number of probes used, while keeping the flooding sphere size similar. Benzene probes were used to identify aromatic and hydrophobic interactions, ethane probes were used to distinguish hydrophobic interactions from aromatic, and methanol probes were used to identify hydrogen-bonding interactions. Low-temperature Monte Carlo minimizations were performed by the Multi-Unit Search for Interacting Conformers (MUSIC) routine in BOSS,[28] using the OPLS force field.[29] In MUSIC, the protein is held fixed, while the probe molecules undergo simultaneous multiple gas-phase minimization. The probes cluster into local minima which define complementary binding regions. Clusters containing eight or more probes are identified with an automated procedure, based on Jarvis-Patrick methodology. Each cluster is represented by its "parent", the lowest-energy probe in the group.

Each snapshot was overlaid to the final equilibration structure using a Gaussian-weighted rmsd alignment[30] to give a common frame of reference. Parent probes within 8 Å of the center of the binding site (defined by Phe31 Cγ) were combined and clustered to give "consensus clusters". A consensus cluster must contain parent probes from ≥50% of the protein conformations. Pharmacophore elements were centered on the average position of the probes in each consensus cluster, and the radius was given by the rmsd of the probes in the cluster to the center of the cluster. Protein flexibility is implicitly included because only probes in conserved (rigid) regions are incorporated into consensus clusters. In flexible regions, where no consensus clusters were present, no limitations or requirements are set. This expands the range of chemical and conformational space that can be explored. Small areas of highly conserved steric constraints within the substrate binding pocket were represented by the inclusion of four excluded volume elements. These were each given a radius of 1.5 Å and were centered at the average position of Ile5 C, Ala 7 Cα, Phe31 Cγ, and Ile50 Cα.

Pharmacophore models were created using snapshots from the first 1, 2, and the full 4 ns of sampling dynamics. A total of 10 snapshots were used for the 1 ns model (taken every 100 ps), and 20 snapshots were used for the 2 and 4 ns models (taken every 100 ps or 200 ps, respectively).

Creation of the Ligand Databases. MOE[31] was used to screen three databases against the models: one set of 50 high-affinity inhibitors, one set containing 541 general DHFR inhibitors, and a general set of druglike noninhibitors. A total of 107 ecDHFR inhibitors were taken from the literature, each with an $IC_{50} \leq 1\ \mu M$. The full set of structures and affinity data are provided with references in the Supporting Information (the data set is completely compatible with that from our earlier work). These inhibitors were merged with the database of 493 DHFR inhibitors from our previous work[19] to yield a database of 591 unique inhibitors (there were nine duplicates). From this database, the top 50 ecDHFR inhibitors were selected for the high-affinity data set; the IC_{50} values ranged from 2 to 28 nM. The remainder of 541 inhibitors was used as the general DHFR inhibitor set. Our previous set of 2303 druglike decoy molecules,[19] obtained from the CMC,[32] was also used in this work. These decoy molecules had a molecular weight greater than 100 and contained one hydrogen-bond donor and one aromatic atom; the set did not contain any folate-like molecules.[19] Rule-based torsion driving in OMEGA[33] was used to produce multiple conformations of each molecule, using an energy cutoff of 14 kcal/mol calculated with the MMFF force field and a heavy-atom rmsd criterion of 1 Å. These pregenerated conformations were compared to the pharmacophore models. A compound was counted as a hit if one conformation could be aligned to the pharmacophore coordinates.

Evaluation of Pharmacophore Models. The predictive performance of each pharmacophore model was estimated by generating a receiver operator characteristic (ROC) curve, where the percentage of true inhibitors (true positives) found is plotted against the percentage of druglike noninhibitors (false positives) identified. The optimal model would lie at the point (0,100) identifying no false positives but finding all true positives, whereas models lying on a line that passes through the origin with a slope of 1 would be no better than random.

As in our previous studies, the number of elements required was varied, e.g., requiring n, $n - 1$, or $n - 2$ elements from an n-site pharmacophore model.[19] A multiplication factor of 1−3 in 1/3 increments was used to increase the radius of each pharmacophore element (e.g., radii = 1 × rmsd, 1.3 × rmsd, 1.7 × rmsd). This is in keeping with similar studies on HIV-1 protease.[17,18] However, the much smaller rmsd values of the consensus clusters found in the

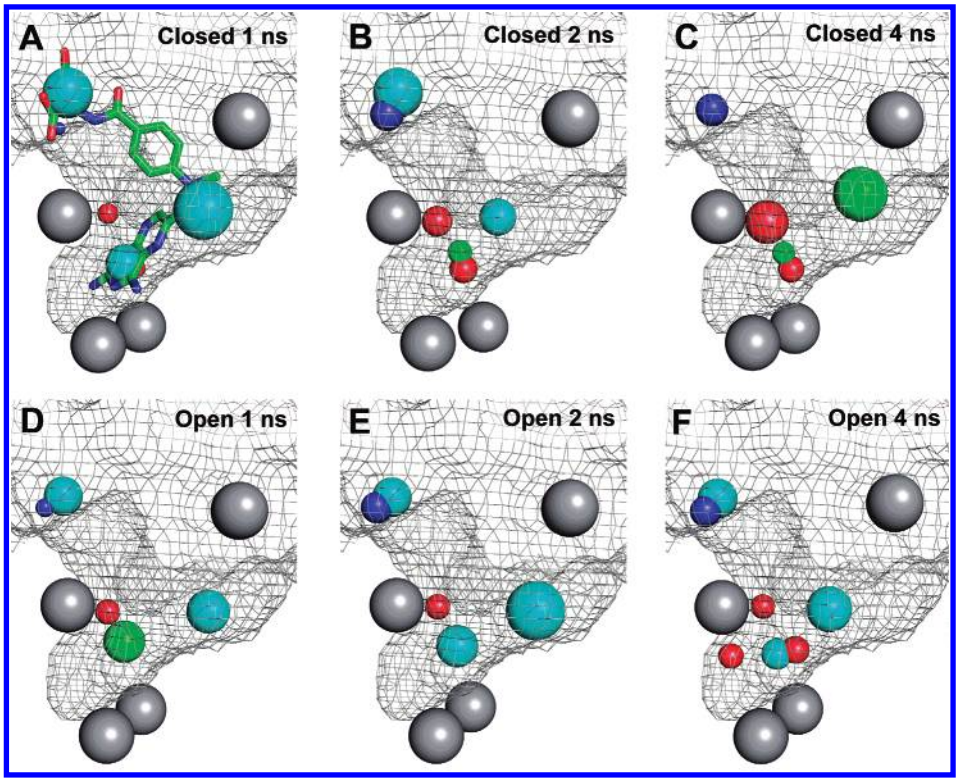

Figure 4. Pharmacophore models are given based on snapshots from (a) 1 ns, (b) 2 ns, and (c) 4 ns of sampling dynamics of the M20 closed-loop conformation of *E. coli* DHFR·NADPH (based on 1RX1). Also given are the models resulting from (d) 1 ns, (e) 2 ns, and (f) 4 ns sampling dynamics of the M20 open-loop conformation of *E. coli* DHFR·NADPH (based on 1RA1). Elements are shown with radii = 1 × rmsd: green spheres map aromatic interactions, cyan spheres require aromatic or hydrophobic groups, blue are hydrogen-bond acceptors, and red are hydrogen-bond donors. The gray spheres are excluded volumes. The molecular surface, from 1RX1 or 1RA1 appropriately, is shown in gray; the bound conformation of methotrexate is superimposed on the 1-ns model to orient readers familiar with the DHFR binding site.

X-ray model required extension to 6 × rmsd to produce reasonable models with radii in the range 0.8−4.1 Å (in our previous work we found optimal models had elements with radii in the range 0.6−4.1 Å).[19]

RESULTS AND DISCUSSION

Comparison of MPS Models Based on the Closed-Loop Conformations. Pharmacophore models resulting from the 1 and 2 ns MD simulations of ecDHFR in the closed-loop conformation have six elements, while the model from the longer 4 ns of sampling has just five. Figure 4 shows that there are five common elements between these three models. An aromatic element (aromatic/hydrophobic in the 1-ns model) makes an interaction with Phe31 at the bottom of the binding site. Two hydrogen-bond donor elements make interactions with NADPH O7 and Asp27 Oδ2 deep in the pocket. The pteridine type moiety of folate and known inhibitors is characteristically mapped by these three elements.

At a region corresponding to the *p*-aminobenzoate linker region of the substrate an aromatic (aromatic/hydrophobic in the 1-ns and 2-ns models) element is present, making an interaction with the cofactor. A hydrogen-bond acceptor in the mouth of the binding site makes an interaction with Arg57; this contact can be fulfilled by a moiety such as the glutamate tail of folate and some known inhibitors such as methotrexate. The 1-ns and 2-ns models also contain an additional sixth element in this area, an aromatic/hydrophobic site, which is lost with the greater flexibility incorporated in the 4-ns model. This sixth element occupies the Cα region of the glutamate tail of folate; however, it is not essential to high-affinity inhibitors of ecDHFR. It is reasonable that incorporating more flexibility shows this site to be nonessential for inhibitors.

Comparison of MPS Models Based on the Open-Loop Conformations. The models from the 1-ns and 2-ns simulations of the open M20 loop both had 5 pharmacophore elements: the aromatic/hydrophobic element at the bottom of the pocket (aromatic in the 1-ns model); the hydrogen-bond donor to Asp27 Oδ2; the aromatic/hydrophobic element midway in the active site; a hydrogen-bond acceptor in the mouth of the binding pocket; and an aromatic/hydrophobic element, also in the mouth of the pocket. The 4-ns model has two additional hydrogen-bond donor sites, one similar to that seen in the closed-loop interacting with NADPH O7 and another interacting with Asp27 Oδ1. The appearance of these sites adds more specificity, which improves the model by reducing the percentage of false positives identified (see performance discussion below). In particular, the appearance of hydrogen-bond donor site with relation to the cofactor is seems to be crucial to the model's performance. In the closed loop simulation, the distance between the Cδ of Phe21 and O7 of the cofactor averages 7.8 Å. When the M20 loop is open, NADPH is not packed in close to the substrate binding site,[11] and this same distance is longer. In the open-loop simulations, the distance between Phe31 Cδ and the cofactor O7 averages 8.3 Å over the first 2 ns, but it is 7.9 Å in the last 2 ns. As more dynamics are incorporated into the pharmacophore model, conformations of the loop more similar to the closed form are sampled, which is consistent with Wright's observation that small populations

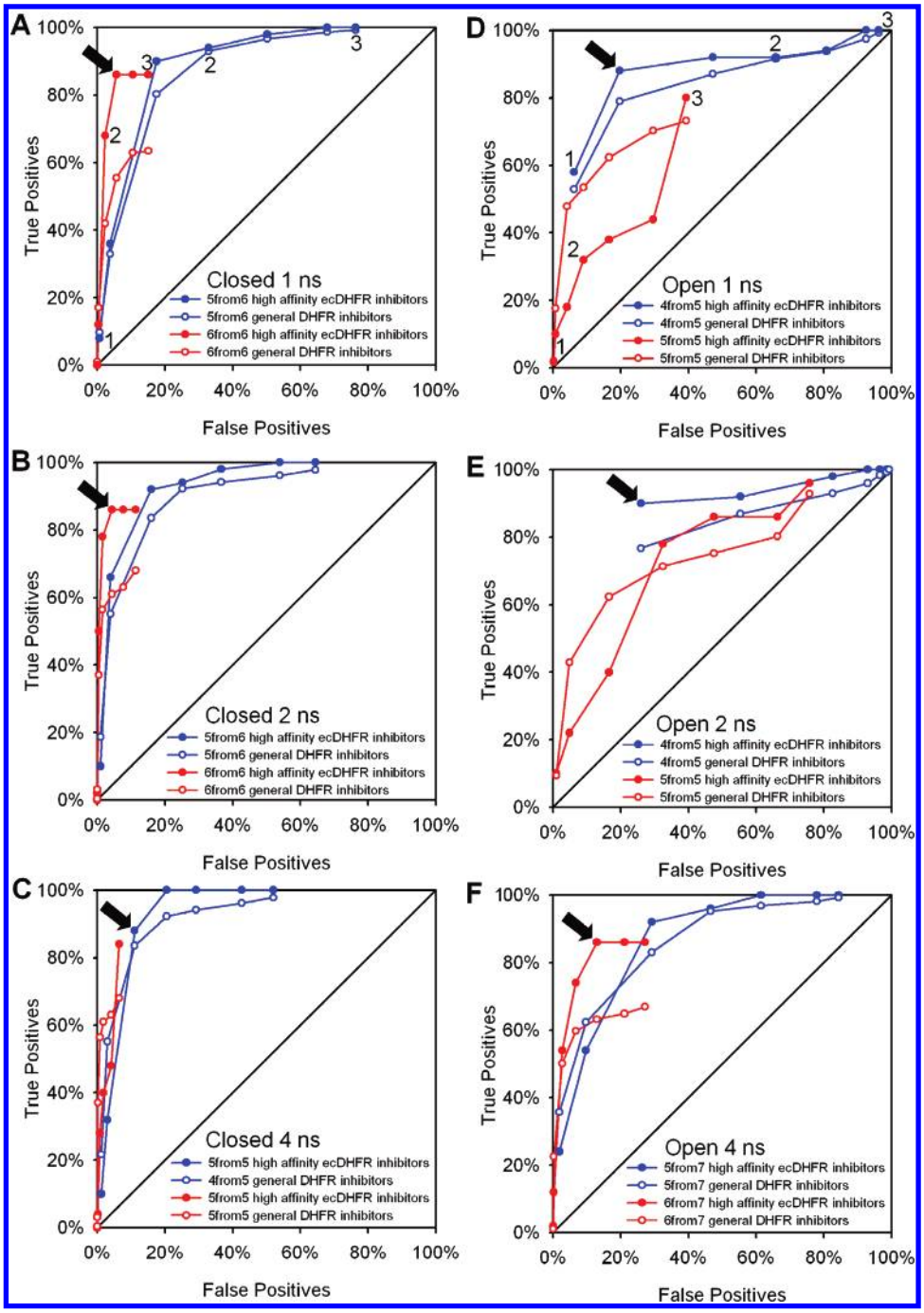

Figure 5. ROC curves for the closed-loop and open-loop pharmacophore models. Series with filled data points are results from screening the high-affinity ecDHFR database; those with open data points are the screening of general DHFR inhibitors. Shown in red are results from models requiring n hits from an n-site pharmacophore model, and shown in blue are results from models requiring $n - 1$ hits from an n-site pharmacophore model. Points along each series represent an increase in pharmacophore element radii of 1 × rmsd to 3 × rmsd. Arrows indicate the optimal model.

of alternate conformations are part of the observerable ensemble.[9] This may explain why the pharmacophore element representing an interaction with NADPH O7 is only present in the 4-ns pharmacophore model based on the open-loop simulation.

Performance of the MPS Pharmacophore Models. Each model was screened against three databases: one set of 50 potent ecDHFR inhibitors, one set of 541 general DHFR inhibitors, and one broad set of 2303 druglike noninhibitors. The number of elements required was set by allowing a partial match in the pharmacophore search. The radius of each element was manually altered in the pharmacophore query editor in MOE[31] to give the required radii multiplication, e.g., 2 × rmsd. The best pharmacophore models hit the maximum of true positives with the least number of false positives being identified; the optimal model is defined as having the shortest distance from (0,100). The performance of pharmacophore models generated from 1, 2, and 4 ns of dynamic sampling of DHFR·NADPH with both the closed (1RX1) and open (1RA1) M20 loop conformations is given in Figure 5. Both sets of models identify DHFR inhibitors, both species-specific and general, over noninhibitors. The optimal pharmacophore model for each system preferentially hits potent ecDHFR inhibitors over other general DHFR

inhibitors. It is a concern that incorporating protein flexibility to expand the chemical space may lead to a more general model with a lack of specificity, but this is not the case with the MPS method.[19]

It is encouraging that using structures with cofactor but no folate mimic, analogous to *apo* structures and hence not possessing conformational bias toward bound inhibitors, produced effective pharmacophore models. Our MPS method is one of very few that can successfully identify inhibitors for unbound, open binding sites. However, it is clear that the models for the closed-loop, the conformation most similar to that of the bound complex prior to reaction, are superior to those of the open loop.

The optimal pharmacophore model from the closed-loop 1-ns simulation (6 from 6, radii = 2.3 × rmsd) identified 86% of the high-affinity ecDHFR inhibitors and 56% of the general DHFR inhibitors with just 6% of the false positives being hit. The optimal 2-ns model (also 6 from 6, radii = 2.3 × rmsd) also identified 86% of the high-affinity inhibitors, a higher percentage (61%) of the less potent DHFR inhibitors were also identified, and even fewer of the druglike noninhibitors were hit (4%). The optimal closed-loop 4-ns model (4 from 5, radii = 1.7 × rmsd) results in an increase of both the high-affinity ecDHFR (88%) and general DHFR (84%) inhibitors identified while maintaining a low false positive hit rate (11%). The bias toward high-affinity ecDHFR inhibitors over other DHFR inhibitors indicates that specificity has not been lost during the incorporation of protein flexibility. In fact, more flexibility results in the identification of more known inhibitors.

Although the optimal pharmacophore model from the open-loop 1-ns simulation (4 from 5, radii = 1.3 × rmsd) identifies 88% of potent ecDHFR inhibitors and 79% of general DHFR inhibitors, it also hits 20% of the false positives. The optimal model from the 2-ns simulation (4 from 5, radii = 1.0 × rmsd) is similar in performance, identifying 90% of the high-affinity ecDHFR inhibitors and 77% of the less potent inhibitors; however, this model also hits 26% of the druglike noninhibitors. It is only when the simulation time is increased to 4 ns that an improvement in open-loop model performance is observed. The optimal 4-ns model (6 from 7, radii = 2.3 × rmsd) hits a comparable number of inhibitors from both the high-affinity ecDHFR (86%) and general DHFR (63%) inhibitor sets, and there is a marked decrease in the percentage of noninhibitors being falsely identified (13%).

It is encouraging to see that increased sampling of conformational space, by extending the simulation time from 2 to 4 ns, can improve the relatively poor open-loop model by reducing the percentage of false positives identified. Recent work by Boehr et al. used NMR relaxation dispersion to study ecDHFR to develop a dynamic energy landscape of catalysis.[9] They found that each intermediate in the catalytic cycle was comprised of a dominate conformation of the protein plus one or two additional conformational states with much lower populations. These higher-energy states resembled the neighboring intermediate in the cycle. This implies that ligand binding occurs by a conformational selection.[34] A small percentage of pre-existing conformations, resembling the bound conformation, coexist in conjunction with the unbound state. A ligand can then bind to the higher-energy conformation, causing an equilibrium shift toward the ligand-bound conformation, which then becomes predominant. The work also suggested that ligand release works in a similar fashion, with the protein adopting a higher-energy conformation resembling the unbound state, before ligand dissociation had occurred.[9] It is possible that extending the dynamic simulations allows more of the "excited states" resembling the closed M20 loop to be sampled, which improves the models performance.

Improvement with increased dynamic sampling was also seen with HIV-1 protease where it was shown that extending the simulation length resulted in identifying more true inhibitors and less false positives.[17] This also adds substantiation to our earlier finding with human, *P. carinii*, and *C. albicans* DHFR where models based on crystal structures with the greatest structural variation identified the most high-affinity inhibitors and the least false positives compared to those with reduced flexibility.[19] The optimal pharmacophore models from the closed-loop simulations are similar in performance to those generated from screening the *P. carinii* and human DHFR models from our previous work,[19] identifying over 86% of true inhibitors and very few false positives. However, the models from the open-loop dynamics are less successful, with the models identifying more false positives. This is not unexpected as the M20 open-loop conformation does not represent a productive bound complex form.

X-ray Pharmacophore Model Comparison. The pharmacophore model resulting from the X-ray structures (Figure 6a) shares five elements with the MD-based models: the aromatic element at the bottom of the pocket, hydrogen-bond donors to Asp27 Oδ1 and Oδ2, the aromatic element midway in the active site, and the hydrogen-bond acceptor in the mouth of the binding pocket. In addition, there are two extra hydrogen-bond elements and one extra aromatic element. One hydrogen-bond element is a result of a second cluster of methanol probes interaction with Asp27 Oδ1, and the other reflects an interaction with the backbone carbonyl of Ile94. The additional aromatic element corresponds to the *p*-aminobenzoate linker region of the substrate interacting with Phe31.

The performance of pharmacophore models generated from X-ray crystallographic structures is given in Figure 6b. The model does identify general DHFR inhibitors over noninhibitors, but the species-specificity is lost, with a higher percentage of general DHFR inhibitors being identified over high-affinity, species-specific inhibitors. The overall performance is similar to models we previously derived for *C. albicans* DHFR,[19] which also showed little conformational flexibility over the collection of crystal structures. In both cases, although the rate of false positive identification is low, less than 50% of the high-affinity true inhibitors are identified. Again, we see that limited flexibility in the active site leads to modest performance. Even though conformational variability is seen in the M20 loop, it only contributes a small part of the binding site; all other regions of the binding pocket are nearly identical.

The optimal X-ray model (5 from 8, radii = 4.3 × rmsd) identified 42% of the high-affinity ecDHFR inhibitors and 65% of the less potent DHFR inhibitors; only 3% of the false positives were hit. Although this model is able to identify true positives over decoy molecules, it lacks the species-

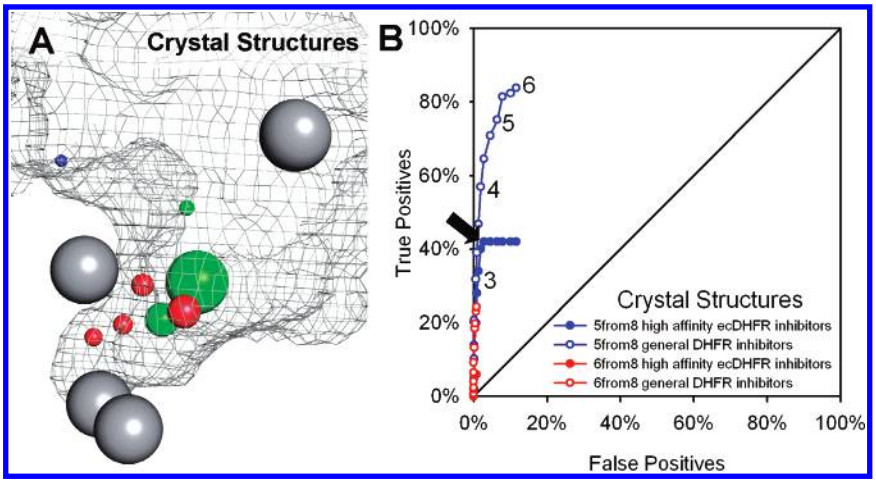

Figure 6. (**a**) Pharmacophore model resulting from an ensemble of ecDHFR crystal structures. Coloring is the same as in Figure 4. Elements are shown with radii = 2 × rmsd. (**b**) Associated ROC curves for the X-ray pharmacophore model. Series with filled data points are results from the screening of a high-affinity ecDHFR database; those with open data points are results from the screening of general DHFR inhibitors. Shown in red are results from models requiring 6 hits from an 8-site pharmacophore model, and shown in blue are results from models requiring 5 hits from an 8-site pharmacophore model. Points along each series represent an increase in pharmacophore element radii from 1 × rmsd to 6 × rmsd in increments of 1/3.

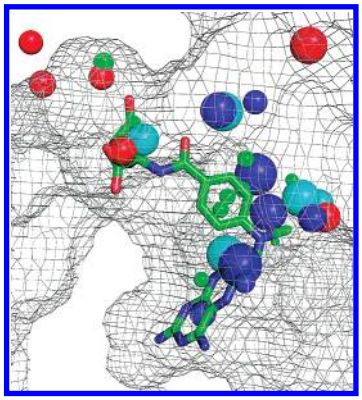

Figure 7. Static pharmacophore model for the equilibrated closed-loop structure (based on 1RX1). Coloring is the same as in Figure 6. The radius of an element is defined as the rmsd of the probe cluster that it represents, analogous to the rmsd of an MPS consensus cluster. Elements are shown with radii of 1 × rmsd. The bound conformation of methotrexate is superimposed on the model to orient the reader.

specificity seen in from the MD pharmacophore models. The low level of conformational variance in the active site may make the pharmacophore model resemble a static pharmacophore model. The number of extraneous sites and the small radii (low rmsds) of the elements is characteristic of models produced from a single rigid structure. This again highlights the importance of including sufficient protein flexibility in order to produce selective pharmacophore models for structure-based drug discovery.

To our knowledge, the models presented here are the first receptor-based pharmacophore models of *E. coli* DHFR. Joseph-McCarthy and Alvarez have made one based on a single crystal structure of *L. casei* DHFR with MCSSS2-PTS.[35,36] MCSS2PTS uses Multiple Copy Simultaneous Search (MCSS),[37] a technique similar to MUSIC, to generate pharmacophore points. For comparison, we have created a model based on a single, static structure taken at equilibration of the closed-loop simulation (Figure 7). In direct analogy to our MPS models, the radius of an element is defined as the rmsd of the probe cluster that it represents, rather than the rmsd of the consensus cluster. Although the two techniques are not exactly the same, we find that our model looks very similar to the MCSS2PTS model. In both our static model and the one proposed by Joseph-McCarthy and Alvarez, there are an overwhelming number of pharmacophore sites, only some of which overlap with the features of known ligands. Joseph-McCarthy and Alvarez highlighted the elements that overlap with known binding features of methotrexate. Our use of the MPS naturally highlights these same elements over the many extraneous sites. This is a clear demonstration of how the addition of protein flexibility helps us to identify the key features of the binding site.

CONCLUSION

From MD simulations of *E. coli* DHFR·NADPH, we have developed receptor-based pharmacophore models that take advantage of protein flexibility. Models resulting from ecDHFR with the closed conformation of the M20 loop were highly selective and identified relevant *E. coli* specific, high-affinity inhibitors over other general DHFR inhibitors. A marked improvement was seen in the open-loop dynamics with an increase in simulation length, indicating that including more flexibility can enhance the pharmacophore models. It is important to note that this inclusion of greater protein flexibility does not result in a loss of specificity, as the optimal models preferentially identify potent ecDHFR inhibitors over other general DHFR inhibitors. These results are particularly encouraging given that the DHFR·NADPH structure does not bias the folate-binding site toward a conformation associated with any particular bound ligand. We have also developed pharmacophore models based on ligand-bound crystal structures. While these models are very selective for DHFR inhibitors, they lack the preference for species-specific, high-affinity inhibitors seen in our models from an ensemble of MD snapshots. Again, this demonstrates the importance of flexibility. The current interest in ecDHFR dynamics in relation to catalysis can only emphasize the importance of including protein flexibility in the quest for identifying novel inhibitors.

ACKNOWLEDGMENT

This work has been supported by the National Institutes of Health (Grant GM65372) and the Beckham Young Investigator Program. M.G.L. would like to thank the University of Michigan's Molecular Biophysics Training Program for support (GM008270). For their generous donation of software we thank OpenEye for OMEGA and CCG for MOE. We also thank Allen Bailey for maintaining the computers used in this work.

Supporting Information Available: *E. coli* inhibitor data set used in this work and the coordinates and rmsd of the pharmacophore elements for all models. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Voet, D.; Voet, J. Amino acid metabolism. In *Biochemistry*, 2nd ed.; John Wiley & Sons, Inc.: New York, 1995; p 762.
(2) Schweitzer, B. I.; Dicker, A. P.; Bertino, J. R. Dihydrofolate Reductase as a Therapeutic Target. *FASEB J.* **1990**, *4*, 2441−2452.
(3) Parker, C. N. McMaster University Data-Mining and Docking Competition - Computational Models on the Catwalk. *J. Biomol. Screen.* **2005**, *10*, 647−648.
(4) Lang, P. T.; Kuntz, I. D.; Maggiora, G. M.; Bajorath, J. Evaluating the High-Throughput Screening Computations. *J. Biomol. Screen.* **2005**, *10*, 649−652.
(5) Radkiewicz, J. L.; Brooks, C. L., III Protein Dynamics in Enzymatic Catalysis: Exploration of Dihydrofolate Reductase. *J. Am. Chem. Soc.* **2000**, *122*, 225−231.
(6) Agarwal, P. K.; Billeter, S. R.; Rajagopalan, P. T. R.; Benkovic, S. J.; Hammes-Schiffer, S. Network of coupled promoting motions in enzyme catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2794−2799.
(7) Agarwal, P. K.; Billeter, S. R.; Hammes-Schiffer, S. Nuclear Quantum Effects and Enzyme Dynamics in Dihydrofolate Reductase Catalysis. *J. Phys. Chem. B* **2002**, *106*, 3283−3293.
(8) Feeney, J. NMR Studies of Ligand Binding to Dihydrofolate Reductase. *Angew. Chem. Int. Ed.* **2000**, *39*, 290−312.
(9) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. *Science* **2006**, *313*, 1638−1642.
(10) Osborne, M. J.; Schnell, J.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. Backbone Dynamics in Dihydrofolate Reductase Complexes: Role of Loop Flexibility in the Catalytic Mechanism. *Biochemistry* **2001**, *40*, 9846−9859.
(11) Sawaya, M. R.; Kraut, J. Loop and Subdomain Movements in the Mechanism of Escherichia Coli Dihydrofolate Reductase: Crystallographic Evidence. *Biochemistry* **1997**, *36*, 586−603.
(12) Hammes-Schiffer, S.; Benkovic, S. J. Relating Protein Motion to Catalysis. *Annu. Rev. Biochem.* **2006**, *75*, 519−541.
(13) Miller, G. P.; Benkovic, S. J. Stretching Exercises - Flexibility in Dihydrofolate Reductase Catalysis. *Chem. Biol.* **1998**, *5*, R105−R113.
(14) Schnell, J. R.; Dyson, H. J.; Wright, P. E. Structure, Dynamics, and Catalytic Function of Dihydrofolate Reductase. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 119−140.
(15) Carlson, H. A.; Masukawa, K. M.; McCammon, J. A. Method for Including the Dynamic Fluctuations of a Protein in Computer-Aided Drug Design. *J. Phys. Chem. A* **1999**, *103*, 10213−10219.
(16) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. Developing a Dynamic Pharmacophore Model for HIV-1 Integrase. *J. Med. Chem.* **2000**, *43*, 2100−2114.
(17) Meagher, K. L.; Carlson, H. A. Incorporating Protein Flexibility in Structure-Based Drug Discovery: Using HIV-1 Protease as a Test Case. *J. Am. Chem. Soc.* **2004**, *126*, 13276−13281.
(18) Meagher, K. L.; Lerner, M. G.; Carlson, H. A. Refining the Multiple Protein Structure Pharmacophore Method: Consistency Across Three Independent HIV-1 Protease Models. *J. Med. Chem.* **2006**, *49*, 3478−3484.
(19) Bowman, A. L.; Lerner, M. G.; Carlson, H. A. Protein Flexibility and Species Specificity in Structure-Based Drug Discovery: Dihydrofolate Reductase as a Test System. *J. Am. Chem. Soc.* **2007**, *129*, 3634−3640.
(20) Damm, K. L.; Carlson, H. A. Exploring Sources of Multiple Protein Conformations in Structure-Based Drug Design. *J. Am. Chem. Soc.* **2007**, *129*, 8225−8235.
(21) DeLano, W. L. *The PyMOL Molecular Graphics System*, *v0.99*; DeLano Scientific: Palo Alto, CA, 2002.
(22) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; Debolt, S.; Ferguson, D. M.; Seibel, G. L.; Kollman, P. A. AMBER, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(23) Holmberg, N.; Ryde, U.; Bulow, L. Redesign of the Coenzyme Specificity in L-Lactate Dehydrogenase from Bacillus Stearothermophilus using Site-Directed Mutagenesis and Media Engineering. *Protein Eng.* **1999**, *12*, 851−856.
(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(25) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 6*; University of California, San Francisco: San Francisco, CA, 1996.
(26) Summerfield, R. L.; Daigle, D. M.; Mayer, S.; Mallik, D.; Hughes, D. W.; Jackson, S. G.; Sulek, M.; Organ, M. G.; Brown, E. D.; Junop, M. S. A 2.13 Angstrom Structure of E-coli Dihydrofolate Reductase Bound to a Novel Competitive Inhibitor Reveals a New Binding Surface Involving the M20 Loop Region. *J. Med. Chem.* **2006**, *49*, 6977−6986.
(27) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(28) Jorgensen, W. L. *BOSS, Version 4.2*; Yale University: New Haven, CT, 2000.
(29) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(30) Damm, K. L.; Carlson, H. A. Gaussian-Weighted RMSD Superposition of Proteins: A Structural Comparison for Flexible Proteins and Predicted Protein Structures. *Biophys. J.* **2006**, *90*, 4558−4573.
(31) *MOE*, *v2005.06*; Chemical Computing Group: Montreal, Canada, 2005.
(32) *Comprehensive Medicinal Chemistry*; MDL Information Systems Inc.: San Leandro, CA, 2005.
(33) *OMEGA*, *Version 1.8.b1*; OpenEye Scientific Software, Inc.: Santa Fe, NM, 2004.
(34) Tsai, C. J.; Kumar, S.; Ma, B. Y.; Nussinov, R. Folding Funnels, Binding Funnels, and Protein Function. *Protein Sci.* **1999**, *8*, 1181−1190.
(35) Joseph-McCarthy, D.; Alvarez, J. C. Automated Generation of MCSS-Derived Pharmacophoric DOCK Site Points for Searching Multiconformation Databases. *Proteins: Struct., Funct., Bioinf.* **2003**, *51*, 189−202.
(36) Joseph-McCarthy, D.; Thomas, B. E.; Belmarsh, M.; Moustakas, D.; Alvarez, J. C. Pharmacophore-Based Molecular Docking to Account for Ligand Flexibility. *Proteins: Struct., Funct., Bioinf.* **2003**, *51*, 172−188.
(37) Miranker, A.; Karplus, M. Functionality Maps of Binding-Sites - a Multiple Copy Simultaneous Search Method. *Proteins: Struct., Funct., Bioinf.* **1991**, *11*, 29−34.

CI700167N