# Comparing Ligand Interactions with Multiple Receptors via Serial Docking

Miguel X. Fernandes, Visvaldas Kairys, and Michael K. Gilson*

Center for Advanced Research in Biotechnology, U. Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, Maryland 20850

Received June 16, 2004

Standard uses of ligand−receptor docking typically focus on the association of candidate ligands with a single targeted receptor, but actual applications increasingly require comparisons across multiple receptors. This study demonstrates that comparative docking to multiple receptors can help to select homology models for virtual compound screening and to discover ligands that bind to one set of receptors but not to another, potentially similar, set. A serial docking algorithm is furthermore described that reduces the computational costs of such calculations by testing compounds against a series of receptor structures and discarding a compound as soon as it fails to satisfy specified bind/no bind criteria for each receptor. The algorithm also realizes substantial efficiencies by taking advantage of the fact that a ligand typically binds in similar conformations to similar receptors. Thus, once detailed docking has been used to fit a ligand into the first of a series of similar receptors, much less extensive calculations can be used for the remaining structures.

## INTRODUCTION

Applications of ligand−receptor docking typically seek compounds that bind tightly to a single targeted receptor, but many actual applications require a ligand that will bind multiple targeted receptors, while not binding another set of "untargeted" receptors. The importance of selective targeting is epitomized by the selective cyclooxygenase (COX) inhibitors, which bind COX-2 in preference to COX-1 and thus provide antiinflammatory benefits while avoiding undesired gastric side-effects.[1] Conversely, the potential for broadly active inhibitors is supported by the existence of compounds that bind the various types of adrenergic receptors[2] and of a synthetic ligand that binds multiple enzymes of the folate pathway[3]—not to mention the undesirable broad specificity of the first generation COX inhibitors. Compounds have also been designed that inhibit the proteases of feline, human, and simian immunodeficiency viruses (FIV, HIV, and SIV) along with clinically relevant mutants of HIV protease.[4] These examples highlight the need, and the potential, for methods of discovering compounds that will bind one set of receptors while sparing another set.

Several existing docking programs address the problem of maximizing the affinity of ligands for multiple closely related targets. FlexE allows for local variations in protein structure, including point-mutations, and gains efficiency by limiting attention to structural variants that possess highly similar backbone traces.[5] The programs DOCK,[6] AutoDock,[7] and FLOG[8] also have been applied to multiple receptor conformations by docking candidate ligands into a single averaged representation of the targeted receptors. The potential for using docking to identify specific ligands has recently been considered in a thoughtful analysis of kinase inhibitors, using AutoDock.[9] However, to our knowledge, there are no docking procedures designed to support systematic selection of compounds that bind one set of targeted structures but not another. The present paper describes such a method and illustrates its utility and efficiency in several sample applications.

* Corresponding author phone: (301)738-6217; fax: (301)738-6255; e-mail: gilson@umbi.umd.edu.

## METHODS

**Serial Docking Algorithm.** The present method docks a single candidate ligand serially to a set of different receptor structures, where the results obtained for each structure determine whether to continue on to the next receptor or to reject the compound. The decision whether to advance the ligand to the next structure is based upon the computed binding energy, which must be stronger than some threshold for each targeted receptor or weaker than a second threshold for the nontargeted receptors. Figure 1 illustrates this procedure for a case in which one is seeking compounds that bind two receptor structures but not a third. The serial docking method also uses the fact that a compound will usually adopt a similar low-energy conformation when fitted into similar receptors. (Data in the Results section support the reliability of this concept.) It is therefore possible to reduce the extent of the conformational search for all but the first receptor of a given type. For example, if receptors 1 and 2 in Figure 1 are structurally similar, and docking into receptor 1 uses a 10 Å range for each translational variable, then docking into receptor 2 could limit the translational variables to a 5 Å cube centered on the most stable conformation found when docking to receptor 1. Since this secondary search region is reduced in size, fewer conformations are required to identify the global energy minimum for the compound in receptor 2. It should be noted that the search ranges for overall rotation of the compound are also reduced, in addition to the translational search ranges (see below). The flowchart in Figure 2 diagrams the use of this approach to dock a ligand to multiple similar receptors.

**Implementation of Serial Docking with Vdock.** Docking is carried out with the program Vdock,[10,11] which minimizes the energy of the ligand−receptor complex as a function of

10.1021/ci049803m CCC: $27.50 
Published on Web 10/09/2004

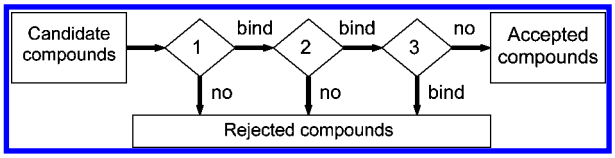


**Figure 1.** Diagram of the serial docking for three different receptor structures, labeled 1, 2, and 3. To be accepted, a compound must bind the first and second receptors, but not the third.

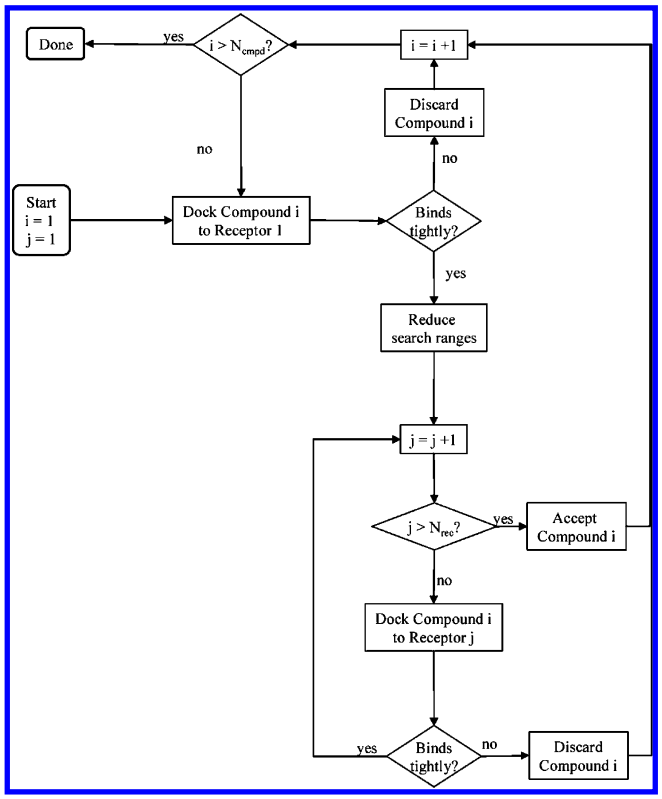


**Figure 2.** Illustrative flowchart of serial docking algorithm applied to $N_{cmpd}$ compounds for case in which compounds are sought that bind all $N_{rec}$ receptors. Whether a compound "binds tightly" is decided based upon a user-specified criterion applied to the pseudobinding energy (see Methods).

three overall rotation angles, three overall translation variables, and $N_{dih}$ torsion angles which can belong to the ligand and to receptor side-chains. The translational range is established by constraining a ligand atom to lie within a rectangular parallelepiped defined with respect to the binding site. The ligand atom used as the center of translation is chosen to be central to the molecule as follows. Initially, all atoms are candidates for the translational center, and each rotatable bond of the ligand that is not in a ring is considered to separate the molecule into two parts. For each such dihedral angle, the number of atoms belonging to each part of the molecule is counted, and the atoms on the side with the smaller number of atoms are disqualified as candidates for the translational center. Only a few candidate atoms remain after all dihedrals have been processed in this way. The translational center is set to the remaining candidate atom closest to the center of geometry of the molecule.

In the present implementation of Vdock, the energy is computed with the CHARMM22 force field,[12−14] where electrostatic screening by the solvent is modeled with the distance dependent dielectric model, $\epsilon_{ij} = 4r_{ij}$, and ligand-protein interactions are computed via 3 precalculated energy grids that account for electrostatic, attractive, and repulsive van der Waals interaction contributions.[15] As previously detailed,[10,11] low-energy conformations are sought by randomly generating new sets of the $N_{dih} + 6$ variables, while gradually narrowing the search around the lowest energy conformation found to date. Typically, $N_{trial} = 4$ narrowing-in procedures are carried out, and the most stable conformation found is advanced to a fine-tuning procedure, after which the docked conformation is reported. This search process is repeated until the desired number of docked conformations, $N_{dock}$, has been generated. To prevent rediscovery of previously docked conformations, successive searches are blocked from generating conformations within exclusion zones of predefined size around each docked conformation that has already been generated. The search is also accelerated by occasionally recombining trial variables that define the position and orientation of the ligand with variables drawn from previously docked conformations, in the manner of a genetic algorithm.[16,17] Normally about $N_{hunt} = 3000$ conformations are tested in each narrowing-in procedure, and another $N_{fine} = 3000$ conformations are tested during the fine-tune phase. It should be stressed that there is no correlation between any two successive conformations, so there is no definable starting conformation, but only a starting range for each variable. In a conventional application of Vdock, each translational variable has a starting range of about 10 Å, the ligand is allowed full overall rotation ranges, and each torsion angle is allowed to vary continuously between 0 and 360°. As discussed above and diagrammed in Figure 2, however, search ranges are reduced in serial docking calculations to multiple similar receptors. In the present study, 250 docked conformations were generated for each ligand in five independent runs of 50 dockings, each with different random number seeds. Except as otherwise noted, $N_{hunt} = 3000$ for the first similar receptor in a series, and 1000 for the subsequent receptors. The reported binding energy of a ligand is a pseudobinding energy computed based entirely upon its interactions with the three precomputed energy grids (see above). All calculations were carried out on a cluster of computers with dual 2000+ Athlon processors running Red Hat Linux 8.0 and connected with a 100 Mb ethernet TCP/IP network. The code is written mainly in FORTRAN 90 and was compiled under the Linux operating system with the pgf90 compiler from the Portland Group Inc.

**Molecular Structures and Parameters.** Crystallographic receptor structures were obtained from the Protein Data Bank (PDB)[18,19] for factor Xa (1g2l[20]), cyclin dependent kinase 2 (CDK2) (1h1p[21]), HIV-1 protease (1a9m[22], 1fb7[23], 1hvr[24], 1meu[25]), FIV protease (1fiv[26]), COX-1 (1pth[27]), and COX-2 (6cox[28]). Theoretical models of CDK 1 (1lc9) and CDK 4 (1ld2) also were downloaded from the Internet [Narayana, K.; Surendra, R. P. K.; Sarma, J. A. R. P. (2002) in http://www.biochem.ucl.ac.uk/bsm/pdbsum]. Homology models of factor Xa were constructed from the crystal structures of factor VIIa (1kli[29]), factor IXa (1rfn[30]), protein C (1aut[31]), and thrombin (1doj[32]). These proteins are all structurally related coagulation factors belonging to the chymotrypsin-like peptidase subfamily S1A.[33] Carboxypeptidase A is a member of the M14A peptidase subfamily,[33] and homology models of this protein were built from the crystal structures of other proteins in the same subfamily: procarboxypeptidase A (1jqg[34]), procarboxypeptidase A2 (1aye[35]), procarboxypeptidase B (1kwm[36]), and carboxypeptidase T (1obr[37]).

Homology models were built with the main-chain conformations of the respective templates after aligning sequences with FASTA,[38] and side-chain conformations were built with the program SCWRL.[39] This well-established method[40−42] uses a knowledge-based library of rotamers to identify the conformations of side-chains of residues that are not conserved in the template relative to the target sequence proteins. Regions of the model containing gaps or insertions of amino acids residues, relative to the template, were energy-minimized with the program Quanta 2000 [Accelrys, San Diego, CA, U.S.A.] to relieve strongly unfavorable interactions. During the minimizations, the model structures were kept fixed except for two residues bracketing each gap or insertion. CHARMM22 force field potential parameters were assigned using Quanta2000. All solvent molecules were removed from the receptors. Missing polar hydrogen atoms were added to the protein structures, and their positions were optimized using a gradient energy minimization method until the energy change between steps was lower than 0.01 kcal/mol. All hydrogen atoms were added to ligand molecules, and their positions were optimized using the same method as above.

**Virtual Screening of Compounds.** The serial docking method was tested for its ability to separate compounds known to bind a targeted protein from a background of decoy compounds. The active compounds for each target (Supporting Information) were chosen to represent a broad range of chemistries and thus provide a meaningful test set. The decoy compounds were a randomly picked subset of the NCI open database, a diverse[43] collection that has been successfully mined for inhibitors of varied therapeutic targets, including the serine protease matriptase,[44] HIV-1 protease,[45] anthrax lethal factor,[46] SARS-CoV protease,[47] *P. falciparum* dihydrofolate reductase,[48] and tyrosine phosphatase-1B.[49]

## RESULTS

This section first shows that serial docking can achieve a substantial speedup without loss of accuracy by using an abbreviated docking search for all but the first of a series of multiple similar receptors. It then documents the applicability and usefulness of comparative docking to multiple receptor structures in applications to compound screening with homology models and discovery of selective targeted ligands.

**Validity of Narrowing Search Ranges in Serial Docking.** As noted in Methods, it should be possible to accelerate docking of a compound to a receptor structure by narrowing the search around low-energy conformations that have already been identified for a similar receptor structure. To assess the validity of this approach, we compared the results of independent docking calculations to multiple receptors with results from calculations in which the searches were narrowed and abbreviated for all but the first receptor in a series. Agreement between the two calculations is then the measure of success.

The test system consists of variants of HIV-1 protease and some of its known ligands. This system is convenient because multiple structure and sequence variants are available in the PDB, and it is of interest because of the clinical significance of both the wild-type and mutant forms. Four protease structures with different, clinically relevant point mutations—G48H, G48V/L90M, and V82F/I84V—were selected, as were

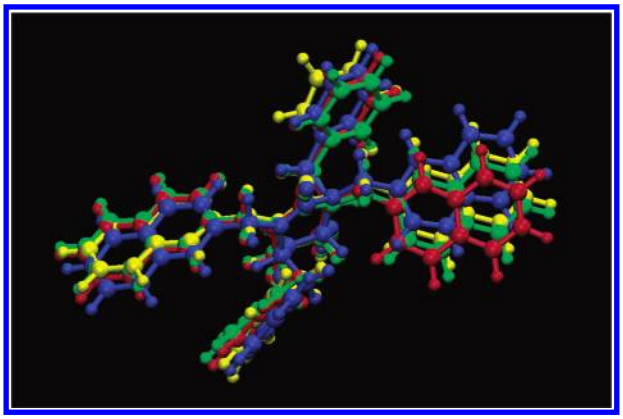

**Figure 3.** Structures of ligand XK263 docked independently to 4 different structures of HIV-protease.

**Table 1.** Observed Ranges of Overall Translational and Rotational Variables of Ligands Docked Independently to 4 Different Structures of HIV Protease (1a9m, 1fb7, 1hvr, 1meu)

| ligand | sb203386 | xk263 | dmp323 | bms182193 |
|---|---|---|---|---|
| rotational range (degrees) | 70 | 23 | 43 | 78 |
| translational range (Å) | 1.7 | 0.4 | 1.2 | 2.4 |

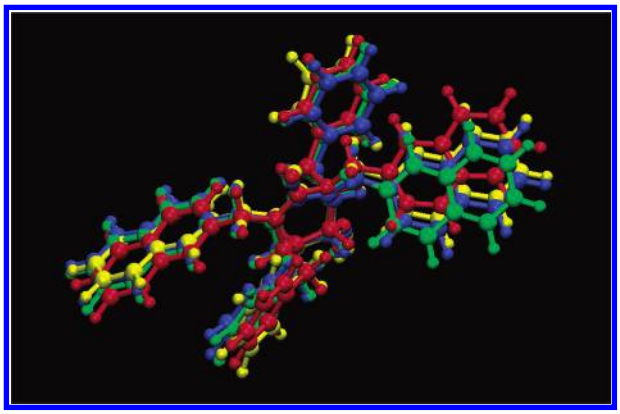

**Figure 4.** Structures of ligand XK263 docked serially to 4 different structures of HIV-protease.

4 ligands of differing size and chemical type (Figure 3). As shown in Table 1, when the ligands are docked independently rather than serially, the maximum observed variation in the rotation angles of the docked structures is about 80°, and that of the translational variables is about 2.5 Å. The variations of the dihedral angles are considerably larger, so they are not listed in Table 1, and the ranges of the dihedrals will not be reduced for the serial docking runs. Overall, the ligands adopt similar conformations in all 4 receptors, as illustrated in Figure 4 for one of the ligands. Because the computed binding energies of the ligands are of central importance in virtual screening, Table 2 presents the mean lowest energy of each ligand, averaged across the 4 receptors, for these independent docking runs, for $N_{hunt} = 1000$, 1500, 2000, and 3000; these data will be compared with those obtained by serial docking.

The same ligands were then serially docked into the same receptors. For the initial receptor, termed the "primary receptor", the search parameters are exactly the same as those used in the baseline tests described in the previous paragraph. However, for the remaining 3 receptors, termed "secondary receptors", the translational and rotational search ranges are reduced around the lowest energy conformation found for

**Table 2.** Average Binding Energies of Ligands Docked Independently to 4 Different Structures of HIV Protease (1a9m, 1fb7, 1hvr, 1meu), as a Function of the Number of Conformations Tried during the "hunt" Phase of Vdock ($N_{hunt}$)

| | ligand average binding energy (kcal/mol) of ligands | | | |
|---|---|---|---|---|
| $N_{hunt}$ | sb203386 | xk263 | dmp323 | bms182193 |
| 3000 | −53.3 | −67.9 | −65.6 | −46.4 |
| 2000 | −51.0 | −67.7 | −65.0 | −45.5 |
| 1500 | −50.7 | −67.3 | −64.7 | −45.0 |
| 1000 | −49.3 | −66.9 | −63.9 | −43.4 |

the primary receptor. The reduced sampling ranges are termed "secondary ranges". Table 3 summarizes the results in terms of the mean binding energy of each compound averaged over the 4 receptors, as a function of the secondary ranges and the value of $N_{hunt}$ used in the secondary dockings. These mean energies are suitable for comparison with those obtained by conventional docking, in Table 2. It should be noted that the serial docking results were found to be insensitive to which receptor is used as the primary.

Comparison of Tables 2 and 3 indicates that the secondary searches with reduced search ranges and $N_{hunt} = 1000$ can yield energies as low as those obtained with conventional docking, i.e., with full sampling ranges and $N_{hunt} = 3000$. See, for example, the results in Table 3 for translational and rotational ranges of 2.5 Å and 90 degrees, respectively. As expected, the conformations generated by this procedure also are similar to those obtained by conventional docking. For example, the serial docking results in Figure 4 may be compared with the conventional results in Figure 3. Thus, the serial docking procedure yields a 3-fold speedup for docking to the secondary receptors in this case. We conclude that serial docking can generate significant speedup by taking advantage of the fact that a given ligand will adopt similar conformations in similar receptors.

A potential concern with the present approach is that reducing the search range for the secondary dockings might make it difficult to generate a bound conformation significantly different from that in the primary docking. This issue is addressed here by examining the results of serial docking calculations using two receptors that are believed to bind the same ligand in significantly different conformations. Structural studies have shown that FIV protease has a smaller $S_3$ subsite than HIV protease[50] because it is formed by bulkier residues: I57, I98, and Q99 as opposed to G48, P81, and V82 in wild-type HIV-1 protease. As a consequence, FIV protease does not accommodate bulky ligand $P_3$ groups, and ligands with such groups must adopt relatively extended conformations. Bulkier $S_3$ side-chains also occur in some drug-resistant HIV protease mutants, such as G48V and V82F, and FIV has been considered a model for drug-resistant HIV protease.[4] As shown in Figure 5, docking compound A77003 (1hvi[51]) primarily to HIV protease and then secondarily to FIV protease yields a more extended conformation in FIV protease than HIV protease, as expected.[50] This illustrates that serial docking does allow discovery of significantly different bound conformations in the secondary and primary receptors, despite the narrowing of the search space in the secondary dockings. This result is traceable in part to the fact that the search ranges of dihedral angles are not reduced in secondary dockings; only translations and overall rotations are narrowed. Thus, the conformation of the ligand is not restricted at all; and it might, in fact, be possible to achieve even greater acceleration by imposing conformational restrictions.

**Comparative Docking to Multiple Homology Models.** When the 3D structure of a drug target has not been solved experimentally, it is often possible to build a structural model based upon a template, a similar protein whose structure is known. Although it would be natural to assume that the template with highest sequence identity to the target will yield the best results, this might not in fact be the optimal choice. Here, we test this assumption by building multiple homology models of Factor Xa and carboxypeptidase A and assessing the models for their ability to enrich the yield of known binders from among a database of compounds that are not expected to bind the targets. Serial docking lends itself well to this strategy because the candidate models are similar in structure.

Three-dimensional models of coagulation factor Xa were constructed based upon 4 different templates: coagulation factor VIIa, coagulation factor IXa, thrombin, and protein C. The root-mean-square deviations (RMSD) over all Cα atoms range between 1.8 and 2.3 Å. Comparison with the performance of various modeling techniques for a range of structural targets [http://eva.compbio.ucsf.edu/~eva/cm/res/accuracy.html] indicates that these models of Factor Xa are quite accurate for the levels of sequence identity (35−44%) of the templates. A virtual screen was set up by spiking 16 known binders of factor Xa into 10000 randomly chosen compounds of the National Cancer Institute database and serially docking all 10016 compounds into the 4 homology models. (The homology model built from factor IX was used as the primary receptor, but the results did not change materially when the structure used was built upon factor VII was used instead; data not shown.) For comparison, all compounds were also docked independently into the crystal structure of factor Xa. All 10016 compounds were then ranked according to their binding energies, and the percent of known actives in the top $x$% of compounds was graphed as a function of $x$ (Figure 6). Here a leftward shift of the curves, or a smaller average rank of the binders in the tables, indicates greater enrichment of actives in the top-scoring compounds; random selection of compounds would yield graphs lying along the diagonal. Table 4 lists a measure of enrichment, the mean percentage rank of the known binders, along with various measures of similarity of the templates with the target, Factor Xa.

The true crystal structure of factor Xa produces the greatest enrichment of known actives: all 16 known binders are ranked within the top 9% compounds. Surprisingly, however, the degree of enrichment proves to be uncorrelated with the sequence identity of the template used to build the homology model as well as with the other standard measures of sequence and structural similarity listed in Table 4. The reason appears to be that small structural details in the active site can significantly influence the docking results without strongly influencing the measures of similarity. For example, in the crystal structure of factor Xa, the side chains of Phe 177 and Tyr 100 line part of the binding pocket, whereas in the model based upon Factor VIIa, these side chains are crossed, effectively restricting access to this part of the binding site (see Figure 7).

**Table 3.** Average Binding Energies (kcal/mol) of Ligands Docked Serially to 4 Different Structures of HIV Protease (1a9m, 1fb7, 1hvr, 1meu) as a Function of the Number of Conformations Tried during the "hunt" Phase of Vdock ($N_{hunt}$) and of the Translational and Rotational Search Ranges Used in the Three Secondary Dockings[a]

| half ranges of translational and rotational variables | | $N_{hunt}$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2000 trials | | | | 1500 trials | | | | 1000 trials | | | |
| trans | rot | sb203386 | xk263 | dmp323 | bms182191 | sb203386 | xk263 | dmp323 | bms182191 | sb203386 | xk263 | dmp323 | bms182191 |
| 5 | 180 | −51.9 | −67.6 | −65.0 | −46.5 | −51.1 | −67.7 | −65.3 | −45.1 | −51.5 | −67.7 | −64.4 | −45.5 |
| 2.5 | 180 | −52.1 | −67.8 | −64.8 | −45.9 | −52.2 | −67.9 | −64.3 | −46.0 | −52.3 | −67.5 | −65.4 | −45.8 |
| 1.25 | 180 | −52.4 | −67.8 | −65.2 | −45.2 | −52.2 | −67.6 | −65.3 | −43.9 | −52.1 | −67.7 | −64.5 | −46.1 |
| 5 | 90 | −52.1 | −68.1 | −65.5 | −44.9 | −52.3 | −68.3 | −65.4 | −44.8 | −52.0 | −68.1 | −65.3 | −44.9 |
| 2.5 | 90 | −52.7 | −68.2 | −65.0 | −46.6 | −53.0 | −68.1 | −65.5 | −45.0 | −53.3 | −68.0 | −65.1 | −46.3 |
| 1.25 | 90 | −52.0 | −68.2 | −65.0 | −46.6 | −50.9 | −68.3 | −65.0 | −46.1 | −51.8 | −67.9 | −64.5 | −45.9 |
| 5. | 45 | −52.9 | −67.7 | −65.0 | −45.4 | −52.8 | −67.7 | −64.4 | −43.5 | −52.7 | −67.8 | −64.9 | −45.7 |
| 2.5 | 45 | −52.9 | −67.5 | −64.6 | −45.4 | −52.6 | −67.6 | −65.2 | −45.3 | −52.4 | −67.7 | −65.1 | −45.7 |
| 1.25 | 45 | −51.9 | −67.7 | −64.4 | −46.1 | −52.7 | −67.6 | −65.1 | −45.0 | −52.8 | −67.6 | −64.8 | −44.0 |

[a] The maximal standard deviations of these data vary between 0.8 kcal/mol for XK263 and 2.1 kcal/mol for BMS182191.

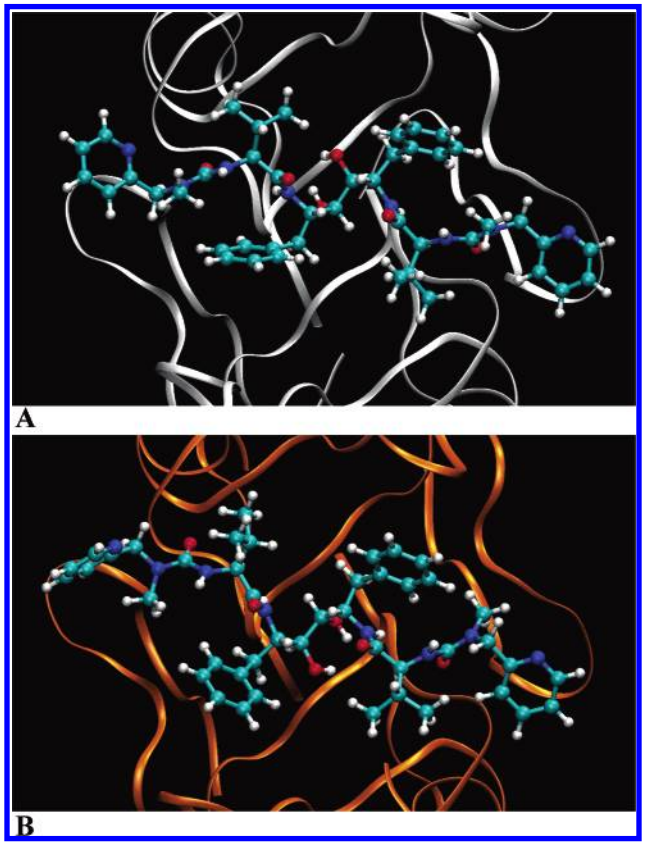


**Figure 5.** Structures of ligand A77003 docked serially to HIV protease (A) and FIV protease (B). Flap residues from both receptors are removed to make the ligand more visible.

Similarly, 3D models of carboxypeptidase A were constructed from 4 different templates: procarboxypeptidase A, procarboxypeptidase A2, procarboxypeptidase B, and carboxypeptidase T. The root-mean-square deviations (RMSD) over all Cα atoms range between 1.2 and 1.4 Å. Thus, these models also are quite accurate for the levels of sequence homology of the templates (21−67%). A virtual screen was set up by mixing 13 known binders of carboxypeptidase A into 10000 randomly chosen compounds of the National Cancer Institute database and serially docking all 10013 compounds into the 4 homology models; the homology model from procarboxypeptidase A2 was used as the primary receptor in the series. For comparison, all compounds were also docked independently into the crystal structure of

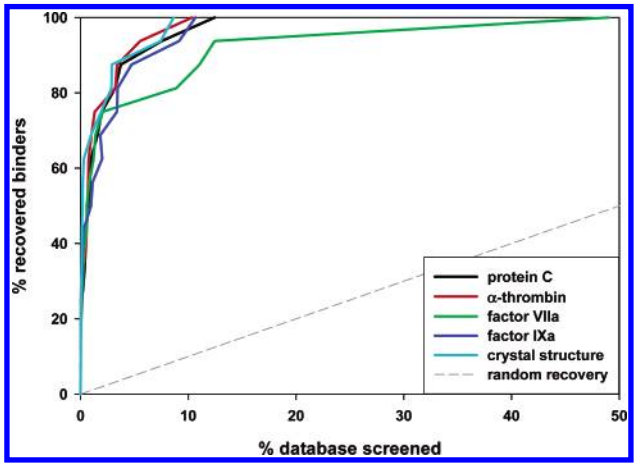


**Figure 6.** Recovery of 16 known binders mixed with 10000 NCI database compounds by independent docking to coagulation Factor Xa and serial docking to homology models of coagulation Factor Xa based upon various templates.

**Table 4.** Virtual Compound Screening Using Homology Models of Coagulation Factor Xa Based upon Four Different Templates of Known Structure[a]

| template | sequence identity (%) | E-score | RMSD (Å) | active site RMSD (Å) | av rank of binders (%) |
|---|---|---|---|---|---|
| crystal str. | 100 | | 0.0 | 0.0 | 1.6 |
| factor IXa | 43.8 | $1.4 \times 10^{-43}$ | 1.8 | 2.1 | 2.4 |
| α-thrombin | 40.7 | $3.4 \times 10^{-17}$ | 1.9 | 1.3 | 1.8 |
| factor VIIa | 38.8 | $9.0 \times 10^{-41}$ | 2.2 | 2.0 | 5.6 |
| protein C | 34.9 | $2.8 \times 10^{-35}$ | 2.3 | 1.0 | 2.2 |

[a] Sixteen known binders and 10000 compounds from the NCI database were serially docked into all four models and independently to crystal structure. Listed are the sequence identity and E-score (in the FASTA alignment) of each template, the overall root-mean-square structural deviation of $C_\alpha$ atoms (RMSD) for the aligned model relative to the crystal structure of Factor Xa, the RMSD of the residues (all atoms) that define the active site, and the average binding-energy rank of any of the 16 binders among the full 10000 compounds.

carboxypeptidase A. The results for these structures are listed in Table 5 and Figure 8. Once again the quality of the screening results correlates poorly with sequence identity and the other measures of similarity. In fact, the model built from template procarboxypeptidase A2, which has the highest sequence identity, gives the worst results. Again, this is a consequence of rather subtle details of the active site conformation. Whereas side chains in the active site of car-

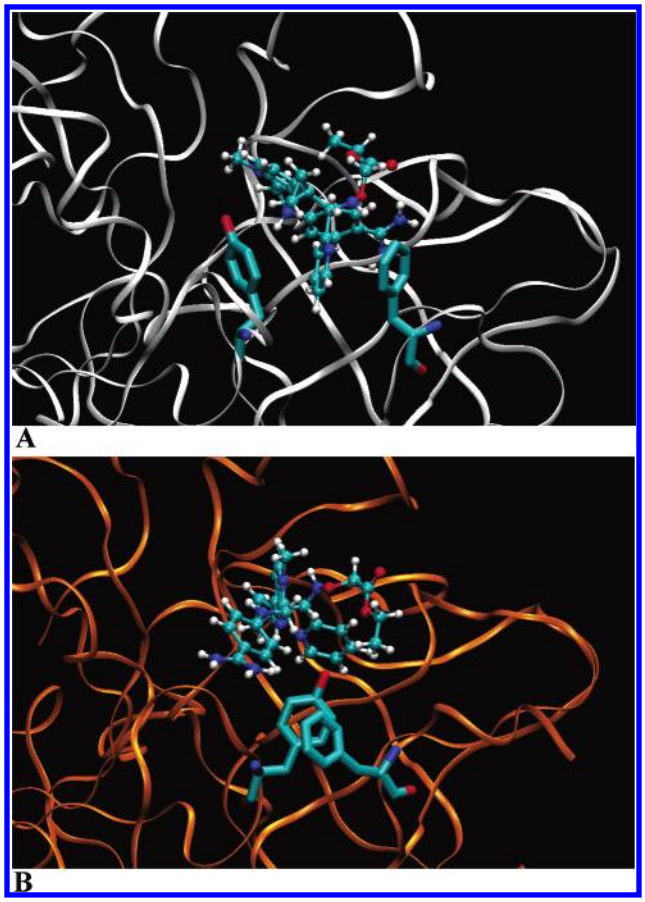


**Figure 7.** (A) Docked structure of compound 1g2m to crystal structure of factor Xa and (B) docked structure of compound 1g2m to model structure of factor Xa built upon template factor VIIa.

**Table 5.** Virtual Compound Screening Using Homology Models of Carboxypeptidase A Based upon Four Different Templates of Known Structure[a]

| template | sequence identity (%) | E-score | RMSD (Å) | active site RMSD (Å) | av rank of binders (%) |
|---|---|---|---|---|---|
| crystal str. | 100 | | 0.0 | 0.0 | 3.1 |
| procarboxypeptid. A2 | 66.6 | $2.9 \times 10^{-87}$ | 1.2 | 2.6 | 6.6 |
| procarboxypeptid. B | 46.5 | $2.1 \times 10^{-58}$ | 0.7 | 2.3 | 5.6 |
| carboxypeptidase T | 31.8 | $1.4 \times 10^{-16}$ | 0.6 | 2.6 | 3.3 |
| procarboxypeptid. A | 30.9 | $3.4 \times 10^{-28}$ | 1.4 | 2.4 | 3.7 |

[a] Thirteen known binders and 10000 compounds from the NCI database were serially docked into all four models and independently to crystal structure. Listed are the sequence identity and E-score (in the FASTA alignment) of each template, the overall root-mean-square structural deviation of $C_\alpha$ atoms (RMSD) for the aligned model relative to the crystal structure of carboxypeptidase A, the RMSD of the residues (all atoms) that define the active site, and the average binding-energy rank of any of the 13 binders among the full 10000 compounds.

boxypeptidase A converge to form multiple stabilizing interactions with the ligands, the active site of the model based on procarboxypeptidase A2 is more open and thus establishes fewer interactions with the ligands (see Figure 9).

The present results violate the expectation that the best homology model is the one built from the template having the greatest sequence identity. Thus, when multiple templates are available and confirmed ligands are known, it is preferable to build all possible homology models and select the one that performs best in a mock virtual screen, as done here, rather than simply using the model based upon the most

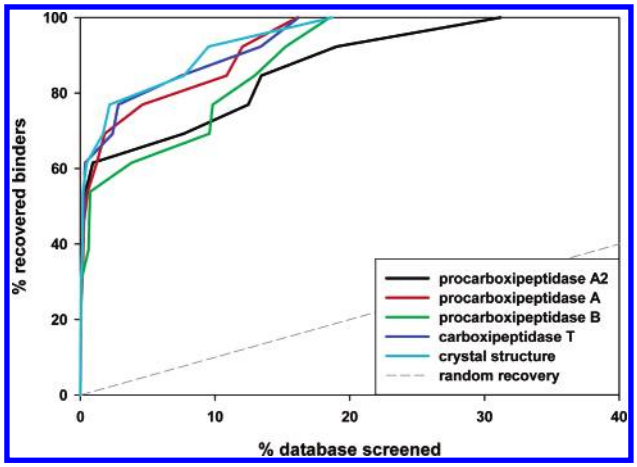


**Figure 8.** Recovery of 13 known binders mixed with 10000 NCI database compounds by independent docking to carboxypeptidase A and serial docking to homology models of carboxypeptidase A based upon various templates.

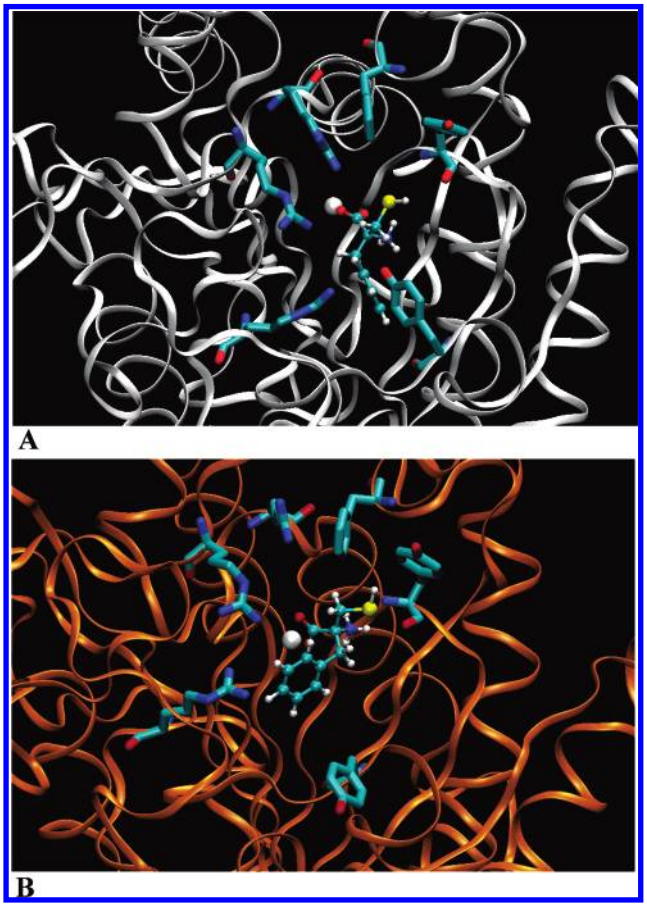


**Figure 9.** (A) Docked structure of compound 8cpa to crystal structure of carboxypeptidase A and (B) docked structure of compound 8cpa to model structure of carboxypeptidase A built upon template procarboxypeptidase A2.

similar template. Since this procedure can be time-consuming, it is worth noting that the serial docking algorithm can substantially speed this analysis. For example, it provides a 2-fold speedup in the examples shown here because only 1000 ligand conformations, rather than the usual 3000, were used in the hunt phase for the 3 secondary receptor structures.

**Screening Compounds for Target Selectivity.** Many medical applications require compounds that selectively bind one or a few targeted receptors, while sparing other, poten-

**Table 6.** Experimental Dissociation Constants ($K_d$) and Computed Pseudobinding Energies (kcal/mol) of Four Ligands for CDK1, CDK2, and CDK4[27]

| | recep | | | | | |
|---|---|---|---|---|---|---|
| | $K_d$ ($\mu$M) | | | binding energy | | |
| ligand | CDK1 | CDK2 | CDK4 | CDK1 | CDK2 | CDK4 |
| NU2058 | 5 | 12 | >100 | −34.8 | −36.0 | −27.2 |
| NU6086 | 2.2 | 2.3 | >100 | −43.4 | −45.9 | −38.2 |
| NU6094 | 1.6 | 1.0 | 16 | −42.1 | −45.4 | −36.8 |
| NU6102 | 0.009 | 0.006 | 1.6 | −45.8 | −45.5 | −38.3 |

tially similar, proteins. This section demonstrates the feasibility of incorporating multireceptor specificity into computational ligand screening, and the applicability of the serial docking algorithm for this purpose, via sample applications to the cyclin-dependent kinases and the cyclooxygenases.

The CDKs are a well-conserved family of enzymes involved in regulation of the cell cycle, whose activity is regulated by transcriptional and translational control, phosphorylation, binding of catalytic and regulatory subunits, and interactions with CDK inhibitory proteins.[52] Malfunction of these control mechanisms leads to abnormal CDK activities and hence abnormal cell cycling, and exogenous reduction in the activity of specific CDKs is of considerable interest as a basis for cancer chemotherapy. As an illustrative application of serial docking for specificity, four CDK inhibitors known to be more active against CDK1 and CDK2 than against CDK4[21] were serially docked to CDK1, CDK2, and CDK4. As shown in Table 6, although the calculated binding energies do not accurately reflect the affinities of the different ligands, the relative affinities of each ligand for the 3 receptors are consistent with their experimentally observed selectivity for CDK1 and CDK2 over CDK4. This result suggests that serial docking can be applied to improve the yield of compounds with desired selectivity across a family of related receptors and motivates the following more extensive illustration for the cyclooxygenases.

The enzymes COX-1 and COX-2 are both involved in the synthesis of prostanoids but produce different products and play different physiologic roles. COX-1 is produced constitutively in a number of tissues and is involved in baseline physiologic functions, including maintenance of the gastric mucosa and regulation of renal function. In contrast, COX-2, which is normally present at low levels, is produced during and contributes to the inflammatory process,[1] and many widely used antiinflammatory agents act by inhibiting COX-2. The earliest such compounds, aspirin and the first generation nonsteroidal antiinflammatory agents, also inhibit COX-1 and therefore can cause major side-effects, such as bleeding of the upper gastrointestinal tract. More recently, compounds such as celecoxib[53] and rofecoxib[54] have been developed that inhibit COX-2 while sparing COX-1 and that therefore treat inflammation with a lower risk of side-effects. Crystal structures of the cyclooxygenases are available, so this system offers a prototypical problem in structure-based design of specific enzyme inhibitors.

The applicability of serial docking to this problem was explored first by serially docking 20 nonselective inhibitors and 20 COX-2 selective inhibitors, to COX-2 (primary receptor) and COX-1 (secondary receptor). Encouragingly, all but two of the nonselective inhibitors yield negative (stabilizing) interactions with both enzymes, and all but two of the selective inhibitors yield positive (destabilizing) interactions with COX-1 and negative interactions with COX-2 (see Table 7). The computed selectivity is accounted for by the presence of Arg and Val in the active site of COX-2, in place of His and Ile in COX-1. Thus, COX-2 can accommodate ligand S558,[28] for example, but COX-1 cannot, due to a steric clash between the ligand and the Ile side chain. (See Figure 10.) Next, a mock compound screen was carried out by serially docking the 20 selective cyclooxygenase inhibitors along with the NCI diversity set (1844 molecules) to COX-2 (primary receptor) and COX-1 (secondary receptor). Of the 1864 molecules, 437 were found to have a positive interaction (destabilizing) with COX-2 and were therefore eliminated from consideration rather than being tested against COX-1. The remaining 1427 ligands were advanced to secondary docking to COX-1. Of these, 1026 had a negative binding energy with COX-1 and were therefore eliminated, leaving only 401 compounds which included 18 of the 20 selective inhibitors. As hoped, the known inhibitors are strongly enriched against the background of 401 based upon their scores against COX-2 (Figure 11), and it should be emphasized that a given percentage enrichment against a background of 400 compounds is far more helpful, resource-wise, than the same enrichment against a background of 1864. Indeed, the removal of over half of the compounds (1026 of 1864) based upon their favorable interaction with COX-1 is of considerable help in narrowing the search for compounds of interest. This example shows that imposing a simple specificity criterion can markedly reduce the number of compounds that deserve further consideration. Note, too, that we have used a rather loose criterion for binding to COX-2, binding energy < 0 kcal/mol. The number of accepted compounds would fall considerably if a more stringent criterion were applied, thus further increasing computational efficiency and decreasing the number of compounds needing further evaluation.

**Table 7.** Computed Pseudobinding Energies for Nonselective Inhibitors of COX-1 and COX-2 and for COX-2 Selective Inhibitors[a]

| | nonselective inhibitors binding energy (kcal/mol) | | COX-2 selective inhibitors binding energy (kcal/mol) | |
|---|---|---|---|---|
| ligand | COX-1 | COX-2 | COX-1 | COX-2 |
| 1 | −21.9 | −27.5 | 1.4 | −46.1 |
| 2 | −2.1 | −36.5 | 27.4 | −40.6 |
| 3 | 30.3 | −43.4 | 26.8 | −43.0 |
| 4 | −0.3 | −27.0 | 25.5 | −44.8 |
| 5 | −34.2 | −43.1 | −1.1 | −44.4 |
| 6 | −36.3 | −42.1 | 27.2 | −39.1 |
| 7 | −36.0 | −40.7 | −7.0 | −40.5 |
| 8 | −35.7 | −41.7 | 63.9 | −38.9 |
| 9 | −5.4 | −38.9 | 26.3 | −39.2 |
| 10 | −19.2 | −12.6 | 244.9 | −19.8 |
| 11 | −35.5 | −41.7 | 0.3 | −41.4 |
| 12 | −35.7 | −38.2 | 5.0 | −44.7 |
| 13 | −12.9 | −34.2 | 4.4 | −43.8 |
| 14 | −26.7 | −38.0 | 18.1 | −42.7 |
| 15 | −5.9 | −11.6 | 8.0 | −45.3 |
| 16 | −31.0 | −32.4 | 30.2 | −37.7 |
| 17 | −33.5 | −22.6 | 16.5 | −40.1 |
| 18 | −11.7 | −21.5 | 103.8 | −44.4 |
| 19 | −36.3 | 27.6 | 16.1 | −32.2 |
| 20 | −13.0 | −28.9 | 6.3 | −8.0 |

[a] Structures of all inhibitors used can be found in the Supporting Information.

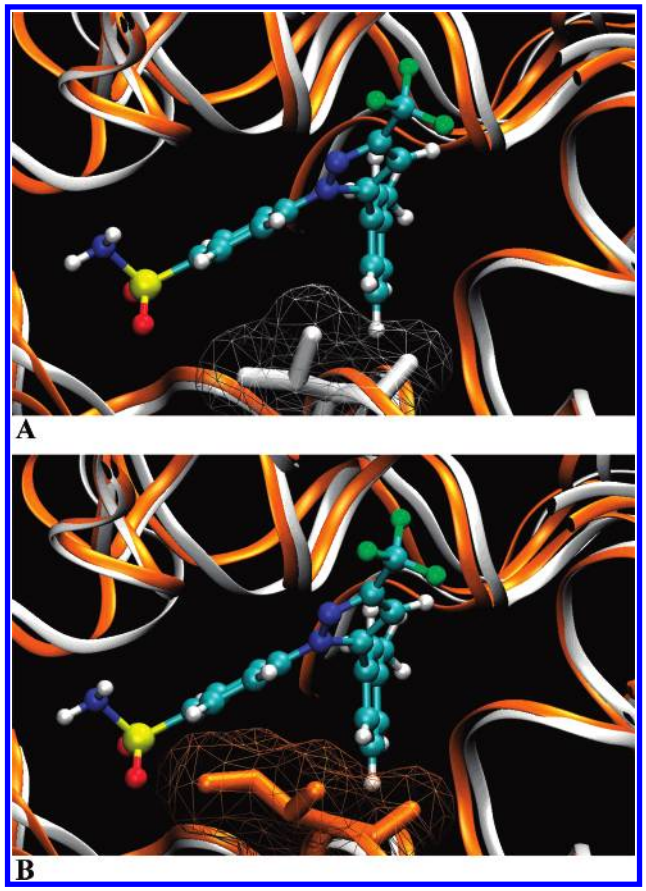


**Figure 10.** (A) Docked structure of ligand S558 along with crystal structures of COX-1 (orange) and COX-2 (white) with VAL523 highlighted and (B) ILE523 highlighted.

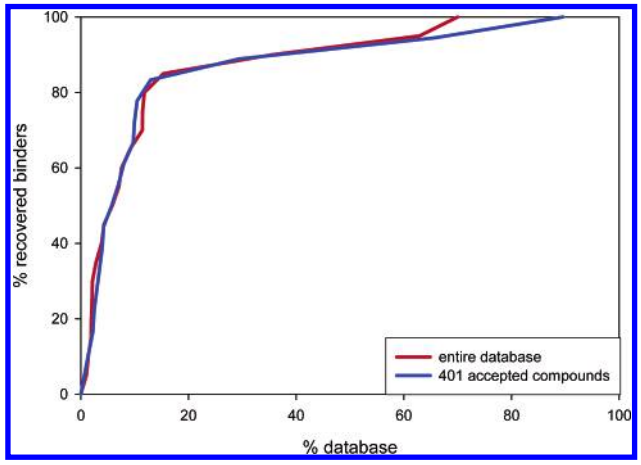


**Figure 11.** Recovery of 20 known COX-2 selective binders mixed with the NCI diversity set (red) and of the 18 known COX-2 selective inhibitors from among the 401 compounds classified as being COX-2 selective after performing a serial docking run (blue).

## DISCUSSION

This paper presents several novel results, the most general of which is that comparative docking of ligands against multiple receptors is an effective and useful strategy in a number of applications. Whereas previous publications have described the incorporation of multiple receptor structures in docking procedures, as discussed in the Introduction, they have typically sought not to compare results across receptors but instead to merge them. To facilitate comparative docking calculations, we have also introduced the serial docking algorithm. This procedure simplifies comparative docking by automating the procedure, and it accelerates the calculations by two methods. First, it saves time by immediately discarding a compound when it fails to satisfy a bind or no-bind criterion for a given receptor structure, rather than advancing it to the next receptor. Second, it reduces the number of energy evaluations required during docking by focusing the search zone for secondary receptors around the optimal docked conformation for the structurally similar primary receptor. Model calculations presented here indicate that this focusing procedure does not harm the results, but it should be noted that optimal ranges for secondary docking may vary from one system to another.

The present study also shows that the ability of a homology model to select known binders from a background of decoy compounds correlates poorly with the sequence identity of the template and target structures as well as with other standard measures of protein similarity. This observation, which runs contrary to naïve expectations, indicates that the docking results are sensitive to details of the active site conformation that are not well reflected in the similarity measures. The practical implication is that the sequence identity of the template structure should not be relied upon as a means of choosing which homology model to use in compound screening. Instead, much better results can be achieved by testing the available homology models in a mock virtual screen based upon known ligands (if available) and a background of presumably inactive decoy compounds, as done here. It should be noted that some very recent studies also have explored the use of homology models in compound screening.[55−58]

Finally, the present study shows that comparative docking across multiple receptors can be used to enrich the yield of selective ligands that bind one set of receptors but not another, as illustrated especially for the case of the cyclooxygenases. It is interesting that even a rather simple docking energy model suffices to distinguish quite reliably (90% success rate) between COX-2 selective inhibitors and nonselective COX inhibitors. The serial docking algorithm is particularly well suited for such applications, since it can efficiently select for compounds that *do* bind one set of receptors but *do not* bind another set. To our knowledge, this represents a simple and novel approach to the design of compounds with a desired receptor selectivity. It worth noting that the quality of the results for systems beyond the cyclooxygenases will, of course, be limited by the accuracy of the scoring function or energy model. However, given that virtual screening with existing methods is known to enrich the yield of binders across a range of drug targets, it is likely that the present method will similarly enrich the yield of compounds with a desired specificity.

## ACKNOWLEDGMENT

This publication was made possible by Grant Number GM61300 and Grant Number GM66524 from the National Institutes of Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

**Supporting Information Available:** Inhibitors of HIV-1 protease, CDK1, CDK2, and CDK4, Factor Xa, caraboxypep-

tidase A, and COX-2. This material is available free of charge via the Internet at http://pubs.acs.org.

## REFERENCES AND NOTES


(1) Vane, J. R.; Bakhle, Y. S.; Botting, R. M. Cyclooxygenases 1 and 2. *Annu. Rev. Pharmacol. Toxicol*. **1998**, *38*, 97−120.

(2) Klabunde, T.; Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **2002**, *3*, 928−944.

(3) Shih, C.; Chen, V. J.; Gossett, L. S.; Gates, S. B.; MacKellar, W. C.; Habeck, L. L.; Mendelsohn, K. A. S. L. G.; Soose, D. J.; Patel, V. F.; Andis, S. L.; Rayl, J. R. B. E. A.; Moroson, B. A.; Beardsley, G. P.; Kohler, W.; Ratnam, M.; Schultz, R. M. LY231514, a pyrrolo[2,3-d]pyrimidine-based antifolate that inhibits multiple folate-requiring enzymes. *Cancer Res.* **1997**, *57*, 1116−1123.

(4) Lee T.; Laco, G. S.; Torbett, B. E.; Fox, H. S.; Lerner, D. L.; Elder, J. H.; Wong, C. H. Analysis of the S3 and S3′ subsite specificities of feline immunodeficiency virus (FIV) protease: development of a broad-based protease inhibitor efficacious against FIV, SIV, and HIV in vitro and ex vivo. *Proc. Natl. Acad. Sci*. **1998**, *95*, 939−944.

(5) Claussen, H.; Bunning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol*. **2001**, *308*, 377−395.

(6) Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of proteins. *J. Mol. Biol*. **1997**, *266*, 424−440.

(7) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **2002**, *46*, 34−40.

(8) Broughton H. B. A method for including protein flexibility in protein−ligand docking: improving tools for database mining and virtual screening. *J. Mol. Graph. Model.* **2000**, *18*, 247−257, 302−304.

(9) Rockey, W. M.; Elcock, A. H. Progress toward virtual screening for drug side effects. *Proteins* **2002**, *48*, 664−671.

(10) David, L.; Luo, R.; Gilson, M. K. Ligand−receptor docking with the Mining Minima optimizer. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 157−171.

(11) Kairys, V.; Gilson, M. K. Enhanced docking with the mining minima optimizer: acceleration and side-chain flexibility. *J. Comput. Chem*. **2002**, *23*, 1656−1670.

(12) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem*. **1983**, *4*, 187−217.

(13) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of protein. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(14) Momany, F. A.; Rone, R. Validation of the general-purpose Quanta3.2/CHARMM force-field *J. Comput. Chem*. **1992**, *13*, 888−890.

(15) Pattabiraman, N.; Levitt, M.; Ferring, T. E.; Langridge, R. Computer graphics in real-time docking with energy calculation and minimization. *J. Comput. Chem*. **1985**, *6*, 432−436.

(16) Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Reading, MA, 1989.

(17) Holland, J. J. *Adaptation in Natural and Artificial Systems*; MIT Press: Cambridge, MA, 1975.

(18) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol*. **1977**, *112*, 535−542.

(19) 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.

(20) Nar, H.; Bauer, M.; Schmid. A.; Stassen, J. M.; Wienen, W.; Priepke, H. W.; Kauffmann, I. K.; Ries, U. J.; Hauel N. H. Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors. *Structure* **2001**, *9*, 29−37.

(21) Davies, T. G.; Bentley, J.; Arris, C. E.; Boyle, F. T.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Hardcastle, I. R.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E.; Tucker, J. A.; Wang, L.; Whitfield H. J. Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. *Nat. Struct Biol*. **2002**, *9*, 745−749.

(22) Hong, L.; Zhang, X. J.; Foundling, S.; Hartsuck, J. A.; Tang, J. Structure of a G48H mutant of HIV-1 protease explains how glycine-48 replacements produce mutants resistant to inhibitor drugs. *FEBS Lett*. **1997**, *420*, 11−16.

(23) Hong, L.; Zhang, X. J.; Hartsuck, J. A.; Tang, J. Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: insights into the mechanisms of drug resistance. *Protein Sci*. **2000**, *9*, 1898−1904.

(24) Lam, P. Y.; Jadhav, P.; Eyermann, C.; Hodge, C. N.; Ru, Y.; Bacheler, L.; Meek, J.; Otto, M.; Rayner, M.; Wong, Y.; Chang, C. H.; Weber, P.; Jackson, D.; Sharpe, T.; Erickson-Viitanen, S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, 380−384.

(25) Ala, P. J.; Huston, E. E.; Klabe, R. M.; McCabe, D. D.; Duke, J. L.; Rizzo, C. J.; Korant, B. D.; DeLoskey, R. J.; Lam, P. Y.; Hodge, C. N.; Chang C. H. Molecular basis of HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with cyclic urea inhibitors. *Biochemistry* **1997**, *36*, 1573−1580.

(26) Wlodawer, A.; Gustchina, A.; Reshetnikova, L.; Lubkowski, J.; Zdanov, J. A.; Hui, K. Y.; Angleton, E. L.; Farmerie, W. G.; Goodenow, M. M.; Bhatt, D.; Zhang, L.; Dunn, B. M. Structure of an inhibitor complex of the proteinase from feline immunodeficiency virus. *Nat. Struct. Biol*. **1995**, *2*, 480−488.

(27) Loll, P. J.; Picot, D.; Garavito, R. M. The structural basis of aspirin activity inferred from the crystal structure of inactivated prostaglandin H2 synthase. *Nat. Struct. Biol*. **1995**, *2*, 637−643.

(28) Kurumbail, R. G.; Stevens, A. M.; Gierse, J. K.; McDonald, J. J.; Stegeman, R. A.; Pak, J. Y.; Gildehaus, D.; Miyashiro, J. M.; Penning, T. D.; Seibert, K.; Isakson, P. C.; Stallings, W. C. Structural basis for selective inhibition of cyclooxygenase-2 by antiinflammatory agents. *Nature* **1996**, *384*, 644−648.

(29) Sichler, K.; Banner, D. W.; D'Arcy, A.; Hopfner, K. P.; Huber, R.; Bode, W.; Kresse, G. B.; Kopetzki, E.; Brandstetter, H. Crystal structures of uninhibited factor VIIa link its cofactor and substrate-assisted activation to specific interactions. *J. Mol. Biol*. **2002**, *322*, 591−603.

(30) Hopfner, K. P.; Lang, A.; Karcher, A.; Sichler, K.; Kopetzki, E.; Brandstetter, H.; Huber, R.; Bode, W.; Engh, R. A. Coagulation factor IXa: the relaxed conformation of Tyr99 blocks substrate binding. *Struct. Fold Des.* **1999**, *7*, 989−996.

(31) Mather, T.; Oganessyan, V.; Hof, P.; Huber, R.; Foundling, S.; Esmon, C.; Bode W. The 2.8 A crystal structure of Gla-domainless activated protein C. *EMBO J.* **1996**, *15*, 6822−6831.

(32) Recacha, R.; Costanzo, M. J.; Maryanoff, B. E.; Carson, M.; DeLucas, L.; Chattopadhyay. D. Structure of human alpha-thrombin complexed with RWJ-51438 at 1.7 A: unusual perturbation of the 60A-60I insertion loop. *Acta Crystallogr. D Biol. Crystallogr*. **2000**, *56*, 1395−400.

(33) Rawlings, N. D.; Tolle, D. P.; Barrett, A. J. MEROPS: the peptidase database. *Nucleic Acids Res.* **2004**, *32 Database issue*: D160−164.

(34) Estebanez-Perpina, E.; Bayes, A.; Vendrell, J.; Jongsma, M. A.; Bown, D. P.; Gatehouse, J. A.; Huber, R.; Bode, W.; Aviles, F. X.; Reverter, D. Crystal structure of a novel mid-gut procarboxypeptidase from the cotton pest *Helicoverpa armigera*. *J. Mol. Biol*. **2001**, *313*, 629−638.

(35) Garcia-Saez, I.; Reverter, D.; Vendrell, J.; Aviles, F. X.; Coll, M. The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. *EMBO J.* **1997**, *16*, 6906−6913.

(36) Barbosa-Pereira, P. J.; Segura-Martin, S.; Oliva, B.; Ferrer-Orta, C.; Aviles, F. X.; Coll, M.; Gomis-Ruth, F. X.; Vendrell, J. Human procarboxypeptidase B: three-dimensional structure and implications for thrombin-activatable fibrinolysis inhibitor (TAFI). *J. Mol. Biol*. **2002**, *321*, 537−547.

(37) Teplyakov, A.; Polyakov, K.; Obmolova, G.; Strokopytov, B.; Kuranova, I.; Osterman, A.; Grishin, N.; Smulevitch, S.; Zagnitko, O.; Galperina, O.; et al. Crystal structure of carboxypeptidase T from Thermoactinomyces vulgaris. *Eur. J. Biochem*. **1992**, *208*, 281−238.

(38) Pearson, W. R. Rapid and Sensitive Sequence Comparison with FASTP and FASTA. *Methods Enzymol*. **1990**, *183*, 63−98.

(39) Bower, M. J.; Cohen, F. E.; Dunbrack, R. L., Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol*. **1997**, *267*, 1268−1282.

(40) Zhao, S.; Goodsell, D. S.; Olson, A. J. Analysis of a data set of paired uncomplexed protein structures: new metrics for side-chain flexibility and model evaluation. *Proteins* **2001**, *43*, 271−279.

(41) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L., Jr. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci*. **2003**, *12*, 2001−2014.

(42) Mendes, J.; Nagarajaram, H. A.; Soares, C. M.; Blundell, T. L.; Carrondo, M. A. Incorporating knowledge-based biases into an energy-based side-chain modeling method: application to comparative modeling of protein structure. *Biopolymers* **2001**, *59*, 72−86.

(43) Milne, G. W.; Nicklaus, M. C.; Driscoll, J. S.; Wang, S.; Zaharevitz, D. National Cancer Institute drug information system 3D database *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 1219−1224.

(44) Enyedy, I. J.; Lee, S. L.; Kuo, A. H.; Dickson, R. B.; Lin, C. Y.; Wang, S. Structure-Based Approach for the Discovery of Bis-

benzamidines as Novel Inhibitors of Matriptase. *J. Med. Chem*. **2001**, *44*. 1349−1355.
(45) Wang, S.; Milne, G. W.; Yan, X.; Posey, I. J.; Nicklaus, M. C.; Graham, L.; Rice, W. G. Discovery of novel, non-peptide HIV-1 protease inhibitors by pharmacophore searching. *J. Med. Chem*. **1996**, *39*, 2047−2054.
(46) Panchal, R. G.; Hermone, A. R.; Nguyen, T. L.; Wong, T. Y.; Schwarzenbacher, R.; Schmidt, J.; Lane, D.; McGrath, C.; Turk, B. E.; Burnett, J.; Aman, M. J.; Little, S.; Sausville, E. A.; Zaharevitz, D. W.; Cantley, L, C.; Liddington, R. C.; Gussio, R.; Bavari, S. Identification of small molecule inhibitors of anthrax lethal factor. *Nat. Struct. Mol. Biol*. **2004**, *11*, 67−72.
(47) Sirois, S.; Wei, D. Q.; Du, Q.; Chou, K. C. Virtual screening for SARS-CoV protease based on KZ7088 pharmacophore points. *J. Chem. Inf. Comput. Sci*. **2004**, *44*, 1111−1122.
(48) Rastelli, G.; Pacchioni, S.; Sirawaraporn, W.; Sirawaraporn, R.; Parenti, M. D.; Ferrari, A. M. Docking and database screening reveal new classes of Plasmodium falciparum dihydrofolate reductase inhibitors. *J. Med. Chem*. **2003**, *46*, 2834−2845.
(49) Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem*. **2002**, *45*, 2213−2221.
(50) Lee, T.; Le, V. D.; Lim, D.; Lin, Y. C.; Morris, G. M.; Wong, A. L.; Olson, A. J.; Elder, J. H.; Wong, C. H. Development of a new type of protease inhibitors efficacious against FIV and HIV variants. *J. Am. Chem. Soc*. **1999**, *121*, 1145−1155.
(51) Hosur, M. V.; Bhat, T. N.; Kempf, D. J.; Baldwin, E. T.; Liu, B. S.; Gulnik, S.; Wideburg, N. E.; Norbeck, D. W.; Appelt, K.; Erickson, J. W. Influence of stereochemistry on activity and binding modes for C(2) symmetry-based diol inhibitors of HIV-1 protease. *J. Am. Chem. Soc*. **1994**, *116*, 847−855.
(52) Vermeulen, K.; van Bockstaele, D. R.; Berneman, Z. N. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Prolif*. **2003**, *36*, 131−149.
(53) Penning, T. D.; Talley, J. J.; Bertenshaw, S. R.; Carter, J. S.; Collins, P. W.; Docter, S.; Graneto, M. J.; Lee, L. F.; Malecha, J. W.; Miyashiro, J. M.; Rogers, R. S.; Rogier, D. J.; Yu, S. S.; Anderson G. D.; Burton, E. G.; Cogburn, J. N.; Gregory, S. A.; Koboldt, C. M.; Perkins, W. E.; Seibert, K.; Veenhuizen, A. W.; Zhang. Y. Y.; Isakson, P. C. Synthesis and biological evaluation of the 1,5-diarylpyrazole class of cyclooxygenase-2 inhibitors: identification of 4-[5-(4-methylphenyl)-3-(trifluoromethyl)-1H-pyrazol-1-yl]benze nesulfonamide (SC-58635, celecoxib). *J. Med. Chem*. **1997**, *40*, 1347−1365.
(54) Chan, C. C.; Boyce, S.; Brideau, C.; Charleson, S.; Cromlish, W.; Ethier, D.; Evans, J.; Ford-Hutchinson, A. W.; Forrest, M. J.; Gauthier, J. Y.; Gordon, R.; Gresser, M.; Guay, J.; Kargman, S.; Kennedy, B.; Leblanc, Y.; Leger, S.; Mancini, J.; O'Neill, G. P.; Ouellet, M.; Patrick, D.; Percival, M. D.; Perrier, H.; Prasit, P.; Rodger, I.; Tagari, P.; Therien, M.; Vickers, P.; Visco, D.; Wang, Z.; Webb, J.; Wong, E.; Xu, L. J.; Young, R. N.; Zamboni, R.; Riendeau, D. Rofecoxib [Vioxx, MK-0966; 4-(4′-methylsulfonylphenyl)-3-phenyl-2-(5H)-furanone]: a potent and orally active cyclooxygenase-2 inhibitor. Pharmacological and biochemical profiles. *J. Pharmacol. Exp. Ther*. **1999**, *290*, 551−560.
(55) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? *Proteins: Struct., Funct., Genet*. **2003**, *50*, 5−25.
(56) Johnson, M. A.; Hoog, C.; Pinto, B. M. A novel modeling protocol for protein receptors guided by bound-ligand conformation. *Biochemistry* **2003**, *42*, 1842−1853.
(57) Evers, A.; Gohlke, H.; Klebe, G. Ligand-supported homology modeling of protein binding-sites using knowledge-based potentials. *J. Mol. Biol*. **2003**, *334*, 327−345.
(58) Oshiro, C.; Bradley, E. K.; Eksterowicz, J.; Evensen, E.; Lamb, M. L.; Lanctot, J. K.; Putta, S.; Stanton, R.; Grootenhuis, P. D. Performance of 3D-database molecular docking studies into homology models. *J. Med. Chem*. **2004**, *47*, 764−767.

CI049803M