PROTEINS: Structure, Function, and Genetics 51:172–188 (2003)

Pharmacophore-Based Molecular Docking to Account for Ligand Flexibility

Diane Joseph-McCarthy,* Bert E. Thomas IV, Michael Belmarsh, Demetri Moustakas, and Juan C. Alvarez
Wyeth Research, Biological Chemistry Department, Cambridge, Massachusetts

ABSTRACT **Rapid computational mining of large 3D molecular databases is central to generating new drug leads. Accurate virtual screening of large 3D molecular databases requires consideration of the conformational flexibility of the ligand molecules. Ligand flexibility can be included without prohibitively increasing the search time by docking ensembles of precomputed conformers from a conformationally expanded database. A pharmacophore-based docking method whereby conformers of the same or different molecules are overlaid by their largest 3D pharmacophore and simultaneously docked by partial matches to that pharmacophore is presented. The method is implemented in DOCK 4.0. Proteins 2003;51:172–188.**

Key words: virtual screening; database searching; PhDOCK; pharmacophoric ensembles

INTRODUCTION

The rapid growth of small-molecule databases available for virtual screening, both commercial and proprietary, necessitates that molecular docking algorithms should be able minimally to handle hundreds of thousands of molecules. Thus, an average upper limit of seconds, rather than minutes, for evaluation on a per-molecule basis is desired. Further, it is necessary to account for ligand flexibility to screen a molecular database accurately. Because the conformation that a small molecule adopts when bound to a given target is not known a priori, allowing for ligand flexibility is crucial to being able to predict its binding to the target.

Docking methods that incorporate ligand flexibility fall into two basic categories: (1) those in which the ligand molecule is either incrementally built or flexed during the search and (2) those in which rigid precomputed conformers from a database are oriented in the target binding site.[1] The first category includes methods that employ Monte Carlo sampling,[2] genetic and evolutionary algorithms,[3–5] simulated annealing,[6] and incremental construction.[7–9] The second group includes rigid docking of flexibases, in which individual conformers are separately docked and scored and only the best-scoring conformer of each molecule is saved,[10] and rigid docking of conformational ensembles[1,11] generated by overlaying related conformers. Methods that treat ligand flexibility on the fly, as the docking proceeds, suffer from a redundant sampling of torsions, that is, conformers must be generated at every anchor position. Incremental construction methods in addition suffer from sensitivity to initial anchor positioning and greedy algorithms that may make incorrect choices for torsional positions during the docking procedure. After the initial anchor placement, greedy algorithms require that each subsequent torsion placement be optimal, which may not result in the most optimal overall molecule position. Because exhaustive searching using incremental construction is not feasible due to time constraints, the final result is a loss in conformational and orientational sampling, in particular with molecules containing multiple rotatable bonds. In contrast, when docking rigid precomputed conformers the cost for generating the conformers is incurred once, during the set-up of the database, thus eliminating the redundancy in torsional sampling. Due to the linear increase in search time with the number of conformers, however, the docking of individual conformers from a flexibase can be prohibitive.[12] Simultaneously, docking ensembles of conformers instead of individual conformers can dramatically speed the search time, allowing for the screening of databases of several hundred thousands of compounds. In one approach that has been described, conformers are overlaid on their common largest rigid fragment, which is the only part of the molecule used in the initial placement of the ensemble into the target site.[11] This method circumvents issues with greedy algorithms because all the conformers are scored; however, it still suffers from sensitivity to anchor (rigid fragment) placement because only a small contiguous subset of the atoms of the molecule are used during the initial docking step. The same method has recently been extended to conformers of different molecules sharing identical anchors.[13]

PhDOCK is a novel pharmacophore-based molecular docking approach described herein. In a PhDOCK database, conformers of the same and different molecules are overlaid by their largest 3D pharmacophore. During the docking, the pharmacophore points (or a subset of them) are matched to predefined DOCK site points in the binding region of the target structure to orient the ensemble. The use of ensembles allows for a large sampling of conformer space with a minimal number of docking events. In addition, the reduction of the query from a molecule to a

*Correspondence to: Diane Joseph-McCarthy, Biological Chemistry Department, Wyeth Research, 87 Cambridge Park Drive, Cambridge, MA 02140. E-mail: djoseph@wyeth.com

Received 7 December 2001; Accepted 2 August 2002

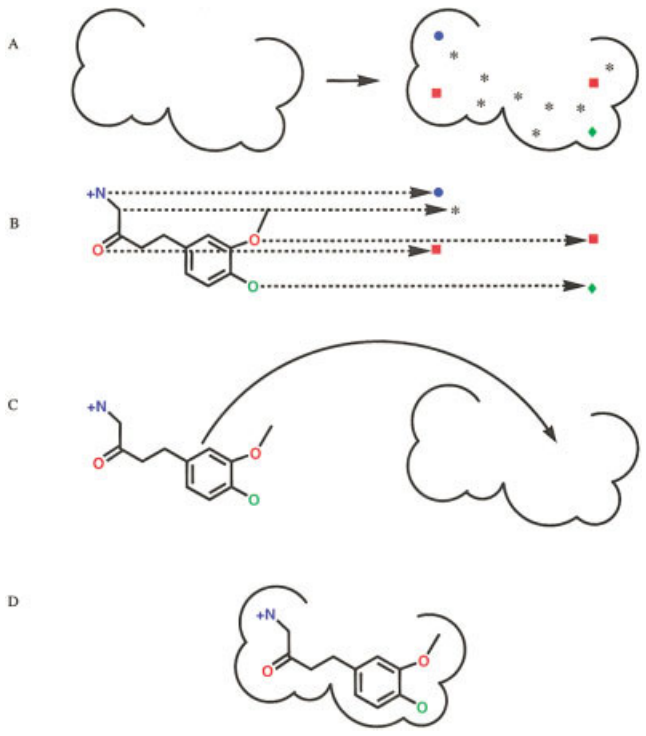


Fig. 1. Schematic of DOCK methodology. In the first step (**A**), the target binding site is filled with site points that may be colored. Then (**B**), distances between pairs of atoms in a database molecule are matched to distances between pairs of site points. A transformation matrix is calculated for the orientation (**C**), the molecule is docked into the binding site (**D**), and the fit of that conformer of that molecule is scored.

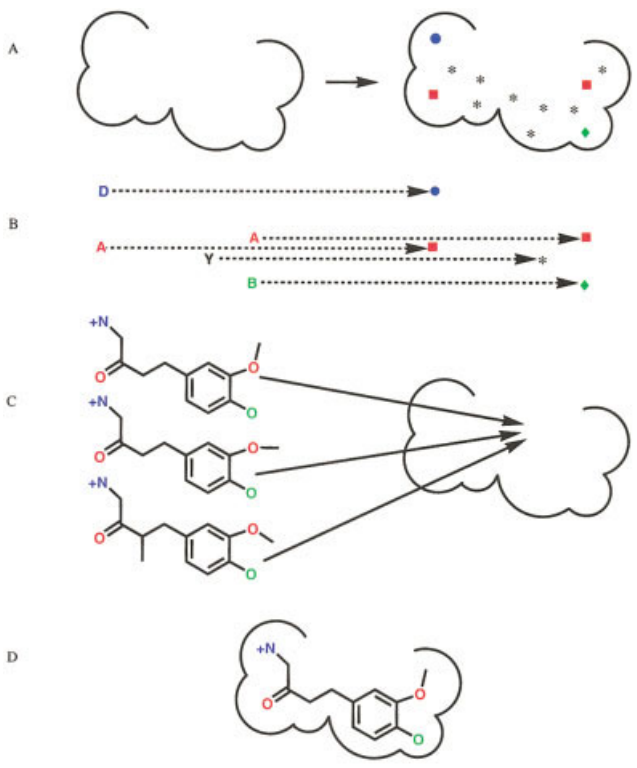


Fig. 2. Schematic of PhDOCK methodology.

pharmacophore reduces the combinatorial matching required to determine all allowed orientations. Thus, the speed-up over standard, multi-conformation DOCK is due to the fact that the orientations are only determined once for each ensemble (and not for the individual conformers) and that many fewer ligand points are used for the matching. Once docked, the interaction of each individual conformer with the target molecule is scored. PhDOCK is implemented in DOCK 4.0.1.[14]

Examples that validate the PhDOCK approach, as well as the use of MCSS2SPTS[15] in conjunction with PhDOCK, are presented. In particular, we show that the method is effective in reproducing ligand–protein complex crystal structures. Further, the use of PhDOCK to screen a subset of the Available Chemicals Directory (ACD) database (a database of more than 110,000 compounds) (MDL Information Systems Inc., 1997) for DHFR is discussed. Comparisons of pharmacophore-based ensemble docking using PhDOCK with rigid docking of the individual database conformers using DOCK 4.0 are given.

METHODS

DOCK Methodology

DOCK is one of the more widely used computational docking programs.[16–18] There are two major steps to the docking process; orienting the ligand into the target site and scoring the resulting complex. In the orienting step, the DOCK program systematically attempts to fit each compound or conformer from a database into the binding site of the target structure such that three or more of the atoms in the database molecule overlap with a set of predefined site points (or a clique) in the target binding site (Fig. 1). The orienting of the ligand can be further guided (or restricted) through the use of "coloring" or chemically labeling. If, for example, it is known from experimental information of bound ligands or by examination of the target structure that a particular site point should only be occupied by a ligand donor atom, that can be specified.

Each acceptable orientation of a ligand in the binding site is then scored on a grid throughout the macromolecular target using precalculated values for the protein component of the interaction energy. A number of different scoring functions can be employed, including the Amber[19] molecular mechanics force field, a contact scoring function, and a chemical score. In customized versions of DOCK, other force fields such as CHARMM[20,21] have been employed. Further, solvation correction factors that account for the desolvation energy of the ligand can be added to the score.[22]

Docking Pharmacophoric Ensembles

In the related PhDOCK (Fig. 2), conformers of the same and different molecules are overlaid based on their largest 3D pharmacophore. Pharmacophore points include hydrogen bond acceptors, hydrogen bond donors, dual atoms that can simultaneously donate and accept hydrogen

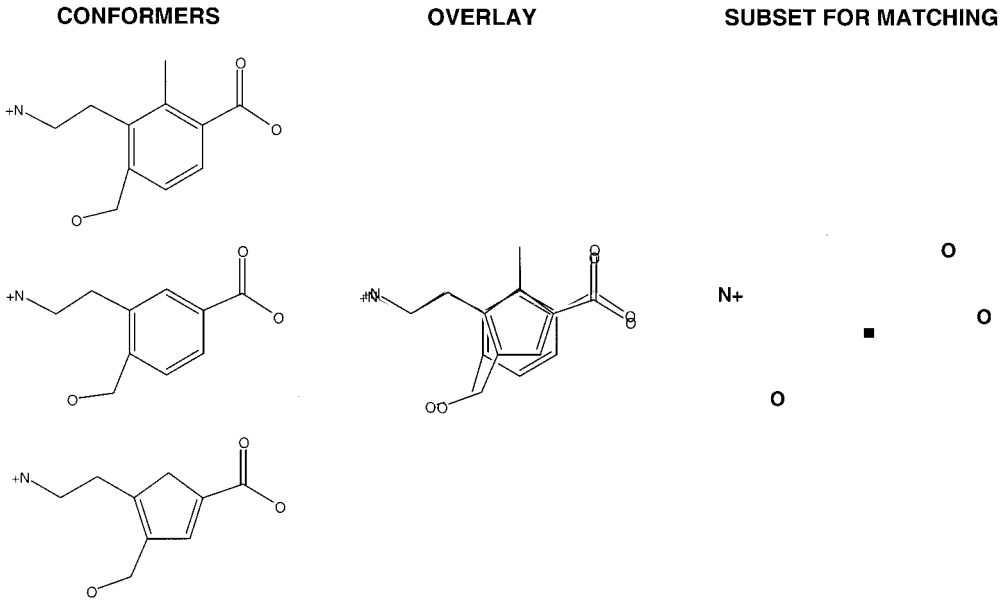


Fig. 3. Schematic example of a PhDOCK database overlay.

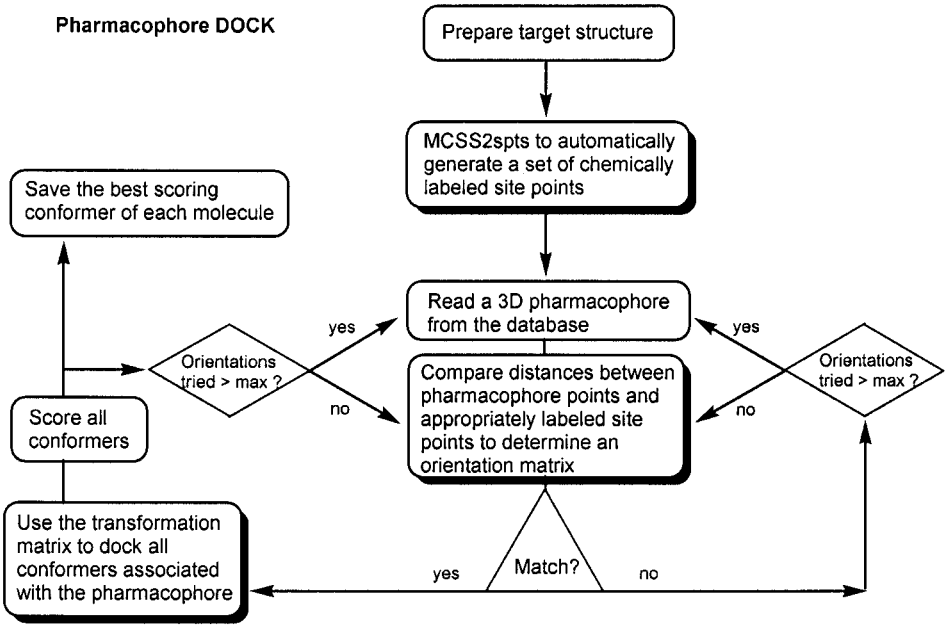


Fig. 4. Flowchart of the PhDOCK algorithm.

bonds, and ring centroids. Only the pharmacophoric points of the ensemble are used to determine the "matching" orientations. For each acceptable orientation, the ensemble of conformers is simultaneously docked into the binding site, and all members of the ensemble are individually scored. All atoms of a given conformer contribute to its score. In the simplified example given in Figure 3, the largest 3D pharmacophore of the set of molecules would have five pharmacophore points. A PhDOCK search could specify, for example, that all combinations of four of these pharmacophore points be matched to all combinations of four of the predefined site points in the binding site, that is, partial matches of the largest ensemble pharmacophore with the site points occur during the docking.

PhDOCK Algorithm

The procedure for carrying out a PhDOCK search (Fig. 4) is largely analogous to that for DOCK 4.0.[14] Briefly, the target structure is prepared and the site points used to guide the docking are generated. Then, a 3D pharmacophore is read from the database (the database format is

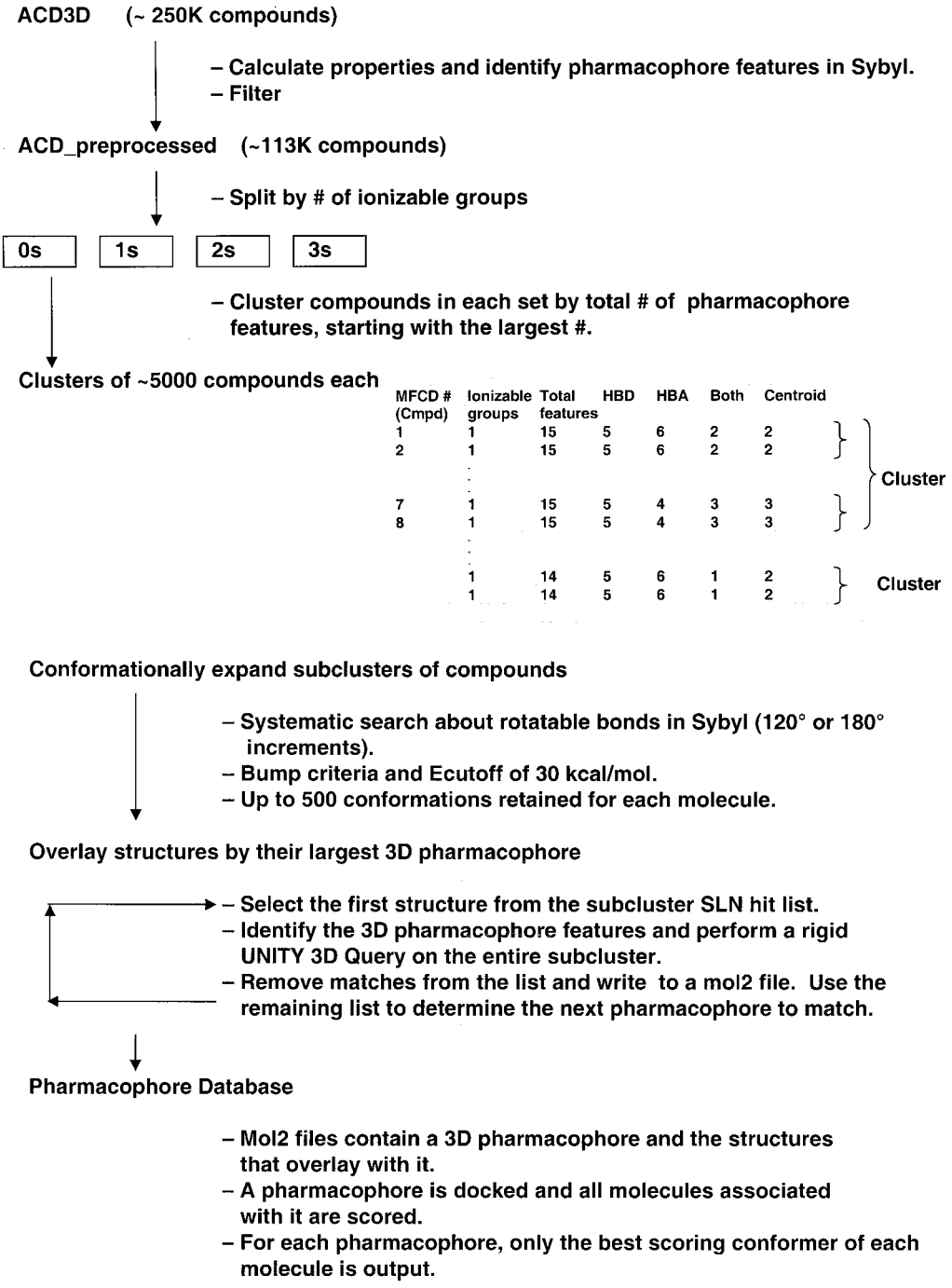


Fig. 5. Outline of PhDOCK database generation.

described in the next section). Distances between a set of pharmacophore points and a set of appropriately labeled site points are compared to determine an orientation matrix. An orientation is accepted if both the differences between matched distances are within the set tolerance and the number of pharmacophore points that clash with the target structure does not exceed the user-set bump maximum. Each acceptable transformation matrix is used to dock all conformers associated with the pharmacophore. The conformers are scored and in the final ranked list only the best-scoring conformer of each molecule is saved. To facilitate an analysis of pharmacophore enrichment in the final hit list, the identity of the pharmacophore with which the conformer is associated in the database is retained. As with standard DOCK, a maximum number of orientations is typically set and a fully exhaustive search is not always done. With PhDOCK, manual matching with a minimum and maximum number of nodes to match must be specified.

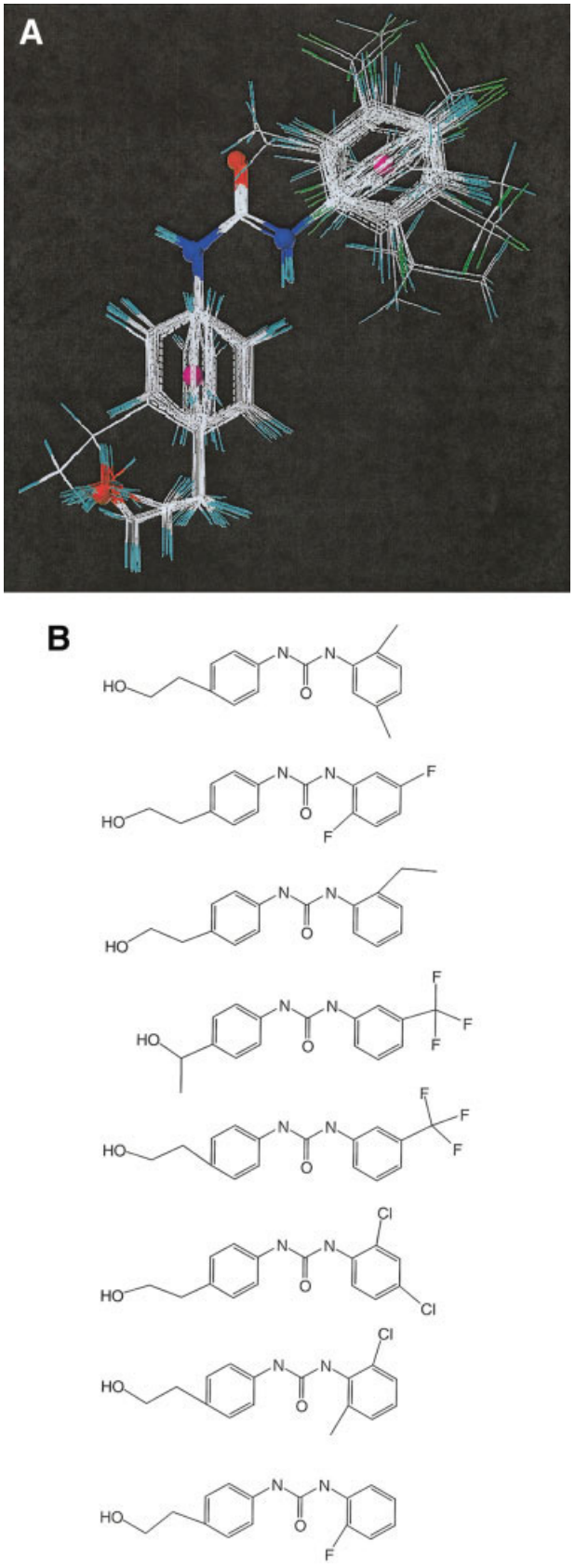


TABLE I. ACD Pharmacophore Database Statistics

Number of molecules	~113,000
Number of conformers	~15 million
Number of pharmacophores	~1.8 million
Average conformers/molecule	~130
Average conformers/pharmacophore	~8.3
Average pharmacophore points/pharmacophore	~8.6
Total size of database	~69.9 Gb

PhDOCK Database Generation and Format

The preparation of a PhDOCK database involves determining pharmacophore feature lists for each compound, generating molecular conformations, and overlaying the conformers onto 3D pharmacophores (Fig. 5). Specifically, a database SD file is processed using the DOCK utilities sdf2mol2 and sybdb, which convert the molecules to Sybyl mol2 format and assigns Sybyl atom types so that the molecules are protonated appropriately at physiological pH. For each compound, a Sybyl SPL script calculates the number of rotatable bonds and ionizable groups and a pharmacophore feature list (a ID pharmacophore). Features include hydrogen bond acceptors, hydrogen bond donors, duals that can act as both hydrogen bond acceptors and donors, and ring centroids. Non-carbon atoms with lone pairs and no attached hydrogens are considered acceptors, those with attached hydrogens and no lone pairs are considered donors, and those with lone pairs and hydrogens are duals. Centroids are derived from five- and six-membered rings.

Next, the database is imported into ISIS (MDL Information Systems Inc., 1997), where it is filtered by the number of rotatable bonds ($\leq$ 14), ionizable groups ($\leq$ 3), and pharmacophore features ($\geq$ 4) and then exported as 3D SD files partitioned by the number of ionizable functional groups per molecule. Because currently a solvation correction is not included in the PhDOCK scoring function, it is desirable only to compare directly similarly charged molecules in the final hit lists. Otherwise, the current lack of a solvation correction may arbitrarily bias the ordering of charged molecules when using a force field scoring function. Further, the molecules are clustered by their 1D pharmacophore; for example, all molecules containing two acceptors, two donors, three centroids, and two duals would reside in the same cluster. The latter is done to speed the overlay process.

Each cluster of molecules is conformationally expanded using the systematic search method in Sybyl (Tripos Associates, 2000, St. Louis, MO). An upper limit of 500 conformations per molecule is employed. The bond rotation increments are set to 120° as a default value and adjusted to 180° when the bond has sp^2 character (e.g., amide, aniline, etc.). Bonds that produce degenerate conformations due to symmetry are not allowed to rotate (e.g.,

Fig. 6. Example ACD overlay for PhDOCK. (**A**) Total of 105 conformers corresponding to 8 unique molecules. Acceptor points are red and ring centroids purple. (**B**) 2D structures of the overlaid conformers.

TABLE II. Pharmacophore DOCK vs. Multiconformation DOCK

Database partition	Pharmacophores	Conformers	Conformers/ Pharmacophore[a]	Avg. no. Pharm. centers	Unique Molecules	Orients/ Conformer[b]	CPU hours[c]	Best score[d]
1								
PhDOCK	6,413	19,890	3.10	12.2	57	464.8	1.0	−123
DOCK	—	19,890	—	—	57	2500	4.7	−104
2								
PhDOCK	24,640	75,360	3.06	8.1	529	79.5	1.4	−111
DOCK	—	75,360	—	—	529	2500	14.2	−118
3								
PhDOCK	7,848	92,152	11.74	4.7	1198	16.1	0.6	−87
DOCK	—	92,152	—	—	1198	2500	15.4	−117

[a]For the ACD, conformers/pharmacophore is on average about 8.3.
[b]The average number of orientations tried/conformer; for PhDOCK, this is the average number of orientations tried/pharmacophore.
[c]On an SGI R10000 195-MHZ processor.
[d]Gridded contact score. No bump filter.

attached methyl groups, *t*-butyl groups, triple bonds, bonds adjacent to triple bonds, aryl amines, nitros, and carboxylates, sulfates, sulfonates, etc.). No minimization is done when the conformations are generated, but there is a bump check to eliminate conformers with steric clashes and a 30-kcal/mol internal energy cutoff is applied. The 30-kcal/mol conformational energy limit is large but probably necessary when generating conformations in vacuo, where electrostatic interactions can be overestimated. The clusters are then overlaid based on their largest 3D pharmacophore using a sphere tolerance in UNITY (Tripos Associates). Specifically, the first molecule of a given cluster is read into Sybyl from an SLN hit list. The largest 3D pharmacophore is extracted from the molecule. The pharmacophore query is used to perform a rigid 3D Unity search of the entire cluster to find conformers of the same and other molecules that match the pharmacophore within a tolerance of 0.25 Å at each point. The pharmacophore query and its associated molecules are saved. Once a molecule has been associated with a pharmacophore, it is deleted from the cluster and the process continues until all molecules have been associated with a pharmacophore. The pharmacophore points represent the parts of the molecule most likely to interact directly with the target. Figure 6 shows an example of an actual overlay from the PhDOCK ACD database.

The final PhDOCK database is a multi-MOL2 format file organized such that each pharmacophore is immediately followed by its associated conformers. The "atom types" for the pharmacophore coordinates are A (acceptor), D (donor), B (dual), and Y (centroid). The uncompressed database is large (~4.7 kb per conformer). To save disk space, a g-zipped version of the database can be read by PhDOCK; this reduces the database size to about 5% of the original with about a 10% increase in the overall search time. More desirable results have recently been obtained using a binary-format PhDOCK database (Kenneth Foreman, unpublished data); this database is about 20% the size of the standard multi-MOL2 format database and can reduce the search time (due to the faster binary reads) by up to 30%.

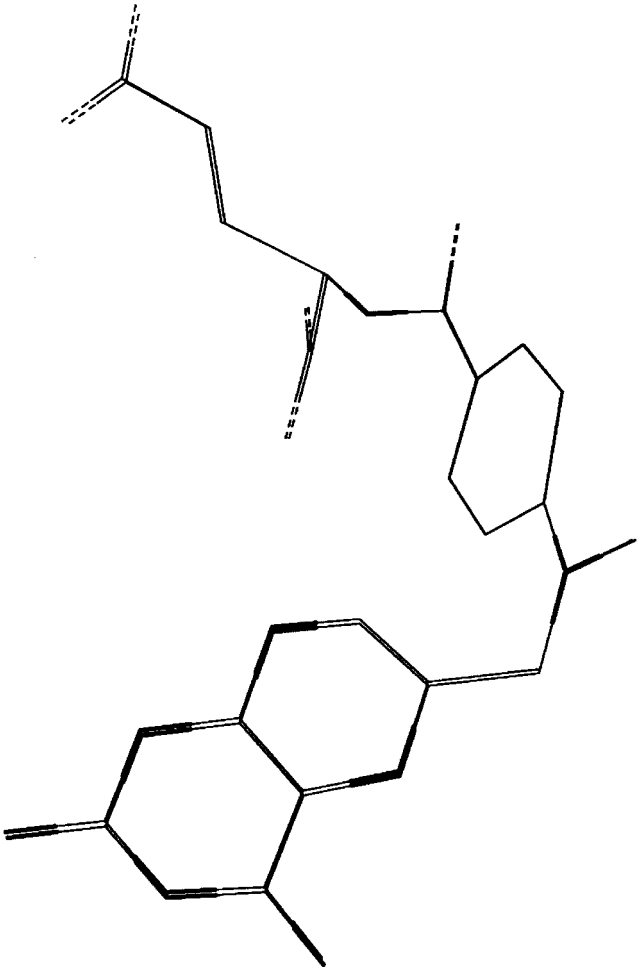


Fig. 7. Superposition of MTX as docked and MTX as bound in the DHFR structure. The PhDOCK MTX database was searched using the exact pharmacophore centers from MTX in the complex structure with DHFR as site points. Nodes minimum 4, distance tolerance 0.25 Å, chemical matching, and a gridded contact score were specified. On average, 688.2 orientations/pharmacophore were tried. Carbons are drawn in thin solid lines, nitrogen in thick solid lines, and oxygens in dashed lines.

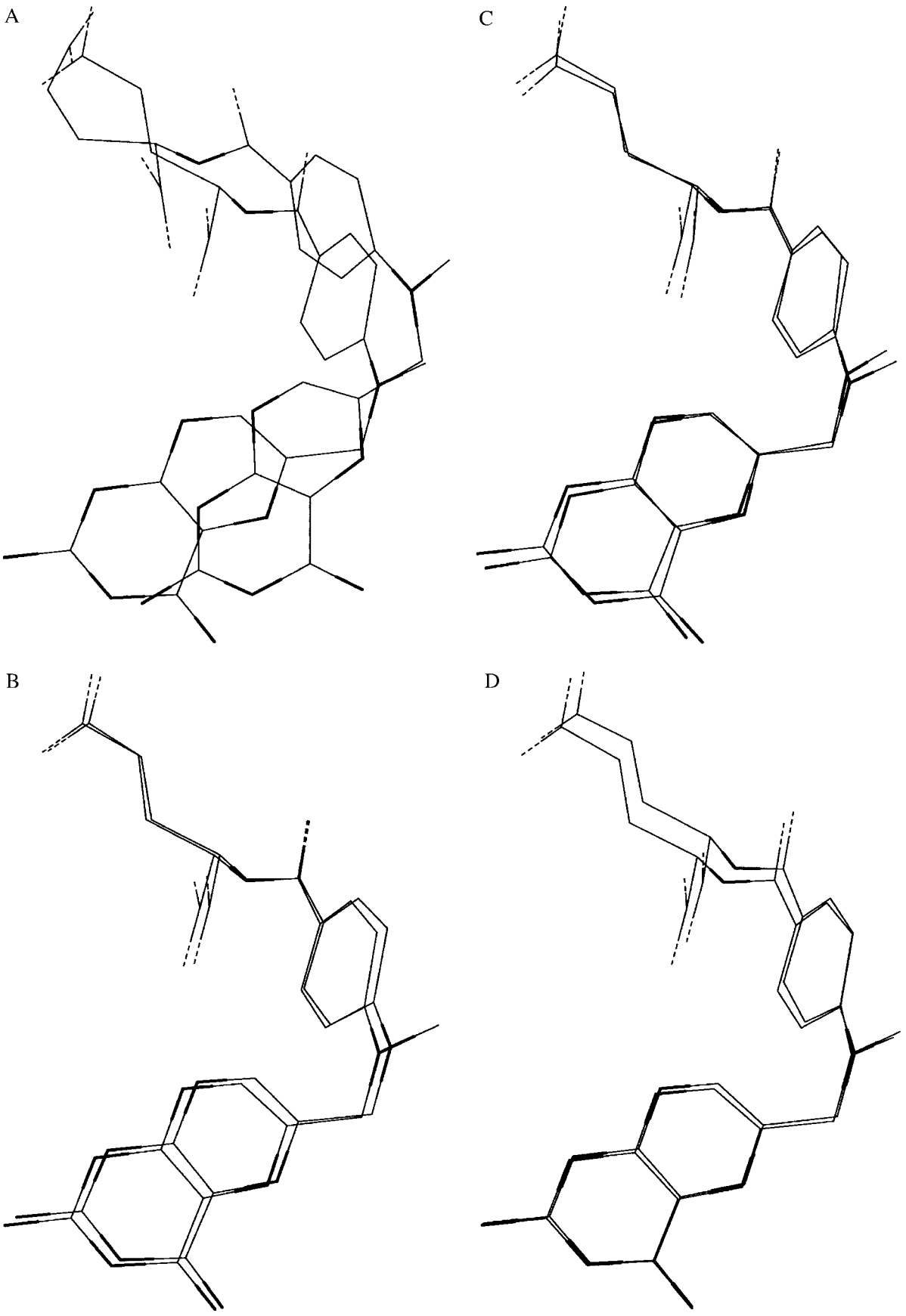


Fig. 8. Superposition of MTX as docked and MTX as bound in the DHFR structure. The PhDOCK MTX database was searched using an MCSS2SPTS-generated site point set. (**A**) Best-scoring conformer from the calculation run using a gridded contact score. (**B**) Second best-scoring conformer with a gridded contact score. (**C**) Best-scoring conformer using a gridded contact score with minimization. (**D**) Best-scoring conformer using a gridded energy score with minimization. Nodes minimum 4, distance tolerance 0.5 Å, and chemical matching were specified. On average, 1120.3 orientations/pharmacophore were tried.

TABLE III. PhDOCK Validation Results†

	DHFR–MTX			MIF–ENO			HIV1–A79285		
Score type[a]	RMSD[b]	Time[c]	Score	RMSD[b]	Time[c]	Score	RMSD[b]	Time[c]	Score
cont	2.01	0.36	−112	1.02	0.015	−68	1.37	0.14	−192
ener		Not docked			Not docked		0.29	0.14	−59.1
cont + min	0.46	0.40	−115	—	—	—	1.37	0.18	−192
ener + min	0.40	0.44	−34.7	0.58	0.16	−34.0	0.18	0.27	−62.7
cont + min + ener + min	—	0.46	—	—	0.58	—	—	0.31	—

†MCSS2SPTS generated site point sets were used.
[a]All scores were calculated on a grid. The following score types were employed: contact (cont), force field energy (ener) in kcal/mol, contact with minimization (min), and energy with minimization. In the last case, both contact and energy scoring with minimization were employed during the same search.
[b]The heavy-atom RMSD in Å.
[c]The CPU time in s/conformer.

In future implementations, the conformers may be generated using an energy function with a solvation component; this should allow the use of a smaller (~5 kcal/mol) conformational energy cutoff and result in the selection of conformers more likely to exist in a physiological environment. Further, to speed the process, the software Omega (OpenEye, 2001, Santa Fe, NM) is now being used to generate the conformers and in-house software used to cluster and overlay the conformers more efficiently. These enhancements are now being implemented to generate a PhDOCK version of our corporate database.

MCSS2SPTS to Generate DOCK Site Points

Standard approaches for generating DOCK site points for a target include the use of the SPHGEN utility that accompanies DOCK and the conversion of atomic positions from crystallographically determined ligands (e.g., water molecules, inhibitors, cofactors, etc.). SPHGEN creates an inverse surface of the binding site that is defined by the set of all overlapping spheres that fill the binding site and are tangent to the molecular surface at only two points. The sphere centers are used as site points. Alternatively, chemically labeled site points can be generated in an automated fashion using the script MCSS2SPTS, which is described in the accompanying article.[15] MCSS-derived site points that are based on theoretical pharmacophores are expected to yield better results with our pharmacophore-based DOCK.

PHDOCK VALIDATION CASES

PhDOCK Run Parameters

In general, default DOCK 4.0 parameters were used with a few exceptions. All test runs described in this article specified maximum orientations of 5000, manual matching with a minimum number of nodes of 4 and a maximum of 15, a distance tolerance of 0.5 Å, and a bump filter with a maximum of three allowed bumps unless otherwise stated. For the gridded contact score, a contact_cutoff_distance of 4 was used, while for the energy score an all-atomic model was specified. When minimization is done, a maximum of 15 iterations of rigid-body simplex minimization is performed on each acceptable orientation. All CPU timings given in this article are for an SGI R10,000 250-MHz processor unless otherwise stated.

Target Structures

High-resolution ligand–protein complex X-ray structures were used as test systems to validate the methodology. In general, all hydrogens were added to Protein Data Bank (PDB) files to reflect the ionization states at pH 7.0, the complexed ligands were removed, and the hydrogens were minimized using CHARMm (Molecular Simulations Inc., 2000, San Diego, CA).[23] For any bound cofactors, partial charges for the MCSS and DOCK calculations were taken from the CHARMm all-atom force field.[21]

The following structures were used: dihydrofolate reductase (DHFR) complexed with reduced nicotinamide adenine dinucleotide (NADPH) and methotrexate (MTX) solved to 1.7-Å resolution (PDB3DFR)[20]; macrophage migration inhibitory factor (MIF) complexed with hydroxyphenylpyruvate (ENO) at 2.5-Å resolution (PDBICA7)[24]; and human immunodeficiency virus 1 (HIV1) protease complexed with a symmetrical difluoroketone containing inhibitor A79285 at 1.7-Å resolution (PDB1DIF).[25]

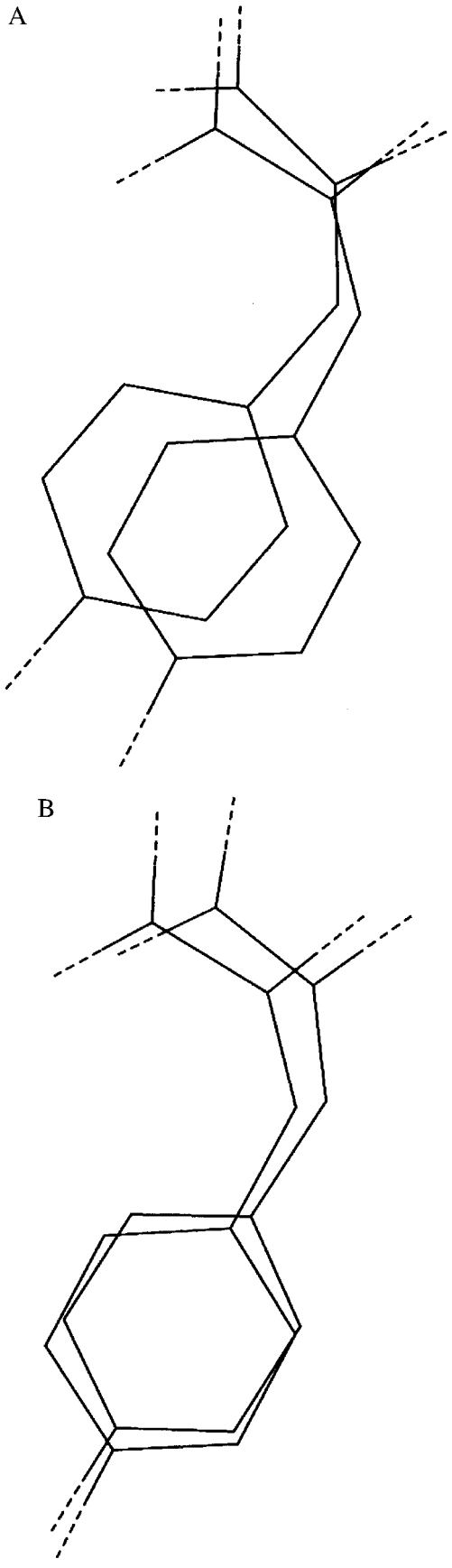
Site Point Sets

Site points for each test system were generated using MCSS2SPTS.[15] The set of 67 site points for DHFR consists of 15 acceptors, 16 donors, 17 duals, and 19 centroids. The set of 35 site points used for MIF consists of 9 acceptors, 6 donors, 6 duals, 10 centroids, and 4 neutrals. The set of 90 site points for HIV1 consists of 17 acceptors, 19 donors, 17 duals, 21 centroids, and 16 neutrals.

For comparison, the exact pharmacophore centers derived from MTX in the crystal structure with DHFR were used as site points. In addition, a set of site points was generated for the MIF structure using SPHGEN; this set consists of 18 uncolored site points from the SPHGEN cluster 0 that are within 1.6 and 2.8 Å from the protein and located in the known binding pocket.

Trial Databases

Small, conformationally expanded databases were constructed for the ligands MTX, ENO, and A79285 as described in the database generation section above. A database consisting of 500 conformers of MTX was constructed. In this case, each conformer is associated with a different 3D pharmacophore. As a second test case, a



PhDOCK database was created that consists of 3 pharmacophores and 54 conformers of ENO; each pharmacophore has 18 conformers associated with it. Finally, a database consisting of 115 pharmacophores and 1000 conformers of A79285 was constructed; the average number of conformers per pharmacophore is 8.7, with a minimum of 1 and a maximum of 19.

PhDOCK ACD Database

The ACD database was generated as described in the Methods section above except that before determining the 1D pharmacophores the database was filtered to remove compounds considered undesirable and duplicate structures. This filtering step is expected to eliminate compounds that are clearly not "drug-like".[26] Table I gives statistics on this PhDOCK ACD database.

For tests of the sampling and speed-up that can be achieved with PhDOCK, three intermediate-size database partitions were created from the part of the ACD database containing ligands with one ionizable group. The original ACD database is ordered from largest to smallest pharmacophore, and the three partitions were randomly selected from three distinct regions of the original database such that their characteristics (e.g., number of conformers per pharmacophore and average number of pharmacophore centers) would be distinct while maintaining similar search times. Statistics for these three partitions are given in Table II.

Modified PhDOCK ACD Database

To determine the speed-up due solely to the overlay of the conformers, for PhDOCK versus DOCK 4.0, a modified-format database was also created for each of the three partitions described above and in Table II. The standard PhDOCK database in Sybyl mol2 format consists of a pharmacophore (as a molecule) followed by all of its associated conformers (as separate molecules). In the modified database, after each pharmacophore the pharmacophore points were added as atoms to each associated conformer. For the DOCK 4.0 calculations, the pharmacophores were deleted from the database and only pharmacophore atom types were allowed to match to the site points. This ensured that exactly the same orientations would be tried for each conformer with DOCK 4.0 as with PhDOCK. To ensure that the total time was not being significantly increased by reading in the additional "atoms" with each conformer during the DOCK search, these same "atoms" were also retained with each conformer in the modified PhDOCK database for the comparison search.

Fig. 9. Superposition of ENO as docked and ENO as bound in the MIF structure. The PhDOCK ENO database was searched using an MCSS2SPTS-generated site point set. (**A**) Best-scoring conformer from the calculation using a gridded contact score. (**B**) Best-scoring conformer using a gridded energy score with minimization. Nodes minimum 4, distance tolerance 1.0 Å, and chemical matching were specified. On average, 49.0 orientations/pharmacophore were tried. Carbons are drawn in solid lines and oxygens in dashed lines.

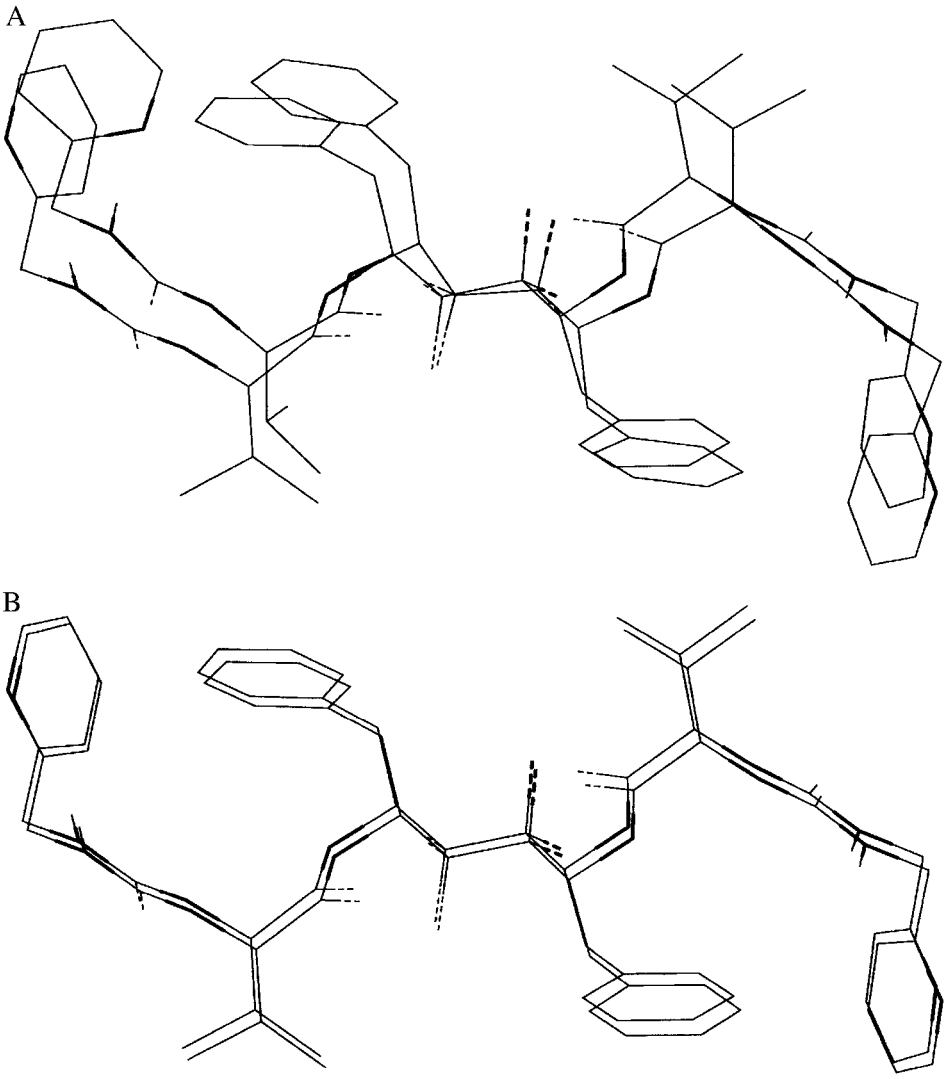


Fig. 10. Superposition of A79285 as docked and A79285 as bound in the HIV1 protease structure. The PhDOCK A79285 database was searched using an MCSS2SPTS-generated site point set. (**A**) Best-scoring conformer from the calculation run using a gridded contact score. (**B**) Best-scoring conformer from the calculation run using a gridded energy score. Nodes minimum 4, distance tolerance 0.5 Å, and chemical matching were specified. On average, 2570.5 orientations/pharmacophore were tried. Carbons are drawn in thin solid lines, nitrogens in thick solid lines, fluorines in thick dashed lines, and oxygens in thin dashed lines.

DHFR Actives Database for Enrichment-Factor Calculations

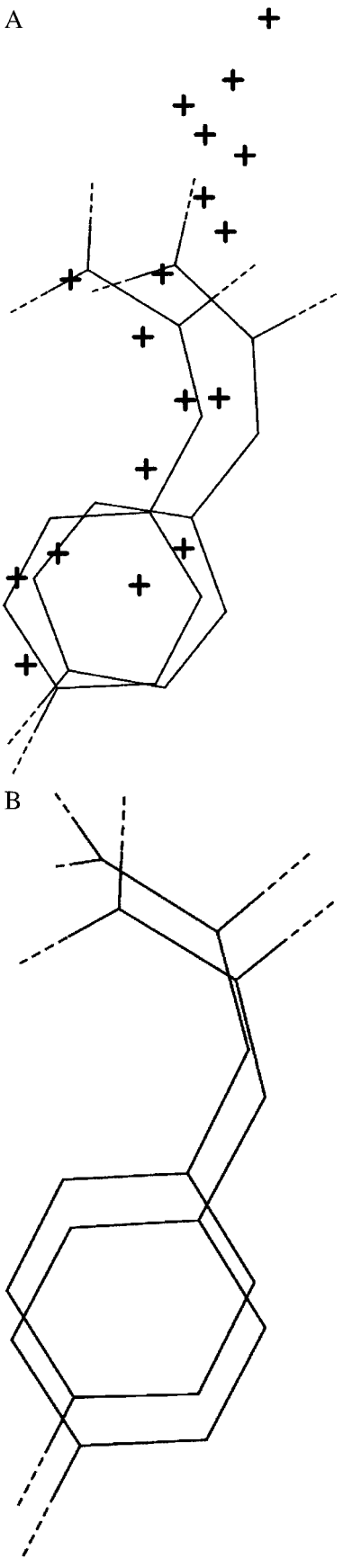
The structures of 12 known binders to *Lactobacillus casei* DHFR were conformationally expanded and organized into pharmacophoric ensembles as described above for the other databases. This set of binders is composed of the ligands from the following protein–ligand complex structures: 3dfr, 7dfr, 1rf7, 1dyh, 1jol, 1dis, 1bzf, 1dyj, 1boz, and 1dyr, as well as brodimoprim and piritrexim. Each of these ligands is known to bind to this strain of DHFR.[27–35] This DHFR actives database was screened for binding to DHFR using both PhDOCK and DOCK 4.0 under the identical conditions to those used for the three large ACD partitions.

RESULTS AND DISCUSSION

PhDOCK Validation

As a first test of PhDOCK and MCSS2SPTS,[15] a database consisting of 500 conformers of MTX was screened, initially using the pharmacophore centers from MTX in the crystal structure with DHFR as site points. As expected, the search correctly docks MTX into the binding site (Fig. 7). The root-mean-square deviation (RMSD) of the best-scoring conformer, as docked, with the X-ray structure is 0.38 Å, the gridded contact score is −113, and the total CPU time was 33.6 s or 0.06 s/conformer.

Next, the same search was carried out except that the set of MCSS-derived site points (without the neutral site points) was used. This set of site points is fairly large, 67 site points, so the timings for this test case represent



somewhat of an upper bound. Using a gridded contact score with no minimization, the best-scoring conformer that PhDOCK returns has an RMSD as docked of 2.01 Å from the X-ray structure [Fig. 8(A)]. While this RMSD is relatively high, the ligand is largely in the correct orientation and position. Further, the second-ranked hit has an RMSD of only 0.31 Å [Fig. 8(B)]. Its score is −106 versus −112 for the best-ranked conformer. Thus, in this case it may be more of a reflection of a deficiency in the scoring function rather than the sampling that the X-ray conformer was ranked second best. This search took 2.96 min of CPU time or 0.36 s/conformer. Inclusion of minimization during the search resulted in the X-ray conformation of MTX being top ranked. Its RMSD as docked is 0.46 Å with a score of −115 [Fig. 8(C)] and a total search time of 3.27 min or 0.40 s/conformer. In this case, the minimization increases the search time by about 10%.

If this calculation is repeated using the gridded energy score instead of the contact score, without minimization, MTX is never docked into the binding site with a favorable (negative) score. All saved orientations place much of the ligand outside of the binding site, presumably because this is a restrictive binding site, and the force field score is more sensitive than the contact score to small deviations from ideal van der Waals contact distances. When minimization is added during the gridded energy evaluation, the RMSD as docked is 0.40 Å [Fig. 8(D)] with a total time of 3.63 min or 0.44 s/conformer (see Table III for additional details).

When the above search is run simultaneously using both the contact score and the energy score with minimization of the energy score only, the total time is 3.66 min, still only 0.44 s/conformer. The output for the contact hits is identical to that for the run specifying only contact score and no minimization, and the output for the energy hits is identical to the results for the run with only energy scoring and minimization. If minimization is specified simultaneously for both the contact and the energy score, the calculations take 3.88 min or 0.46 s/conformer. The results in this case are not exactly the same as when the two calculations are run separately; the difference is due to the sequence of the random number generator used for the minimization steps. Calculations using a receptor-based contact score that is not on a grid result in a more accurate score and therefore the RMSD is sometimes a little lower, but these calculations also require about 10 times longer to run.

As a second example, the ENO PhDOCK database, which consists of 3 pharmacophores and 54 conformers, was screened using the set of MCSS-derived site points for the X-ray structure of MIF.[15] Because the full site point set was relatively small, neutral site points were included and could match any database atom. Using contact scoring and

Fig. 11. Superposition of ENO as bound in the MIF structure and ENO as docked using (**A**) PhDOCK and (**B**) standard DOCK 4.0. (A) SPHGEN-generated site points used for both searches are shown as crosses. Nodes minimum 3, distance tolerance 0.5 Å, and no chemical matching were specified. On average, 163.0 orientations/pharmacophore were tried for (A) and 10280.8 orientations/conformer for (B).

no minimization, the RMSD of the best scoring conformer with the X-ray structure [Fig. 9(A)] was 1.02 Å, and it took 0.80 s to dock the 54 conformers or 0.015 s/conformer. Again, with the energy score the search did not dock ENO into the binding site with the ring deepest in the pocket unless minimization was specified. With the energy score and minimization, the RMSD as docked of the best-scoring conformer was 0.58 Å [Fig. 9(B)] and the total time was 8.58 s or 0.16 s/conformer. A more permissive distance tolerance of 1.0 Å was used for this calculation; this is reasonable given that 1.2 Å was used for the clustering of the site points with MCSS2SPTS, there is some question about the resonance structure of the bound ligand, and the binding site is relatively small.

As a third test case, the A79285 ligand database was screened using the MCSS2SPTS-generated site points for the HIV1 protease structure. With a gridded contact score and no minimization, the RMSD of the best-scoring conformer is 1.37 Å from the X-ray conformation [Fig. 10(A)] and the CPU time is 0.14 s/conformer. With a gridded energy score and no minimization, the RMSD of the best-scoring conformer is 0.29 Å and the time is also 0.14 s/conformer [Fig. 10(B)]. In this case, the search did correctly dock the ligand using the energy score and no minimization, potentially due to the less restrictive nature of the binding site. When the search is carried out minimizing the contact score, the RMSD of the best-scoring conformer is still 1.37 Å, while with the energy score and minimization the RMSD of the best-scoring conformer reduces to 0.18 Å. The time for the contact score search is 0.18 s/conformer while that for the energy score search is 0.27 s/conformer. If the contact and energy searches are carried out simultaneously minimizing both scoring functions, the total time is 0.31 s/conformer. See Table III for a summary of the PhDOCK validation results.

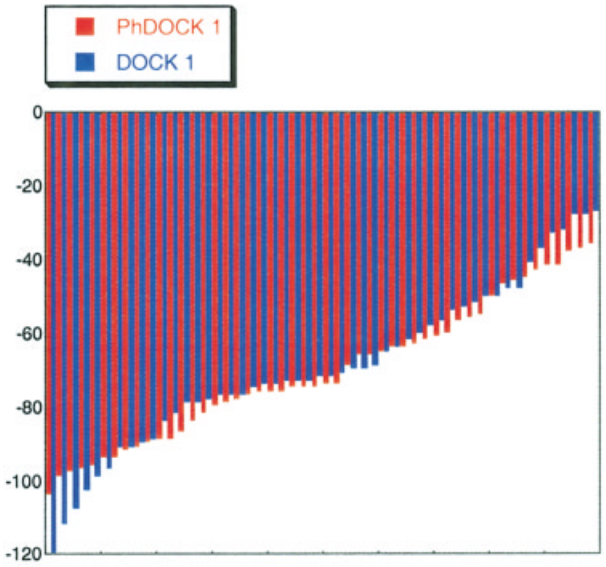
Comparison of PhDOCK and DOCK 4.0

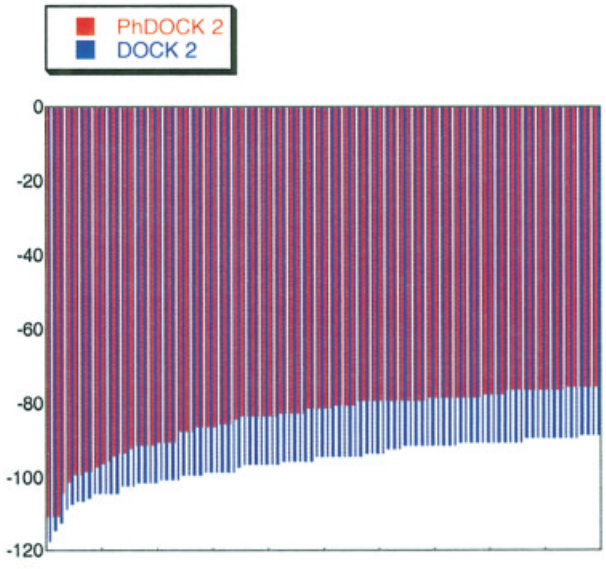
The MIF system was also used as an initial comparison of PhDOCK with standard DOCK 4.0. PhDOCK was used with the same ENO database and a set of 18 uncolored site points generated using SPHGEN [Fig. 11(A)]. A large number of closely spaced site points had been selected in the region of the ligand. If the ligand bound structure did not exist, it is likely that fewer site points would have been manually selected in that region. Even given this, a minimum of three nodes to match (vs. four) had to be specified for PhDOCK to be able to dock ENO correctly using this site point set. The RMSD of ENO as docked was 1.06 Å, the gridded contact score was −74, and the total CPU time was 0.82 s. Thus, the equivalent search for MIF using MCSS2SPTS-generated site points [Fig. 9(A)] yielded better results than this search with the SPHGEN site points [Fig. 11(A) and Table III].

For comparison, the 54 ENO conformers were separately and rigidly docked into the MIF binding site using the same SPHGEN site point set and standard DOCK 4.0 with identical run parameters [Fig. 11(B)]. In this case, the RMSD of ENO as docked was 0.94 Å, the gridded contact score was −78, and the total CPU time was 47.1 s. It is clear from the timings that a large speed-up can be achieved with PhDOCK. In this case, without chemical matching, the speed-up was about 50-fold.

Large-scale tests of MCSS2SPTS and PhDOCK were carried out with the PhDOCK ACD database. Table II summarizes the results of screening molecules containing one ionizable group for binding to DHFR. PhDOCK was compared to standard DOCK 4.0. For DOCK 4.0, the pharmacophores were removed from the database and each conformer was separately docked. The average number of conformers per pharmacophore for each of the database partitions was about 3, 3, and 12, with an average number of pharmacophore centers per pharmacophore of 12, 8, and 5, respectively. A maximum of 2500 orientations per database ligand was specified and no minimization of the orientations was performed with either method. In this example, the speed-up for the three partitions is about 4.5-fold, 10-fold, and 25-fold (see Table II). A high ratio of conformers to pharmacophores combined with a relatively low average number of pharmacophore centers results in a dramatic speed-up. For the entire ACD database in PhDOCK format (Table I), the average number of conformers per pharmacophores is ~8.3 and the average number of pharmacophore centers per pharmacophore is ~8.6, and so the overall speed-up with similar run parameters is expected to be in the 10 to 25-fold range. Inclusion of a bump filter during the run results in an even more dramatic speed-up. This is due to the fact that many more orientations fail due to the bump check when all atoms of a molecule are used for the check compared to only the pharmacophore points, and therefore many more orientations are tried (and not scored) using standard DOCK. Further, in future versions of the database with more conformers per molecule (due to lower torsional step increments) we expect an even greater degeneracy and in databases of only "lead-like" molecules with low molecular weight[36] the average number of pharmacophoric centers will be low.

A comparison of the contact score distributions for the top hits from the three partitions (Fig. 12) shows that overall the distributions, and therefore the effective sampling, is similar with both methods. For the first partition, where PhDOCK achieves a 4.5-fold speed-up over DOCK 4.0, the two methods identify similarly scoring orientations. For the second and in particular third partitions, however, for which the more dramatic speed-ups (10- and 25-fold, respectively) are achieved, there appears to be some loss in sampling with PhDOCK. Examination of the enrichment rates given below suggests, however, that this loss may not be significant and may be due to the limitations of using a contact score that is devoid of chemical knowledge. In these runs, the PhDOCK methodology required that acceptable orientations match at least four pharmacophoric centers to the chemically labelled site points, yielding chemically matched orientations even though the scoring function employed here is purely shape based. Allowing DOCK 4.0 to match any carbon atom to a "hydrophobe" site point allows for greater sampling and therefore potentially better "scoring" orientations using





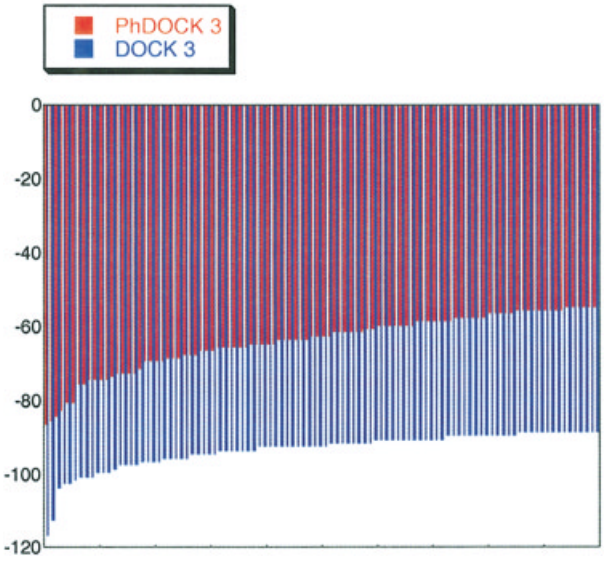
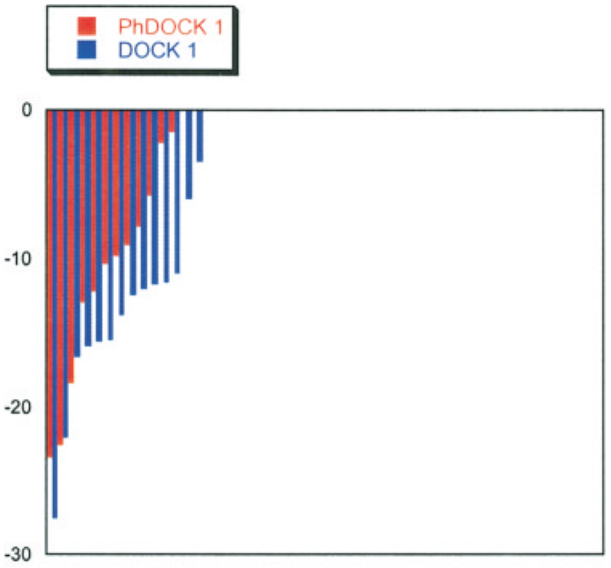
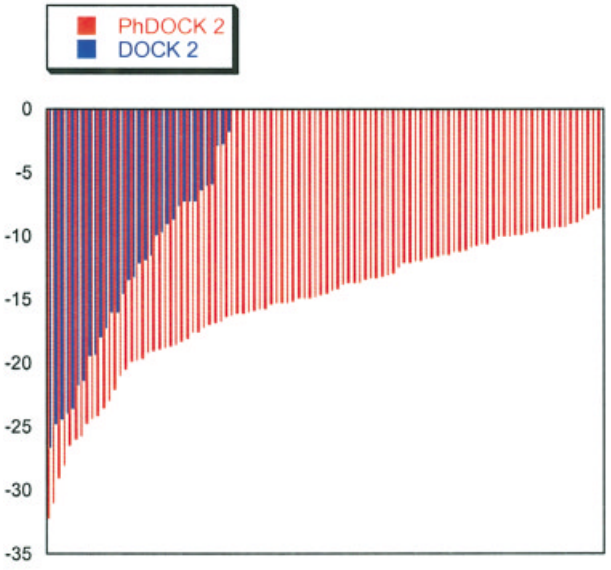


Figure 12.





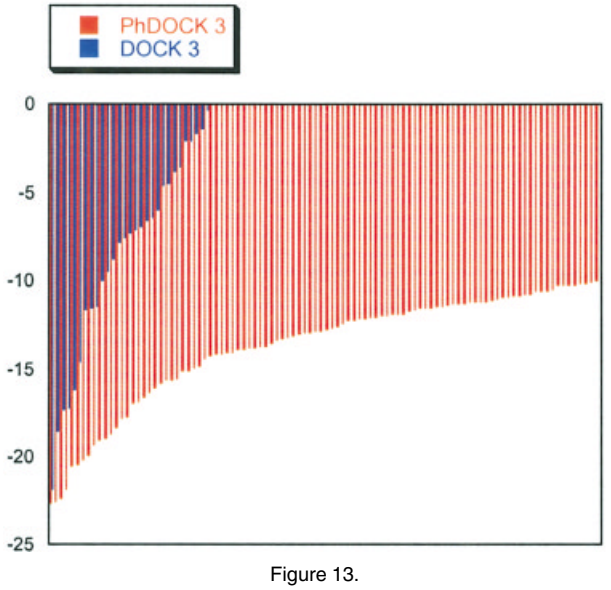


Figure 13.

TABLE IV. Further Analysis of ACD Search Results for DHFR

Database partition	% Identical MFCDs[a]		% Identical Pharmacores[a,b]	% Related pharmacores[a,c]
2	11/53 = 21%	44/106 = 42%	0%	26%
3	38/120 = 32%	108/240 = 45%	6%	45%

[a]Comparison of PhDOCK and DOCK 4.0 results; examination of the top 50 hits for partition 2 (with 529 unique molecules) and the top 100 hits for partition 3 (with 1198 unique molecules). Because partition 1 only has 57 unique molecules, the sample size is not large enough for these statistics to be significant.
[b]Largest 3D pharmacophore.
[c]Minimal % identical 1D pharmacophore (fingerprint).

the shape-based function; however, these additional orientations appear to be less relevant once their chemical complementarity is examined.

Further examination of the hit lists reveals the overlap in terms of the molecules selected using the two methods. When the top 10% of ranked molecules are compared in terms of their compositions and not only their energy distributions, the list of molecules is 21% identical for partition two and 32% for partition three. For partitions two and three, these numbers increase to 42 and 45%, respectively, when the top-ranked 20% are considered (see Table IV). When the top hits for each partition are rescored using the energy function, it becomes clear the relevant sampling is significantly greater with PhDOCK than with DOCK 4.0 (see Fig. 13).

TABLE V. Analysis of Timings for ACD Searches for DHFR

Database partition	Speed-up relative to DOCK 4.0[a]	Estimate of speed-up due to overlay[a,b]	Speed-up with minimization[c]
1	4.5×	5.4×	2.8×
2	10.3×	6.7×	13.9×
3	25.4×	8.0×	263.4×

[a]Comparison of PhDOCK and DOCK 4.0 results using gridded contact score and no minimization.
[b]Speed-up of PhDOCK relative to DOCK 4.0 due solely to the overlay of the conformers for the calculations referred to in column 1.
[c]Comparison of PhDOCK and DOCK 4.0 results using gridded contact score with 15 iterations of rigid-body simplex minimization for each acceptable orientation. The same search as in column 1 except with minimization during docking.

Analysis of Large-Scale PhDOCK Searches

Somewhat surprisingly, the lists of 3D pharmacophores associated with the top-ranked molecules are 0 and only 6% identical, respectively, for partitions two and three (Table IV). For partition one, while 88% of the 50 top-ranked molecules (of 57 unique molecules) from the PhDOCK and DOCK searches are identical, the two methods selected the same conformer of each molecule for only 1 of the 50. The overlap in terms of related pharmacophores, those derived from the same 1D pharmacophore, is slightly higher than for the molecule comparisons, as expected. These results suggest that in the database the 3D pharmacophores may be somewhat overdetermined. The maximum number of unique molecules associated with any of the docked pharmacophores is 2, while the average ranges from 1.06–1.32 for the partitions. The average number of conformers per docked pharmacophore, however, is 20, 9, and 23 for partitions one, two, and three, respectively. Nonetheless, the ranked list of 3D pharmacophores resulting from a PhDOCK search can be used to carry out secondary 2D searches of additional databases such as Catalyst (Molecular Simulations Inc., 2000),[37] ISIS (MDL Information Systems Inc., 1997), or UNITY (Tripos Associates, 2000) databases. In subsequent work, the relationship between partial pharmacophores among the top-ranked molecules will be fully analyzed.

As stated earlier, the speed-up with PhDOCK relative to DOCK 4.0 is due to the facts that many fewer ligand points are used for the matching step (pharmacophore points vs. ligand atom positions) and the ensemble is docked instead of the individual conformers. To determine the speed-up due solely to the overlay, a modified PhDOCK database was created, as described in the Methods section, for each of the three partitions. The PhDOCK and DOCK 4.0 searches were rerun with this modified database. For the three database partitions, the speed-up due to the overlay is estimated to be 5-, 7-, and 8-fold, respectively (see Table V). In addition, the calculations discussed above were carried out using a gridded contact score with no minimization. If minimization is done during the docking, the overall speed-up of PhDOCK relative to DOCK 4.0 is 3-, 14-, and 263-fold, respectively, for the three partitions (cf. columns 3 and 1 in Table V).

DHFR Enrichment-Factor Results

The DHFR actives database was screened for binding to DHFR using both PhDOCK and DOCK 4.0 and the resulting ranked poses were interleaved with the rankings for each of the three large ACD partitions. Thereby, enrich-

Fig. 12. Histograms comparing the contact scores of the top hits from the three database partitions for the PhDOCK search (blue) and the standard DOCK 4.0 search (red) of DHFR summarized in Table II. For partition one, the scores for the top 50 hits are plotted because with both methods not more than 55 molecules were docked with a negative score. For partitions two and three, the scores for the top 100 hits are plotted.

Fig. 13. Histograms comparing the energy scores after rescore of the top contact hits shown in Fig. 12. For each of the three database partitions, the PhDOCK results are blue and the DOCK 4.0 results red.

ment rates for the knowns relative to each database partition were determined. As with the three partitions, for the DOCK 4.0 calculations the pharmacophores were removed from the database and each conformer was separately docked and scored. For partition 1, with the larger, more complex randomly selected ACD molecules (having an average of ~12 pharmacophore centers/molecule) ~10 % of the DHFR knowns would be found in the top 2% of the ranked database (an enrichment factor of ~5) using PhDOCK and DOCK 4.0 [Fig. 14(A)]. Overall, there is a somewhat greater enrichment factor with PhDOCK relative to DOCK 4.0 (~4.1 vs. ~3.8 for 10% of the ranked database). For partition 2, 42% of the knowns are in the top 2% of the ranked database using PhDOCK and only 25% are when using DOCK 4.0 [Fig. 14(B)]. For partition 3, 83% of the knowns are in the top 2% of the ranked database with PhDOCK and 42% are with DOCK 4.0 [Fig. 14(C)]. Thus, for all three database partitions, using PhDOCK yielded a greater enrichment of DHFR knowns relative to the results obtained with DOCK 4.0. In fact, the partitions that showed the most dramatic speed-up using PhDOCK compared to DOCK 4.0 also showed the most significant relative enrichment of DHFR knowns. Again, this result suggests that the poses selected with DOCK4.0 are less chemically relevant.

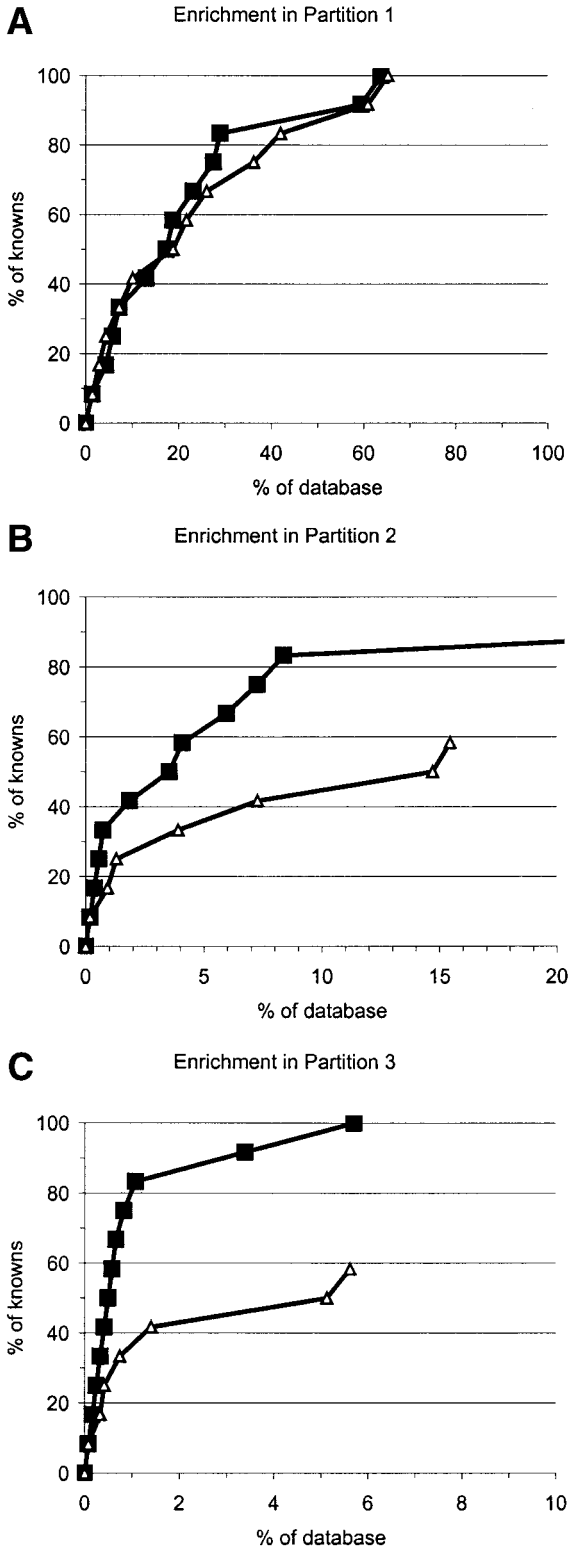


Fig. 14. Enrichment of 12 DHFR knowns seeded into the three "random" database partitions (the same ACD database partitions referred to in Table II and Fig. 12). In each plot of % knowns vs. % ranked database, the enrichment obtained using PhDOCK is shown in the curve with black squares while that obtained using DOCK 4.0 is in the curve with white triangles. For partition 1, with 57 random molecules plus 12 knowns, the graph (**A**) is extended to 100% of the ranked database. For partition 2 with 529 unique molecules, the graph (**B**) extends to 20% and for partition 3 with 1198 unique molecules it extends to 10% (**C**).

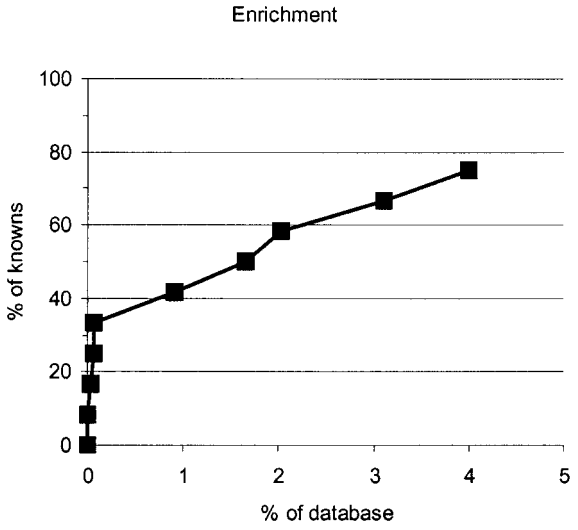


Fig. 15. Enrichment of 12 DHFR knowns seeded into the entire processed ACD database of over 110,000 compounds.

The DHFR actives database was also seeded into the entire processed ACD database of over 110,000 molecules. The calculations were run as for the three partitions except that a bump filter with a bump maximum of three was employed to speed the search. Figure 15 shows that 42% of the knowns are found in the top 1% of the ranked database, 58% in the top 2%, and 75% in the top 5%.

In summation, PhDOCK together with MCSS2SPTS has been used to reproduce a number of ligand–protein

complex structures. Using a gridded contact score with no minimization, PhDOCK is able to place the ligand close to the X-ray position, and these searches are fast. These results and the others presented in this article suggest that a reasonable, general docking strategy might involve screening a large database using only a gridded contact score without minimization, given that the site points are fully colored and pharmacophore points are being matched. Then, a subset of the hits can be rescored using a force field energy score with minimization.

The strengths of PhDOCK are that it is extremely fast and thus enables a large sampling of conformational and orientational space. Also, coloring and critical clustering of site points, which further speed the calculation, are fully functional. Further, it should be possible to analyze the results of a PhDOCK search to identify pharmacophores that are enriched in the hit list.

CONCLUSIONS

PhDOCK enables docking by the largest 3D pharmacophore of a molecule. In PhDOCK, conformers of the same and different molecules are overlaid by their largest 3D pharmacophore. The ensemble of conformers is simultaneously docked into the target structure and then the individual conformers are scored. The use of an ensemble allows for a large sampling of conformer space with a minimal number of docking events. In addition, the reduction of the query from a molecule to a substructure reduces the combinatorial matching that takes place in determining the docked orientations, that is, only the pharmacophore points (or a subset of them) are matched to predefined DOCK site points in the binding region of the target to dock the ensemble. In the related Ensemble DOCK, all 3D conformers of a given database molecule are overlaid by their largest rigid substructure.[11] One weakness of this ensemble method[11] is that the ligand atoms that make specific interactions with the target are often not part of the largest rigid substructure of the ligand. This reduces the ability of the user to utilize colored (chemically labeled) site points or critical clusters, which can greatly enhance the docking speed. With PhDOCK, coloring and critical clustering of site points are fully functional. In addition, the opportunity to define relevant pharmacophores based on retroanalysis should allow for rapid 3D pharmacophore searches of other databases that may not be set up for virtual screening.

In conclusion, PhDOCK is a fast, reliable pharmacophoric method for structure-based molecular docking. Its speed is partly due to the fact that it accounts for ligand flexibility in the database generation. The database creation process is being optimized and a binary format has been developed that dramatically reduces the overall size requirements. Future improvements to PhDOCK will also involve including solvation corrections in the scoring and incorporating target flexibility by utilizing multiple receptor scoring grids. In its present implementation, PhDOCK is able to screen a molecular database accurately and more rapidly than standard existing methods.

ACKNOWLEDGMENTS

The authors thank Iain McFadyen for setting up the DHFR knowns database, Anang Shelat for initial testing of the PhDOCK ACD, and Brian Shoichet (Northwestern University), David Lorber (Northwestern University), John Swanson (Tripos Inc.), and Matt Stahl (OpenEye) for helpful discussions.

REFERENCES

1. Thomas BE IV, Joseph-McCarthy, D, Alvarez JC. In: Guner OF, editor. Pharmacophore-based molecular docking International La Jolla, CA: University Press; 2000. p 351–367.
2. McMartin C, Bohacek RS. QXP: Powerful, rapid computer algorithms for structure-based drug design. J Comp Aided Mol Design 1997;11:333–344.
3. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–748.
4. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. Chem Biol 1995;2:317–324.
5. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 1998;19:1639–1662.
6. Morris GM, Goodsell DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. J Comp Aided Mol Design 1996;10:293–304.
7. Kramer B, Rarey M, Lengauer T. CASP2 experiences with docking flexible ligands using FLEXX. Proteins 1997;Suppl 1:221–225.
8. Rarey M, Wefing S, Lengauer T. Placement of medium-sized molecular fragments into active sites of proteins. J Comp Aided Mol Design 1996;10:41–54.
9. Makino S, Kuntz ID. Automated flexible ligand docking method and its application for database search. J Comput Chem 1997;18: 1812–1825.
10. Kearsley SK, Underwood DJ, Miller MD. Flexibases: A way to enhance the use of molecular docking methods. J Comp Aided Mol Design 1994;8:565.
11. Lorber DM, Shoichet, BK. Flexible ligand docking using conformational ensembles. Protein Sci 1998;7:938–950.
12. Knegtel RMA, Bayada DM, Engh RA, von der Saal W, van Geerestein VJ, Grootenhuis PD J Comparison of two implementations of the incremental construction algorithm in flexible docking of thrombin inhibitors. J Comp Aided Mol Design 1999;13:167–183.
13. Su AI, Lorber DM, Weston GS, Baase WA, Matthews BW, Shoichet BK. Docking molecules by families to increase the diversity of hits in database screens: computational strategy and experimental evaluation. Proteins 2001;42:279–293.
14. Ewing T, Makino S, Skillman A, Kuntz I. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. J Comp Aided Mol Design 2001;15:411–428.
15. Joseph-McCarthy D, Thomas BE IV, Alvarez JC. Automated generation of MCSS-derived pharmacophoric DOCK site points for searching multi-conformation databases. Proteins 2003;51:189–202.
16. Kuntz ID, Blaney JM, Oarley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule–ligand interactions. J Mol Biol 1982;161:269–288.
17. Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. J Comput Chem 1992;13:505–524.
18. Shoichet BK, Stroud RM, Santi DV, Kuntz ID, Perry KM. Structure-based discovery of inhibitors of thymidylate synthase. Science 1993;259:1445–1450.
19. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. J Am Chem Soc 1995;117:5179–5197.
20. Neria E, Fischer S, Karplus M. Simulation of activation energies in molecular systems. J Chem Phys 1996;105:1902–1921.
21. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck

JD, Field MJ, Fischer S, Gao I, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.

22. Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. Proteins 1999;34:4–16.
23. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4: 187–217.
24. Lubetsky JB, Swope M, Dealwis C, Blake P, Lolis E. Pro-1 of macrophage migration inhibitory factor functions as a catalytic base in the phenylpyruvate tautomerase activity. Biochemistry 1999;38:7346–7354.
25. Silva AM, Cachau RE, Sham HL, Erickson JW. Inhibition and catalytic mechanism of HIV-1 aspartic protease. J Mol Biol 1996;255:321–340.
26. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 1997;23:3–25.
27. Bolin JT, Filman DJ, Matthews DA, Hamlin RC, Kraut J. Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution. I. General features and binding of methotrexate. J Biol Chem 1982;257:13650.
28. Basran J, Casarotto MG, Barsukov IL, Roberts GCK. Role of the active site carboxylate in dihydrofolate reductase: kinetic and spectroscopic strudies of the aspartate 25 → asparagine mutant of the *Lactobacillus casei* enzyme. Biochemistry 1995;34:2872–2882.
29. Birdsall B, Burgen AS, Hyde EL, Roberts GC, Feeney J. Netgative cooperativity between folinic acid and coenzyme in their binding to *Lactobacillus casei* dihydrofolate reductase. Biochemistry 1981; 20:7186–7195.
30. Morgan WD, Birdsall B, Polshakov VI, Sali D, Kompis I, Feeney J. Solution structure of a brodimoprim analogue in its complex with *Lactobacillus casei* dihydrofolate reductase. Biochemistry 1995;34: 11690–11702.
31. Martorell G, Gradwell MJ, Birdsall B, Bauer CJ, Frenkiel TA, Cheung HT, Polshakov VI, Kuyper L, Feeney J. Solution structure of bound trimethoprim in its complex with *Lactobacillus casei* dihydrofolate reductase. Biochemistry 1994;33:12416–12426.
32. Then RL, Hermann F. Properties of brodimoprim as an inhibitor of DHFR. Chemotherapy 1984;30:18–25.
33. Reyes VM, Sawaya MR, Brown KA, Kraut J. Isomorphous crystal structures of *Escherichia coli* dihydrofolate reductase complexed with folate, 5-deazafolate, and 5, 10-dideazatetrahydrofolate: mechanistic implications. Biochemistry 1995;34:2710–2723.
34. Woolley JLJ, Ringstad JL, Sigel CW. Competitive protein binding assay for piritrexim. J Pharm Sci 1989;78:749–752.
35. Gangjee A, Vidwans AP, Queener SF, Kisliuk RL. 2,4-Diamino-5-deaza-6-substituted pyrido[2,3-*d*]pyrimidine antifolates as potent and selective non-classical inhibitors of dihydrofolate reductases. J Med Chem 1996;39:1438–1446.
36. Oprea TI, Davis AM, Teague SJ, Leeson PD. Is there a difference between leads and drugs? A historical perspective. J Chem Inf Comput Sci 2001;41:1308–1315.
37. Sprague PW. Automated chemical hypothesis generation and database searching with Catalyst(R). Perspect Drug Disc Design 1995;3:1–20.