QUASI: A Novel Method for Simultaneous Superposition of Multiple Flexible Ligands and Virtual Screening Using Partial Similarity

Nikolay P. Todorov,*,† Ian L. Alberts,†,§ Iwan J. P. de Esch,‡ and Philip M. Dean†

De Novo Pharmaceuticals Ltd., Vision Park, Histon, Cambridge CB49ZR, U.K., and Department of Pharmacochemistry, Vrije Universiteit, Amsterdam, The Netherlands

Received August 1, 2006

The structure of many receptors is unknown, and only information about diverse ligands binding to them is available. A new method is presented for the superposition of such ligands, derivation of putative receptor site models and utilization of the models for screening of compound databases. In order to generate a receptor model, the similarity of all ligands is optimized simultaneously taking into account conformational flexibility and also the possibility that the ligands can bind to different regions of the site and only partially overlap. Ligand similarity is defined with respect to a receptor site model serving as a common reference frame. The receptor model is dynamic and coevolves with the ligand alignment until an optimal self-consistent superposition is achieved. When ligand conformational flexibility is permitted, different superposition models are possible and consistent with the data. Clustering of the superposition solutions is used to obtain diverse models. When the models are used to screen a database of compounds, high enrichments are obtained, comparable to those obtained in docking studies.

INTRODUCTION

Experimental information about the structure of receptors, available from X-ray crystallographic and NMR studies, is invaluable for the rational design of new drugs. Nevertheless, for many targets of therapeutic interest, the structure of the receptor is unknown, and only information about diverse compounds binding to the receptor is available. In such cases, pharmacophore and QSAR methods are used to extract characteristics of the ligands that may be important for binding and correlate activity with those descriptors.[1−5] Once such a model is created it can be used in a predictive manner on new compounds. The dependence of the quality of the model on accurate ligand alignment has been a recurring problem in traditional QSAR approaches.[6] New developments in methodology hold much promise for alleviating this dependency.[7,8]

In this paper some of the outstanding issues in the molecular alignment methodology are investigated, namely the simultaneous superposition of multiple ligands taking into account flexibility and also how to include in the analysis the fact the ligands can bind to different regions of the site and only partially overlap with each other.

A number of molecular superposition methods have been presented in the literature and have been reviewed.[9] Normally, pairwise superpositions are performed,[10−13] and the simultaneous superposition of all ligands in the set has been rarely considered. When this has been conducted, usually, a base molecule is selected, and the rest of the molecules are superimposed on it without consideration for the overall optimality. Perkins and Dean[14] presented a method where pairwise matches are performed first, and results from all pairs are analyzed by using a consensus match. Johnson et al.[15] described a related technique that is also based on careful analysis of pairwise matches. There are two difficulties facing methods relying on this approach. First, all pairwise matches have to be performed, and this can be quite time-consuming especially when conformational flexibility is taken into account. Second, unless all ligands are considered simultaneously it is possible that superpositions of subsets of ligands may not be sufficient to generate the optimal alignment of all ligands at the same time, because information about the rest of the ligands is ignored. In order to cope with the second problem, a final optimization considering the similarities between all pairs of compounds could be performed once the matches have been analyzed and a putative superposition has been suggested.[16] There are several methods that perform simultaneous superposition of multiple ligands. DISCO,[17] GASP,[18] and Catalyst[19] automatically identify pharmacophore features common to all ligands in the set and optimize the superposition. Compass[20] conducts an alignment using a common substructure or pharmacophore to position the ligands. A neural network derives a model that relates biological activities to features representing the distances to the surfaces of the ligands from a set of sampling points. The model is used to realign the ligands. The alignment is iteratively improved using these steps. The requirement for a common substructure or pharmacophore touches on another important issue, namely how to deal with sets of ligands that do not have a common pharmacophore but explore interactions with different residues and regions in the active site and only partially overlap. This issue has been addressed in the case when two molecules are matched. Barakat and Dean,[21] Mills et al.,[22] and Perkins et al.[23] use the method of null correspondences that allows badly matching regions of

* Corresponding author phone: +44 1223 238000; fax: +44 1223 238088; e-mail: nikolay.todorov@denovopharma.com.
† De Novo Pharmaceuticals Ltd.
‡ Vrije Universiteit.
§ Current address: Schrodinger, Inc., 120 West 45th Street, 29th Floor, New York, NY 10036.

10.1021/ci6003338 CCC: $37.00
Published on Web 05/12/2007

the molecules to be ignored. In the approach of Perkins et al.,[24] which maximizes the surface-volume overlap between two molecules, portions of the molecular surfaces that do not match well are excluded. In a recent study, De Esch et al.,[25] making extensive use of the procedures described in refs 22 and 23, proposed a binding model for histamine H_3 receptor antagonists. This is a problem that requires careful consideration of partial similarity. A stepwise procedure was used by fixing the superposition of the compounds for which the bioactive configuration had been ascertained and gradually adding more of the compounds in the set. The model led to the discovery of a series of novel extremely potent histamine H_3 receptor antagonists.

The success of that study highlighted the necessity and prompted the investigation of the problem of multiple-ligand superposition, particularly for the case where the available ligands overlap only partially, but when considered simultaneously provide a picture of the complete active site.

The strategy for superposing ligands adopted in this work is as follows. Each ligand is characterized by a set of descriptor points. Initially, the ligand conformations and orientations are randomized. A *receptor site model* is created based on the descriptor-point positions. The number of points in the receptor site model is larger, or equal to, the number of points in any of the ligands but much smaller than the sum of the number of points in all ligands. Because the receptor model can be larger than any of the ligands, it is possible that different ligands fit well in different parts of the receptor site model, and thus the concept of partial similarity is an integral part of the method. After the receptor model is created, the ligands are rigidly fitted onto it. This initial model is likely to be far from optimal since it is based on randomized ligand configurations and most of the ligands may not fit it well. The ligand positions, orientations, and conformations as well as the receptor model are gradually modified in the search for an optimal self-consistent alignment.

Site models defined as sets of points have been used extensively in molecular docking algorithms to characterize the shape and the chemical composition of the receptor site.[26−28] While these models are static, the site models used in this ligand superposition method are dynamic and coevolve with ligand alignment. Ligand similarity is defined with respect to the receptor site model, rather than between the ligands themselves. When ligand conformational flexibility is explored, different superposition models are possible and consistent with the data. Clustering of the superposition solutions is used to obtain diverse models.

The main interest in the generation of receptor site models is that they can be used to search for novel, active ligands from compound databases. The method presented in this work can be applied to screen databases of available compounds, by flexibly superposing each database molecule individually onto the static site model. The use of receptor site models in virtual screening leads to high enrichments, comparable to those obtained in docking studies.

METHODS

This section describes the steps of the superposition method: molecular descriptors used, generation of receptor-site models, ligand refitting, scoring of superpositions and their optimization. A flowchart diagram of the method is presented in Figure 1.

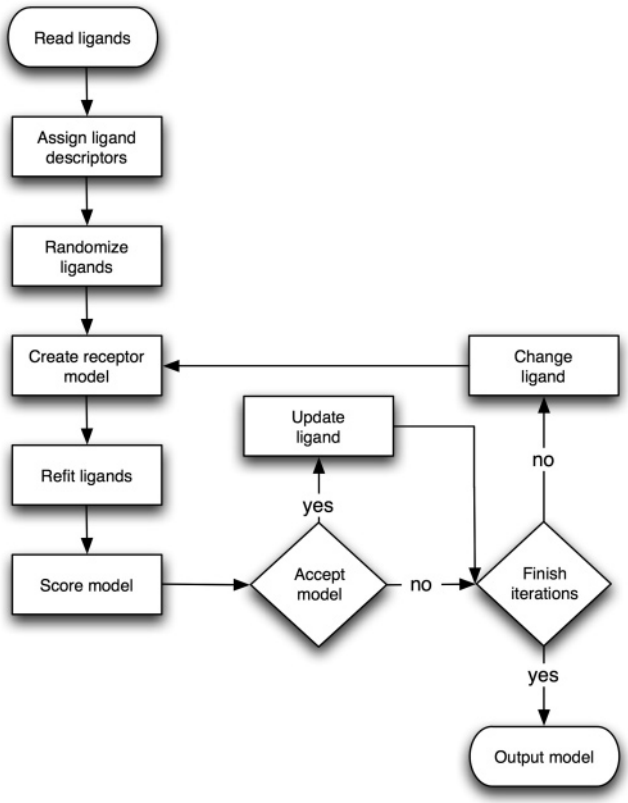

Figure 1. Flowchart diagram of the method.

Ligand Descriptors. Each ligand is described by three sets of points: *hydrogen-bond donor*, *hydrogen-bond acceptor*, and *steric* points. For each hydrogen-bond donor group in the ligand, a hydrogen-bond donor point is defined and positioned at 3.0 Å from the donor atom along the vector defined by the donor atom and the attached hydrogen atom. Hydrogen-bond acceptor points coincide with the hydrogen-bond acceptor atoms in the ligand. Amphoteric groups in the ligands, like a hydroxyl group, will have one acceptor point coincident with the oxygen atom and one donor point positioned 3.0 Å along the oxygen-hydrogen vector. Steric points coincide with all atoms that are not hydrogen-bond acceptors. These assignments are based on results of crystal surveys of hydrogen-bonding interactions, which demonstrate a pronounced directional distribution of donor atoms toward more positionally invariant acceptor atoms.[22]

Initial Configurations of the Ligands. Initially, random conformations and orientations are generated for the ligands and their geometric centers are made coincident. It should be stressed that this initial positioning is not crucial for the success of the method; it is just a starting point for global optimization of the superposition.

Receptor Site Models. The ligand descriptors and their coordinates are used to create receptor site models. The receptor site model consists of a certain number of points from each of the descriptor types. The number of points can be fixed or a range of possible values can be specified. In the latter case, during evolution of the ligand alignment, the receptor site model exhibits dynamic behavior by varying between a minimum and maximum number of points. The minimum and maximum number of points forming the site

model are determined on the basis of the number of descriptors in the ligands. The minimum number of points of each type included in the site model is set equal to the largest number of points of that type among all ligands. For example, if there are two ligands, the first one with 30 steric, 3 donor, and 2 acceptor points and the second one with 20 steric, 5 donor, and 4 acceptor points, then the receptor site model should contain at least 30 steric, 5 donor, and 4 acceptor points. The rationale is that all ligands should be able to fit into the receptor site. The default for the maximum number of points is set equal to the minimum number of points (e.g., 30 steric, 5 donor, and 4 acceptor points in the example). Even with this conservative setting, the size of the receptor site model can still be larger than any of the ligands and could be further increased if no consistent superposition is achieved. If the maximum number of points is increased, the size of the receptor site model also increases, and this enhances the exploration of superpositions with partial similarity in which different ligands explore different parts of the receptor site and only partially overlap.

In order to create the receptor site model, the ligand with the largest number of descriptor points (weighed by type) is selected. All points from the selected ligand are made part of the receptor site model. The rest of the points in the receptor site model are selected for each type of descriptor (steric, donor, acceptor) from the other ligands in a stepwise manner. At each step, the point is selected that is nearest but not closer than 1.0 Å and no farther than a certain distance (4.0 Å for steric and 5.0 Å for donor and acceptor points) to the points already included in the receptor site model (the use of a maximum distance prevents consideration of receptor models with isolated disconnected pockets). The selected point is added to the receptor site model. This step is repeated until either no more points could be selected, or the maximum number of points allowed for that type of descriptor is reached. This calculation is repeated for each descriptor type. The outcome is a set of receptor model points with their coordinates and descriptor types. It provides a description of the putative active site to which all ligands bind and serves as a common frame of reference.

Ligand Refitting. Once a receptor site model is available, each ligand is refitted onto it. This procedure mimics docking of the ligands into the current model of the active site. The current conformations of the molecules are used in this step, and, thus, this is analogous to a rigid docking procedure. The refitting procedure works as follows. For each atom in the molecule undergoing refitting, the site model point is found which is nearest to that atom. This yields a correspondence list from which a least-square superposition matrix is calculated[29] and the molecule is transformed to the new position and orientation. A functional value is calculated corresponding to the weighted rmsd of the fit. After this step, the site model point closest to a particular atom may be different from the one previously identified. The correspondences are, thus, reassigned, and the refitting is performed again. This is repeated several times until convergence is reached or a maximum of ten iterations have been conducted.

Scoring of Superpositions. A scoring function not only takes into account how well the ligands fit to the site model but also encourages the derivation of site models with relatively few points. The scoring function comprises three terms:

$$S = F_{\mathrm{rmsd}} + F_{\mathrm{vdw}} + F_{\mathrm{pnt}}$$

First, F_{rmsd} represents the weighted rmsd as calculated during ligand refitting is used to score the fit. Second, F_{vdw} corresponds to intramolecular steric clashes—if the distance d_{ij} between two nonbonded atoms i and j in a single ligand is less than 70% of the sum of van der Waals radii, the difference is added to the score

$$F_{\mathrm{vdw}} = \sum_{\mathrm{lig}} \sum_{i>j} \theta(0.7 * d_{ij}^{\mathrm{vdw}} - d_{ij})$$

where $\theta(x)$ is the Heaviside step function (defined as equal to zero for x less than zero and one for x greater or equal to zero). Third, F_{pnt} is a term that favors smaller site models—the number of points of each type in the site model multiplied by weight factors given in the section Parameters.

Optimization of the Superposition and the Receptor Site Model. A Monte Carlo (MC) stochastic tunneling optimization procedure[30] is used to find the best superposition and the corresponding receptor site model. This novel method has been shown to perform better than more established optimization methods like simulated annealing[31] on complex multimodal functions. Stochastic tunneling replaces the original energy surface by the following nonlinear transformation: $f_{\mathrm{STUN}} = 1 - \exp(-\gamma(f - f_0))$, where f_0 is the current best value of the scoring function f, and γ is a problem dependent parameter. The original functional space is compressed onto the interval [0,1], and thus it is possible to tunnel through functional barriers present on the original functional landscape if a Monte Carlo search at a finite temperature is performed on the transformed landscape. The positions of all minima of the original function are preserved. Here γ was set to 1.0 and the Monte Carlo temperature to 0.005. The default number of Monte Carlo steps used for site model generation is $5*10^4$.

Several types of transition are used to generate new ligand configurations: (i) rigid body translations, (ii) rotations and (iii) conformational changes of a randomly selected ligand, and (iv) change in the number of points in the receptor site model. After each transition is performed, a receptor model is generated using the updated ligand positions, the ligands are refitted onto the model, the fit is scored, and a decision is made whether to accept or reject the transition (Figure 1).

Transition type (iii) allows for the implementation of several strategies to handle conformational flexibility and protonation states in the ligands. Each ligand can be represented as several conformers and/or tautomers given as input. One of those is selected when transition type (iii) is tried. The conformation is modified by a random rotation around a selected rotatable bond. If the transition is accepted, then the conformation is stored and used for the next transition (iii) involving the same conformer. Thus flexibility can be handled either by creating static conformers prior to the superposition process and using them as input or by allowing dynamic conformational changes (default). The latter is the simpler and more general technique.

Analysis of Receptor Site Models. Different site models are generated in several runs of the algorithm using different random number seeds. These solutions are analyzed by additional scoring, filtering, and clustering the models. In

addition to the rmsd score, rescoring is also provided in terms of the total volume of the superposed ligands by considering ligand points weighted by the type of feature (steric, donor or acceptor). Each ligand point is associated with a sphere of radius 1.5 Å, and the union of the volume from all points is calculated.

Low-score solutions (within 10% of the lowest score, rmsd, or volume score) are retained and clustered to identify representative receptor site models. The final ligand configurations from each run are concatenated, and the resulting multiligand structures are clustered. First, the two sets of superposed structures that are furthest apart in terms of the total rmsd are selected as cluster centers. Next, more cluster centers are added sequentially, each time selecting the structure that has the largest nearest-neighbor distance to a cluster center already included in the list. The procedure terminates when the rmsd of the last added cluster center to its nearest-neighbor center becomes less than a specified threshold (3.0 Å). Each of the remaining structures is assigned the cluster number of the nearest cluster center, and the structure with the lowest superposition score is selected as a representative from each cluster.

Virtual Screening. The main application of the receptor site models generated by the above process is that they can be used to screen databases of commercially available compounds for novel active ligands. For this purpose, each putative database ligand is considered individually by flexible superposition onto the static site model. Since only one ligand is superimposed onto the site model at a time, the number of steps in the MC procedure can be drastically reduced (the default is 500), in comparison with that for the multiligand superposition involved in site-model generation, without deterioration of the results to obtain a very fast (under 1 s per ligand on an SGI R10000 workstation) screening time.

The common volume of the screened ligand with the fixed ligands used to generate the site model, weighted by the type of ligand point (steric, donor, acceptor), is used to assess the fitness of the superposition. Again, as discussed above, each ligand point is associated with a sphere of radius 1.5 Å. The superposition can be repeated several times, and the best result used. In this case, the best scores have the largest common volume.

When a database of compounds is screened, a ranked list is produced according to the superposition scores and used to calculate *enrichment graphs* (i.e., the percentage of ranked compounds screened versus the percentage of active compounds) and *hit rates* at 5% (% of known active ligands among the top 5% ranking compounds).

Parameters. Several parameters and weighting factors appear in the methodology and can be specified and controlled by the user. The values for these parameters were selected empirically based on the results of test superpositions.

The main parameters are the weighting factors for the rmsd fit of the ligands onto the site model. A weighting factor of 0.2 is used for the steric points and 1.0 for the donor and acceptor points. These values were chosen because when the pteridine rings of the dihydrofolate ligands methotrexate and folate were superposed (see below), any lower hydrogen-bonding/steric weight ratio yielded an incorrect stacked alignment of the rings compared to the crystal alignment. The same values, namely 0.2 for the steric points and 1.0 for the donor and acceptor points, are used to select the ligand with the largest number of descriptor points for the creation of the receptor site model and for the additional scoring of the superposition by volume.

The method can be used in different ways in terms of the number of descriptor points in the site model. Multiple runs can be conducted specifying a different number of donor, acceptor, and steric points. The site model with the minimum number of site points that gives a ligand superposition with a low rmsd should be selected. Alternatively, if the site-model selection is to be performed in a single run in which the number of descriptor points could change, a penalty factor could be specified for each point added to the model in order to discourage the use of unnecessarily large active-site models. The penalties would specify the amount by which the ligand fit should improve in order to justify the addition of a new point to the site model. Values of 0.005 for steric and 0.1 for donor and acceptor points were selected empirically, reflecting the predominance of steric points compared to hydrogen-bonding points.

The frequency of ligand conformational changes is set to five times the frequency of the other transitions. The maximum steps for ligand translation, orientational, and conformational changes are set to 3 Å, 180°, and 180°, respectively.

RESULTS

There are two types of objective tests that could be used to assess the performance of a molecular superposition method. First, if protein−ligand cocrystal structures are available for a series of ligands binding to the same protein, the protein backbone or active site residues could be used to align the ligands in the same reference frame. This is referred to as crystal superposition. The superposition generated without reference to the crystal structures could then be compared to that reference model in terms of rmsd. The closer the two superposition models are, the more accurate the active site representation. This implies that the molecular superposition method is valuable in building site models in the absence of a crystal structure. Second, and more importantly, how useful is the generated model in virtual screening applications? If a database containing random small molecule structures is seeded with several known active ligands, then the effectiveness of the generated active site model in retrieving the ligands as high-ranking candidates needs to be assessed. This type of analysis is frequently used in molecular docking applications to gauge the performance of the methodology. Of course, if crystal structures are available for a set of ligands in complex with the protein of interest, these could be used directly in screening applications without recourse to the active site model superposition.

Our molecular superposition method was applied to three families of ligands. The first application involves the alignment of dihydrofolate reductase (DHFR) ligands in comparison with available crystal structures to explore how closely the method could reproduce the crystal superposition and how sensitive a screening application is to the exact details of the superposition. The two other examples utilize standard data sets for molecular docking applications, namely

Table 1. DHFR Complexes

PDB code	1aoe	1diu	1dyi	1lud	4dfr
reference	33	34	35	36	32
resolution	1.6 Å	NMR	1.9 Å	NMR	1.7 Å
organism	*C. albicans*	*L. casei*	*E. coli*	*L. casei*	*E. coli*
residues	Ile-8, Ile-9, Val-10, Glu-32, Phe-36, Ile-112	Phe-3, Leu-4, Trp-5, Asp-26, Phe-30, Ala-97	Leu-4, Ile-5, Ala-6, Asp-27, Phe-31, Ile-94	Phe-3, Leu-4, Trp-5, Asp-26, Phe-30, Ala-97	Leu-4, Ile-5, Ala-6, Asp-27, Phe-31, Ile-94

estrogen receptor (ER) and thymidine kinase (TK) ligands, and will be a useful benchmark to assess the performance of the screening application in comparison with molecular docking that relies on available protein structures. These applications demonstrate the potential utility of the method in drug design projects. Unless explicitly stated, we have always used the default parameter settings specified in the Methods section, and conformational flexibility has been handled dynamically on the fly.

DHFR Ligands. Five ligands cocrystallized with DHFR[32−36] were used. The ligands were taken from the following complexes in the Protein Databank: methotrexate (MTX) from 4dfr, 1,3-diamino-7 (1-ethylpropyl)-7H-pyrrolo[3,2-F]-quinazoline (GW345) from 1aoe, brodimoprim-4,6-dicarboxylate from 1diu, folate from 1dyi, 1lud. Table 1 and Figure 2 show details of the complexes and the ligand structures. In order to compare how well the ligand superpositions correspond to the experimental data, the five complexes were aligned in the reference frame of 4dfr (Figure 3). For this purpose, the protein backbone atoms were used of six corresponding residues listed in Table 1. The selected residues were positioned in the active site, and their interactions with the ligands were conserved among the five complexes.

Three experiments were performed to evaluate the performance of the method. First, several (up to ten) MTX molecules were superposed. Second, MTX was split into two fragments, and superpositions were generated using the two fragments and MTX relying extensively on partial similarity optimization. Third, the five DHFR ligands were superposed, and the generated models were compared to the crystal superposition.

Superposition of MTX Molecules. The performance of the method was first investigated by flexibly superimposing several copies of MTX (from 2 to 10 copies) and varying the number of Monte Carlo steps in the optimization procedure from 10^3 to 10^4 in steps of 10^3. Figure 4 shows the median of the fit, corresponding to a weighted rmsd, to the receptor model (over ten repeats) for all combinations of number of ligands and Monte Carlo steps (900 runs in total). The graph shows that as expected the performance of the method improves as the number of Monte Carlo

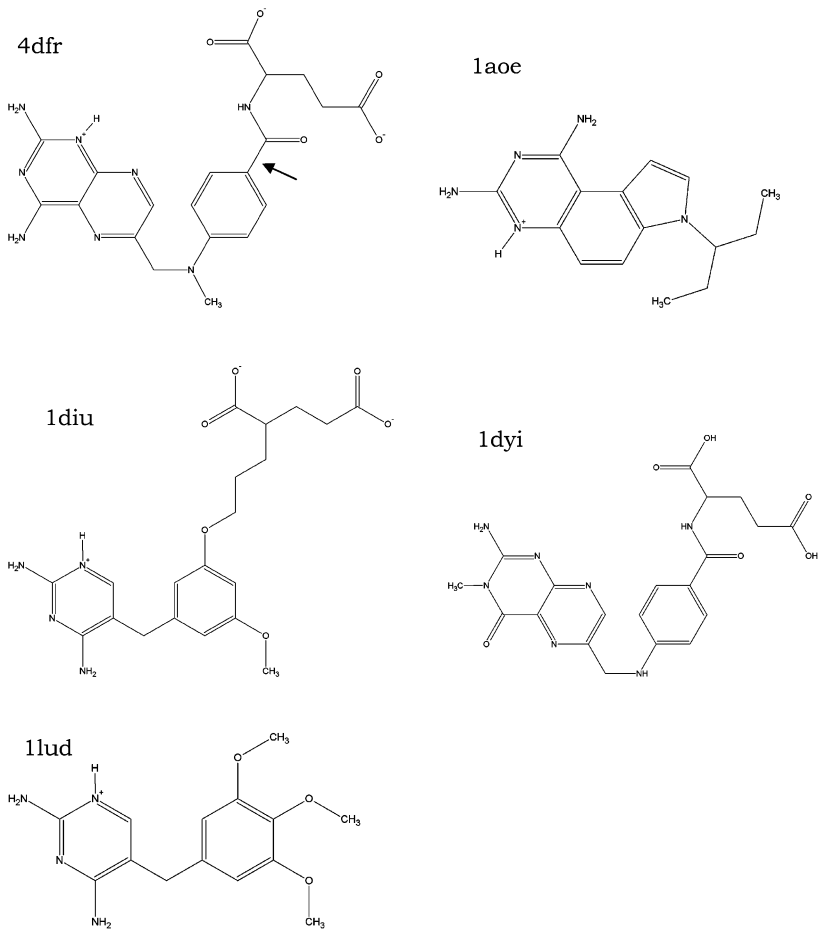

Figure 2. Dihydrofolate reductase ligands from cocrystals with pdb codes 4dfr, 1aoe, 1diu, 1dyi, and 1lud.

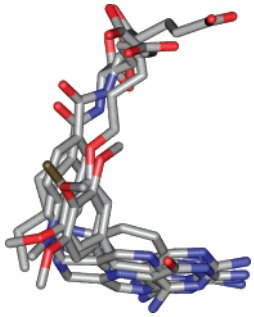

Figure 3. Crystal superposition of the five methotrexate ligands.

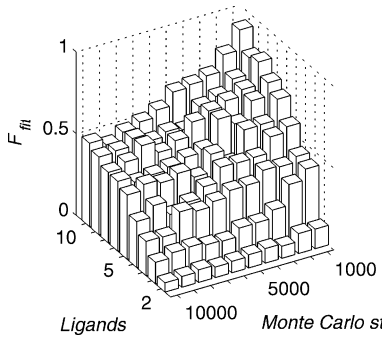

Figure 4. The median fit score over 10 runs as a function of the number of methotrexate ligands and the number of Monte Carlo steps (in 10^3).

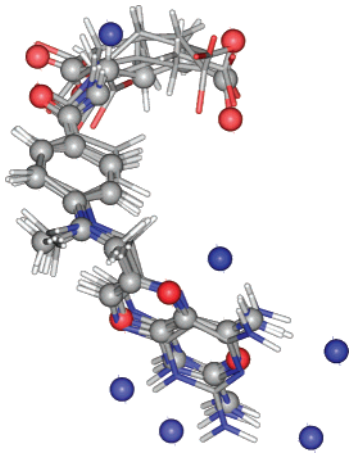

Figure 5. Superposition of five MTX molecules shown as sticks and the corresponding site model shown as spheres: gray spheres − steric points, blue spheres − donor points, and red spheres − acceptor points.

optimization steps increases and the number of ligands decreases. Figure 5 shows one of the superpositions and the corresponding receptor model of five flexible MTX molecules, obtained in 5000 steps. Different runs produced superpositions of MTX in different but self-consistent conformations and correspondingly different receptor models. The diversity of the ligands should be increased and the ligand flexibility decreased in order to restrict the number of putative receptor models. It should be noted however that if the conformation of one of the MTX molecules is fixed in the conformation found in the crystal structure, then the correct crystal superposition is produced. This corresponds to a situation where there is a ligand available with a known active conformation.

MTX and Two Fragments. The MTX molecule was split into two fragments by deleting the bond indicated with an arrow in Figure 2, and superpositions were obtained of the fragments and MTX itself. In this test there are no atoms common to all three molecules, and consideration of partial similarity is clearly important. The correct superposition, where the two fragments are aligned to their counterparts in MTX, was produced with the lowest score (Figure 6). In fact, this corresponds to a family of superpositions with varying but self-consistent conformations of the participating molecules. This is analogous to the previous test case where identical MTX molecules were matched; the solutions differ in conformation due to lack of constraints from different diverse ligands. Interestingly, an alternative solution is also found (with higher score), in which both fragments are placed onto the pteridine ring of MTX. Unsurprisingly, the match for one of the fragments is not as good as that found in the best solution; still the atoms are positioned close to receptor points without following the bond network of the pteridine ring system.

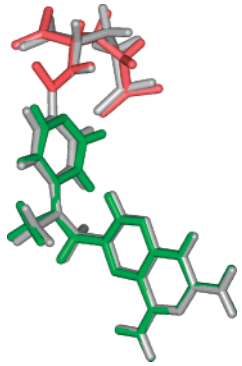

Figure 6. Superposition of methotrexate (gray) and two fragments (red and green).

Five DHFR Ligands. After these initial tests, the five DHFR ligands described in the beginning of this section (Figures 2 and 3) were superimposed. The crucial question to investigate was how closely the structure of the binding site could be predicted based on the ligand superpositions. Not only are the ligands quite flexible but also there is some diversity in the set that allows the ligands to constrain each other in order to produce a good superposition. The balance between ligand flexibility and ligand diversity will determine how well the actual crystal superposition is reproduced. If flexibility dominates then there will be multiple self-consistent superpositions, but no objective way of selecting between them. This was the case when identical MTX ligands were superposed. On the other hand, if there is available a set of diverse rigid ligands spanning the whole of the active site, one solution closely corresponding to the crystal superposition will be expected to dominate.

One hundred runs of the algorithm were performed using different random number seeds to flexibly superpose the five ligands. Low score solutions were retained and clustered. Three clusters and cluster representatives were identified. The representative solutions of the three clusters are shown in Figure 7.

It is interesting to note in this example that the pteridine ring system in folate (1dyi) is rotated compared to the other ligands in the crystal superposition (e.g., 4dfr). This is due to the different patterns of the hydrogen-bonding groups, dominating the interaction of the pteridine system with the receptor. The generated superposition that is closest to the

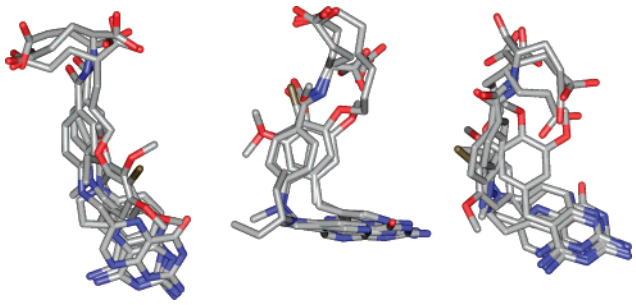

Figure 7. Representatives of the three solution clusters for superposition of the ligands in the *dhfr5* set.

crystal superposition yielded a total rmsd of 1.94 Å from all five ligands, and the closest cluster representative had an rmsd of 2.20 Å (the second structure in Figure 7).

Virtual Screening. In the virtual screening experiments performed with this data set as well as with the TK and ER ligand data sets we followed the evaluation methodology of Bissantz et al.[37] A set of 990 random compounds and 10 known ligands was used for each target to assess the performance.

The set of random compounds was obtained from Bioinfo 2004 − Release A, which is a library of 550 000 druglike compounds selected from a total of 2.5 million compounds from commercial screening collections (e.g., Asinex, Biospecs, Leadquest, etc.), filtered for druglikeness (>100 filters) and uniqueness. The molecules are ionized at the most likely ionization site, and the library is regularly updated by Rognan and co-workers. A set of 1000 random compounds can be downloaded from the group Web site (http://bioinfo-pharma.u-strasbg.fr, accessed June 16, 2005); the first 990 compounds of this set were used for the screening experiments in this work under the name of *bioinfo990*.

DHFR Ligands. Ten DHFR ligands (referred to as the *dhfr10* set, see Figure 8), different from the five used for the generation of the active site model (*dhfr5*), were extracted from the World Drug Alerts (WDA) database[38] and added to the 990 random compounds.

Several screening experiments were performed. First, we used the crystal superposition of the *dhfr5* set as a query to screen the data set of 1000 compounds (*bioinfo990+dhfr10*). Since the superposition algorithm is stochastic and only a limited amount of sampling could be afforded in a screening application, the screen was repeated five times in default mode (500 Monte Carlo steps) to check how the enrichment curves vary. The results of these experiments are summarized in Figure 9. The enrichment curves in Figure 9 do show some (but acceptable) variation over different runs. It is possible to further reduce this variation, by increasing (i) the number of Monte Carlo steps used to refit the structures onto the model and (ii) the number of times the refit is attempted. Enrichment rates are very high compared to random selection; the hit rates (% of known DHFR ligands) among the top-ranked 5% of the database is 100%, i.e. 100% of the ligands will be retrieved if only the top 5% of the compounds are selected.

Second, we used each of the five *dhfr5* ligands individually as queries to check if a query with multiple ligands enhances the screening enrichment compared to individual-ligand queries. Results are shown in Figure 10a. The enrichment graphs and the hit rates vary substantially if individual ligands are used as queries and are lower than the corresponding values for the multiligand-superposition query. This suggests that it is beneficial to use a superposition of multiple ligands rather than individual ligands as queries for screening. The enrichments are poorer and less inconsistent if the similarity toward individual ligands is assessed using simple 2D Tanimoto similarity measure based on Daylight fingerprints (Figure 10b) suggesting that more elaborate 2D similarity measures would be required.

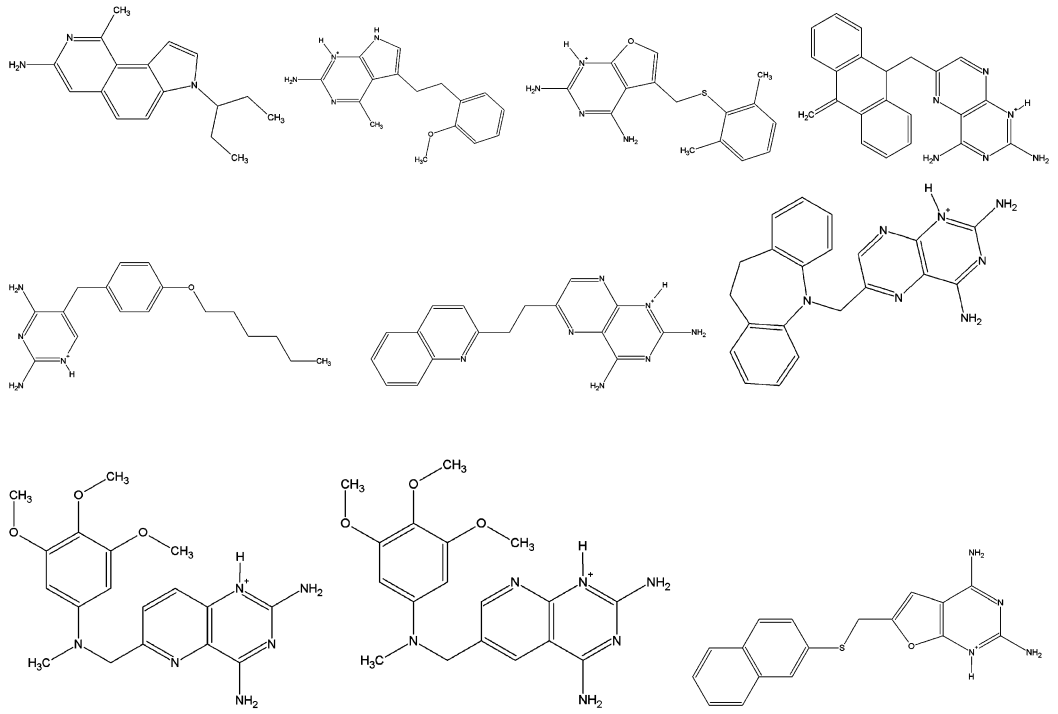

Figure 8. Ten DHFR ligands used for screening.

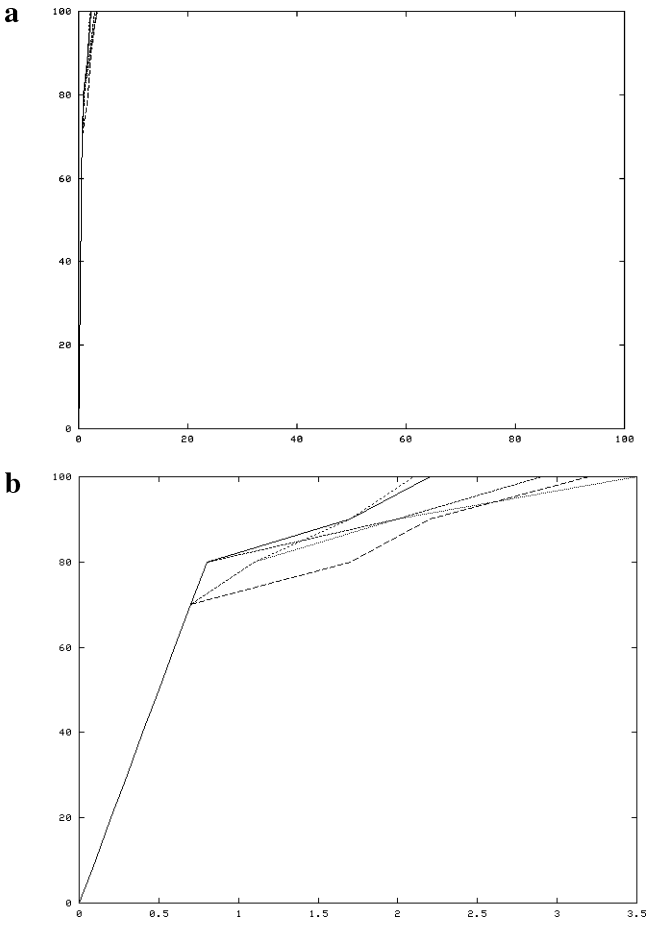

Figure 9. a. Enrichment plot of five screening DHFR experiments and b. focus on the 1−5% region.

Third, several superposition models were generated by the method and the enrichments obtained using those models were compared to the enrichment from the crystal superposition. This is the most important test of all since this will be the situation faced in a project with no available 3D protein information. Figure 11 shows the enrichment curves generated using as queries the three representatives of the clustered superpositions of the *dhfr5* set. The enrichment rates are also listed in Table 2. It appears that despite the differences in the geometry of the superposition, the queries result in similar enrichment ratios and hit rates at 5%. These are also similar (80%, 80%, and 90%, respectively) although slightly lower compared to the results achieved when the crystal superposition (100%) is used. Another useful metric is the percentage of the database required to retrieve 50% of the hits. As shown in Table 2, these values are all small and identical (0.5%) for the X-ray model and the three cluster representatives. These results suggest that the superposition algorithm captures the essential features responsible for the recognition of the receptor although the relationship between these features can be expressed in several different ways.

TK Ligands. The *bioinfo990* data set and 10 TK ligands (denoted as *tk10*) described by Bissantz et al.[37] were used in this screening experiment. This is directly compatible with the virtual screening by molecular docking experiment performed by Bissantz et al. Five TK ligands (assigned the name *tk5* and shown in Figure 12), different from the previous *tk10*, were extracted from the WDA database and used to generate an active site model.

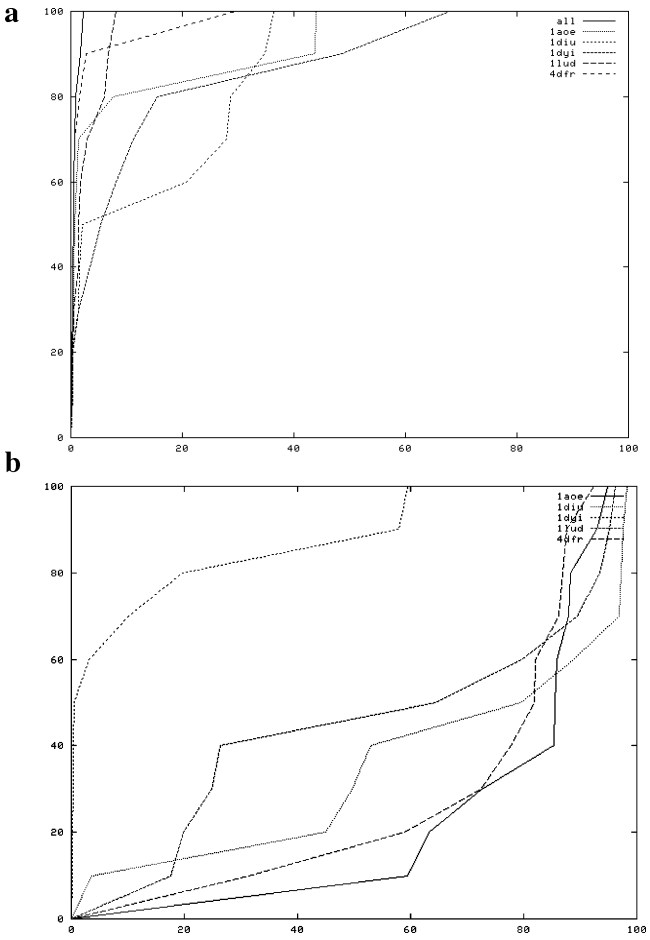

Figure 10. a. Enrichment plots of screening with individual ligands used as queries and b. enrichment plots of screening using Daylight fingerprints.

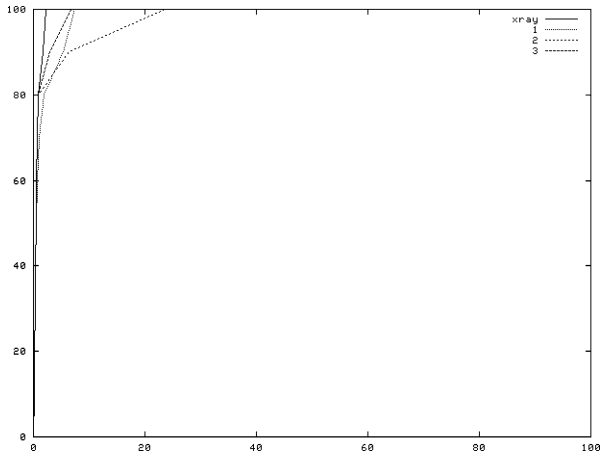

Figure 11. Enrichment plots for the three representative models for the *dhfr5* set generated by the superposition method.

Table 2. Results from Virtual Screening for DHFR

model	ranks of actives	hit rate at 5% (%)	%DB 50% hits
Dhfr X-ray	1−8, 11, 15	100	0.5
Dhfr 1	1, 5−9, 11, 25	80	0.5
Dhfr 2	1−8, 63, 238	80	0.5
Dhfr 3	1−7, 9, 30, 67	90	0.5

One hundred superposition runs were performed with the *tk5* data set, followed by the clustering procedure. Three representative models were identified (Figure 13). The

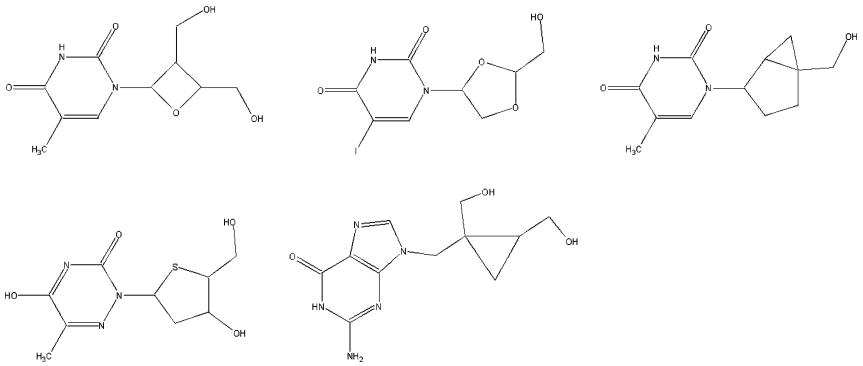

Figure 12. The structures of the five TK ligands used to generate site models (*tk*5 set).

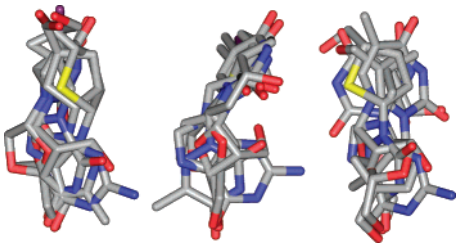

Figure 13. Representatives of the three solution clusters for superposition of the ligands in the *tk*5 set.

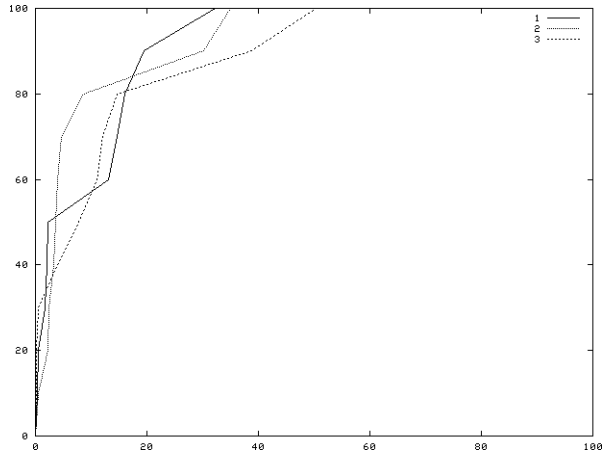

Figure 14. Enrichment plots for the three representative models for the *tk*5 set generated by the superposition method.

Table 3. Results from Virtual Screening for TK[a]

model	ranks	hit rate at 5%	%DB 50% hits
TK 1	4,5,18,21,22,131,146,161,195,322	50	2.2
TK 2	5,22,24,32,36,39,46,86,302,350	70	3.6
TK-3	1,2,5,38,78,110,119,146,388,503	40	7.8

[a] *Hit rate at 5%* specifies what percent of the actives are retrieved when 5% of the ranked compounds in the database are examined, and *%DB 50% hits* specifies what percent of the ranked database has to be examined to retrieve 50% of the active ligands.

selected representatives were used as queries to screen the database of 1000 compounds (*bioinfo990*+*tk10*). The enrichment graphs using the three models are shown in Figure 14, and enrichment rates are listed in Table 3. The values show some variation depending on which model is used; the hit rates at 5% were 50%, 70%, and 40% for the three models. This result compares very favorably to the best result (25%) obtained by Bissatz et al. (37) using the 3D protein structure with molecular docking and consensus scoring. There was also some variation in the percentage of the database (2.2, 3.6, and 7.8%) to retrieve 50% of the hits.

ERα Ligands. This enrichment experiment was conducted in a manner similar to the TK study above. The *Bioinfo990* set of ligands was spiked with 10 ERα ligands (set *er10*) used by Bissantz et al.[37] Another set of five ligands (denoted *er5* and shown in Figure 15) were extracted from the WDA (different from the *er10* set) and used to generate a receptor site model.

The same procedure as with the TK ligands was followed: 100 superposition runs were performed with the *er5* data set, followed by clustering. In this case, 11 representative models were identified (the two most dissimilar representatives are depicted in Figure 16) and used to screen a database composed of the *bioinfo990* and *er10* compound sets. The enrichment curves are shown in Figure 17, and enrichment rates are listed in Table 4. The percent of known ER compounds in the top-ranking 5% of the database was around 80% for all models (70% for two of the models, 80% for five of the models, and 90% for four of the models). This is comparable to the result obtained by Bissantz et al.[37] using molecular docking and consensus scoring. The percentage of the database to retrieve 50% of the hits was very low, less than 1%, for all the representatives.

Partial Similarity: Histamine H_3 Receptor Ligands. A pharmacophore model for the binding of antagonists to the H_3 receptor has been proposed on the basis of the superposition of 14 ligands.[25] The model reveals two hydrophobic pockets and several hydrogen-bond binding positions. The main feature of the model is the two hydrophobic pockets, the deduction of which is nontrivial since all ligands are roughly linear and have at most one hydrophobic group. Consideration of partial overlaps between ligands is essential for the elucidation of this type of pharmacophore. As already mentioned in the Introduction, a careful stepwise strategy had to be employed to derive this pharmacophore, and some truncation of the ligands was also found to be beneficial.[25] The pharmacophore was proposed based on analysis of pairwise superpositions between ligands 1, 2, and 3 (following the nomenclature of ref 25), and it was also suggested that a fourth ligand, number 9 (structures shown in Figure 18), would be helpful to constrain the pharmacophore. The rest of the ligands were refitted using pairwise superpositions with one of the ligands that already had an established alignment. In this work, we attempt a simultaneous superposition of H_3 ligands to generate a site model using our method, which incorporates full ligand flexibility and partial similarity.

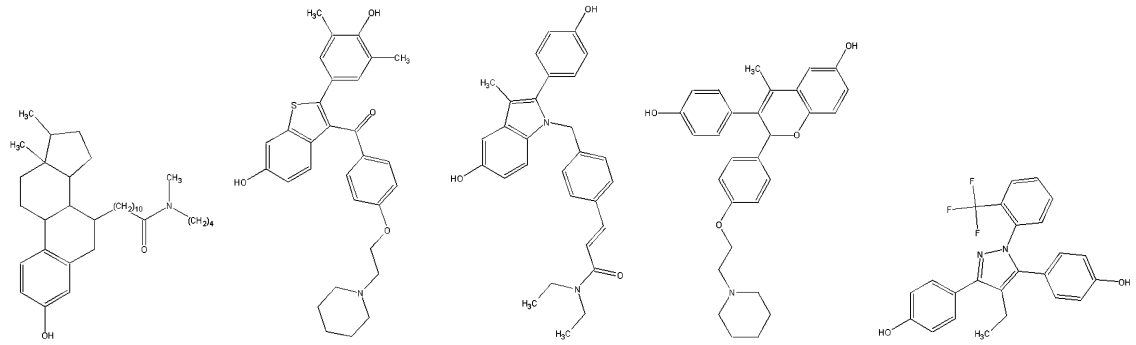

Figure 15. The structures of the five ER ligands used to generate site models (*er*5 set).

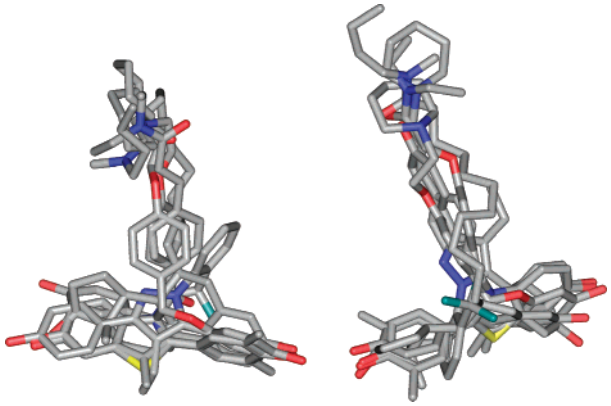

Figure 16. Representatives of the two most disssimilar solution clusters for superposition of the ligands in the *er*5 set.

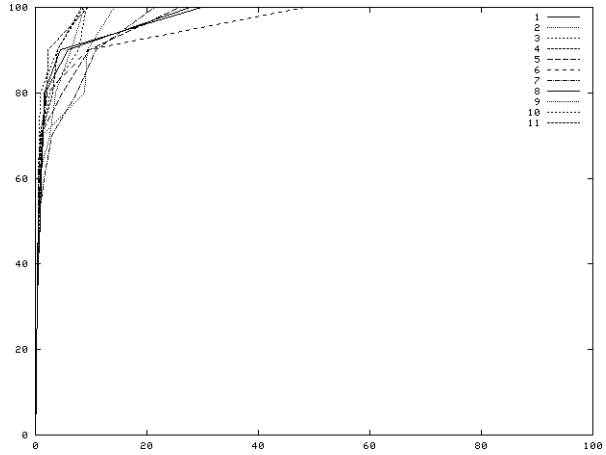

Figure 17. Enrichment plots for the eleven representative models for the *er5* set generated by the superposition method.

Superposition and Site Model. The four ligands, 1, 2, 3, and 9, were used to generate superpositions and site models in this work. The ligands were superimposed simultaneously. Site models were considered with between 18 and 27 steric points, 4 and 6 donor points, and 1 acceptor point. This is based on the minimum number of features in the site model determined from the highest number of points of that type in the four molecules (e.g., 18 steric features); the maximum number of features is by default set to 50% more features (e.g., 18 + 9 steric features). Each superposition was created using 20 000 Monte Carlo transitions.

One of the generated superpositions is shown in Figure 19. The corresponding site model aiding the superposition is shown in the same figure. The two-pocket structure of the superposition and the site model is clearly observable.

Table 4. Results from Virtual Screening for ER

model	ranks	hit rate at 5%	%DB 50% hits
ER 1	1−3,6,9,10,14,17,57,280	80	0.9
ER 2	1−6,25,36,62,86	80	0.5
ER-3	1−4,7,9,12,26,78,92	80	0.7
ER-4	1−6,10,31,40,83	90	0.5
ER-5	1−6,9,49,96,258	80	0.5
ER-6	1−4,6−7,12,20,96,482	80	0.6
ER-7	1−3,5−6,15,29,74,109,213	70	0.6
ER-8	1−5,9,14,16,44,298	90	0.5
ER-9	1−4,8−9,12,88,93,141	70	0.8
ER-10	1−7, 9, 40, 87	90	0.5
ER-11	1−6, 9, 21−22, 94	90	0.5

We note that if we restrict the number of points in the site model, superpositions and site models with one pocket are generated with a poorer fit of the ligands. Variations of basic two-pocket models were also generated from different runs. For example, the imidazole ring which is common to all ligands can be rotated with respect to the bond that connects the ring to the rest of the structure. Other variations, in different structural regions, are also observed with superposition scores similar to the score of the model shown in Figure 19. The implications of these variations were further explored.

Superposition and Site-Model Refinement. One of the H_3 ligands considered in ref 20 did not fit the proposed pharmacophore despite the significant similarities in structure

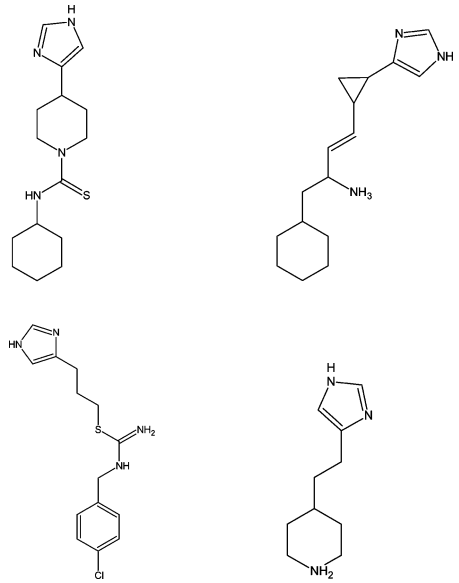

Figure 18. Structures of the four H_3 receptor ligands 1, 2, 3, and 9. CACTVS[41] was used to create the images.

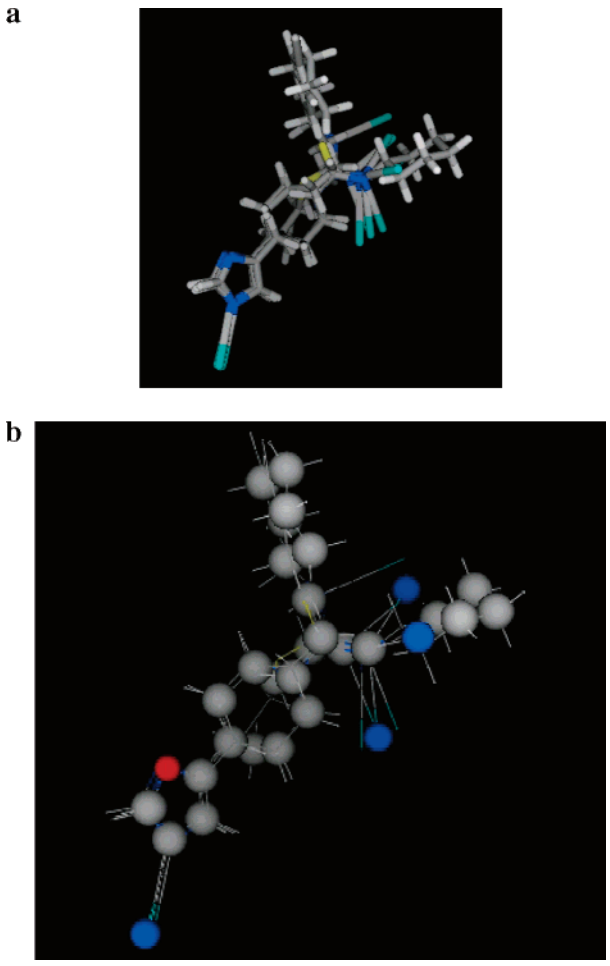

Figure 19. a. Superposition of four H_3 ligands. Hydrogen-bond donor projections are in cyan and are connected to the corresponding H atoms. b. The corresponding receptor site model.

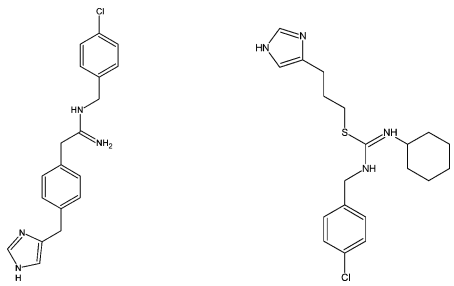

Figure 20. Structures of H_3 receptor ligands 13 and 24.

and activity with the other ligands in the set. Could a modification of the model explain the activities of all ligands? We investigated this question by adding two more ligands, namely ligands 13 and 24 (structures shown in Figure 20), to the set of four ligands used already. The addition of more ligands could also help in selecting one of the competing superpositions and receptor-site models. Ligand 13 is the ligand that did not fit to the suggested model, and ligand 24 is the ligand synthesized to prove the validity of the two-pocket model pharmacophore. The objective here is to identify alternative models that retain the two-pocket feature, which is considered important to explain structure−activity data[39,40] and to which all H_3 ligands can be successfully fitted.

All six ligands were superimposed simultaneously, starting from random positions, orientations, and conformations. A superposition model was obtained (Figure 21a,b) that suggests that it is possible to fit structure 13 in a manner that utilizes hydrogen-bonding groups common to the other ligands in the set and maintains the two-pocket model. There are three projected hydrogen-bond donor positions in the model (apart from the one on the imidazole ring). The

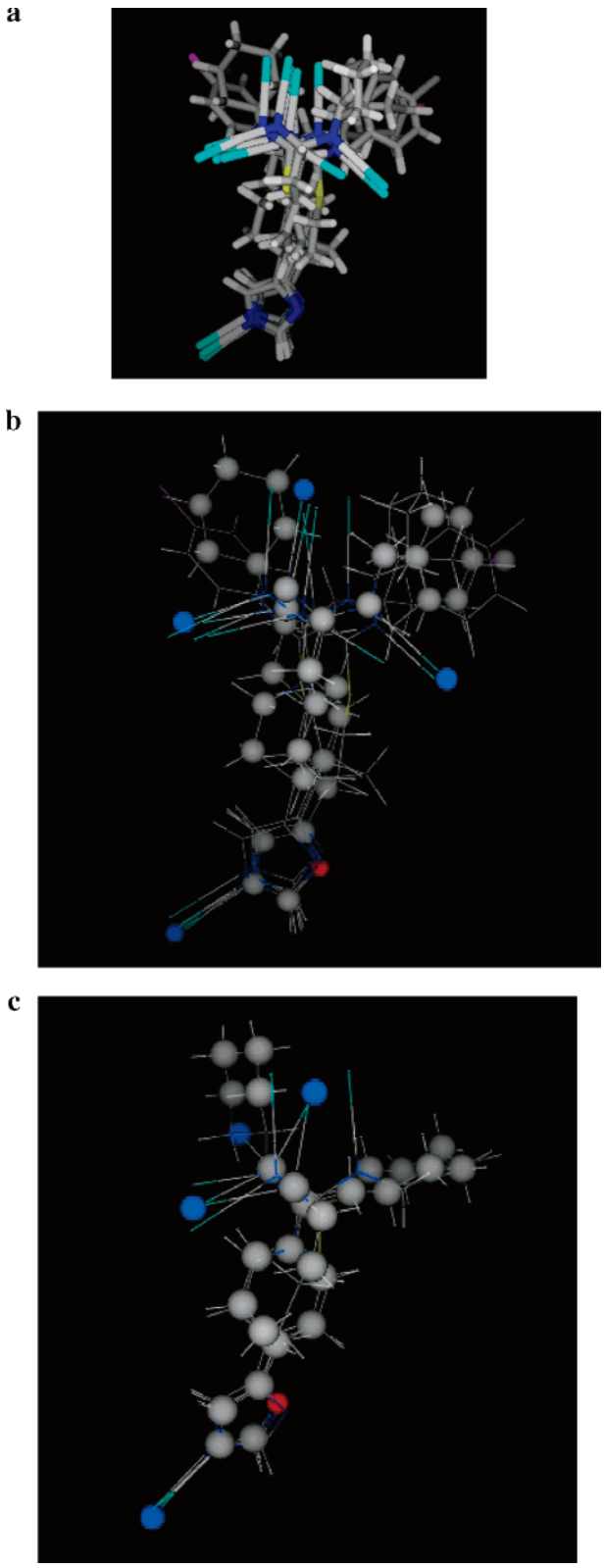

Figure 21. a. Superpositions of the six H_3 receptor ligands, b. the corresponding receptor-site model, and c. the receptor model based on four ligands (Figure 19b) in an orientation that facililtates comparison with the current model (part b).

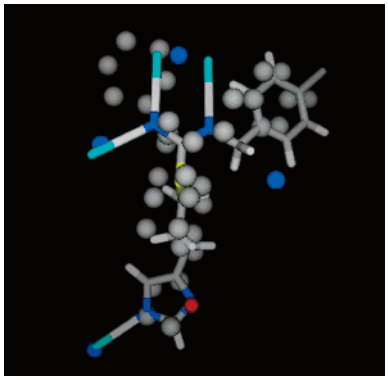

Figure 22. Ligand 3 and the receptor site model.

projections of the protonated amine group in ligand 2 match the projections of the amidino group of ligand 13, and the other ligands fit into that pattern by orienting the projections of the hydrogen-bond donor groups they contain toward one of the corresponding points in the receptor-site model. The basis for this superposition is illustrated in Figure 21a,b. The model based on four ligands is shown in a similar orientation in Figure 21c for comparison. The main difference is the position of one of the hydrogen-bond donor projection points which is roughly symmetrically positioned with respect to the axis defined by the other two projection points. This allows ligand 13 to fit the site model. One comment after examining the model is that for some of the ligands the projections from two different hydrogen-bond donor groups map onto the same donor point in the receptor-site model. An example is shown in Figure 22 where ligand 3 is shown with the receptor model of Figure 21b. This observation however applies to both the previously proposed and to the current modification of the model. It could be the case that in the actual receptor model there is a carboxyl rather than a carbonyl (for example) group complementary to the group in the ligand and that perhaps more donor points have to be included in the model.

Refitting Ligands onto the Site Mode. We used the modified site model (Figure 21b) as a basis to refit the rest of the ligands considered in ref 25. Some of the other ligands contain more hydrogen-bond acceptor groups which are not present in the ligands used to derive the model. In this procedure, we restricted the number of steric and hydrogen-bond donor points in the site model to the numbers already derived but allowed for more hydrogen-bond acceptor points to be added if required by the ligand to be superimposed. The fit of all remaining ligands was good. Most of the ligands gave only one superposition mode, judged by the closeness of the solutions generated in several runs. Other ligands, generally the more flexible ones, suggested several alternative fits to the model. The results confirmed that the modified model is able to successfully accommodate all ligands.

DISCUSSION

This work describes a novel method for superposing ligand structures, deriving receptor site models and virtual screening that uses stochastic tunneling optimization. The new method addresses several major problems in QSAR studies, namely how to align ligands exhibiting only partial similarities with each other, defining and exploring similarity among more than two ligands, and conformational flexibility. The proposed solution relies on the definition of an auxiliary receptor site model with respect to which the superposition is performed. The site model is defined as a set of points of different descriptor types taken from different ligands. The receptor site model is dynamic and coevolves with the ligand superposition. Ligand similarity is defined with respect to the receptor site model serving as a common reference frame rather than between the ligands themselves. The superposition models could also be scored according to the total volume occupied by the ligand atoms, weighted by feature type, and clustered. Once a superposition is derived, it can be used to screen a compound database in the search for novel ligands for a particular target protein.

The utility of the method was demonstrated on three data sets: DHFR, TK, and ER ligands. Two measures were used to evaluate objectively the performance of the methodology. First, as in the case of DHFR, where 3D crystal structures are available of complexes of several different ligands binding to the same protein, a crystal superposition for the ligands could be derived (based on protein backbone and active site atoms) and compared to the superposition generated independently of the protein structure. Second, and more important, since the main application of the generated ligand superpositions and active site models would be in high throughput screening of compound databases, enrichment curves and hit rates were calculated for a benchmark database composed of a mixture of random compounds and known active ligands. The results can subsequently be compared to other virtual screening methods. We note the high enrichment ratios achieved by the superposition method even in comparison with procedures that exploit the protein crystal structure. This makes the methodology particularly valuable in cases where no 3D protein information is available and molecular docking methods cannot be used.

The method was applied to the simultaneous superposition of up to ten MTX ligands. The superpositions resulted in very low rmsd values between the different copies. The matching of two MTX fragments with MTX itself highlights the importance of integrating partial similarity in ligand alignment protocols. In this case, the three molecules do not have corresponding atoms in common, and complementary molecular superposition methods, lacking the automatic inclusion of partial similarity, would tend to stack the molecules on top of each other. Partial similarity allows for the correct overlap of the fragments with the appropriate section of the whole MTX molecule.

The superposition of the five DHFR ligands exemplified the influence of ligand flexibility and diversity. In this case, three clustered sets of diverse superposition models were generated with similar scores, including a solution that resembles the known crystal superposition. In the absence of corresponding crystal structures, there is no objective approach for suitably differentiating between these solutions, and without additional information all of them should be considered as valid models.

In cases where the ligand crystal structures are known, reproduction of these structures by superposition of the ligands is nontrivial when ligand flexibility is invoked, and many alternative solutions with similar scores are generated. If the ligands are held in their crystal structure configurations, then a solution close to the crystal superposition will be obtained. However, in the absence of receptor data, and only

information concerning ligand activity, then the active ligand conformation is unknown and flexibility must be invoked. Increasing the number of diverse ligands helps in the identification of the best receptor site model.

In the absence of receptor data, the best scoring receptor-site models from the clustered solutions were used for the virtual screening of molecular databases. These models showed significant geometrical diversity and, therefore, may be different to the actual receptor site. This is a consequence of the ability of the superposition method to identify important ligand binding features, even though their orientation in space may not be ideal. One approach to identify and exploit the receptor model that is most likely to produce the highest enrichment is to screen a minidatabase against all selected models and choose the one that yields the highest enrichment for use in screening the actual database. This procedure is frequently employed when molecular docking is utilized as the virtual screening method.

The method proposed in this work extends previous studies and should be of general utility in drug design projects. One of the main results from this study to note is that ligand-based virtual screening without utilization of a 3D receptor structure could in favorable cases produce enrichment results comparable to or better than molecular docking.

ACKNOWLEDGMENT

We thank Nicholas Perry and David Lloyd for fruitful discussions and ideas and Anthony Williams for making available the software for 2D similarity calculations.

REFERENCES AND NOTES

(1) Debnath, A. S. Pharmacophore Mapping of a Series of 2,4-Diamino-5-deazapteridine Inhibitors of Mycobacterium avium Complex Dihydrofolate Reductase. *J. Med. Chem.* **2002**, *45*, 41−53.

(2) Boström, J.; Gundertofte, K.; Liljefors, T. A Pharmacophore Model for Dopamine D4 Receptor Antagonists. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 769−786.

(3) Bernard, D.; Coop, A.; MacKerrel, A. D., Jr. 2D Conformationally Sampled Pharmacophore: A Ligand-Based Pharmacophore to Differentiate δ Opioid Agonists from Antagonists. *J. Am. Chem. Soc*. **2003**, *125*, 3010−3107.

(4) Smith, P. A.; Sorich, M. J.; McKinnon, R. A.; Miners, J. O. Pharmacophore and Quantitative Structure-Activity Relationship Modeling: Complementary Approaches for the Rationalization and Prediction of UDP-Glucuronosyltransferase 1A4 Substrate Selectivity. *J. Med. Chem.* **2003**, *46*, 1617−1626.

(5) Hahn, M.; Rogers, D. Receptor Surface Models. 2. Application to Quantitative Structure-Activity Relationship Studies. *J. Med. Chem.* **1995**, *38*, 2091−2102.

(6) *Pharmacophore Perception, Development, and Use in Drug Design;* Guner, O. F., Ed.; International University Line: La Jolla, CA, 2000.

(7) *7*. Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi. S. Grid-Independent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3243.

(8) Stiefl, N.; Baumann, K. Mapping Property Distributions of Molecular Surfaces: Algorithm and Evaluation of a Novel 3D Quantitative Structure-Activity Relationship Technique. *J. Med. Chem.* **2003**, *46*, 1390−1407.

(9) Lemmen, C.; Lengauer, T. Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215−232.

(10) Lemmen, C.; Lengauer, T.; Klebe, G. FlexS: A Method for Fast Flexible Ligand Superposition. *J. Med. Chem.* **1998**, *41*, 4502−4520.

(11) Pitman, M. C. FLASHFLOOD: A 3D Field-based similarity search and alignment method for flexible molecules. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 587−612.

(12) Feher, M.; Schmidt, J. M. Multiple flexible alignment with SEAL: a study of molecules acting on the colchicines binding site. *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 495−502.

(13) Miller, M. D.; Sheridan, R. P.; Kearsley, S. K. SQ: a program for rapidly producing pharmacophorically relevant molecular superpositions. *J. Med. Chem.* **1999**, *42*, 1505−1514.

(14) Perkins, T. D. J.; Dean, P. M. An exploration of a novel strategy for superposing several flexible molecules. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 155−172.

(15) Cosgrove, D. A.; Bayada, D. M.; Johnson, A. P. A novel method of aligning molecules by local surface shape similarity. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 573−591.

(16) Mestres, J.; Rohrer, C. D.; Maggiora, G. M. A molecular-field-based similarity study of non-nucleoside HIV-1 reverse transcriptase inhibitors. 2. The relationship between alignment solutions obtained from conformationally rigid and flexible matching. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 39−51.

(17) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83−102.

(18) Jones, G.; Willett, P.; Glen, R. C. A Genetic Algorithm for Flexible Molecular Overlay and Pharmacophore Elucidation. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532−549.

(19) Greene, J.; Kahn, S.; Savoj, H.; Sprague, P.; Teig, S. Chemical Function Queries for 3D Database Search. *J. Med. Chem.* **1994**, *37*, 1297−1308.

(20) Jain, A. N.; Koile, K.; Chapman, D. Compass: Predicting Biological Activities from Molecular Surface Properties. Performance Comparisons on a Steroid Benchmark. *J. Med. Chem*. **1994**, *37*, 2315−2327.

(21) Barakat, M. T.; Dean, P. M. Molecular Structure Matching by Simulated Annealing. III The Incorporation of Null Correspondences into the Matching Problem. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 107−117.

(22) Mills, J. E. J.; de Esch, I. J. P.; Perkins, T. D. J.; Dean, P. M. SLATE: A method for the superposition of flexible ligands. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 81−96.

(23) Perkins, T. D. J.; Mills, J. E. J.; Dean, P. M. Molecular surface-volume and property matching to superimpose flexible dissimilar molecules. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 479−490.

(24) Perkins, T. D. J.; Mills, J. E. J.; Dean, P. M. Molecular partial similarity using surface-volume comparisons. In *Computer-assisted lead finding and optimization*; van de Waterbeemd, H., Testa, B., Folkers, G., Eds.; VCH: Weinheim, 1997; pp 421−432.

(25) De Esch, I. J. P.; Mills, J. E. J.; Perkins, T. D. J.; Romeo, G.; Hoffmann, M.; Wieland, K.; Leurs, R.; Menge, W. M. B.; Nederkoorn, P. H. J.; Dean, P. M.; Timmerman, H. Development of a Pharmacophore Model for Histamine H3 Receptor Antagonists, Using the Newly Developed Molecular Modeling Program SLATE. *J. Med. Chem.* **2001**, *44*, 1666−1674.

(26) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.

(27) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G.; A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(28) Welch, W.; Ruppert, J.; Jain, A. N.; Hammerhead: fast, fully flexible automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **1996**, *3*, 449−462.

(29) Kabsch, W.; A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. B: Struct. Sci.* **1976**, *A32*, 922−923.

(30) Wenzel, W.; Hamacher, K. A stochastic tunneling approach to global optimization. *Phys. Rev. Lett*. **1999**, *82*, 3003−3007.

(31) Kirkpatrick, S.; Gellat C. D.; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983**, *220*, 671−680.

(32) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. Crystal structures of Escherichia coli and Lactobacillus casei dihydrofolate reductase refined at 1.7 Å resolution. I. General features of binding of methtrexate. *J. Biol. Chem*. **1982**, *257*, 13650−13562.

(33) Kuyper, L. F.; Baccanari, D. P.; Jones, M. L.; Hunter, R. N.; Tansik, R. L.; Joyner, S. S.; Boytos, C. M.; Rudolph, S. K.; Knick, V.; Wilson, H. R.; Caddell, J. M.; Friedman, H. S.; Comley, J. C.; Stables, J. N. High-Affinity Inhibitors of Dihydrofolate Reductase: Antimicrobial and Anticancer Activities of 7,8-dialkyl-1,3-diaminopyrrolo[3,2-F]-quinazolones with Small Molecular Size. *J. Med. Chem*. **1996**, *39*, 892−903.

(34) Martorell, G.; Gradwell, M. J.; Birdsall, B.; Bauer, C. J.; Frenkiel, T. A.; Cheung, H. T. A.; Polshakov, V. I.; Kuyper, L.; Feeney, J. Solution structure of bound trimethoprim in its complex with Lactobacillus casei dihydrofolate reductase. *Biochemistry* **1994**, *33*, 12416−12426.

(35) Reyes, V. M.; Sawaya, K. A.; Brown, K. A.; 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.

(36) Polshakov, V. I.; Bierkofski, R. B.; Birdsall, B.; Feeney, J. Towards understanding the origins of the different specificities of binding the

reduced (NADPH) and oxidized (NADP+) forms of nicotinamide adenine dinucleotide phosphate coenzyme to dihydrofolate reductase. *J. Mol. Struct.* **2002**, *602*, 257−267.
(37) Bissantz, C.; Folkers, G.; Rognan, D. Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.
(38) World Drug Alert, Thompson/Derwent. http://thomsonderwent.com/products/pca/worlddrugalerts/ (accessed Jan 15, 2007).
(39) Leurs, R.; Vollinga, R. C.; Timmerman, H. The medicinal chemistry and therapeutic potentials of ligands of the histamine H3-receptor. *Prog. Drug Res.* **1995**, *45*, 107−165.
(40) Lipp, R.; Stark, H.; Schunack, W. Pharmacochemistry of H3 Receptors: The Histamine Receptor. In *Receptor Biochemistry and Methodology*; Schwartz, J.-C., Haas, H. L., Eds.; Wiley-Liss Inc.: New York, 1992; Vol. 16, pp 57−72.
(41) Ihlenfeldt, W. D.; Takahashi, Y.; Abe, H.; Sasaki, S. Computation and Management of Chemical Properties in CACTVS: An Extensible Networked Approach toward Modularity and Compatibility. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 109−116.

CI6003338