ELSEVIER

A method for including protein flexibility in protein-ligand docking: Improving tools for database mining and virtual screening

Howard B. Broughton

Merck, Sharp & Dohme Neuroscience Research Centre, Essex, United Kingdom

Second-generation methods for docking ligands into their biological receptors, such as FLOG, provide for flexibility of the ligand but not of the receptor. Molecular dynamics based methods, such as free energy perturbation, account for flexibility, solvent effects, etc., but are very time consuming. We combined the use of statistical analysis of conformational samples from short-run protein molecular dynamics with grid-based docking protocols and demonstrated improved performance in two test cases. Our statistical analysis explores the importance of the average strength of a potential interaction with the biological target and optionally applies a weighting depending on the variability in the strength of the interaction seen during dynamics simulation. Using these methods, we improved the number of known dihydrofolate reductase ligands found in the top-ranked 10% of a database of drug-like molecules, in searches based on the three-dimensional structure of the protein. These methods are able to match the ability of manual docking to assess likely inactivity on steric grounds and indeed to rank order ligands from a homologous series of cyclooxygenase-2 inhibitors with good correlation to their true activity. Furthermore, these methods reduce the need for human intervention in setting up molecular docking experiments. © *2000 by Elsevier Science Inc.*

Keywords: molecular dynamics, docking, flexible, cyclooxygenase, prostaglandin H2 synthase, dihydrofolate reductase, database search, rank order

Color plates for this article are on pages 302–304.

Corresponding author: H.B. Broughton, Merck, Sharp & Dohme Neuroscience Research Centre, Terlings Park, Eastwick Road, Harlow, Essex CM20 2QR, United Kingdom. Tel.: 01279 440000.

E-mail address: hbro@merck.com (H.B. Broughton)

INTRODUCTION

Prediction of the binding geometry and energy (and hence affinity) of ligands at their biological targets remains a very difficult problem, but one of great interest in the design of novel medicines. In the modeling of protein-ligand interactions, it is normal practice to make drastic, simplifying assumptions: the protein usually is treated as rigid; side-chain orientations, tautomeric forms, and protonation states often are selected based on a subjective view of likely hydrogen bonding patterns; any crystallographically observed water molecules may or may not be included (or may be supplemented by programs that add solvent to the molecule) on a largely arbitrary basis; and inorganic counterions generally are ignored. Despite such simplifications, the methods remain useful in areas such as database mining (virtual or real), combinatorial library design, and lead optimzsation, leading to considerable recent interest in the field.[1–3] Programs such as DOCK,[4] FlexX,[5] GOLD,[6] and AutoDOCK[7] have become regularly used tools for molecular modelers.

Although the "holy grail" of molecular docking is to reproduce the experimentally observed "pose" of the ligand in the binding site, our objective in this study was oriented more directly to the problems of drug discovery: we wanted to establish whether the inclusion of protein flexibility in a docking protocol would improve its utility as a "virtual screen" either in a database mining context or for the prediction of rank-order activity in a homologous series. There is a clear advantage in finding lead compounds in the first few thousand compounds screened from typical corporate sample collections comprising a few hundred thousand to a few million compounds. Combinatorial libraries are best designed to explore structure-activity relationships (SAR) fully in interesting areas, not to rediscover conclusions that would have been obvious from consideration of the steric capacity of the protein target site. Hence, there is a demand for a rapid method to assess automatically whether a candidate library member is likely to

Journal of Molecular Graphics and Modelling 18, 247–257, 2000

655 Avenue of the Americas, New York, NY 10010

1093-3263/00/$–see front matter
PII S1093-3263(00)00036-X

be active or not. In lead optimization, methods that can suggest where a particular drug candidate may interact with other biological targets than the intended one will enable efficient use of those screening resources that are available. Models based on such methods can even help to avoid errors in the definition of the generic scope of a chemical invention, by ensuring that the compounds covered in a patent, beyond the experimentally tested examples, are likely to be active against that particular biological target. Manual docking of ligands into proteins often is sufficient for these purposes, and human intervention can avoid some pitfalls of automated methods. However, automation is necessary for those applications where a more objective assessment is required or where large numbers of molecules are to be examined.

The FLOG method for automatically docking multiple diverse conformations of candidate ligands into a site on a protein surface is typical of "second-generation" docking programs in that it allows the ligand to be flexible while making simplifying assumptions about the protein.[8] Recent literature has described improvements in methods that allow ligands to be flexible,[9,10] but more significantly has begun to describe methods that provide for some flexibility of the protein, for example, by considering ensembles of experimentally determined structures[11] or by allowing for some limited side-chain flexibility.[12,13] We believed that it would be valuable to try to incorporate protein flexibility and water mobility into the FLOG method in the hopes that this would improve FLOG's ability to select active compounds from databases This view appears to be supported by a recent publication on the effect of induced fit with at least one docking program.[14] The objective of improved selection from a database would be met if a modified FLOG were able to find more and better actives faster (i.e., would rank the compounds higher following a database search) than FLOG based solely on the crystal structure. Because many of the potential applications of docking in medicinal chemistry are oriented toward work within a homologous series of compounds, a modified FLOG also would ideally surpass the original version in its ability to distinguish inactive, weakly active, and active ligands in such series. It is possible to make a crude, somewhat subjective assessment of quality of fit manually, by docking each compound into the target site and seeing whether there are sufficient unfavorable contacts with the protein to suggest that it is highly likely that the molecule will be inactive, and we sought to reproduce this ability to distinguish potentially active from inactive compounds based on steric fit using FLOG. We also hoped that the enhancements would be sufficient to lead to an improved rank correlation between the FLOG score and the true activity rank of compounds in a homologous series.

In order to implement this in practice, we decided to use statistically weighted descriptors of the target active site in single FLOG runs, rather than trying to solve the combinatorial problem of docking multiple conformations of multiple molecules into multiple protein conformations. Although this approach has some similarities with literature methods using ensembles of experimentally determined protein conformations,[11] we chose to use molecular dynamics simulation as the source of conformational samples. This is an important choice because only *one* crystal structure is available for many proteins of pharmacological interest, and because it ensures a wider exploration of the accessible conformational space of the protein than might be expected for the crystallographic case, where many solid-state constraints may exist. We also used a novel method of producing an "average" model of the active site into which docking was to be performed, based on statistically weighting the mean potential protein-ligand interaction energies computed from the molecular dynamics runs. Molecular dynamics simulations (of much greater length and complexity) also are the method of choice in free energy perturbation methods of calculating the interaction energy between ligand and receptor.[2] FLOG has been described in detail elsewhere,[8] but the methods used here could be adapted to work with other grid-based molecule docking programs.

The test beds chosen for each of these objectives were designed to be representative of the various levels of structural data available in typical cases. Thus, for the database mining problem, the well-studied and crystallographically refined case of *Lactobacillus casei* dihydrofolate reductase/methotrexate complex[15] (DHFR/MTX, PDB[16] code 3DFR, 1.7 Å structure) was used. This enables direct comparison with an extensive literature, based on a very good crystal structure complete with water molecules. The database used was the MINDEX12 flexibase,[17] based on the 12th edition of the *Merck Index*,[18] which contains about 7000 drug-like molecules including known DHFR ligands and substrates. For the rank ordering of a series of candidate molecules to a protein target, the murine cyclooxygenase-2 (COX-2, prostaglandin H2 synthase) structure[19] (PDB code 1CX2, resolution 3.0 Å) was used in conjunction with a range of structures mostly of the *vic*-diaryl heterocyclic class. This protein was chosen for its medicinal importance, its membership in the class of membrane-associated proteins, and because its structure has been solved at a poorer level of resolution than was the case for 3DFR. COX-2 also is known to have limited flexibility in one small part (a section of helix D, near the CF_3 group of SC-558 in the 1CX2 crystal structure), whereas the remainder of the active site is quite rigid and unaffected by ligand binding.[20] These complexities test more fully the ability of the modified FLOG methods to work with structural data more typical of real-world drug discovery. Furthermore, it is known that the COX-2 inhibitory activity of members of the *vic*-diaryl heterocyclic class of ligands is exquisitely sensitive to structural variation in some parts of the molecule while being tolerant of considerable changes elsewhere.[21---26] Hence, a reliable method for predicting rank-order binding for these ligands should be able to distinguish to some degree whether or not the effect of changes of each of these kinds would be expected to have an effect on the enzyme inhibitory properties of the molecule.

METHODS

Initial Structure Preparation

The crystal structures 3DFR and 1CX2 were retrieved from the Brookhaven database using the WWW interface and read into the Sybyl molecular modeling package.[27] In the case of 1CX2, a single protein chain, with associated ligand and haem, was obtained by deletion of all atoms in chains B, C, and D. Atom types and charges were assigned for 3DFR using the Tripos implementation of the original Kollman AMBER united-atom set,[28] with atoms in the ligands being set by hand to the nearest equivalent type. Atom types and charges were assigned for 1CX2 using the Tripos implementation of all-atom AMBER95,[29] with supplementary parameters and the charges

for the haem based on those provided on the AMBER website.[30] The water model used for 3DFR was TIP3P, using water molecules produced by adding hydrogen atoms to the oxygen atoms whose coordinates were present in the PDB entry using Sybyl's "fillvalence" function. No water was present in the PDB entry, and none was added to the model of COX-2. Ligand and cofactor charges were obtained by calculating MOPAC[31] 6 AM1 ESP-fit charges on the crystallographically observed geometry of MTX and NADPH (NDP) and MNDO ESP-fit charges on the optimized geometry for the COX-2 ligand SC-558, after addition of hydrogens in idealized positions where necessary. N-acetylglucosamine was used as a model system for the NAG groups on COX-2, and the charges were scaled after attachment to the protein to maintain overall neutrality.

DHFR Model: Refinement and Simulation

The following sequence, which was chosen for its ability to adequately handle the relaxation of poorly oriented water molecules, was used:

1. Protein, MTX, and cofactor (NDP) frozen
2. Molecular dynamics simulation, 1 fs time step, 10 fs nonbonded reset with velocity scaling
   - 100 fs at 50K
   - 100 fs at 150K
   - 100 fs at 200K
3. 1000 fs without velocity scaling at 300K
4. Energy minimization for 100 steps of Powell algorithm
5. Side chains, NDP, and MTX "unfrozen," protein backbone held rigid
6. Energy minimization for 1000 steps of Powell algorithm
7. All constraints removed and minimization for another 1000 steps of Powell to a gradient <0.5.

After this process of annealment, the final structure showed a C_α fit of 0.55 Å root-mean-square (RMS) and maximum C_α displacement of 1.34 Å, with respect to 3DFR.

Dynamics simulation of DHFR followed a heating regime in which simulations of 100 fs were carried out at temperatures of 50K, 150K, and 250K, followed by a 700 fs interval at 300K, all with velocity scaling and nonbonded update every 10 fs. Velocity scaling then was turned off and simulation continued at 300K for another 50 ps with a structure written to disk every 1 ps.

COX-2 Model: Refinement and Simulation

The COX-2 model was partially refined by energy minimization, initially 21,000 iterations using the Powell algorithm with a nonbonded cutoff of 8 Å and a dielectric function of 1/4r, followed by a final optimization of 200 iterations with a nonbonded cutoff of 12 Å. The Sybyl "anneal" function was used to set up "hot," "warm," and "cold" regions of the protein, these being:

- Hot: regions in which the atoms are included in the energy calculation and can move in minimizations or dynamics simulations
- Warm: regions in which interactions are considered for energy calculation but the atoms are held rigid
- Cold: regions in which the atoms are held rigid and are not considered in evaluating the energy of the system.

Initially, an "Active_Site" region was defined as any residue with at least one atom within 8 Å of the ligand SC-558. The "hot" region then was defined as all those residues with at least one atom within an 8 Å radius of any atom of any Active−Site residue, the "warm" region included residues that had at least one atom within 12 Å of any Active_Site residue but which were not in the "hot" region, and the remaining atoms were in the "cold" region. Minimization of the structure with these zones defined proceeded to convergence (gradient <0.05), providing the base structure for the dynamics simulation. The structure obtained in this way is shown in Color Plate 1, superimposed on the C_α atoms of the A chain of the homodimer formed by the A and B chains in the crystal structure, with the annealment regions shown in different colors.

The dynamics simulation on COX-2 started at 10K and proceded for 1000 fs with a 1 fs time step and a nonbonded update with velocity scaling and momentum removal every 10 fs. The temperature was incremented 10K and the simulation continued under otherwise identical conditions another 1000 fs. This cycle was repeated until a temperature of 300K was reached. Velocity scaling then was turned off and the simulation continued for another 76 ps, with the structure being written every 500 fs. After this dynamics simulation, the RMS fit of the C_α atoms to those of the crystal structure was 0.82 Å.

Generation of the FLOG Grids and Match Points for Both Models

The last 40 structures recorded for the DHFR case and the last 150 recorded from the COX-2 case were used for the next step. The DHFR structures were aligned on the inhibitor MTX and the residues and water molecules with atoms within 8 Å of the ligand, but excluding the ligand itself, were extracted for the generation of FLOG grids. The COX-2 structures were aligned on the inhibitor SC-558 and residues with atoms within 10 Å of the ligand were similarly extracted.

Although FLOG has been described elsewhere, a brief summary of the steps involved is included here:

1. Assign atom types to the protein. Normally, the side-chain orientations, decision to keep or exclude water molecules etc., also are done at this stage.
2. Set up a three-dimensional lattice of equally spaced points separated by a distance d ($d = 0.3$ was used for these experiments except as noted below).
3. At each point in the lattice, calculate the interaction energy with the protein for each of five (hydrogen-bond acceptor, hydrogen-bond donor, polar, hydrophobic and "other" [vander-Waals interactions only]) or seven (the above plus cation and anion) prototypical atom types to give five or seven "grids" (respectively).
4. Find the maxima (in FLOG, more positive scores represent more favorable interactions) in each grid. Place a "match point" at each maximum subject to the condition that the maximum must represent an interaction energy of >1.0 FLOG score unit and that match points must not be closer

than 1.2 Å. Each match point acquires the type of the grid in which it represents a maximum.

5. For each conformation of each candidate ligand, find orientations such that atom-atom distances in the ligand approximately correspond to match point-match point distances (clique detection method as used in SQ[32]) between atoms and match points of corresponding types.
6. Score each orientation of each conformation of each molecule by summing the interaction energy for each atom of the ligand, reading the value from the grid of the type corresponding to the property-type of the atom (H-bond donor, acceptor, etc.).
7. Optionally apply some optimization, including simplex rigid-body translation and rotation, to the candidate orientation, and calculate final best score.

Atom type assignment was carried out on each sample conformation from the dynamics runs according to the standard FLOG procedure. For these experiments, the amino-acid side-chain orientations and water positions were not modified, with no human intervention. Generation of the interaction energy grids used by FLOG was done in the standard, fully automated way, producing five three-dimensional grids for each sample from the molecular dynamics run. For comparison purposes, the same techniques also were applied to the original crystal structures without further refinement except that in the case of DHFR, for this "control" experiment only, the water molecules were all removed from the crystal structure before grid generation, for reasons explained later.

The grids obtained in this way from the structures sampled from molecular dynamics were subjected to statistical analysis. Each grid point can be specified by four indices—i, j, and k specifying the offsets from the grid origin in each dimension, and l specifying the type of the grid (donor, acceptor, etc.)—and each was considered in turn and statistics computed across all conformers of the protein sampled by dynamics. The values computed for each grid point were the minimum value, maximum value, mean value, standard deviation, and a weighted mean. The latter was calculated for a grid point W_{ijkl}, based on the mean (M_{ijkl}) and standard deviation (S_{ijkl}) observed at that point, thus:

If

$$M_{ijkl} = 0,$$

$$W_{ijkl} = 0$$

Else

$$W_{ijkl} = M_{ijkl} * \exp - (S_{ijkl}^2/M_{ijkl}^2)$$

This weighted grid emphasizes areas where the "interaction energy" varies little for a particular atom type between the conformers sampled, giving a good score for those cases where there is "reliably" a good interaction for the candidate atom type across all conformers sampled. It also down-weights the effect of occasional occupation of a region of space by a part of the protein or a water molecule, because, as has been observed by others[11], the magnitude of repulsive terms can otherwise unrealistically dominate the "average" interaction energy.

The grid points calculated in this way were written to grid files with the same dimensionality as the original input grids. Color Plate 2c shows the difference between the simple mean van-der-Waals grid and the corresponding weighted mean grid (both contoured at a level corresponding to a very slightly unfavorable interaction, −0.1) for the DHFR case.

The statistically derived grids were used for comparison with those grids directly generated from the crystal structures, all the remaining steps in the FLOG process being those used as standard, with a relative weighting for van-der-Waals and electrostatic effects of 0.3 and 1.0, respectively.

Match points for the clique-detection step of FLOG were generated as normal from these grids. FLOG normally automatically chooses the match point with the largest interaction energy as an "essential" match point, requiring it to be present in every clique. For this study, the "essential" match point was chosen manually to ensure that a comparable point was used in each run; the point type selected was "+"[32] (i.e., any ligand atom could be matched to the point). For the DHFR case, the point was chosen (as previously[8]) to lie close to the binding site of the 2-amino group of the pteridine of MTX. For the COX-2 case, the match point with highest interaction energy that lay within the part of the binding pocket occupied by both selective and nonselective cyclooxygenase inhibitors was used in order to allow comparison with nonselective compounds such as flurbiprofen.

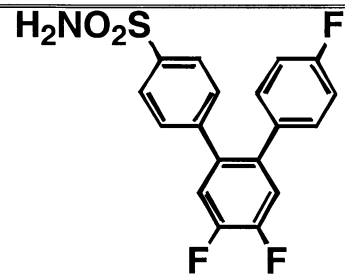
Databases and Searches

FLOG was used to orient conformations of molecules from appropriate databases on the match points, followed by scoring using the corresponding grids. For COX-2, a selection of 28 ligands selected from the literature or evaluated against human COX-2 in house were chosen, the majority from the *vic*-diaryl heteroaryl class. A flexibase[17] (COXBASE) was constructed containing 250 diverse conformations of each of these and was searched using FLOG. The structures of molecules in COXBASE are given in Figure 1. The choice of compounds in COXBASE was made to include compounds that would clearly be expected to be inactive, based on their massively increased size over SC-558 and the evidently limited space available in the enzyme, compounds that contained more subtle effects such as halogen substitution, and compounds where conformational effects due to juxtaposition of substituents could be expected. Compounds also were chosen to reflect the wide variety of core structures from the literature. Thus, COXBASE contains compounds with large changes in structure that are expected to remove activity and compounds with more subtle changes that may vary inhibitory activity up or down with respect to SC-558. The nature of the bioassays used makes rank order of potency rather than explicit IC_{50} values the most appropriate measure of activity for these compounds.

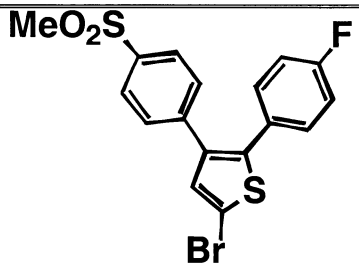
RESULTS AND DISCUSSION

Dynamics Studies

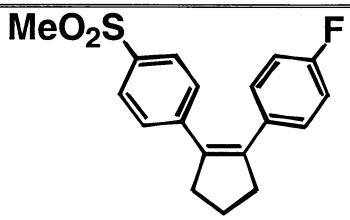
The initial dynamics simulations were routine, short simulations. This is in some contrast to alternate methods of predicting binding affinity such as free energy perturbation methods, which generally require lengthy and specialized calculations and very careful choice of force-field parameters for atoms in prosthetic groups and ligands. In both simulations, graphs of the local temperature in the region of the ligand *vs* the overall temperature, and of energy and temperatures *vs* time, showed



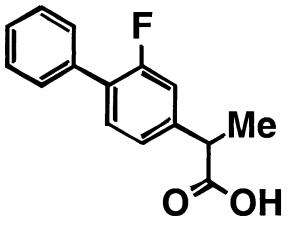
1



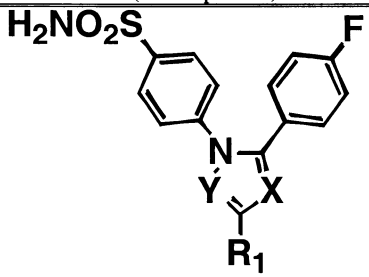
4 (DUP-697)



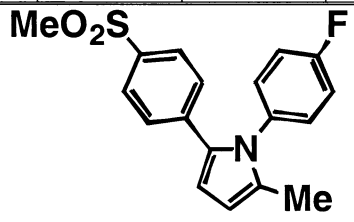
7



13 (Flurbiprofen)



#	X	Y	R_1
3	N	CH	CF_3
5	CH	CH	H
9	CH	N	CF_3



11

#	X	R_1	R_2	R_3	R_4
2	S	H	H	p-F	p-SO_2NH_2
6	S	H	H	p-F	p-SO_2Me
8	S	H	Br	p-F	p-SO_2Me
10	S	H	H	H	p-SO_2Me
12	O	H	H	p-F	p-SO_2Me
14	O	H	H	H	p-SO_2Me
15	S	H	H	p-F	H
16	S	H	SO_2-NHMe	p-F	p-SO_2NH_2
17	S	H	CMe_2-OH	p-F	p-SO_2Me
18	S	EtO-CO	H	p-F	p-SO_2Me
19	S	H	H	H	m-Br, p-SO_2Me
20	S	Br	H	p-F	p-SO_2Me
21	S	H	H	p-SO_2Me	p-SO_2Me
22	S	I	H	p-F	p-SO_2Me
23	O	H	H	2,6-di-Cl	p-SO_2Me
24	O	H	H	H	m-SO_2Me
25	O	H	H	H	o-SO_2Me
26	O	H	H	p-$NHCMe_2{}^nBu$	p-SO_2Me
27	S	H	H	o-$OCMe_2{}^nPr$	p-SO_2Me
28	S	H	Br	o-Br	p-SO_2Me

Figure 1. Compounds in COXBASE.

the expected behavior. The conformations used for later analysis were selected based on these graphs and the point at which the simulations appeared to have reached equilibrium. Figure 2 shows these graphs for the COX-2 case, confirming acceptable local behavior even though the protein was partially constrained and certain possible longer-range interactions were ignored.

Initially, an unconstrained molecular dynamics study had

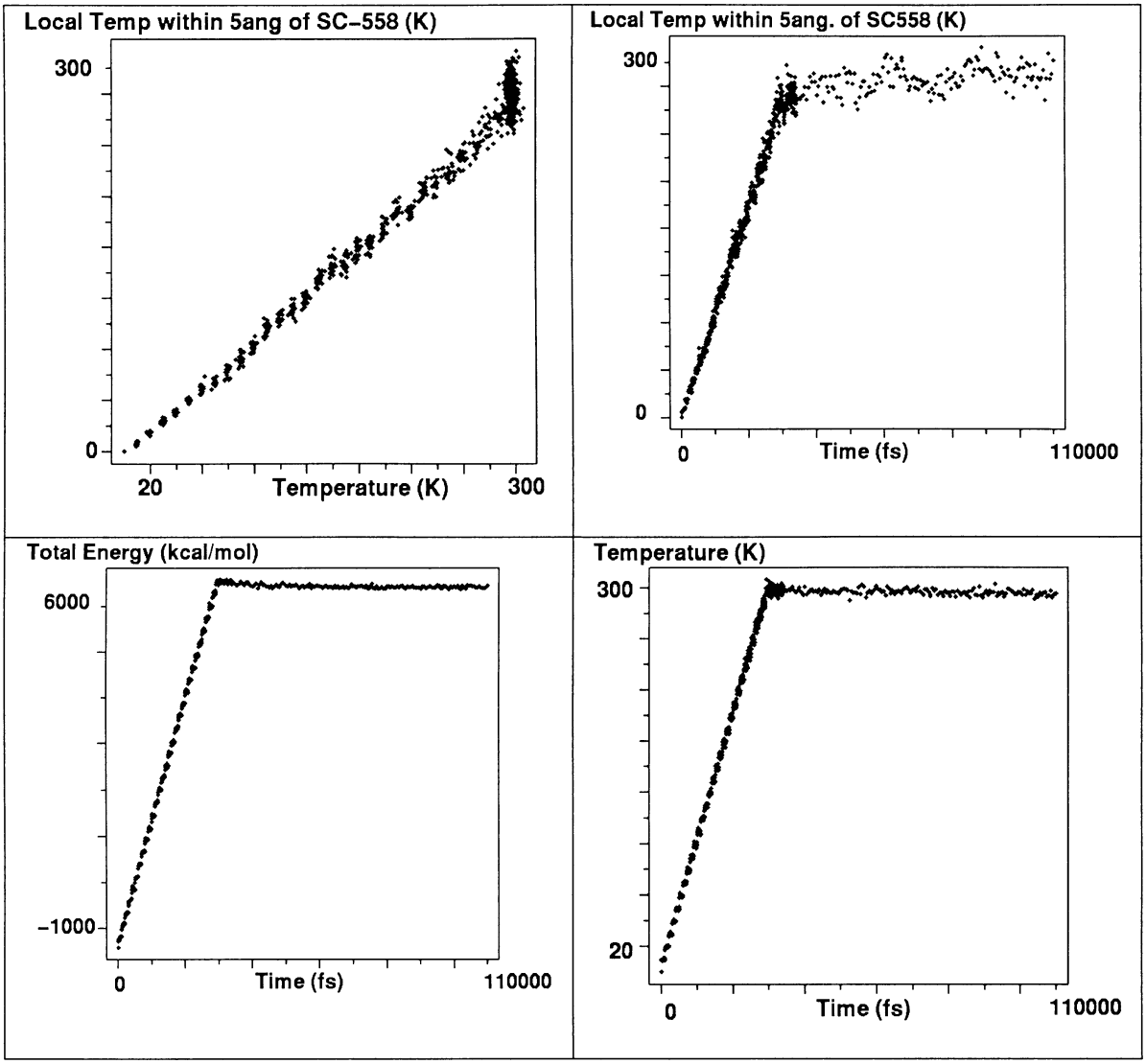


Figure 2. COX-2 simulation: temperatures and energies.

been tried on COX-2, but the structure unfolded significantly and this problem could not be circumvented even by extensive preminimization and very slow heating schedules. Therefore, it was felt that this was likely to be due to one or more of the following:

1. Only one chain of the homodimeric enzyme was being considered, particularly because the atoms with the largest forces on them following energy minimization were all close to where the dimer interface should have been.
2. The membrane and interaction of the membrane with the protein were completely absent from the model
3. There was no water of solvation to reduce longer-range electrostatic interactions and/or to form stabilizing bridges.

As can be seen from Color Plate 1, the parts of the protein that were treated as rigid were indeed largely those that are normally close to the dimer interface and hence the artificial rigidification of this part of the protein can be seen as a crude model for the effect of dimerization.

FLOG: DHFR Against MINDEX12

In an attempt to approximate somewhat more closely the typical state of knowledge during the database-mining lead-discovery phase of a project, the grids generated here from the crystal structure of DHFR do not directly correspond with those used in the original FLOG report.[8] In that case, the side chains were carefully oriented by hand and a deliberate decision was taken to exclude all but two of the water molecules in the crystal structure prior to grid generation. It is extremely unlikely that, in a "real" situation, information would be available to decide which water molecules to retain or which side-chain orientation should be chosen for those many cases where the x-ray data are ambiguous. Leaving all the water molecules in their crystallographically observed positions also is not a practical option, because the mobility of the water is not modeled and therefore the water excludes volumes that are accessible and are needed for the binding of some active compounds. Here, therefore, the remaining option was to gen-

erate the grid after simple removal of the water molecules from the crystal structure, which is a common approach in such studies.[14] In turn, this simplification gives rise to other problems: visualization of the grids by contouring reveals a complex energy surface (Color Plate 2a), which unsurprisingly contains many local maxima, and hence gives rise to many match points during the match point generation step. Due to the large amounts of time required to try every ligand conformation in every orientation on so many match points, several steps had to be taken to simplify the grids and match points in this case. First, the grid resolution *d* was set to 0.5 instead of 0.3. Two hundred match points generated further than 5Å away from the crystallographically observed position of the ligand were removed (this distance was that used in the original FLOG report[8]), along with points found to be isolated from the remainder (e.g., in shallow surface depressions). These steps left 116 match points. The corresponding grids for the "Weighted Average" case (Color Plate 2b) are notably simpler and the surfaces are smoother, and there was no need to use lower resolution or artificially reduce the number of grid points. The "Average" and "Weighted Average" grids show a number of differences, most clearly in the van-der-Waals grid. Color Plate 2c shows an area close to the carboxylate-binding area for MTX where the occasional presence of a water molecule during the dynamics run occludes the space in the "Average" grid case, but not in the "Weighted Average" model. Both these models do handle water mobility, albeit in different ways, and therefore there was no need to remove any of the crystallographically observed water molecules from these runs.

In order to assess the relative capacity of the methods to find active compounds, after completion of the searches of both enantiomers of compound conformers in MINDEX12, the lists of molecules oriented successfully by FLOG were ordered by FLOG score and the resulting list searched for 16 compounds identified from the Merck Index as having activity at DHFR. In Table 1 each compound is listed together with the proportion of the database (as a percentage) that would have had to be screened to find it had the FLOG score by that method been used as a guide to screening order. Results from the crystallographically derived (Crystal_%), mean (Average_%), and weighted mean (Weighted_%) grids/match points are shown, and the data are given for the "standard" FLOG method with simplex optimization of the ligand into the receptor cavity turned on.

It is worth noting that all the methods succeeded in finding not only compounds of the folate/pterin class, but also managed to score reasonably highly other DHFR inhibitors such as trimethoprim, brodimoprim, and tetroxoprim. It is possible within FLOG to "turn off" the simplex optimizer (step 7 in the FLOG summary given earlier), reducing search times significantly. We felt that grids and match points that better represented the receptor might reduce the need for the simplex optimization step. In order to study the effect of the dynamics sampling on this aspect of FLOG, a more summary form of results is needed, and, together with the amount of CPU used, is given in Table 2. Some idea of the overall performance of a particular variant of the FLOG method can be obtained by simply summing the percentile positions (as given in Table 1) of these known ligands in the database to give the SUM parameter. A correction factor is applied for the compounds that had negative FLOG scores and therefore were not rank ordered; these are assumed to be found at the point in the database halfway between the last compound scored by FLOG and the end of the database. Although the simple SUM gives an idea of the utility of a method, it does not adequately reflect the fact that it is often more useful to find one lead after screening 5% of the database and a second after screening 35% than to find two leads at the 20% point. To reflect this, the SQUARE-SUM is calculated as

Table 1. FLOG rank of known DHFR ligands in MINDEX12

Name	Monograph no.	Crystal_%	Average_%	Weighted_%
Denopterin	I2944	0.01	0.47	0.10
A-denopterin	I151	0.03	1.13	0.27
Methotrexate	I6065	0.04	0.17	0.41
Folinic acid	I4254	0.07	29.86	9.19
Folic acid	I4253	0.13	0.93	0.19
Methopterin	I6064	0.39	0.07	0.74
Aminopterin	I493	0.68	0.03	0.07
Ninopterin	I6647	1.04	0.04	0.14
Trimetrexate	I9851	2.85	1.70	3.25
Piritrexim	I7654	4.92	2.43	6.43
Rhizopterin	I8348	5.05	0.30	0.80
Tetroxoprim	I9386	6.51	4.15	6.01
Brodimoprim	I1401	12.83	5.12	11.24
Sapropterin	I8515	16.15	2.77	6.93
Trimethoprim	I9840	21.93	36.55	42.70
Edatrexate	I3553	—	—	62.23

FLOG does not output orientations of structures with negative scores (unfavorable interaction expected with the receptor), and so compounds not found are listed with a dash.

Table 2. Overall efficiency of various modifications to the FLOG method

Grid/match point	Optimizer on or off	No. in top 10%	No. in top 1%	SUM	SQUARE SUM	CPU hours
Ideal	—	16	16	1.94	0.003	—
Random	—	2	0	799.77	532.66	—
Crystal	On	12	7	204.74	184.66	32.3
Average	On	13	7	168.94	92.21	13.8
Weighted average	On	13	8	150.70	60.45	14.9
Crystal	Off	10	6	266.98	220.77	15.3
Average	Off	11	5	187.33	81.67	4.9
Weighted average	Off	12	7	190.94	74.93	5.0

$$\text{Square\_sum\_\%} = 100 * \Sigma(<\text{rank}>^2/<\text{database\_size}>^2)$$

Both the SUM and the SQUARE-SUM should ideally be as small as possible. For database mining purposes, further quantities of interest are the number of desired compounds that would be found by screening 1% or 10% of the database; these numbers should be as large as possible. For comparison, the equivalent figures for a "Random" method (compounds picked distributed evenly throughout the database) and an "Ideal" method (all compounds picked at the very top of the database) are given.

"Crystal" (standard, crystal-structure based) FLOG performance is, as expected, markedly improved on "Random," although still well short of "Ideal." The "Average" method improves the performance of FLOG on all counts; with the optimizer in use, there is a marked decrease in both the SUM and SQUARE-SUM measures, reflected in one additional "lead compound" being found in the top 10% of the database. The "Weighted Average" method improves the figures still further, reducing yet more the SUM and SQUARE SUM parameters and bringing a further "lead compound" into the top 1%. Because the potential surface is "smoothed" by the dynamics sampling and thus there are fewer maxima and correspondingly fewer match points, there is also a marked improvement in the CPU time required (on a Cray J-90 16-processor machine) by the two averaged methods when compared to the crystal-structure based method. This improvement in CPU time vastly more than offsets the initial investment of computational effort required to carry out the molecular dynamics simulations ($\sim$ 8 hours of much cheaper Silicon Graphics R10000 Octane workstation CPU), even when the database is a small one, as here. Furthermore, this time difference was obtained even after lowering the resolution of the "Crystal" grid and manually eliminating two thirds of the automatically generated match points. Despite the much simpler matching, the conformations and orientations of MTX selected by the averaged methods are at least as good as those picked by the crystal-structure based method; Color Plate 2 shows both the experimentally observed and docked positions of MTX in the protein from the "Crystal" (Color Plate 2a, root-mean-square deviation (RMSD) = 2.3 Å) and "Weighted Average" (Color Plate 2b, RMSD = 2.1 Å) runs. Whereas the former achieves a good alignment of the carboxylic acids, the latter performs better in the arguably more interesting region of the pteridine.

Turning off the simplex optimization function has the expected effect on "Crystal" FLOG: both the SUM and SQUARE SUM parameters increase markedly over their values with the optimizer in play, and noticeably fewer compounds are found within the useful 1% and 10% ranges. The "Average" method improves the SUM and more noticeably the SQUARE SUM parameters, and brings one more compound into the 10% range. However, in the absence of the optimizer, it actually performs slightly worse than "Crystal" in recruiting compounds to the top 1% range. In contrast, the "Weighted Average" method improves all parameters. The performance of the "Weighted Average" method with no optimization step surpasses that of the "Crystal" method *with* the simplex optimization turned on in terms of SQUARE-SUM and is equivalent on all the other measures, despite consuming less than one sixth the CPU time.

FLOG: COX-2 Against COXBASE

In this experiment, we sought to evaluate the abilities of the modified FLOG against two criteria, based on practical needs within a typical drug discovery program. First, could FLOG's ability to automatically rank order active, weakly active, and inactive compounds be improved to perform at least as well as a manual evaluation of quality of fit? Ideally, FLOG should provide as good a screen-out of compounds that could never fit the active site as a manual docking, but should be far quicker. Initial experiments showed that unmodified FLOG would be likely to fail to recognize some molecules that were too big, so the experiment included molecules able to test whether the modified methods were improved in this regard. Second, could FLOG be improved to achieve the more difficult goal of a good rank correlation between the FLOG score and the rank order of potency of compounds in a homologous series? Rank correlation is essential here, because many compounds were completely inactive, and it probably is more relevant as there is no *a priori* reason to suppose that the modified FLOG scores and the affinity constants would be *linearly* related. Nonparametric statistics are entirely suited to such cases. The compounds were rank ordered by IC_{50} value recorded in comparable assays, and the rank of each compound is listed in the "Actual Rank" column of Table 3.

To obtain manual dockings for comparison with the automated methods, each of the structures in Figure 1 was manually docked into the protein (except for flurbiprofen for which x-ray crystallographic data were available) using SC-558 as a template for initial orientation. The active site cavity was repre-

Table 3. Rank of compounds from COXBASE

ID	Manual fit	Actual rank	Calculated rank with optimizer on			Calculated rank with optimizer off		
			Crystal	Average	Weighted	Crystal	Average	Weighted
1	Good	1	4	3	3	4	2	3
2	Good	2	2	1	1	2	1	1
3	Good	3	9	5	5	7	5	4
4	Good	3	18	13	11	23	14	10
5	Good	5	3	2	2	3	3	2
6	Good	6	15	12	10	10	11	9
7	Good	7	22	11	7	9	8	7
8	Good	8	20	14	16	16	13	12
9	Good	9	10	4	6	6	4	6
10	Good	10	12	10	8	14	10	8
11	Good	11	19	18	17	11	16	14
12	Good	12	14	6	13	15	7	13
13	Good*	13	6	7	4	8	12	5
14	Good	14	11	8	12	12	6	11
15	Good	15	28	24	23	28	24	23
16	No	16	**1**	19	18	**1**	19	17
17	No	16	**5**	22	22	17	22	20
18	No	16	**7**	21	24	**5**	21	22
19	Possible	16	8	16	9	22	18	19
20	Possible	16	13	15	14	20	15	15
21	Possible	16	16	20	21	13	20	21
22	Possible†	16	17	17	19	18	17	18
23	Possible	16	21	9	15	19	9	16
24	No	16	23	23	20	21	23	24
25	No	16	24	25	27	24	25	27
26	No	16	25	26	26	25	26	25
27	No	16	26	27	28	26	27	28
28	Possible	16	27	28	25	27	28	26
Spearman rank correlation coefficient (actual rank vs predicted rank)			0.32	0.78	0.78	0.53	0.78	0.87
CPU time (minutes, Cray Origin 2000)			50	16	20	35	8	10

* Flurbiprofen, unselective compound not of diaryl heterocyclic class.
† Expected lower rank than compound 20.
Bold face indicates serious errors of the automated methods (see text).

sented using the "Separated Surface" method of the Sybyl MOLCAD unit, which draws a surface at the midpoint between protein and ligand, using the crystal structure coordinates of ligand (SC-558) and protein taken from the PDB entry 1CX2. Compound conformations and orientations were adjusted by constrained energy minimization and/or by hand to reduce to a minimum the amount of the compound visible outside the pocket representation when represented as a ball-and-stick picture with the size of the "balls" set to 0.4 times the van-der-Waals radius of each atom. The "Manual fit" listing in Table 3 indicates by "good," "no," or "possible" whether the structures were considered likely to fit the active site and hence be able, at least on steric grounds, to be active. Although this is necessarily a somewhat subjective judgment, compounds were considered to fit if none of the atoms of the ball-and-stick model showed through the separated surface, to be possible fits if only a small amount of one or two nonhydrogen atoms showed through, and to be unable to fit if more than this showed through. Experience has shown that this is generally a good guide to what is not active. Some examples of each of these subjective categories are given in Color Plate 3. Clearly, compounds that could "fit" the pocket in this way could be inactive for electronic, conformational preference, or other reasons; reasonable steric fit is a necessary, but not a sufficient, condition for activity. However, the null hypothesis for a compound that clearly failed to fit would be that it would be inactive, whereas the null hypothesis for a compound that fit sterically would be that it would be active. The "Manual fit" listing in Table 3 does not consider that for many of the compounds rated "possible" there are close analogues that are known from the literature to be inactive. Such structure-activity information, taken in conjunction with the structural data,

would often permit a clear prediction of "active" or "inactive," but is often not available in the early stages of a reasearch program and so has not been used here.

Grids were generated as before for the unperturbed crystal structure ("Crystal"), and from the dynamics run an "Average" and "Weighted Average" grid were prepared. Match points were generated from each of these grids, providing 101, 26, and 27 match points, respectively, from which the highest-scoring point close to the bromophenyl ring of SC-558 in the crystal structure was selected as the "essential" point.

FLOG was used as before to orient and score conformations of each molecule from COXBASE in the active site, with and without the simplex optimization step, considering both enantiomers of all the conformations in COXBASE. In Table 3, the results of these searches are given as the rank order in which compounds were found, along with the quality of manual fit and the rank of the actual activity from literature reports or as measured in equivalent assays in house. To provide an idea of the scale, actual ranks 15 and 16 represent compounds with an IC_{50} of $\geq$100 μM; the most potent compound (with actual rank of 1) had an IC_{50} of $\sim$4 nM, whereas compounds ranked 1 to 11 all had IC_{50} values better than 100 nM. A test compound ranked above 11 by any of the FLOG methods therefore would be expected to be active, whereas for one ranking further down the list than rank 15 the null hypothesis would be that the compound would be inactive.

The data in Table 3 can be analyzed for important errors made by the automated methods. A compound predicted *not* to fit by the manual method should not be ranked highly by the automated methods (the contrary is not necessarily true: a compound judged to be a "good" or "possible" fit manually may fail to rank highly by the automated methods because of poor electrostatic complementarity or other good reasons). In the case of the unmodified crystal-structure based runs, three compounds (16, 17, and 18) failed on this measure, all of which were substituted on the central heterocyclic core with moderately sized groups. However, the "Average" and "Weighted Average" runs *were* able to correctly assign these and all other compounds to ranks consistent with the manual method and with experimental results. Similarly, the quality of fit of the predicted to the actual, measured rank order of potency can be evaluated for each case by the Spearman rank correlation coefficient provided in the penultimate row of Table 3. The rank correlation coefficients of the method based purely on the crystal structure are rather poor, mainly due to some particularly severe errors such as with compound 16 (ranked first, but really inactive) and compound 4 (ranked 18th or 23rd, depending on use or otherwise of the optimizer, but really a $\sim$10 nM compound). However, the "Average" and "Weighted Average" methods provide a satisfactory rank ordering of the compounds in which the actual rank order of activity is well reproduced, with the majority of the errors being with compounds bearing small substituents such as halogen atoms, or juxtapositions of substituents likely to distort the conformation (such distortions were not accounted for in the building of COXBASE, where the emphasis was on diversity of conformation rather than low energy). These improvements over the "Crystal" method also were achieved with significant savings in the computer resources required. Flurbiprofen was slightly overrated by most of the methods, but it is remarkable that the methods are able to rank order across structural classes even with limited accuracy. Between classes of compounds, there is none of the cancellation of errors that can occur between members of a homologous series, and this is why in the database searching test (where cross-class ranking is important) the best that could be achieved by these methods was to raise the hit rate in the top 10% of the database from $\sim$0.3% ("Random" method) to $\sim$2.2% ("Weighted Average" method).

CONCLUSION

It is clear that the "Average" and particularly the "Weighted Average" methods are better than the unmodified "Crystal" method of using FLOG at database searching. With these improvements, the automated methods also perform well enough to be used for the tasks outlined in the Introduction relating to homologous series, in particular providing a rapid tool for detection of likely inactives that is as good as or better than manual alignment of candidate compounds in the binding site. Although the modifications require an initial molecular dynamics simulation to be run, the time spent on this is more than compensated for by the time saved on FLOG itself, even with a small database of candidate structures. In addition, the modified methods provide a more objective tool that no longer requires the user to specify side-chain orientations or elect whether to retain particular crystallographic water molecules and requires no manual editing of the match points used to align candidate ligands in the binding pocket. The methods could be readily adapted to other docking methods and scoring functions.

ACKNOWLEDGMENTS

We thank Drs. R. Sheridan, M. Miller, and S. Kearsley for assistance in the use and modification of the FLOG, SQ, and related code; Dr. Peter Hunt for helpful discussions; and Mark Oeullet and David Percival for biological results.

REFERENCES

1 Dixon, J.S. Evaluation of the CASP2 Docking Section. *Prot. Struct. Funct. Genet.* 1998, Volume date 1997, Suppl. 1, pp. 198–204, and other papers in the special supplement

2 Kubinyi, H., Folkers, G., and Martin, YC (eds.). 3-D QSAR in drug design: Ligand-protein interactions and molecular similarity. In: *Perspectives Drug Discovery Design*, *Volume 9/10/11*. Kluwer/Escom, Dordrecht, The Netherlands, 1998

3 Stahl, M., and Böhm, H.-J. Development of filter functions for protein-ligand docking. *J. Mol. Graphics Model.* 1998, **16,** 121–132

4 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., and Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* 1982, **161,** 269–288

5 Rarey, M., Kramer, B., Lengauer, T., and Klebe, G.. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 1996, **261,** 470--489

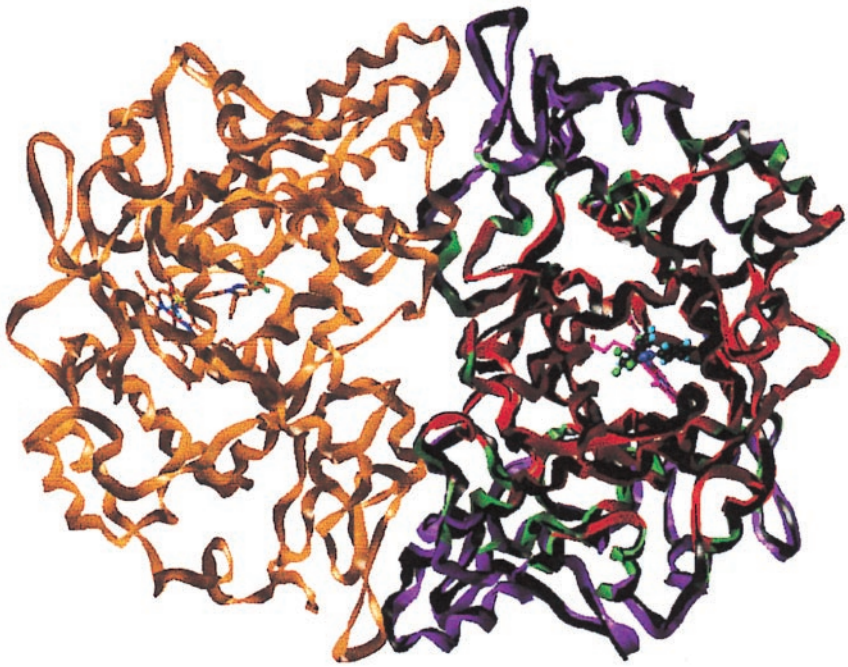
6 Jones, G., Willett, P., Glen, R.C., Leach, A.R., and Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 1997, **267,** 727–748

7 Goodsell, D.S., Morris, G.M., and Olson, A.J. Auto-

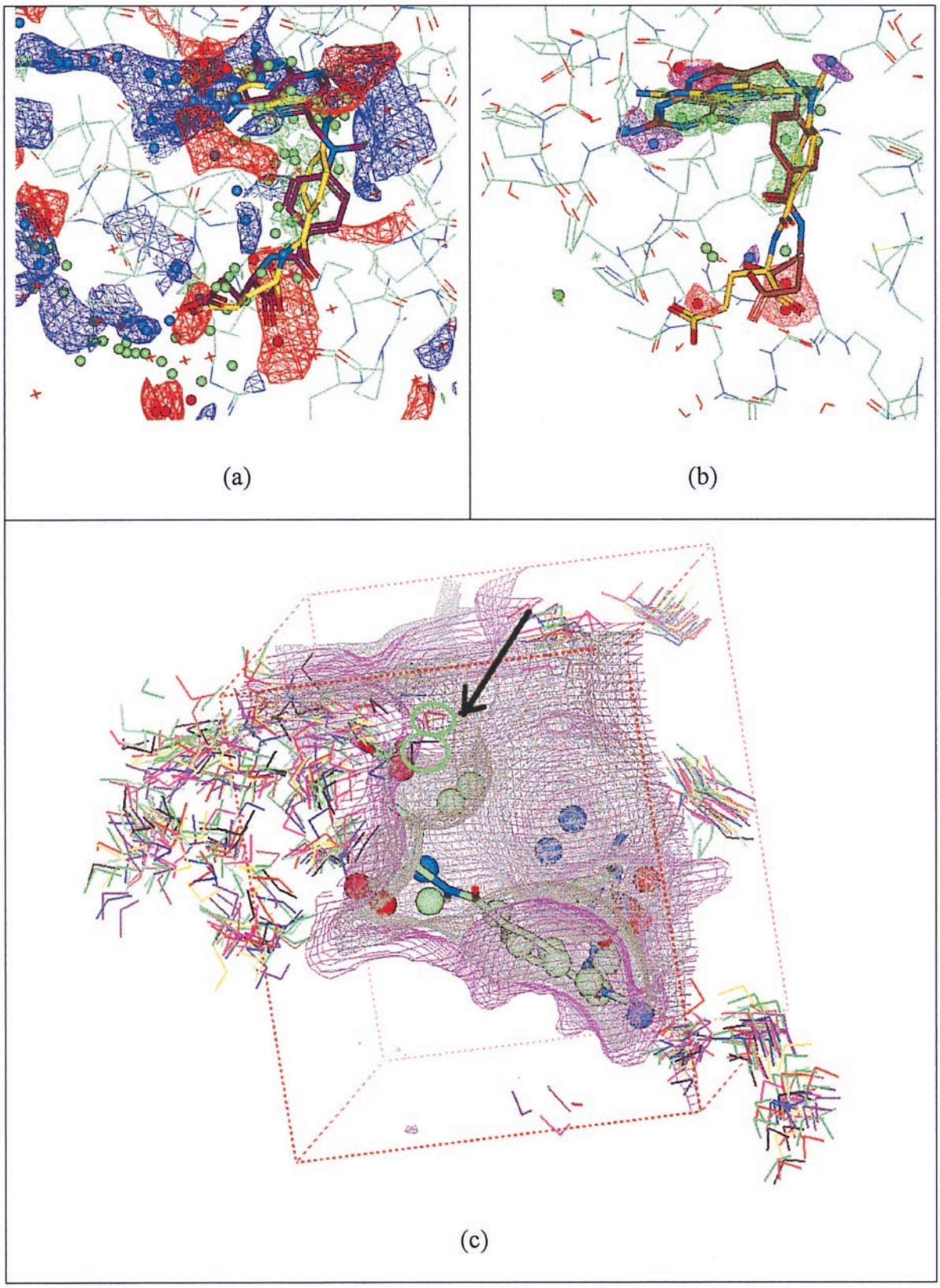
mated docking of flexible ligands: Application of AutoDock. *J. Mol. Recognit.* 1996, **9,** 1–5
8 Miller, M.D., Kearsley, S.K., Underwood, D.J., and Sheridan, R.P. FLOG: A system to select "quasi-flexible" ligands complementary to a receptor of known three-dimensional structure. *J. Comput. Aided Mol. Design* 1994, **8,** 153–174
9 Hoffmann, D., Kramer, B., Washio, T., Steinmetzer, T., Rarey, M., and Lengauer, T. Two-stage method for protein-ligand docking. *J. Med. Chem.* 1999, **42,** 4422–4433
10 Wang, J., Kollman, P.A., and Kuntz, .ID. Flexible ligand docking: A multistep strategy approach. *Prot. Struct. Funct. Genet.* 1999, **36,** 1–19
11 Knegtel, R.M.A., Kuntz, I.D., and Oshiro, C.M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* 1997, **266,** 424–440
12 Wade, R.C., Sobolev, V., Ortiz, A.R., and Peters, G. Computational approaches to modeling receptor flexibility upon ligand binding: Application to interfacially activated enzymes. *NATO ASI Ser., Ser. E (Structure-Based Drug Design)* 1998, **352,** 223–232
13 Leach, A.R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* 1994, **235,** 345–356
14 Murray, C.W., Baxter, C.A., and Frenkel, A.D. The sensitivity of the results of molecular docking to induced fit effects: Application to thrombin, thermolysin and neuraminidase. *J. Comput. Aided Mol. Design* 1999, **13,** 547–562
15 Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C., and Krant, J. Crystal structures of Escherichia coli and Lactobacillus casei dihydrofolate reductase refined at 1.7 Angstroms resolution I. General features and binding of methotrexate. *J. Biol. Chem.* 1982, **257,** 13650–13662
16 Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.E. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977 **112,** 535: *http://www.rcsb.org/pdb/*
17 Kearsley, S.K., Underwood, D.J., Sheridan, R.P., and Miller, M.D. Flexibases: A way to enhance the use of molecular docking methods. *J. Comput. Aided Mol. Design* 1994, **8,** 565–582
18 Budavari, S (ed.). *Merck Index.* Merck & Co. Inc., Whitehouse Station, NJ, 1996
19 Kurumbail, R.G., Stevens, A.M., Gierse, J.K., McDonald, J.J., Stegeman, R.A., Pak, J.Y., Gildehaus, D., Miyashiro, J.M., Penning, T.D., Seibert, K., Isakson, P.C., and Stallings, W.C. Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature* 1996, **384,** 644--648
20 Luong, C., Miller, A., Barnett, J., Chow, J., Ramesha, C., and Browner, M.F. Flexibility of the NSAID binding site in the structure of human cyclooxygenase-2. *Nat. Struct. Biol.* 1996, **3,** 927–933
21 Khanna, I.K., Weier, R.M., Yu, Y., Collins, P.W., Miyashiro, J.M., Koboldt, C.M., Vennhuizen, A.W., Currie, J.L., Seibert, K., and Isakson, P.C. 1,2-Diarylpyrroles as potent and selective inhibitors of cyclooxygenase-2. *J. Med. Chem.* 1997, **40,** 1619–1633
22 Penning, T.D., Talley, J.J., Bertenshaw, S.R., Carter, J.S., Collins, P.W., Docter, S., Graneto, M.J., Lee, L.F., Malecha, J.W., Miyashiro, J.M., Rogers, R.S., Rogier, D.J., Yu, S.S., Anderson, G.D., Burton, E.G., Cogburn, J.N., Gregory, S.A., Koboldt, C.M., Perkins, W.E., Seibert, K., Veenhuizen, A.W., Zhang, Y.Y., and Isakson, P.C. Synthesis and biological evaluation of the 1,5-diarylpyrazole class of cyclooxygenase-2 inhibitors: Identification of 4-[5-(4-methylphenyl)-3-(trifluoromethyl)-*1H*-pyrazol-1-yl[benzenesulfonamide (SC-58635, Celecoxib). *J. Med. Chem.* 1997, **40,** 1347–1365
23 Reitz, D.B., Li, J.J., Norton, M.B., Reinhard, E.J., Collins, J.T., Anderson, G.D., Gregory, S.A., Koboldt, C.M., Perkins, W.E., Seibert, K., and Isakson, P.C. Selective cyclooxygenase inhibitors: Novel 1,2-diarylcyclopentenes are potent and orally active COX-2 inhibitors. *J. Med. Chem.* 1994, **37,** 3878–3881
24 Bertenshaw, S.R., Talley, J.J., Rogier, D.J., Graneto, M.J., Rogers, R.S., Kramer, S.W., Penning, T.D., Koboldt, C.M., Veenhuizen, A.W., Zhang, Y., and Perkins, W.E. 3,4-Diarylthiophenes are selective COX-2 inhibitors. *Bioorg. Med. Chem. Lett.* 1995, **5,** 2919–2922
25 Li, J.J., Norton, M.B., Reinhard, E.J., Anderson, G.D., Gregory, S.A., Isakson, P.C., Koboldt, C.M., Masferrer, J.L., Perkins, W.E., Seibert, K., Zhang, Y., Zweifel, B.S., and Reitz, D.B. Novel terphenyls as selective cyclooxygenase-2 inhibitors and orally active anti-inflammatory agents. *J. Med. Chem.* 1996, **39,** 1846
26 Khanna, I.K., Weier, R.M., Yu, Y., Xu, X.D., Koszyk, F.J., Collins, P.W., Koboldt, C.M., Veenhuizen, A.W., Perkins, W.E., Casler, J.J., Masferrer, J.L., Zhang, Y.Y., Gregory, S.A., Seibert, K., and Isakson, P.C. 1,2-Diarylimidazoles as potent, cyclooxygenase-2 selective, and orally active antiinflammatory agents. *J. Med. Chem.* 1997, **40,** 1634–1647
27 Tripos, Inc., St. Louis, MO, USA. *http://www.tripos.com*
28 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U., Ghio, C., Alagona, G., Profeta, S. Jr., and Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 1984, **106,** 765–784
29 Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M. Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, PA. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* 1995, **117,** 5179–5197
30 *http://www.amber.ucsf.edu/amber/amber.html*
31 Stewart, J.J.P., QCPE program #455
32 Miller, M.D. Sheridan, R.P., and Kearsley, S.K. SQ: A program for rapidly producing pharmacophorically relevant molecular superpositions. *J. Med. Chem.* 1999, **42,** 1505–1514

Howard B. Broughton

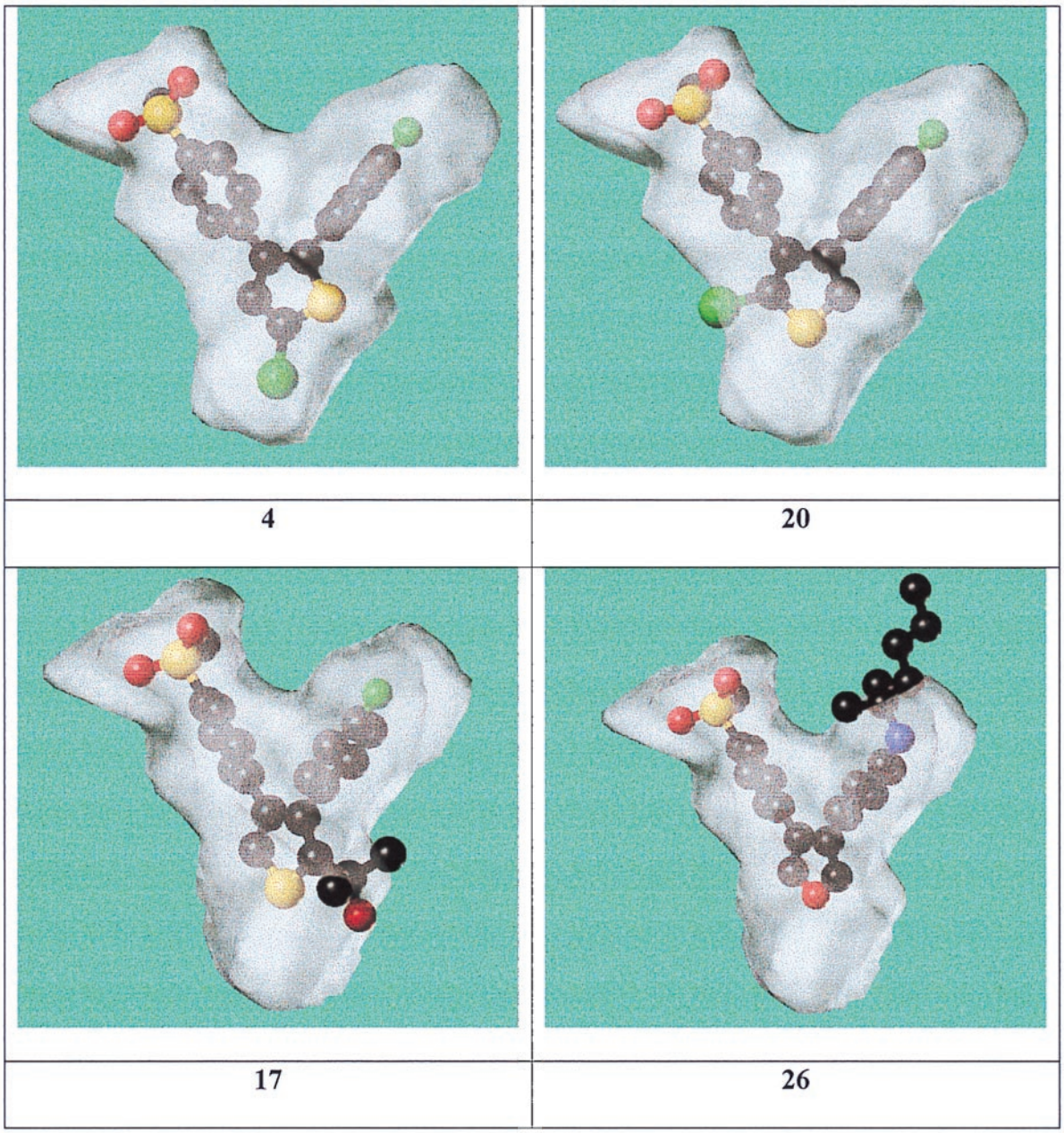
A method for including protein flexibility in protein-ligand docking: Improving tools for database mining and virtual screening



Color Plate 1. Hot (red), warm (green), and cool (magenta) regions of the annealed model of COX-2 superimposed on the A (black) and B (orange) chains of the crystal structure of COX-2. SC-558 from the annealed model is shown in ball-and-stick representation with CPK colors. The haems and B-chain SC-558 from the crystal structure can be seen as stick representations.



Color Plate 2. Grids, match points, and results for DHFR. Panel a shows the hydrogen bond donor (blue) and acceptor (red) grids and the hydrophobic grid (green) contoured at the 2.5 level. The protein is illustrated by "wires" with the carbon atoms light green. The small balls are match points, colored according to type (green = hydrophobic, blue = acceptor, red = donor). Methotrexate as observed in the crystal structure is shown with yellow carbon atoms, and the best orientation from the "Crystal" run (with optimization) is shown with violet carbons. Panel b shows the same information for the "Weighted Average" run. Panel c shows the difference between the van-der-Waals grids (countoured at −0.1) for the average (gray mesh) and weighted average (violet mesh) cases. MTX is shown as sticks with light green carbon atoms, and a major difference produced by the occasional presence of a water molecule (circled with green) is shown by the heavy black arrow, showing that much less space is available for MTX in the "Average" case.



Color Plate 3. Manual docking of ligands in the COX-2 binding pocket. Active compounds such as compound 4 are enclosed entirely within the separated surface generated from the crystal structure. A compound such as compound 20, where only part of one atom protrudes beyond the surface, was considered a "possible" active; that this compound is not active in bioassay strongly suggests that larger protuberances would not be tolerated. The lower panel shows two compounds where several atoms protrude significantly beyond the separated surface and would therefore be considered to be likely to be inactive.