Journal of Computer-Aided Molecular Design, 11 (1997) 333–344.

J-CAMD 400

QXP: Powerful, rapid computer algorithms for structure-based drug design

Colin McMartin* and Regine S. Bohacek**

Research Department, Novartis Pharmaceuticals Corporation, Summit, NJ 07901, U.S.A.

Received 12 June 1996
Accepted 17 March 1997

Keywords: Structure-based design; Pharmacophore; Docking; Pseudo-receptor; Force field; Metropolis; Simulated annealing; Monte Carlo search; Conformational analysis

Summary

New methods for docking, template fitting and building pseudo-receptors are described. Full conformational searches are carried out for flexible cyclic and acyclic molecules. QXP (quick explore) search algorithms are derived from the method of Monte Carlo perturbation with energy minimization in Cartesian space. An additional fast search step is introduced between the initial perturbation and energy minimization. The fast search produces approximate low-energy structures, which are likely to minimize to a low energy. For template fitting, QXP uses a superposition force field which automatically assigns short-range attractive forces to similar atoms in different molecules. The docking algorithms were evaluated using X-ray data for 12 protein–ligand complexes. The ligands had up to 24 rotatable bonds and ranged from highly polar to mostly nonpolar. Docking searches of the randomly disordered ligands gave rms differences between the lowest energy docked structure and the energy-minimized X-ray structure, of less than 0.76 Å for 10 of the ligands. For all the ligands, the rms difference between the energy-minimized X-ray structure and the closest docked structure was less than 0.4 Å, when parts of one of the molecules which are in the solvent were excluded from the rms calculation. Template fitting was tested using four ACE inhibitors. Three ACE templates have been previously published. A single run using QXP generated a series of templates which contained examples of each of the three. A pseudo-receptor, complementary to an ACE template, was built out of small molecules, such as pyrrole, cyclopentanone and propane. When individually energy minimized in the pseudo-receptor, each of the four ACE inhibitors moved with an rms of less than 0.25 Å. After random perturbation, the inhibitors were docked into the pseudo-receptor. Each lowest energy docked structure matched the energy-minimized geometry with an rms of less than 0.08 Å. Thus, the pseudo-receptor shows steric and chemical complementarity to all four molecules. The QXP program is reliable, easy to use and sufficiently rapid for routine application in structure-based drug design.

Introduction

Structural information obtained by modern biophysical techniques has been used very effectively to rapidly discover highly novel potent compounds [1–3]. The role of computational techniques in this process is in a phase of rapid growth. During an active structure-based drug discovery project, there is a need to evaluate a constant stream of candidate molecules. When the project is well resourced with medicinal chemists, this can become a daunting task. Two questions are repeatedly asked about a proposed molecule:

(1) How well does it dock into a binding site?

(2) How does it compare to active molecules or to a design template or pharmacophore?

Ideally, the methods used to obtain this information should be effortless, reliable and rapid, allowing large numbers of design proposals to be routinely evaluated.

QXP (quick explore) was designed to meet these requirements. For template fitting, QXP uses a superposition force field which automatically assigns short-range attractive forces to similar atoms in different molecules [4–6]. Fast Monte Carlo searches are used to match proposed molecules to a template and for flexible docking to

*To whom correspondence should be addressed at: Thistlesoft, 302 Lindsley Drive, Suite 1B, Morristownship, NJ 07960, U.S.A.
**Present address: Ariad Pharmaceuticals, 26 Landsdowne Street, Cambridge, MA 02139, U.S.A.

a binding site. QXP can also be used to make template models from flexible active ligands and to make pseudo-receptors. The template making algorithm was evaluated using ACE inhibitors and compared to three published studies. The performance for docking is compared to the results reported for seven other programs.

Methods

This section starts with a description of the force fields and molecular mechanics routines used by QXP. The general searching algorithm for both template fitting (pharmacophore applications) and docking (ligand-binding site applications) is described next. Specific details of the methods for conformation searching, template fitting and docking are then given, followed by information on the structures used to evaluate the program.

Force field parameters

QXP uses a modified version of the AMBER force field [8]. Special parameters have been added for zinc*. Partial charges are calculated using bond dipole moments [9].

Contact force field

The docking applications in this paper use short nonbonded cutoffs and a distance-dependent dielectric of 4.0r. Smooth distance cutoffs are used. For docking applications, these are 5.0 Å for electrostatic energy and 1.2 times the sum of the van der Waals radii for the Lennard-Jones potential energy. The cutoff distance specifies the point at which the energy becomes zero. A constant increment is added to the energy of the normal potential function so that it is zero at the cutoff point. A smoothing function is then applied so that the gradient is also zero at the cutoff. With a short cutoff, the magnitude of the negative van der Waals energy at the bottom of the potential well is reduced and, as a result, attractive van der Waals forces are very weak. To reward binding in a concave cavity, an extra energy term is added for all nonbonded atom pairs except those involving hydrogen. It has a value of 0.0 at the cutoff distance of 5.0 Å and decreases in an S-shaped curve to –0.5 kJ/mol when the distance is the sum of the van der Waals radii.

Bare force field

For determining global minimum ligand energies, template fitting applications and for generating conformers at the early stage of docking searches, electrostatic energies are turned off. A van der Waals cutoff of 1.05 times the sum of the van der Waals radii is used to eliminate the attractive part of the potential. Internal ligand energies calculated by QXP accurately reflect bonded energy and ignore gas-phase nonbonded energies.

Superposition force field

The superposition force field used by QXP is a slight modification of one described in detail previously [6]. Atoms in separate molecules with similar hydrogen bonding character or hydrophobicity experience a short-range attractive potential. The original potential function is modified to remove the discontinuous gradient at zero. The equation used is

$$E_{sup} = K_{sup} (dist^2 - d_{cut}^2)^2/d_{cut}^4$$

where E_{sup} is the superposition energy, dist is the interatomic distance and d_{cut} is the cutoff distance. K_{sup} is the energy constant for perfect superposition. This function is still convex at distances near the cutoff, but it is concave at short distances and has a gradient of zero at the point of exact superposition.

Energy minimization

Polak–Ribiere conjugate gradient minimization [10] is used with modifications to improve the speed. Three modifications increase the speed: (i) Gradients and energy terms which involve only fixed atoms are not evaluated. (ii) At each step in a conjugate gradient, it is necessary to find the point of minimum energy along a search vector. In QXP, a line search is replaced by estimation of the energy gradients along the search vector at the starting point and at a point 0.1 Å from the starting point. The point on the search vector where the gradient is zero is found by linear interpolation or extrapolation. Occasionally, this estimation fails and the energy is greater at the new position. When this happens, the gradient at the new position is used for a revised estimation (by interpolation) of a zero gradient position. Typically, only two gradient calculations are needed for each step of the minimization cycle. (iii) A gradient and Hessian representation is used for the force field due to fixed atoms in the binding site cavity. At the start of minimization, the sum of the energies, gradients and second derivatives resulting from interactions with fixed atoms are stored for each moving atom. The coordinates for the atom are also stored. Energies and gradients for the atom are computed using these values. Updates for an atom occur when it has moved more than 0.1 Å.

Dynamics

The algorithm of Beeman [11] was implemented. Dynamics is used in current applications in QXP to explore local conformational space. Rescaling of velocities to give an exact average kinetic temperature is carried out at

*The bond angle of two atoms attached to zinc and for carbon oxygen zinc is set to 100° with a 5 kJ/rad^2 force constant. *The force only applies* when the angle is less than 100°. Sulfur–zinc bonds have a length of 2.3 Å and a force constant of 200.0 kJ/Å^2; for oxygen–zinc bonds the corresponding values are 2.15 and 30.0.

each step. The maximum movement of an atom in any single time step is limited to 0.1 Å. Hydrogen vibrations are damped by assigning an atomic weight of 10 to hydrogen.

Rapid search for low-energy conformers of a single molecule (wide-angle Metropolis sampling)

The torsional conformation space of small molecules can be explored very rapidly using Metropolis [12] sampling. Often, small steps are used for Metropolis sampling, e.g. 5°. We use sampling over 360° to allow torsional energy barriers to be readily crossed. With wide-angle sampling, the acceptance rate is reduced. To speed up the procedure, each torsion is sampled several times in succession using a rapid, but exact, relative energy calculation. At the start of sampling a torsion bond rotation, the positions of atoms which can come into contact are described using cylindrical coordinates with the axis of the cylinder coincident with the rotating bond. With cylindrical coordinates, it is possible to rapidly calculate pairwise distances, and hence the nonbonded energy, for any angle of rotation. The process scales with order n, where n is the number of atoms in the ligand. Metropolis sampling is carried out 10 times on each torsion before moving to the next torsion.

Overview of the search strategy of QXP

Searches are carried out using a modified Monte Carlo perturbation/energy minimization method. After an initial Monte Carlo perturbation, a fast search step is introduced to obtain an approximate low-energy structure before energy minimization. This procedure will be referred to as MCFSEM (Monte Carlo perturbation/Fast Search/Energy Minimization).

MCFSEM algorithm

(I) Apply Monte Carlo perturbation to ligand (ring atoms with energy minimization, molecular rotation and translation, torsion bonds).

(II) Fast search for ligand structure with low docked energy (molecular rotation and translation, torsion bonds).

(III) Minimize molecular mechanics energy (Cartesian space).

(IV) Save the minimized structure if energy is low.

(V) Select a saved structure and go to step I, until the required number of cycles is completed.

Monte Carlo perturbation (MCFSEM step I)

To minimize user input, QXP automatically assigns the possible perturbation moves for a molecule after the structure file has been read and before the search starts. Torsion bond perturbations are used for flexible parts of the molecule and Cartesian atom perturbation for cyclic parts. A Cartesian perturbation is carried out by randomly selecting a number of ring atoms. This primary list is augmented by adding atoms directly attached to the primary atoms. When ring atoms have been perturbed, the structure is energy minimized before it is used in the fast search step (step II) of an MCFSEM procedure.

To select the perturbation moves to make at the start of each search cycle, QXP uses a method which proved to be successful for template matching [6]. In early search cycles, the number of perturbations is maximal, i.e. all atoms and torsion bonds are moved. As the search proceeds, the number of perturbations is reduced to a small number such as 2 or 3, the actual torsion bonds or atoms being selected randomly.

Conformation searching on a single molecule

Conformation searching is used to obtain a set of low-energy structures and an estimate of the global minimum energy. The speed of conformation searching for low-energy conformers is increased by including wide-angle Metropolis sampling. The Metropolis procedure only varies one torsion angle at a time, and for some molecules the search might be trapped in a subset of conformation space. To avoid this, in the application CONF, the Metropolis step is embedded in a loop which starts with random perturbation of a number of torsion bonds.

Algorithm for CONF

(1) Randomly perturb n randomly selected torsion bonds up to 360° (n decreases linearly, as the loop repeats, from the total number of rotatable bonds to two).

(2) If ring atoms are present:

(a) apply Cartesian perturbations,

(b) carry out Cartesian energy minimization,

(c) add conformer to store if energy is low.

(3) Apply one pass of wide-angle torsional Metropolis searching through all rotatable bonds, to generate a new low-energy conformer.

(4) Carry out Cartesian energy minimization.

(5) Add conformer to store if energy is low.

(6) Randomly select a structure from the store and go to step 1 until the specified number of loops is carried out.

In step 2b the number of atoms perturbed and the maximum distance decrease linearly as the loop repeats from all atoms and 4.0 Å down to a maximum of 10 atoms and 2.0 Å at the end of the search. After a ring step, even if the new structure is of high energy, the algorithm always carries out a Metropolis torsion step. This enables side-chain clashes to be resolved and decreases the chance of missing a low-energy structure compatible with the ring geometry.

Template fitting

Template fitting is carried out by replacing intramol-

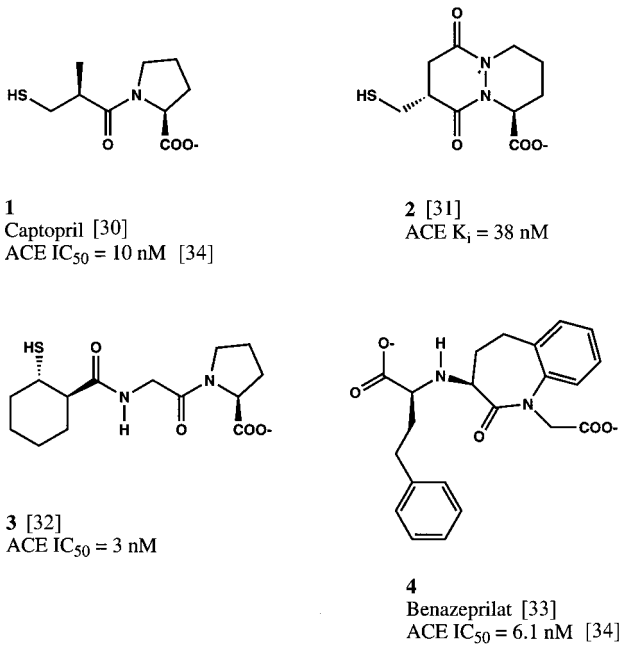


Fig. 1. Four potent, conformationally restricted ACE inhibitors, **1**–**4**, used to construct a composite ACE inhibitor template.

ecular nonbonded energies with the superposition energy, described earlier in this section. Internal energies of the molecules are calculated using a normal molecular mechanics force field. By co-minimizing these two energies, molecules can be matched to simultaneously satisfy the requirements for a good overlap of similar atom types while maintaining low conformational energies of the ligands.

Algorithm for TFIT

(1) Search for global energy minima of moving ligands in isolation; save the energy values; return to starting structure.

(2) Monte Carlo perturbation (steps 1–4 of the procedure CONF, with addition of molecular rotation and translation – up to 360° and 4.0 Å for the first search cycle, decreasing to 30° and 0.5 Å).

(3) Carry out *fast search-templates* (see below).

(4) Energy minimization, store the structure if the energy is <50 kJ + lowest energy found to date.

(5) Randomly select a stored structure and go to step 2 until the requested number of search cycles is complete.

Algorithm for fast search-templates

(1) Carry out *rigid-body alignment* (see below). The structure is treated as the working structure and updated during the search.

(2) Outer loop (20 cycles): Monte Carlo molecular translation and rotation of moving ligands (all ligands, maximum 180.0° and 3.0 Å for the first cycle, to two ligands, maximum 30.0° and 0.5 Å for the last cycle).

(3) Carry out *rigid-body alignment*; update the working structure if the energy is lower.

(4) Inner loop (10 cycles): Wide-angle torsional Metropolis perturbation on the updated structure (temperature 300 K, four randomly selected torsion bonds).

(5) Carry out *rigid-body alignment*; update the working structure if the energy is lower.

(6) Repeat inner loop (10 times).

(7) Repeat outer loop (20 times).

Rigid-body alignment is carried out for each ligand which can move during the search. Each is sequentially superimposed on the ligand immediately preceding it in the structure file. The target for each atom of the ligand being superimposed is the nearest atom of the target ligand.

The relative orientations of ligands during a search can be biased by using zero-order bonds to connect ligand atoms expected to lie close in the final overlap. The constrained atom pairs are introduced into the list of atoms to be superimposed with a relative weight of five. The weight is not critical and was established by trial and error. A quadratic energy potential of 5 kJ/(Å^2 mol) is applied to constrained atoms during energy minimization.

TABLE 1
SPEED AND EFFECT ON STRUCTURE OF ENERGY MINIMIZATION USING DEFAULT FORCE FIELDS OF BATCHMIN AND QXP

	BATCHMIN			QXP		
	Steps[a]	Time[b]	Rms[c]	Steps[a]	Time[b]	Rms[c]
1crn	2000 (0.39)	6:33	1.12	870 (0.05)	2:16	0.66
5pti	2000 (0.87)	9:22	1.19	1330 (0.05)	4:34	0.73
1ifb	2000 (0.71)	30:14	1.41	1900 (0.05)	16:54	0.87
4tmn	550 (0.05)	1:36	1.28	780 (0.05)	0:09	0.95
1hvi	560 (0.05)	1:36	10.65	840 (0.05)	0:14	0.67
4dfr	590 (0.05)	1:52	1.21	600 (0.05)	0:06	0.58

In the first three examples, the whole protein was allowed to move. The other examples show minimization of a ligand in a fixed binding site.
[a] Number of iterations, up to a maximum of 2000, to achieve an energy gradient of 0.05 kJ/(Å mol). The gradient is shown in brackets.
[b] Elapsed time (in min:s); Silicon Graphics Indigo (150 MHz R4400).
[c] Rms movement (Å) of all atoms for the first three proteins and of ligand atoms for the three complexes.

After energy minimization, if all the atoms are less than 0.5 Å from an existing structure, the structures are considered similar. The new structure then replaces the old one if its energy is lower.

Docking a ligand in a binding site (MCDOCK)

Docking is carried out using a molecular mechanics force field without superposition energy. Guide atoms are generated inside the binding site before the search starts. These atoms are in van der Waals contact with the binding site and with each other with a van der Waals radius of 1.5 Å. A grid map representation of the binding site is created as follows. A low-resolution 0.6 Å grid is used for the entire binding site. Each point of this grid that is 3 Å or less from guide atoms is then mapped to a higher $2\times 2\times 2$ 0.3 Å resolution matrix which records van der Waals and electrostatic energies at this resolution. The two different resolutions are used to minimize the number of grid energies which must be calculated and the usage of computer memory. The 0.6 Å grid boxes at distances of greater than 3 Å from the guide atoms are filled to build a smoothly rising wall round the site. The boxes are filled in a series of successive shells, so that the energy increases continuously as the distance from a guide atom increases. The search algorithm for TFIT is used with modification of the fast search subroutine.

Algorithm for fast search-docking

(1) Calculate the grid map energy plus internal ligand energy of the starting structure; the starting structure is treated as a working structure and updated during the fast search.

(2) Outer loop (20 cycles): Perform wide-angle torsional Metropolis perturbation on the updated structure (temperature 300 K, three torsion bonds on the first cycle to one torsion on the final cycle); this step uses internal ligand energy only.

(3) Carry out *rigid-body alignment* onto guide atoms, and update the working structure if the grid map energy plus relative internal ligand energy is lower; if energy is less than 100 kJ, exit *fast search-docking*, i.e. go to energy minimization in main algorithm.

(4) Inner loop (20 cycles): Perform Monte Carlo molecular translation and rotation on the ligands (maximum 30.0° and 2.0 Å on the first cycle of the outer loop to maximum 10.0° and 0.5 Å on the last cycle of the outer loop).

(5) Carry out *rigid-body alignment* onto guide atoms, and update the working structure if the grid map energy plus relative internal ligand energy is lower; if energy is less than 100 kJ/mol, exit *fast search* and go to energy minimization.

(6) Repeat inner loop (20 times).

(7) Repeat outer loop (20 times).

(8) Carry out torsion space minimization using grid map and internal energies; update the working structure.

(9) Carry out a torsional Metropolis search step using the grid map energy plus internal ligand energy.

(10) Update the working structure if the energy is lower.

Experimental data used for testing docking algorithms

The ligands used for testing docking methods are shown in Fig. 1. The binding sites used for docking include all the amino acid residues which have at least one atom less than 7 Å from a ligand atom. Peptide chains were capped with *N*-acetyl and methyl amide groups. Polar hydrogen atoms were added using MacroModel. The Brookhaven Databank access code [13] is followed by the resolution in Å and the reference to the original publication. Published coordinates for thiorphan were used [14] with the protein coordinates for 5tmn [15]; ZFLA, 4tmn, 1.7 [15]; A77003, 1hvi, 1.8 [16]; L700417, 4phv, 2.1 [17]; methotrexate, 4dfr [18]; carboxypeptidase phosphonopeptide inhibitor, 7cpa, 2.0 [19]; tri-*N*-acetylchitotriose, 1hew, 1.75 [20]; 1-deaza-adenosine, 2ada, 2.5 [21]; D-Phe-Pro-Arg-

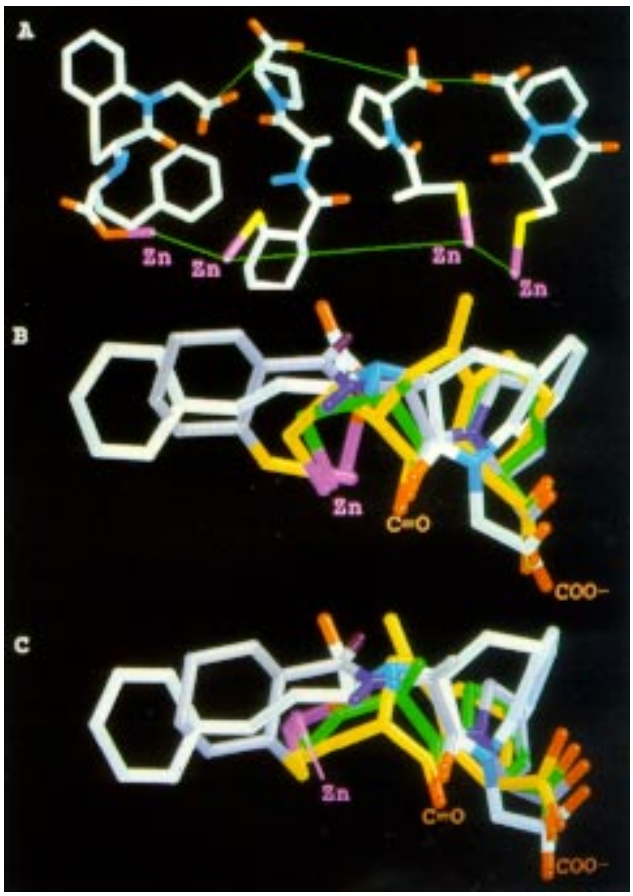
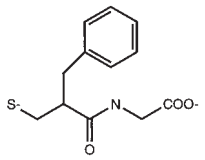
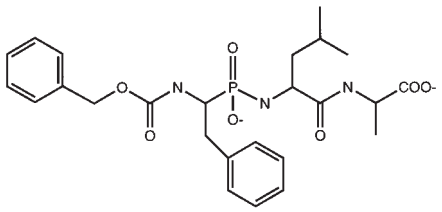


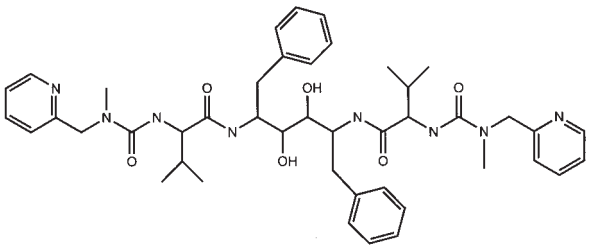
Fig. 2. Flexible template matching of four ACE inhibitors. Five hundred cycles of TFIT were applied to the structure in A (all atoms were colored green). The thin green lines show atoms constrained to overlap by zero-order bonds. Two of the structures appearing in the output are shown in B and C.



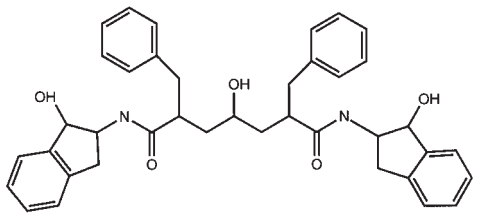
5 Thiorphan [14]



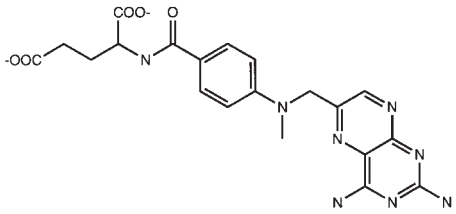
6 ZFLA [15]



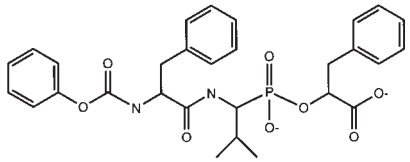
7 A77003 [16]



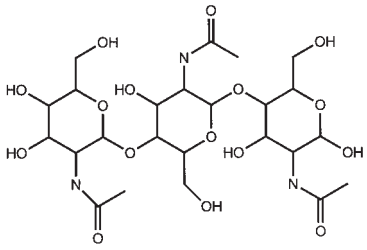
8 L700417 [17]



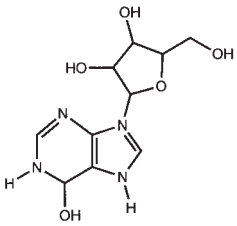
9 Methotrexate [18]



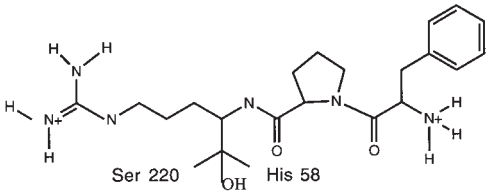
10 Carboxypeptidase phosphonopeptide inhibitor [19]



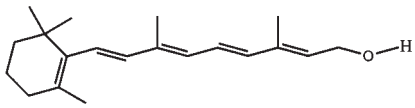
11 Tri-*N*-acetylchitotriose [20]



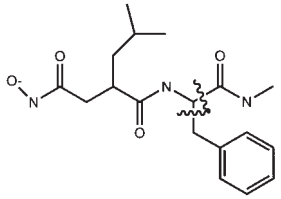
12 1-Deaza-adenosine [21]



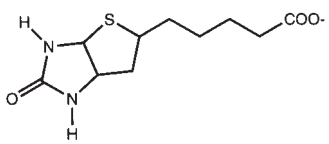
13 D-Phe-Pro-Arg-hemiketal [22]



14 Retinol [23]



15 Hydroxamate collagenase inhibitor [24]



16 Biotin [25]

Fig. 3. Structures of 12 ligands, **5**–**16**, used for docking tests. The part of structure **15** to the right of the wavy line is in solvent and is omitted in some of the rms calculations. The whole structure is used for docking.

hemiketal, 1ppb, 1.9 [22]; retinol, 1rbp, 2.0 [23]; hydroxamate collagenase inhibitor, 1hfc, 1.6 [24]; biotin, 1stp, 2.6 [25]. Waters were removed and the binding site residues were fixed with some exceptions. For thiorphan and ZFLA in thermolysin, Val^{139}, Glu^{143} and Leu^{202} of the enzyme were allowed to move during minimization. For methotrexate in dihydrofolate reductase, a single water which hydrogen bonds to Trp^{30} of the enzyme and to one of the amino groups of the inhibitor was included. The water was allowed to rotate during minimization. For 1hvi and 4phv, a single crystallographic water which forms two hydrogen bonds with the ligand and two with the protein was included and allowed to move during energy minimization. For the energy minimization tests of proteins, the following structures, with all waters removed, were used: crambin, 1crn, 1.5 [26]; ileal fatty acid binding protein, 1ifb, 2.0 [27]; pancreatic trypsin inhibitor, 5pti, 1.5 [28].

User interface

A single command line carries out a complete modeling application on a structure or series of structures. To run the program, the user types the program name, the application name and the name of a structure file which must contain atoms with appropriate atom color assignments. The program can also be run from a graphics interface designed for minimal user effort called LAZYMOUSE.

Applications and Results

Energy minimization

QXP uses short distance cutoffs. It has been shown that, during energy minimization, less distortion of the X-ray coordinates of a protein occurs with short cutoff distances and a distance-dependent dielectric of 4.0r [29].

To evaluate energy minimization and the force field used by QXP, three simple proteins and three protein–ligand complexes were minimized using QXP and BATCHMIN. All atoms of the simple proteins were allowed to move. In the complexes, only ligand atoms were allowed to move. The 'contact' force field of QXP was compared to default parameters of the AMBER implementation in BATCHMIN v. 4.5. The results of this comparison are shown in Table 1.

Although the QXP force field is also based on AMBER, the default version has shorter cutoff distances and a lower dielectric constant than the BATCHMIN defaults. With QXP, the energies converged in fewer steps and a shorter time. Three ligands were energy minimized in a protein binding site with the protein atoms fixed. Two of these moved less with the QXP procedure. The third had an rms difference from the starting structure similar to that obtained with BATCHMIN. For the ligands, convergence to a gradient of less than 0.05 kJ/Å was between 7 and 18 times faster with QXP.

Template fitting

The four potent ACE inhibitors [30–33], shown in Fig. 2, were submitted to 500 steps of TFIT. To guide the superposition, the carboxylic acids at the right-hand end of the line drawings were connected with zero-order bonds. Sulfur atoms and the second carboxylic acid of structure **4** were also connected. The equivalence of these atoms is used in the studies of other workers and is based on the SAR of ACE inhibitors [34–36]. Template making can also be carried out without constraints, but the searches take longer and larger numbers of low-energy solutions may be found.

The ensemble has 22 rotatable bonds and six unsaturated rings. One of the molecules can be regarded as fixed with respect to molecular rotation and translation, leaving 18 degrees of molecular rotational and translational freedom. Figure 3 shows the results of a flexible search for matches of these four ACE inhibitors. The best solution found by template fitting shows the four inhibitors overlapping to occupy a flat volume. The carbonyl groups, known to be critical for activity, are well matched, although these groups were not constrained. The cyclic portions of the molecules are well superimposed and their match is suggestive of a slit-like hydrophobic binding pocket in the active site of ACE.

A number of ACE pharmacophores have been proposed in the literature [34–36]. Each of these extensive studies proposes a single solution. The solutions are, however, all different from each other. The output from the template fitting application of QXP contains low-energy structures similar to all the proposed solutions.

Docking

Table 2 shows a comparison of docking using MCDOCK (QXP) and BATCHMIN [37] with default AMBER parameters. All atoms of the binding sites were constrained. The number of steps was adjusted so that both methods obtained a close result to the X-ray structure. In these two examples, MCDOCK is more rapid and the results of the docking have a smaller rms difference for the conformer closest to the X-ray structure. With MCDOCK, the closest conformer was the lowest energy structure in each case.

Table 3 summarizes the docking results obtained using MCDOCK and X-ray data for 12 protein–ligand complexes. The ligands were randomly rotated and translated and the torsion angles were randomized before docking. Ten different proteins were used and included metalloproteases, an acid protease, a sugar hydrolase and a binding protein. Four of the ligands are bonded to a zinc atom and one has a covalent bond to a histidine. The remaining ligands only interact with the binding site through noncovalent bonds. The interactions range from largely polar (tri-*N*-acetylchitotriose) to entirely hydrophobic (retinol). Many of the ligands are highly flexible. The two

TABLE 2
COMPARISON OF DOCKING USING QXP (MCDOCK) AND THE DEFAULT AMBER VERSION OF BATCHMIN V. 4.5 (BMIN4.5)

	No. of cycles	CPU time (min)	Lowest energy rms (Å)	Best match rms (Å)	Place
Thiorphan					
MCDOCK	30	3	0.91	0.91	1
BMIN4.5	250	859	8.46	1.9	7
ZFLA					
MCDOCK	300	83	1.1	1.1	1
BMIN4.5	2000	1896	4.37	1.1	396

largest ligands, tri-*N*-acetylchitotriose and L700417, contain 20 and 24 rotatable bonds.

Docking gave rms differences between the *lowest energy* docked structure and the energy-minimized ligand of less than 0.76 Å for 10 ligands and less than 0.30 Å for seven ligands. A typical result is shown in Fig. 4. The remaining two ligands had rms values of 1.20 and 1.54 Å. For these ligands and eight of the other ligands, the difference was less than 0.3 Å when part of one of the molecules lying outside the main binding site was excluded from the rms calculation.

When the list of energy-minimized structures for each ligand and binding site was examined, it became apparent that many minima lying close to each other in space were 9sometimes found. For example, for the 1hvi complex, eight structures lay within 0.5 Å of the energy-minimized X-ray coordinates of the ligand. These structures exhibit a wide range of association energies, from –90.6 to –68.5 kJ/mol. (The association energy of docked structures is calculated as the sum of (i) the binding site–ligand non-bonded energy, (ii) the relative internal ligand energy and (iii) the energies, relative to local minima, of the binding site and of any ligand binding site covalent interactions.) This result shows that it is important to extensively search local regions of conformational space in order to find the lowest energy structure.

Pseudo-receptors

When active molecules are known but the binding site is not, it may be useful to convert a template model to a binding site model. Such a model opens up the possibility of applying docking methods to evaluate potential ligands and for using de novo design methods [38].

We have developed a simple method for creating a pseudo-receptor, i.e. a model of a hypothetical binding site which is complementary in steric and chemical properties to a template or pharmacophore. The application PSEUD enables receptors with these properties to be rapidly generated. The first step is carried out interactively using a graphics program. Hydrogen bonding groups in the template are chosen by the user and small complementary hydrogen bonding molecules are placed in the vicinity and connected to the template by zero-order bonds. In the present example, cyclopentanone is used for a hydrogen acceptor and pyrrole for a hydrogen donor. The structure (Fig. 5A) is now submitted to PSEUD. The program aligns the hydrogen bonding receptor molecules and adds propane as a hydrophobic molecule. Dynamics, and energy minimization, are performed to optimize the hydrogen bond geometries while applying the constraints specified by zero-order bonds. The algorithm for adding guide atoms is then used to add layers of pseudo-atoms

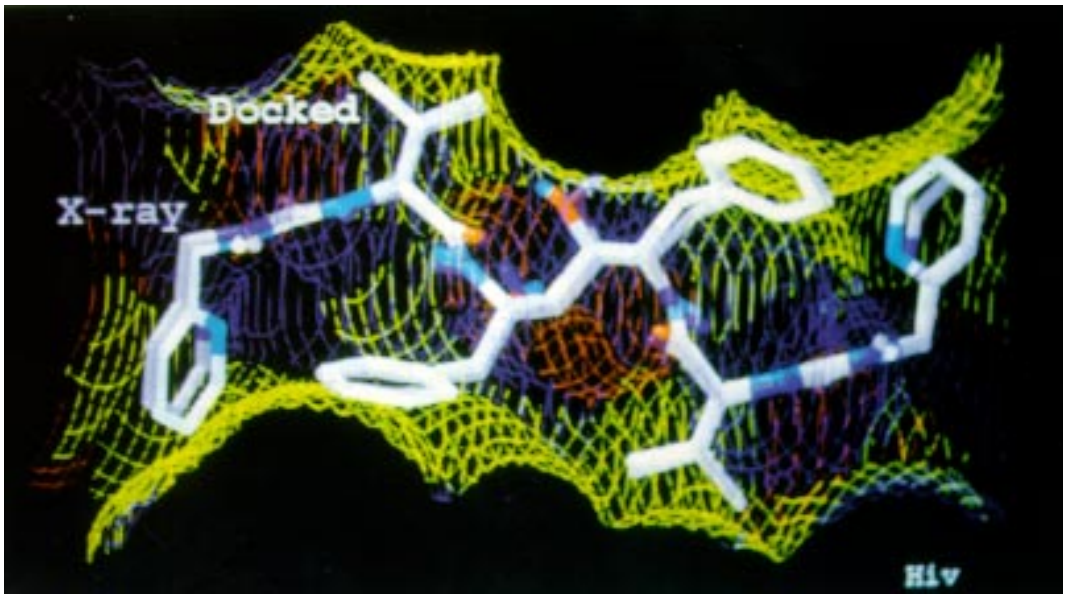


Fig. 4. Docking of A77003 in HIV protease. The lowest energy docked structure is shown with light atoms and the X-ray result with darker atoms. The mesh is an accessible surface of the binding site [50] which is useful for showing steric and chemical complementarity.

TABLE 3
RESULTS FOR DOCKING LIGANDS INTO THEIR RESPECTIVE BINDING SITES USING MCDOCK; RANDOM CONFORMERS AND ORIENTATIONS WERE GENERATED BEFORE DOCKING

Structure[a]	Lowest energy	Closest to X-ray		
	Rms (Å) to X-ray-min[b]	Rms to X-ray	Rms to X-ray-min	Place
5	0.18	0.98	0.18	1
6	0.05	1.10	0.05	1
7	0.23	0.49	0.23	1
8	0.19	1.10	0.13	1
9	0.74	0.96	0.30	5
10	0.14	0.49	0.14	1
11	0.23	0.56	0.23	1
12	0.04	0.52	0.04	1
13	0.03	0.49	0.03	1
14	0.06	0.52	0.06	1
15	1.54 (0.22)[c]	0.37	0.09	4
16	0.74	0.54	0.08	7

[a] Structures and references are given in Fig. 3.
[b] Rms of distance from docked ligand to the ligand minimized in the binding site using DOCKMIN.
[c] Ligand atoms in the side chains in the solvent, shown by wavy lines in Fig. 3, are omitted from the calculation.

around the structure. The guide atoms are next converted to propane to provide a hydrophobic surface for all parts of the ligand not covered by hydrogen bonding molecules. Dynamics, followed by energy minimization, is performed with the ligand held fixed, the hydrogen bond distances constrained and all moving atoms constrained using flat well potentials of radius 3 Å. Outside the flat well, the energy rises as a quadratic function of distance (20 kJ/(Å^2 mol). This step optimizes the geometry of the receptor and its contacts to the template.) Propane was selected since the density of carbon atoms will be similar to that found in side chains such as those of valine or phenylalanine.

Figure 5 shows the result for an HIV protease inhibitor. The accessible surface shows that the pseudo-receptor fits tightly around the ligand and defines its shape precisely with good complementarity. A comparison of the surface with that of the HIV binding site (Fig. 3) shows that the side-chain binding pockets are present in the pseudo-receptor but that the real binding site cavity is larger where it opens into the solvent.

A pseudo-receptor was created for an ACE template made from four inhibitors [39]. The fit of each ligand in the pseudo-receptor model was evaluated by energy minimizing each in the pseudo-receptor, allowing only ligand atoms to move. Each ligand was then randomly perturbed and redocked. Rms differences between the lowest energy docked structures and the template-fitted, energy-minimized structures were less than 0.08 Å (Table 4). The inhibitors differ in size so that the receptor contains space in which misalignment of the smaller ligands might occur. The calculated molecular mechanics association energies varied from –49 to –89 kJ/mol and are in the range typically found for potent ligands in a binding site using QXP. Thus, the pseudo-receptor atoms are in reasonable van der Waals contact with the ligands and do not simply orient the ligands by creating an excessively restricted cavity.

Discussion

Comparison to other methods

The comparisons which follow are made with respect to the goals of QXP, i.e. accurate flexible docking and template fitting, combined with reasonable speed. The comparisons are not intended to reflect the quality of other methods, which may have been designed for other purposes. In a recent review of publications in which structure-based design was used to develop new potent compounds [3], three docking methods were cited. These were BATCHMIN [7], DOCK [40] and AUTODOCK [41]. In addition to these methods, QXP will be compared to a program called LIGIN [42], a method called ADAM

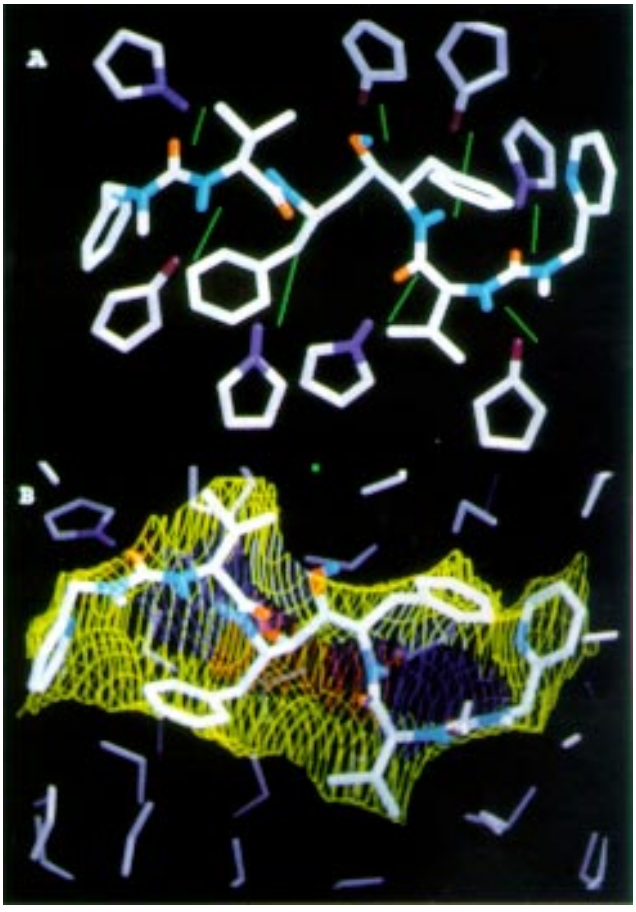


Fig. 5. (A) Starting structure for the generation of a pseudo-receptor. The ligand is A77003 with the coordinates found in the X-ray structure of the complex with HIV protease. (B) The accessible surface of the pseudo-receptor. The mesh shows the accessible surface [50], generated after removing the ligand, of the pseudo-receptor. This surface should be compared to the surface of the actual HIV protease binding site in Fig. 4.

TABLE 4
DOCKING OF ACE INHIBITORS INTO THE ACE PSEUDO-RECEPTOR

Structure	Rms movement (Å) on energy minimization	Rms (Å) after docking (lowest energy structure compared to minimized ligand
1	0.23	0.03
2	0.15	0.08
3	0.24	0.02
4	0.18	0.02

for flexible docking based on hydrogen bonding [43], a multiple start Monte Carlo method [44] and a fast flexible docking method called HAMMERHEAD [45]. A second recent review [46], reported no other methods for flexible docking. Table 5 summarizes the main characteristics of each program.

Apart from QXP, BATCHMIN is clearly the most versatile and powerful method, although its performance is relatively slow. The methods of QXP are an extension of the Monte Carlo perturbation/energy minimization method used by BATCHMIN and rely on molecular mechanics energies. BATCHMIN is the prototype for our methods and the work described here is indebted to this powerful program. A large number of publications using BATCHMIN have appeared. The comparisons to BATCHMIN, in the Applications and Results section, use MacroModel's default values for the AMBER force field and do not necessarily represent the best results which can be obtained with this versatile program. A detailed study of the use of BATCHMIN for docking, where the force field parameters were optimized, has been reported [7]. Four thermolysin inhibitors were docked using perturbation of torsion angles without molecular translation or rotation. BATCHMIN has been successfully applied in a number of structure based design projects [3] and is well suited to working with macrocycles and other complex structures. The searches carried out by QXP included rotation and translation of the ligand and thus cover a wider volume of conformational space than that of the BATCHMIN study [7]. AUTODOCK carries out a search in torsion space and is not suitable for cyclic compounds. Results for AUTODOCK have been reported for small molecules with two or three torsional degrees of freedom [41]. In three out of five cases, the lowest energy solution was not the best match to the X-ray structure. An attractive feature of the flexible docking program ADAM is that it is systematic. It generates all possible ligand conformers and is ideally suited to molecules with a small number of rotatable bonds which form multiple hydrogen bonds to the binding site. For more flexible molecules, the number of conformers will grow exponentially. The method does not search conformers of cyclic compounds and it is not clear how well it will perform when limited hydrogen bonding occurs. HAMMERHEAD is the most rapid, flexible docking method and performs well on the four examples published. It is essentially a torsion space method which constructs ligands using the starting bond angles and lengths and does not include conformational searching of cyclic molecules.

The search times for QXP, for small flexible molecules, are comparable to those reported, sometimes using slower

TABLE 5
COMPARISON OF DOCKING METHODS PUBLISHED IN THE LAST 4 YEARS

Method	Ligand	Protein	Longest time (h:min)	Number of ligands	Maximum number of torsions	Rms range (Å)
LIGIN[a]	Rigid	Rigid	0:20	14	0	0.7–3.0
MSMC[b]	Rigid	Rigid		2	0	0.41–0.65
DOCK[c]	Rigid	Rigid	0:22	4	0	0.9–2.8
A'DOCK[d]	Torsion	Rigid	2:95	5	3	0.9–4.0
ADAM[e]	Torsion	Rigid	0:30	2	6	0.3–0.5
H'HEAD[f]	Torsion	Rigid	400 s	4	11	0.6–1.7
B'MIN[g]	Flexible	Flexible	Days	4	12	0.7–0.9
QXP[h]	Flexible	Flexible	3:20	12	24	0.5–1.1[i]
QXP[j]	Flexible	Flexible	0:03	1	7	0.98
QXP[k]	Rigid	Rigid	20 s	1	1	0.56
B'MIN[l]	Flexible	Flexible	13:00	1	2	0.57
QXP[l]	Flexible	Flexible	0:36	1	2	0.48

[a] Silicon Graphics Indigo (150 MHz IP22) [42].
[b] Multiple start Monte Carlo [44].
[c] Silicon Graphics Iris 4D/25 [40].
[d] AUTODOCK on a Convex C1 [41].
[e] Silicon Graphics Iris 4D (40 MHz R300) [43].
[f] HAMMERHEAD, Silicon Graphics (150 MHz R4400) [45].
[g] BATCHMIN, several days on a Silicon Graphics 4d/280 [7].
[h] Silicon Graphics Indigo 2 (150 MHz R4400), binding site flexibility is allowed in a number of cases as described in the Methods section.
[i] The rms for the carboxypeptidase inhibitor is 1.5 Å when parts of the ligand which are in solvent are included in the rms calculation.
[j] Thiorphan, see Table 2.
[k] Retinol was kept rigid in the fast search and then fully energy minimized in the binding site. The search time was 20 s.
[l] Silicon Graphics (150 MHz R4400). Direct comparison of BATCHMIN and QXP for the docking of 9-benzyl-9-deazaguanine in purine nucleoside phosphorylase (W. Guida, unpublished results). The default parameters were used for QXP. The conditions for BATCHMIN were those previously reported for the docking of thermolysin inhibitors [7].

computers, for rigid docking methods. For comparison, a single study of rigid docking using QXP is reported in Table 5.

The *template fitting procedure* of QXP can be compared to other methods used to generate ACE templates [34–36]. Each study proposed a different single solution. None of these methods allow fully flexible matching of all the atoms which overlap. QXP allows the molecules to adopt low-energy conformers which maximize the overlap. QXP could rapidly generate several template models, including those previously proposed. The procedure of QXP is very convenient for routine evaluation of proposed inhibitors.

Efficiency of searching conformational space

An effective search requires a suitable method for exploring conformational space, and an appropriate potential energy function. The potential energy function should be such that the X-ray structure of the ligand lies close to a local minimum energy. Ideally, the docked structure with the lowest energy should also be close to the X-ray structure. The methods of QXP perform well in exploring conformational space, judging by the number of cases where a structure very close to the energy-minimized X-ray structure is obtained, even when the number of rotatable bonds is large.

The template fitting searches are based on a similar algorithm to that used for docking and their efficiency is confirmed by the ACE template results where a number of previously reported templates were found in a single search.

Unification of pharmacophore and binding site methods

The pseudo-receptor for ACE described in this paper is based only on active ligands and has functional groups similar to those found in protein binding sites. The results reported here show that the pseudo-receptor has properties similar to those of a protein binding site with respect to energy and the ability to orient the ligands (Table 4). This method should allow binding site methods to be applied in the absence of an experimental structure.

Speed and ease of use When a novel structure is being considered in a drug discovery project, the basic question asked is: *Is this compound worth synthesizing?* The methods of QXP often allow this question to be answered by submitting a single job to the computer. Once a template or binding site model has been set up, test ligands can be added to the model and evaluated with no further user intervention.

Conclusions

QXP provides docking and template fitting searches with full flexibility for cyclic and acyclic molecules. The internal energy of the ligands is reported and the ligand conformers are not restricted to local minima. The methods are effective for ligands with at least 24 internal torsional degrees of freedom. For template fitting, at least four flexible ligands can be matched simultaneously, with full searching of torsional and cyclic parts of the molecules. For docking, the accuracy is better than that which has been reported for programs with the more restricted goal of rigid docking or of flexible docking using only torsional perturbations. For 10 out of the 12 docking tests reported in this paper, the main source of rms difference from the published crystal structure was the distance of the nearest local energy minimum from the reported experimental structure.

A recent publication describes in detail the use of TFIT and MCDOCK in the design of dual ACE/neutral endoprotease inhibitors [39]. The results of docking and template fitting were able to distinguish between active and inactive compounds for 23 flexible molecules for ACE and for 19 out of 20 molecules for NEP. Two further recent publications describe the successful use of the program to design potent benzo-fused macrocyclic inhibitors of zinc metalloproteases [47,48].

We have been using QXP for several years and only minor modifications have been required in the last 2 years. The program is well suited for routine application in structure-based drug design projects.

The program QXP, with many additional applications and the graphics interface LAZYMOUSE, is available under the name FLO96 from Colin McMartin [49].

Acknowledgements

The authors wish to thank Bryan Marten and Wayne Guida for their support and helpful discussions during this work.

References

1 Erickson, J.W. and Fesik, S.W., In Venuti, M.C. (Ed.) Annual Reports of Medicinal Chemistry, Academic Press, New York, NY, U.S.A., 1992, p. 271.
2 Greer, J., Erickson, J.W., Baldwin, J.J. and Varney, J., J. Med. Chem., 37 (1994) 1035.
3 Bohacek, R.S., McMartin, C. and Guida, W.C., Rev. Med. Chem., in press.
4 Kearsley, S.K. and Smith, G.M., Tetrahedron Comput. Methodol., 3 (1990) 615.
5 Klebe, G., Mietzner, T. and Weber, F., J. Comput.-Aided Mol. Design, 8 (1994) 751.
6 McMartin, C. and Bohacek, R.S., J. Comput.-Aided Mol. Design, 9 (1995) 237.
7 Guida, W.C., Bohacek, R.S. and Erion, M.D., J. Comput. Chem., 13 (1992) 214.
8 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Algona, G., Profeta, S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
9 Mohamadi, F., Richards, N.G., Guida, W.C., Liskamp, R., Lip-

ton, M., Caufield, C., Chang, G., Hendrikson, T. and Still, C., J. Comput. Chem., 11 (1990) 440.
10 Polak, E. and Ribiere, G., Revue Française Inf. Rech. Oper., 16-R1 (1969) 35.
11 Beeman, D., J. Chem., 20 (1976) 130.
12 Metropolis, N., Rosenbluth, A., Rosaenbluth, M., Teller, A. and Teller, E., J. Chem. Phys., 21 (1953) 1087.
13 Abol, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J., In Allen, F.H., Bergerhoff, G. and Sievers, R. (Eds.) Crystallographic Data-bases – Information Content, Software Systems, Scientific Applications, Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, 1987, pp. 107–132.
14 Roderick, S.L., Fournie-Zaluski, M.C., Roques, B.P. and Matthews, B.W., Biochemistry, 28 (1989) 1493.
15 Monzingo, A.F. and Matthews, B.W., Biochemistry, 21 (1982) 3390.
16 Bone, R., Vacca, J.P., Anderson, P.S. and Holloway, M.K., J. Am. Chem. Soc., 113 (1991) 9382.
17 Hosur, M.V., Bhat, T.N., Kemp, D., Baldwin, E.T., Liu, B., Gulnik, S., Wideburg, N.E., Norbeck, D.W., Appelt, K. and Erickson, J.W., J. Am. Chem. Soc., 116 (1994) 847.
18 Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C. and Kraut, J., J. Biol. Chem., 257 (1982) 13650.
19 Kim, H. and Lipscomb, W.N., Biochemistry, 30 (1991) 8171.
20 Cheetam, J.C., Artymuik, P.J. and Phillips, D.C., J. Mol. Biol., 224 (1993) 613.
21 Wilson, D.K., Rudolph, F.B. and Quicho, F.A., Science, 252 (1991) 1278.
22 Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S.R. and Hofsteenge, J., EMBO J., 8 (1989) 3467.
23 Cowan, S.W., Newcomer, M.E. and Jones, T.A., Proteins, 8 (1990) 44.
24 Spurlino, J.C., Smallwood, A.M., Carlton, D.D., Banks, T.M., Vavra, K.J., Johnson, J.S., Cook, E.R., Falvo, G.J., Wahl, R.C. and Pulvino, T.A., Proteins, 19 (1994) 98.
25 Weber, P.C., Ohlendorf, D.H., Wendoloski, J.J. and Salemme, F.R., Science, 85 (1989) 243.
26 Teeter, M.M., Proc. Natl. Acad. Sci. USA, 81 (1984) 6104.
27 Sacchettini, J.C., Gordon, J.I. and Banszak, L.J., J. Mol. Biol., 208 (1989) 327.
28 Wlodawer, A., Walter, J., Huber, L. and Sjolin, L., J. Mol. Biol., 180 (1984) 301.
29 Whitlow, M. and Teeter, M.M., J. Am. Chem. Soc., 108 (1986) 7163.
30 Cushman, D.W., Cheung, H.S., Sabo, E.F. and Ondetti, M.A., Biochemistry, 16 (1977) 5484.
31 Hassall, C.H., Kroehn, A., Moody, C.J. and Thomas, W.A., J. Chem. Soc. Perkin Trans. I, (1984) 155.
32 Weller, H.N., Gordon, E.M., Rom, M.B. and Plusces, J., Biochem. Biophys. Res. Commun., 125 (1984) 82.
33 Watthey, J.W.H., Stanton, J.L., Desai, M., Babiarz, J.E. and Finn, B.M., J. Med. Chem., 28 (1985) 1511. ACE activity was measured with the substrate hippuryl-His-Leu using the procedure described by Cushman, D.W. and Cheung, H.W., Biochem. Pharmacol., 20 (1971) 1637.
34 Andrews, P.R., Carson, J.M., Caselli, A., Spark, M.J. and Woods, R., J. Med. Chem., 28 (1985) 393.
35 Mayer, D., Naylor, C.B., Motoc, I. and Marshall, G.R., J. Comput.-Aided Mol. Design, 1 (1987) 3.
36 Hausin, R.J. and Codding, P.W., J. Med. Chem., 33 (1990) 1940.
37 MacroModel v. 2.5 was obtained from Professor W. Clark Still, Columbia University, New York, NY, U.S.A.
38 Bohacek, R.S.B. and McMartin, C., J. Am. Chem. Soc., 116 (1994) 5560.
39 Bohacek, R.S., De Lombaert, S., McMartin, C., Priestle, J. and Gruetter, M., J. Am. Chem. Soc., 118 (1996) 8231.
40 Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 505.
41 Goodsell, D.S. and Olsen, A.J., Proteins Struct. Funct. Genet., 8 (1990) 195.
42 Sobolev, V., Wade, R.C., Vriend, G. and Edelman, M., Proteins, 25 (1996) 120.
43 Mitzutani, M.Y., Tomioka, N. and Itai, A., J. Mol. Biol., 235 (1994) 318.
44 Hart, T.N. and Read, R.J., Proteins, 13 (1992) 206.
45 Welch, W., Ruppert, J. and Jain, A.N., Chem. Biol., 3 (1996) 449.
46 Kuntz, I.D., Meng, E.C. and Shoichet, B.K., Acc. Chem. Res., 27 (1994) 117.
47 Ksander, G.M., De Jesus, R., Yuan, A., Ghai, R.D., Trapani, A., McMartin, C. and Bohacek, R., J. Med. Chem., 40 (1997) 495.
48 Ksander, G.M., De Jesus, R., Yuan, A., Ghai, R.D., McMartin, C. and Bohacek, R., J. Med. Chem., 40 (1997) 506.
49 FLO96, Morristownship, NJ, U.S.A., tel.: 201 455 1229, e-mail: cmcma @ix.netcom.com.
50 Bohacek, R.S. and McMartin, C., J. Med. Chem., 35 (1992) 1671.