Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming

Daniel K Gehlhaar[1], Gennady M Verkhivker[1], Paul A Rejto[1], Christopher J Sherman[1], David B Fogel[2], Lawrence J Fogel[2] and Stephan T Freer[1]*

[1]Agouron Pharmaceuticals, Inc., San Diego, CA 92121-1121, USA and [2]Natural Selection, Inc., 1591 Calle De Cinco, La Jolla, CA 92037, USA

Background: An important prerequisite for computational structure-based drug design is prediction of the structures of ligand–protein complexes that have not yet been experimentally determined by X-ray crystallography or NMR. For this task, docking of rigid ligands is inadequate because it assumes knowledge of the conformation of the bound ligand. Docking of flexible ligands would be desirable, but requires one to search an enormous conformational space. We set out to develop a strategy for flexible docking by combining a simple model of ligand–protein interactions for molecular recognition with an evolutionary programming search technique.

Results: We have developed an intermolecular energy function that incorporates steric and hydrogen-bonding terms. The parameters in this function were obtained by docking in three different protein systems. The effectiveness of this method was demonstrated by conformationally flexible docking of the inhibitor AG-1343, a potential new drug against AIDS, into HIV-1 protease. For this molecule, which has nine rotatable bonds, the crystal structure was reproduced within 1.5 Å root-mean-square deviation 34 times in 100 simulations, each requiring eight minutes on a Silicon Graphics R4400 workstation. The energy function correctly evaluates the crystal structure as the global energy minimum.

Conclusions: We believe that a solution of the docking problem may be achieved by matching a simple model of molecular recognition with an efficient search procedure. The necessary ingredients of a molecular recognition model include only steric and hydrogen-bond interaction terms. Although these terms are not necessarily sufficient to predict binding affinity, they describe ligand–protein interactions faithfully enough to enable a docking program to predict the structure of the bound ligand. This docking strategy thus provides an important tool for the interdisciplinary field of rational drug design.

Chemistry & Biology May 1995, **2**:317–324

Key words: AG-1343, flexible docking, HIV-protease inhibitors, molecular recognition, rational drug design

Introduction

Computational prediction of the structures of ligand–protein complexes from the conformations of the unbound ligand and protein molecules, usually referred to as the docking problem, is an exciting subject of long-standing interest in theoretical studies of molecular recognition [1–13]. Docking is now an important component of protein-structure-based drug design [13,14], which has been particularly effective in the search for HIV-1 protease inhibitors [15,16]. More than one hundred crystal structures of HIV-1 protease complexes have been solved. This structural information is a resource that can be used for the testing and refinement of molecular recognition models and docking techniques.

The flexible docking of ligands into HIV-1 protease is a particularly demanding problem. Even when rigid ligands are docked into a structure representing the conformation that the homodimeric protease adopts when ligands are bound, there are two, nearly degenerate, symmetry-related binding modes. For flexible ligands, the number of possible alternative solutions increases dramatically. Hence, an accurate molecular recognition model must be sufficiently sensitive to distinguish between these modes. In addition, the flaps that enclose the bound ligand constrain the HIV-1 protease active site and cause the alternative binding modes to be separated by large energy barriers, so that it may be hard for a search algorithm to cross the energy barriers to find the global minimum.

Underlying any docking strategy is a description of ligand–protein association. In principle, a complete thermodynamic description of this process involves contributions from several opposing sources, including solvent reorganization, conformational entropy and van der Waals and electrostatic interaction energies. For biomolecular systems, it is difficult to measure these terms with sufficient accuracy to permit quantitative predictions. Moreover, the complete energy function necessary for the prediction of binding affinity may not be suitable for docking simulations. The energy function used in

*Corresponding author.

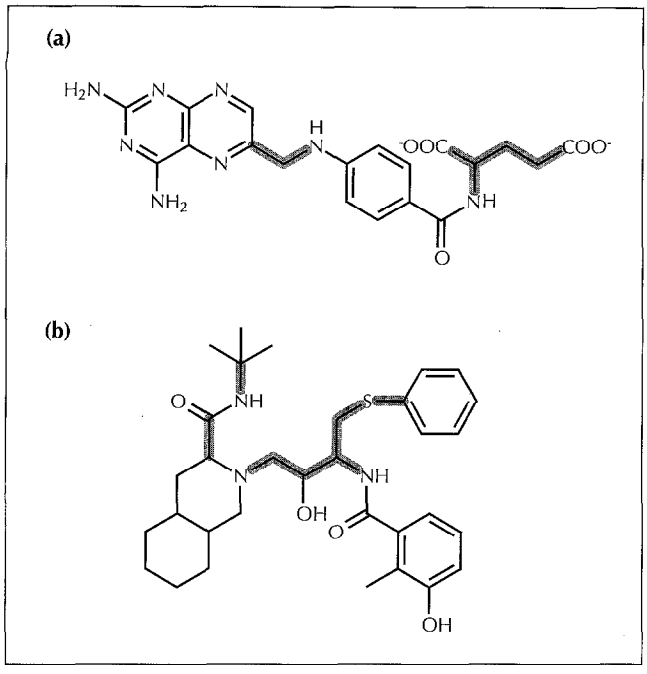

Fig. 1. The structures of MTX and AG-1343. **(a)** A two-dimensional representation of MTX. The freely-rotating bonds are highlighted. **(b)** A two-dimensional representation of AG-1343. The orientation is chosen to be consistent with that shown in Figure 3.

docking studies must be simple enough to permit its rapid evaluation and, more importantly, the resulting energy landscape must be smooth enough to allow the search to proceed efficiently without becoming trapped in local minima [17].

A fundamental component of models for molecular recognition is the steric energy function, based on surface complementarity [1–5]. This term alone, however, is not sufficient to distinguish consistently between alternative binding modes. In general, electrostatic interactions may provide additional specificity to discriminate between 'true' and 'false' solutions [1–5]. They are highly sensitive, however, to assignment of the charges as well as to geometric details of the ligand–protein interface. In an effort to circumvent these complications, we have developed a simple, and possibly minimal, molecular recognition model that incorporates both a steric term and a hydrogen-bond contribution calibrated to reproduce the structure of the ligand in crystal structures of ligand–protein complexes, rather than to determine their binding affinity.

A variety of methods have been applied to docking studies, including Monte Carlo optimization techniques [1–10] and genetic algorithms [11,12]. Because of the high computational cost of flexibly docking ligands, early studies focused on rigid molecules [1]. Unfortunately, this approach is likely to fail when the bound conformation of the ligand is unknown. As a result, attention has turned toward the flexible docking of ligands into proteins held in their bound conformation (see, for example, [8–11]), where the internal degrees of freedom of the ligand are varied to optimize simultaneously the intramolecular ligand interactions and the intermolecular ligand–protein interactions.

We have applied a stochastic search technique, evolutionary programming [18–20], that has been shown to be effective in a variety of optimization problems. A full conformational and positional search of the candidate inhibitor was conducted within the rigid active site of the crystallographic bound-protein conformation. No assumptions regarding either the likely bound conformation of the ligand or any specific hydrogen-bond interactions were made. We refined our strategy by studying the flexible docking of methotrexate (MTX) into dihydrofolate reductase (DHFR), and applied it to the flexible docking of the inhibitor AG-1343, a potential new drug against AIDS, into HIV-1 protease.

Results and discussion

To verify the ability of the molecular recognition model to reproduce crystal structures, and to optimize the parameters in the search procedure, the classic case of

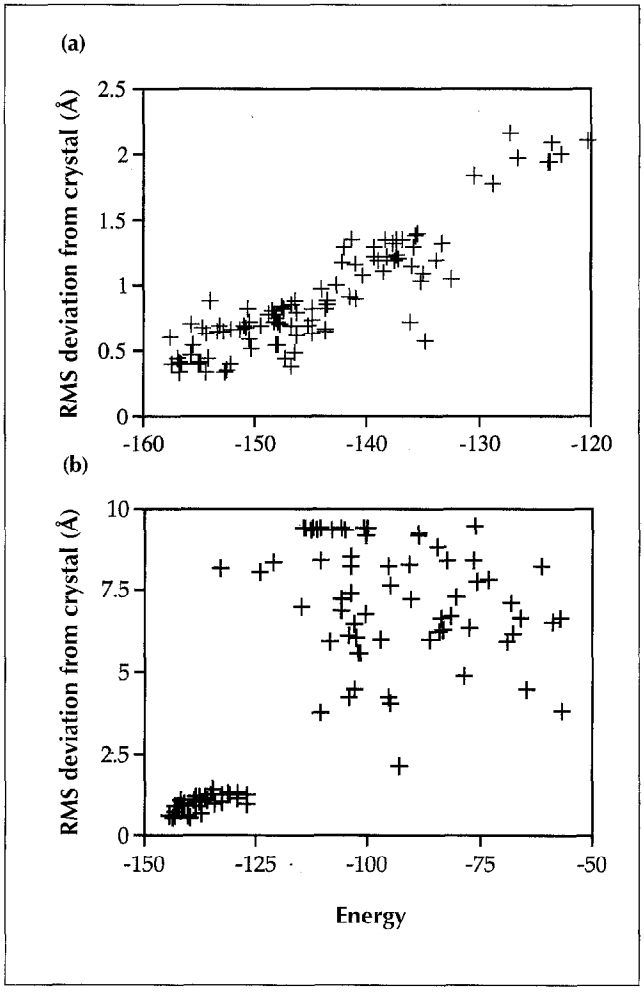

Fig. 2. Deviation of docked structures from crystallographically determined structures **(a)** The results of 100 docking runs of MTX into DHFR; 91 solutions are within 1.5 Å rms of the crystal structure. **(b)** The results of 100 docking runs of AG-1343 into HIV-1 protease; 34 solutions are within 1.5 Å rms of the crystal structure.

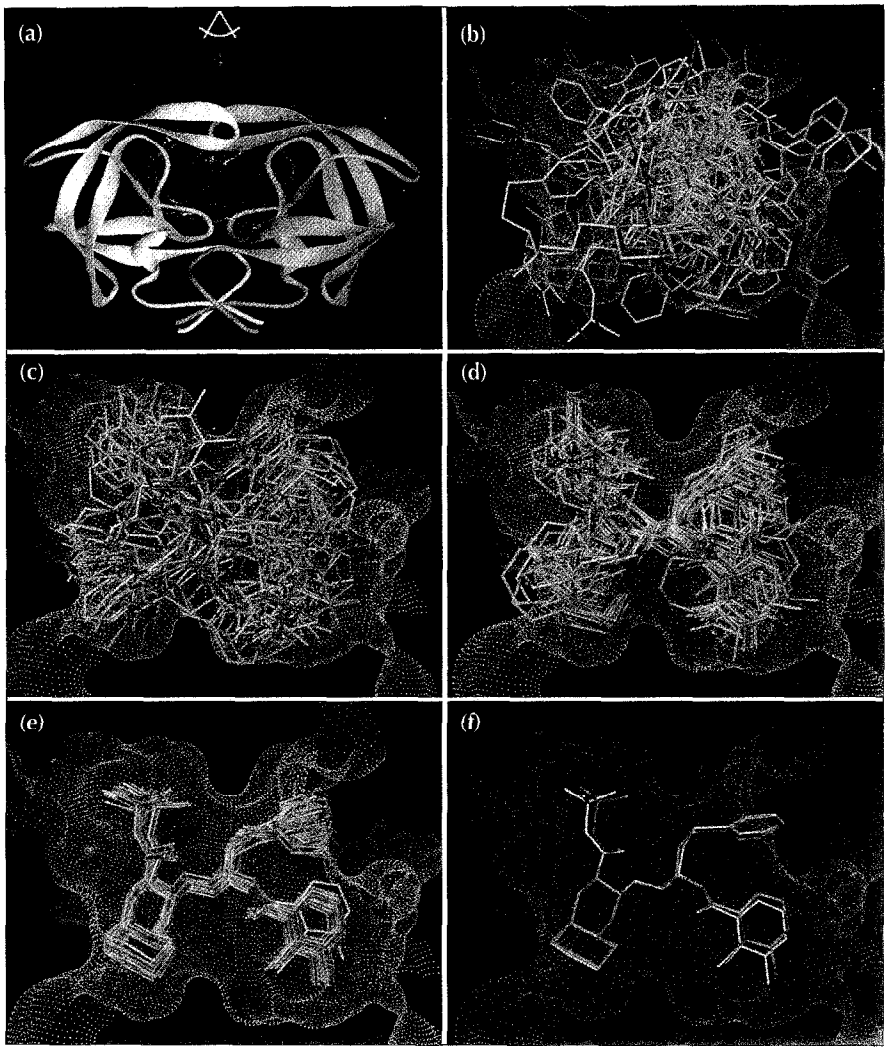

Fig. 3. A representation of the docking process for a typical successful docking of AG-1343 into HIV-1 protease. **(a)** A ribbon diagram of the bound form of the dimeric HIV-1 protease receptor with the Connolly surface of the binding site shown in white. The flaps are closed in this conformation. **(b)** and **(c)** 30 representative structures from the top 15 % of the population at generations 2 and 70, respectively. **(d)** 30 representative structures from the top 15 % of the population at generation 92. By careful inspection, two binding modes can be seen. One corresponds to the experimentally observed binding orientation; the other is related by a two-fold rotation about an axis passing through the ligand center of mass. **(e)** 30 representative structures from the top 15 % of the population at generation 149. Only the experimentally observed binding mode is now populated. **(f)** The most favorable structure after conjugate gradient energy minimization, shown in green. The experimentally determined crystal structure is shown in white. The rms deviation between the docked structure and the crystal structure is 0.4 Å.

docking MTX into the folate-binding region of DHFR was studied [21]. Seven rotatable bonds were considered for the MTX molecule (Fig. 1a). As a quantitative measure of docking performance, we use the percent of docking simulations that identify the conformation of the bound ligand within 1.5 Å root-mean-square (rms) deviation from the crystallographically observed conformation. Given this criterion, the binding mode seen in the crystal structure was observed 91 times out of 100 runs. In the remaining nine structures the pteridine ring is inverted relative to the crystal structure, consistent with the binding mode of the natural substrate [21]. The nearly linear correlation between the rms deviation from the crystal structure and the energy of the docked structure shows that for this system the molecular recognition model is sufficiently accurate to distinguish between alternative binding modes (Fig. 2a).

Docking the potent nonpeptidic inhibitor AG-1343 (Fig. 1b) into the active site of HIV-1 protease is a considerably more challenging problem. The HIV-1 protease active site is relatively large and has a number of specific binding pockets. The symmetry-related and energetically similar binding modes make the search for the crystallographically observed binding mode difficult. In principle, a complete solution of the docking problem implies a flexible protein molecule as well as a flexible ligand. HIV-1 protease, however, undergoes a significant conformational change upon ligand binding that involves backbone rearrangements of up to 7 Å [15,16]. Therefore, we study docking of the flexible ligand into the rigid protease held in its bound conformation.

Previous computational attempts to insert a ligand into the binding site of HIV-1 protease have failed because of insurmountable barriers to penetration of the active site [4]. In our study, the repulsive term of the energy function is linearly increased throughout the docking simulation. The ligand is thereby allowed to penetrate into the protein core during the initial stages of the docking simulation when the population consists mostly of random orientations, enabling the evolutionary search to explore a variety of possible solutions that would otherwise be forbidden by the presence of high energy barriers. Eventually, in later stages of the evolution, scaling of the potential energy function narrows the search to only a few energetically favorable binding modes that serve, in effect, to 'funnel' the search into the global minimum represented by the ligand's crystallographic conformation. This method is analogous to temperature scaling in

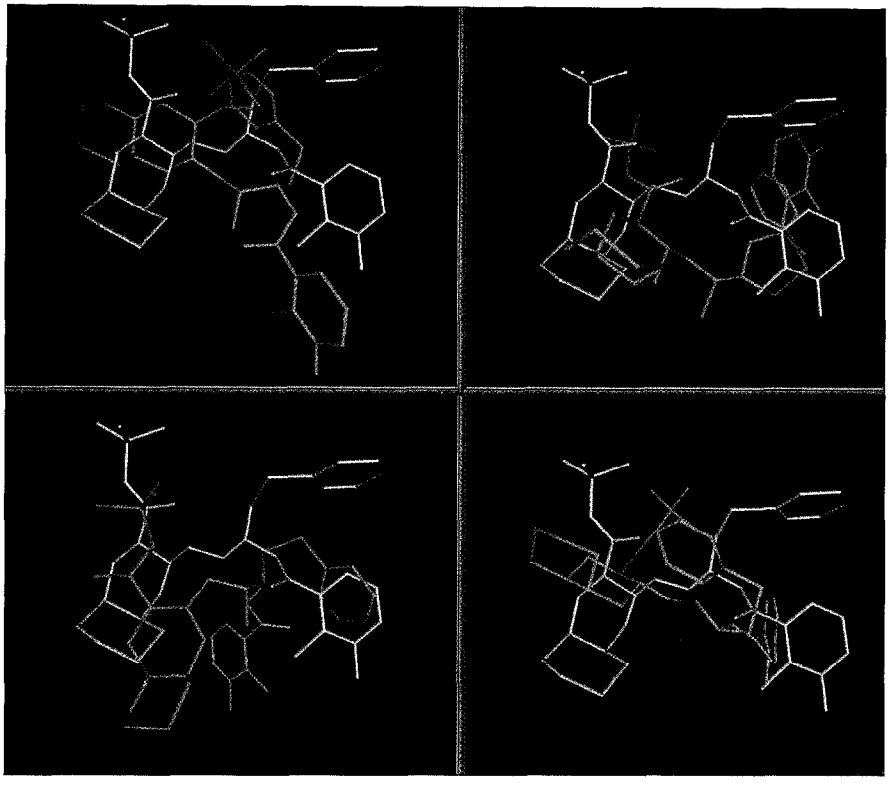

Fig. 4. The four members of the initial population with the smallest rms deviation from the crystal structure; the closest has a rms deviation of 4.0 Å.

Monte Carlo optimization (see, for example, [2,4,6]), in that it employs a combination of evolutionary search and simulated annealing. Optimizing the energy parameters, such as the 'softness' of the energy function repulsion, may further facilitate conformational search both by promoting escape from local minima and by destabilizing alternative solutions [17].

The crystal structure of the AG-1343 complex (K. Appelt, unpublished data) was reproduced within the 1.5 Å rms tolerance 34 times from 100 simulations (Fig. 2b). The success rate of these simulations is remarkable given the large number of freely rotating bonds located in the center of the AG-1343 molecule and the nature of the HIV-1 protease binding site. Assuming only a 0.2 Å resolution for positional variables and 3.6° for angles, the nine conformational angles and six positional and orientational variables in AG-1343 would constitute a search space of ~10^{30} possible solutions. During the simulation, however, the system explores a continuous space of positions and angles. The search of this enormous space during the flexible docking of AG-1343 was achieved within eight minutes of cpu time on a single Silicon Graphics R4400 workstation.

For one of the successful docking simulations (Fig. 3), the final predicted structure of AG-1343 has only 0.4 Å rms deviation from the crystal structure. This conformation was not present as a member of the initial population, and the solution is not merely fortuitous (Fig. 4). The interaction energies for structures near the crystallographic orientation are lower than the energies of structures docked in incorrect orientations (Fig. 2b). The solutions with large (near 8.0 Å) rms deviations from the crystal structure yet with relatively low interaction energy (see the upper left corner of Fig. 2b) are symmetry-related to the crystal structure. As expected for such solutions, their energy is nearly degenerate with the crystal solution. Nevertheless, the energy function correctly ranks their energy as higher than that of the crystal structure.

We found that the crystallographic conformation represents the global minimum of the energy landscape and that the energy decreases gradually in the neighborhood of the crystallographic binding mode (Fig. 2). Because of these properties, the performance of structure prediction can be increased by iterating the simulation. Given that the probability of successfully docking AG-1343 into HIV-1 protease with our method is $p = 0.34$, the crystallographic structure can be found after N simulations with probability $1-(1-p)^N$. Therefore, if one chooses the lowest energy conformer from eight simulations, the probability that it is the crystal structure is increased to greater than 0.95.

The results of this study are encouraging and suggest that the molecular recognition model introduced in this work may have general applicability to other ligand–protein systems. We are currently studying a more diverse set of ligand–protein complexes to determine the limits of the model for structure prediction.

Significance

We have developed a simple model of molecular recognition that allows the reliable docking of conformationally flexible ligands into protein binding sites by evolutionary programming. In

this work, the correct structure of the AG-1343–HIV-1 protease complex was obtained in 34 out of 100 docking simulations, each taking eight minutes on a single Silicon Graphics R4400 workstation. The confidence of finding the correct structure in this case can be increased to 95% by running the simulation only eight times.

The results suggest that solution of the flexible docking problem requires two ingredients: first, a simple energy function that alleviates the frustration of binding-energy landscapes [17], and second, an effective stochastic search procedure. We conclude that steric and hydrogen-bond interactions combined with a rudimentary intramolecular term provide a reasonable framework for this energy function. While a more detailed model that includes, for example, solvation and entropic effects may be required to predict binding affinity, a simple energy function is apparently preferable for structure prediction because it yields an energy landscape with relatively few local minima. We believe, therefore, that more complicated models of ligand–protein interactions may not be generally useful for the structural prediction of ligand–protein complexes.

The strategy of flexible docking introduced in this work can be a valuable tool in a variety of rational drug design applications including rapid database searching for lead compounds [22] and prediction of the structure of protein complexes of novel inhibitors [23].

Materials and methods

Molecular recognition model

As a minimal requirement, the energy function used for docking must account for hydrophilic and hydrophobic surface complementarity [5,7]. In our molecular recognition model, we include steric and hydrogen-bond contributions calculated from a piecewise linear potential (Fig. 5). One advantage of the chosen functional form is that it has a finite value when the interatomic distance approaches zero. Thus, the function is well behaved even when there are severe close contacts between protein and ligand atoms, such as occur during initial stages of the docking simulation when the ligand conformations are largely in random orientations. To permit penetration of the protein core during these early stages, the repulsive parameter of the intermolecular ligand–protein interaction potential was linearly scaled from zero to its final value.

The ligand–protein energy function defined by our molecular recognition model is a pairwise sum over all ligand and protein heavy atoms. The parameters used in the pairwise potential depend on the atom types involved in the interaction. In our model, there are four different atom types; hydrogen-bond donor, hydrogen-bond acceptor, both donor and acceptor, and nonpolar. These atom types interact through steric and hydrogen-bond potentials (Table 1), which have the same functional form but different parameters that reflect the notion that a single hydrogen bond should have a larger weight than a single steric interaction (Table 2) [7]. Every pair of atoms has one and only one type of interaction. Primary and secondary amines are classified as donors while oxygen and nitrogen atoms with no bound hydrogens are defined as acceptors. Hydroxyl groups and crystallographic water molecules are defined as both donor and acceptor. Carbon and sulfur atoms are defined as nonpolar.

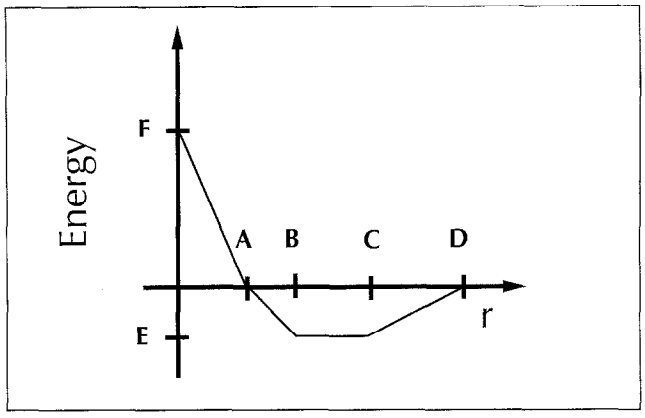

Fig. 5. The piecewise linear pairwise potential function used for the ligand–protein interaction energy. Values are given in Table 2.

The initial potential parameters were extracted from a potential used for *de novo* design of enzyme inhibitors [24]. The values were refined [25] to reproduce consistently the crystal structure of the complex by performing 25 docking runs into three different proteins: MTX into DHFR [21], a proprietary compound from Agouron Pharmaceuticals into FK506-binding protein (FKBP) and AG-1284 into HIV-1 protease [26].

The internal energy of the ligand is given by a torsional potential and a non-bonded term; all bond distances and angles are fixed during the docking simulation. The torsional potential has the form:

$$E = A[1 = \cos(n\phi - \phi_0)]$$

where $A = 3.0$, $n = 3$, $\phi_0 = \pi$ for sp^3–sp^3 bonds, and $A = 1.5$, $n = 6$, and $\phi_0 = 0$ for sp^3–sp^2 bonds. Torsional angles about sp^2–sp^2 bonds are held fixed, as are dihedral angles in rings. The nonbonded term provides a constant penalty of 10 000 when the interatomic distance between any pair of non-bonded ligand atoms is smaller than a threshold of 2.35 Å and vanishes otherwise; it was chosen to prevent internal collapse of the ligand.

Table 1. Pairwise atomic interaction types used in the molecular recognition model.

Ligand atom type	Protein atom type			
	Donor	*Acceptor*	*Both*	*Nonpolar*
Donor	Steric	Hydrogen bond	Hydrogen bond	Steric
Acceptor	Hydrogen bond	Steric	Hydrogen bond	Steric
Both	Hydrogen bond	Hydrogen bond	Hydrogen bond	Steric
Nonpolar	Steric	Steric	Steric	Steric

Table 2. Parameters of the atomic pairwise ligand–protein potentials[a].

Interaction type	A	B	C	D	E	F
Steric	3.4	3.6	4.5	5.5	-0.4	20.0
Hydrogen bond	2.3	2.6	3.1	3.4	-2.0	20.0

[a]A, B, C and D are in Angstroms. E and F are in arbitrary energy units.

No assumptions were made regarding likely ligand conformations or any specific ligand–protein interactions, including hydrogen-bond formation. The ligand is required to remain in a parallelepiped that encompasses the active site, obtained from the crystal structure of a ligand–protein complex. This binding box is defined by adding a 2 Å cushion to the bound ligand. A constant energy penalty of 200 is added to every ligand atom outside the box. For HIV-1 protease, the box is defined using the AG-1343 complex and has dimensions of 15.2 Å x 15.4 Å x 14.0 Å. The intermolecular potential is precalculated on a 0.2 Å grid covering the protein binding site.

The values for the ligand bond distances and bond angles were obtained from the crystal structure of the bound ligand–protein complex, and are indistinguishable from the values obtained after minimization of the ligand in the absence of the protein using the Amber force field [27]. The bond connecting the amide and the aromatic ring in AG-1343 (see the lower right of Fig. 1b) is not allowed to rotate in our model because it involves two sp^2 carbon atoms. It is fixed at the crystallographic value of 51.4°, which changes to 41.6° after minimization of the ligand in the absence of the protein. All crystallographic water molecules, including the critical bridging water, are included in the docking simulation as a part of the protein structure.

Evolutionary programming

In the course of the evolutionary process (Fig. 6), a population of candidate ligand conformations competes for survival. The energy of each member in the population is compared with a fixed number of randomly selected opponents. A win is assigned to the member with the lower energy, and the number of competitions a member wins is used to determine whether it survives into the next generation [19]. All surviving members produce offspring by Gaussian mutation so as to maintain a constant population size. The best member of the final generation is minimized using a conjugate gradient algorithm [28].

In the population of ligand conformations, each member is represented by an encoded vector composed of its six rigid body coordinates and the dihedral angles about its rotatable bonds. The initial ligand conformations are generated by randomizing the encoded vector, given that the center of mass of each ligand conformation must lie anywhere within the rectangular parallelepiped that defines the active site. Rigid rotation and rotatable dihedral angles are uniformly randomized between 0 and 360 degrees.

We have tuned the parameters of the search procedure by systematically varying the population size, the number of competitors in the stochastic competition, and the initial value of

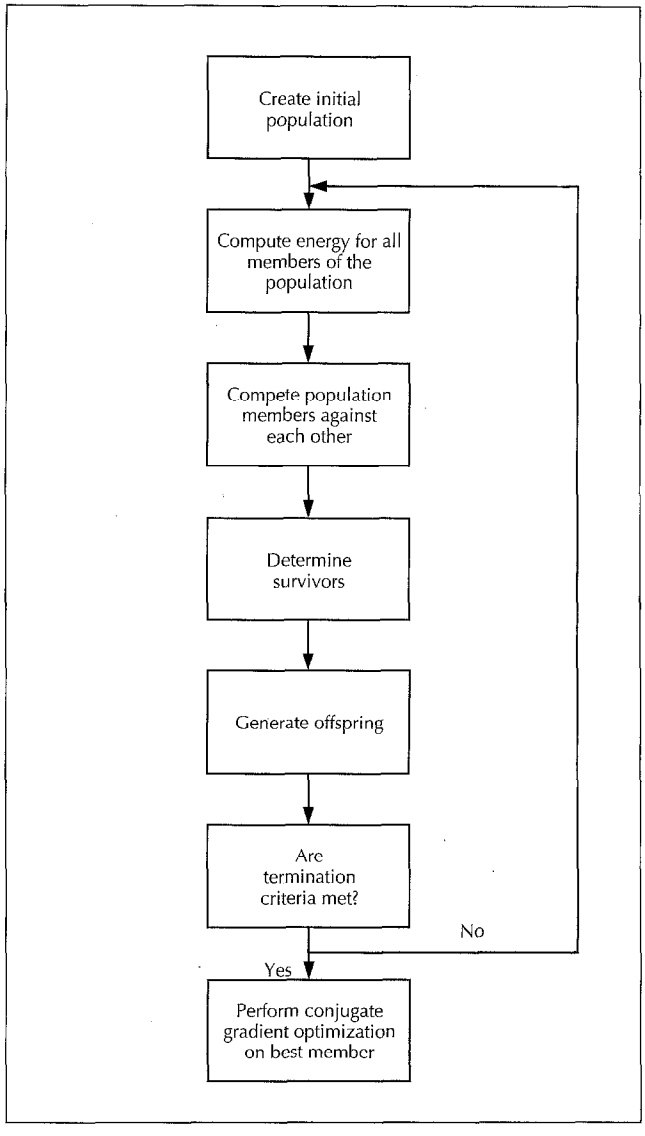

Fig. 6. A flow diagram of the general evolutionary programming protocol.

the standard deviation of the Gaussian mutation used to generate offspring. For each sampled value of these parameters, 100 docking simulations of MTX into DHFR were performed. This system was chosen for the tuning studies because it is a canonical yet non-trivial test case with seven rotatable bonds and several distinct local minima.

A mathematical analysis of simple evolutionary algorithms [29] gives a logarithmic dependence between the energy and the population size, a result that was observed in our experiments (Fig. 7a). The goal of our tuning was to maximize convergence while minimizing computation time, and consequently a population size of 1200 was chosen for MTX. Given the larger number of rotatable bonds in AG-1343, a population size of 2000 was used in this case.

Maintaining diversity in the population throughout the simulation is important to avoid premature convergence to local minima. This can be achieved by using a small number of competitors in the stochastic tournament, which gives less fit members a chance to survive and generate offspring. The smallest number of competitors that gave consistent

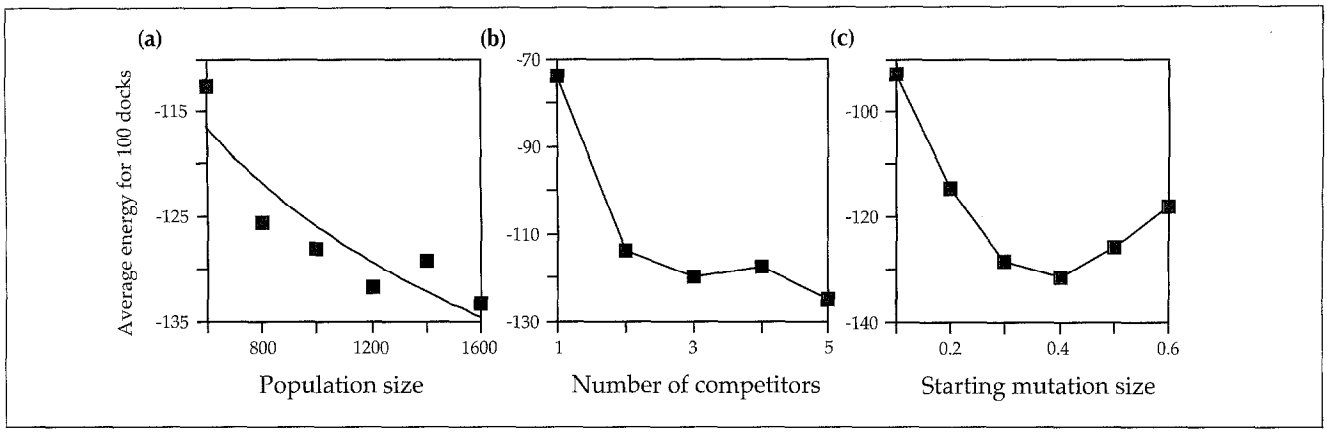

Fig. 7. Docking of MTX into DHFR. Each data point represents the average energy of 100 docks of MTX into DHFR. **(a)** Effect of varying population. A logarithmic fit is shown (r = 0.907). **(b)** Effect of varying the number of competitors in the stochastic tournament. **(c)** Effect of varying the initial standard deviations of the Gaussian mutations used to generate offspring from parents. Units are Angstroms for positional degrees of freedom and radians for rotational degrees of freedom.

convergence was three (Fig. 7b), and this value was used for both systems studied.

The standard deviation of the Gaussian mutations used to generate offspring affects the size of mutations. If the mutations are too small, the system does not efficiently explore the search space. If the mutations are too large, offspring bear little or no resemblance to the parent, and the search is undirected. Although a deterministic method for the calculation of mutation sizes based on the value of the energy function can be effective on some surfaces when the global minimum is known [19], such methods are not generally applicable to arbitrary potentials. Large mutations are likely to be beneficial early in the simulation when the object is to locate the minima, while small mutations will be productive late in the simulation when the algorithm refines solutions near to the global minimum. This process resembles simulated annealing, where large mutations are analogous to high temperature, and small mutations to low temperature. However, it is difficult to predict the most appropriate scaling scheme for the mutations. Consequently, we chose to use a self-adaptive strategy in which the mutation sizes are allowed to vary, with selection pressure determining optimal values as the simulation progresses [29]. Standard deviations of Gaussian mutations σ'_i for each variable were generated from parent values σ_i as follows:

$$\sigma'_i = \sigma_i \exp(\alpha N(0,1) + \beta N_i(0,1))$$

where $\alpha = 1/\sqrt{2n}$, $\beta = 1/\sqrt{2\sqrt{n}}$, $N(0,1)$ is a Gaussian random number with zero mean and unit variance and $N_i(0,1)$ is a different random number for each component of the encoded vector. n is the number of variables; n = 15 in the case of AG-1343. α affects the deviation size of the offspring, while β affects the deviation of individual components of the vector. Based on tuning (Fig. 7c), an initial mutation size of 0.4 Å for positional degrees of freedom and 0.4 radians for rotational degrees of freedom is optimal.

References

1. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. & Ferrin, T.E. (1982). A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* **161**, 269–288.
2. Goodsell, D.S. & Olson, A.J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins: Struct. Funct. Genet.* **8**, 195–202.
3. Cherfils, J. & Janin, J. (1993). Protein docking algorithms: simulating molecular recognition. *Curr. Opin. Struct. Biol.* **3**, 265–269.
4. Caflisch, A., Niederer, P. & Anliker, M. (1992). Monte Carlo docking of oligopeptides to proteins. *Proteins: Struct. Funct. Genet.* **13**, 223–230.
5. Walls, P.H. & Sternberg, M.J.E. (1992). New algorithm to model protein–protein recognition based on surface complementarity. Applications to antibody–antibody docking. *J. Mol. Biol.* **228**, 277–297.
6. Hart, T.N. & Read, R.J. (1992). A multiple-start Monte Carlo docking method. *Proteins: Struct, Funct. Genet.* **13**, 206–222.
7. Knegtel, R.M.A., Antoon, J., Rullmann, C., Boelens, R., & Kaptein, R. (1994). MONTY: a Monte Carlo approach to protein–DNA recognition. *J. Mol. Biol.* **235**, 318–324.
8. Leach, A.R. (1994). Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **235**, 345–356.
9. Totrov, M. & Abagyan, R. (1994). Detailed *ab initio* prediction of lysozyme–antibody complex with 1.6 Å accuracy. *Nat. Struct. Biol.* **1**, 259–263.
10. Zacharias, M., Luty, B.A., David, M.E. & McCammon, J.A. (1994). Combined conformational search and finite-difference Poisson–Boltzmann approach for flexible docking: application to an operator mutation in the lambda repressor–operator complex. *J. Mol. Biol.* **238**, 455–465.
11. Judson, R.S., Jaeger, E.P. & Treasurywala, A.M. (1994). A genetic algorithm based method for docking flexible molecules. *J. Mol. Struct.* **308**, 191–206.
12. Xiao, Y.L. & Williams, D.E. (1994). Genetic algorithms for docking of actinomycin D and deoxyguanosine molecules with comparison to the crystal structure of actinomycin D–deoxyguanosine complex. *J. Phys. Chem.* **98**, 7191–7200.
13. Kuntz, I.D., Meng, E.C. & Shoichet, B.K. (1994). Structure-based molecular design. *Accounts Chem. Res.* **27**, 117–123.
14. Appelt, K., *et al.*, & White, J. (1991). Design of enzyme inhibitors using iterative protein crystallographic analysis. *J. Med. Chem.* **34**, 1925–1934.
15. Appelt, K. (1993). Crystal structures of HIV-1 protease–inhibitor complexes. *Perspect. Drug Discovery Des.* **1**, 23–48.
16. Wlodawer, A., & Erickson, J.W. (1993). Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* **62**, 543–585.
17. Goldstein, R.A., Luthey-Schulten, Z.A. & Wolynes, P.G. (1992). Protein tertiary structure recognition using optimized Hamiltonians with local interactions. *Proc. Natl. Acad. Sci. USA* **89**, 9029–9033.
18. Fogel, L.J., Owens, A.J. & Walsh, M.J. (1966). *Artificial Intelligence Through Simulated Evolution.* Wiley, New York.
19. Fogel, D.B. (1995). *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence.* IEEE Press, Piscataway.
20. Bowie, J. U. & Eisenberg, D. (1994). An evolutionary approach to folding small α-helical proteins that uses sequence information and empirical guiding fitness function. *Proc. Natl. Acad. Sci. USA* **91**, 4436–4440.
21. Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C. & Kraut, J. (1982). Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution. *J. Biol. Chem.*

257, 13650–13662.
22. Blaney, J.M. & Dixon, J.S. (1993). A good ligand is hard to find: automated docking methods. *Perspectives in Drug Design* **1**, 301–319.
23. Li, Z., *et al.*, & Cohen, F.E. (1994). Anti-malarial drug development using models of enzyme structure. *Chemistry & Biology* **1**, 31–37.
24. Gehlhaar, D.K., Moerder, K.E., Zichi, D., Sherman, C.J., Ogden, R.C. & Freer, S.T. (1995). *De novo* design of enzyme inhibitors by Monte Carlo ligand generation. *J. Med. Chem.* **38**, 466–472.
25. Gehlhaar, D.K., Verkhivker, G., Rejto, P.A., Fogel, D.B., Fogel, L.J. & Freer, S.T. (1995). Docking conformationally flexible small molecules into a protein binding site through evolutionary programming. In *Proceedings of the Fourth Annual Conference on Evolutionary Programming*. MIT press, Cambridge, in press.
26. Reich, S.H., *et al.*, & Shetty, B. (1995). Protein structure-based design of potent, orally bioavailable, nonpeptidic inhibitors of human immunodeficiency virus protease. *Proc. Natl. Acad. Sci. USA* **92**, 3298–3302.
27. Weiner, S.J., *et al.*, & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784.
28. Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (1992). *Numerical Recipes in C. The Art of Numerical Computing*. (2nd edn), Cambridge University Press, Cambridge.
29. Schwefel, H.-P. (1981). *Numerical Optimization of Computer Models*. Wiley, Chichester.

Received: **4 Apr 1995**; revisions requested: **13 Apr 1995**; revisions received: **21 Apr 1995**. Accepted **1 May 1995**.