

AutoDock-related material

1. Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. *J. Comput. Chem.*, **1998**, *19*, 1639-1662.
2. Huey, R., Morris, G. M., Olson, A. J. and Goodsell, D. S. A Semiempirical Free Energy Force Field with Charge-Based Desolvation. *J. Comput. Chem.*, **2007**, *28*, 1145-1652.
3. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. and Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785-2791.

Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function

GARRETT M. MORRIS,¹ DAVID S. GOODSSELL,¹
ROBERT S. HALLIDAY,² RUTH HUEY,¹ WILLIAM E. HART,³
RICHARD K. BELEW,⁴ ARTHUR J. OLSON¹

¹Department of Molecular Biology, MB-5, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037-1000

²Hewlett-Packard, San Diego, California

³Applied Mathematics Department, Sandia National Laboratories, Albuquerque, NM

⁴Department of Computer Science & Engineering, University of California, San Diego, La Jolla, CA

Received February 1998; accepted 24 June 1998

ABSTRACT: A novel and robust automated docking method that predicts the bound conformations of flexible ligands to macromolecular targets has been developed and tested, in combination with a new scoring function that estimates the free energy change upon binding. Interestingly, this method applies a Lamarckian model of genetics, in which environmental adaptations of an individual's phenotype are reverse transcribed into its genotype and become heritable traits (*sic*). We consider three search methods, Monte Carlo simulated annealing, a traditional genetic algorithm, and the Lamarckian genetic algorithm, and compare their performance in dockings of seven protein–ligand test systems having known three-dimensional structure. We show that both the traditional and Lamarckian genetic algorithms can handle ligands with more degrees of freedom than the simulated annealing method used in earlier versions of AUTODOCK, and that the Lamarckian genetic algorithm is the most efficient, reliable, and successful of the three. The empirical free energy function was calibrated using a set of 30 structurally known protein–ligand complexes with experimentally determined binding constants. Linear regression analysis of the observed binding constants in terms of a wide variety of structure-derived molecular properties was performed. The final model had a residual standard error of 9.11 kJ mol⁻¹ (2.177 kcal mol⁻¹) and was chosen as the new energy

Correspondence to: A. J. Olson; e-mail: olson@scripps.edu

Contract/grant sponsor: National Institutes of Health, contract/grant numbers: GM48870, RR08065

function. The new search methods and empirical free energy function are available in AUTODOCK, version 3.0. © 1998 John Wiley & Sons, Inc. J Comput Chem 19: 1639–1662, 1998

Keywords: automated docking; binding affinity; drug design; genetic algorithm; flexible small molecule protein interaction

Introduction

A fast atom-based computational docking tool is essential to most techniques for structure-based drug design.^{1,2} Reported techniques for automated docking fall into two broad categories: matching methods and docking simulation methods.³ Matching methods create a model of the active site, typically including sites of hydrogen bonding and sites that are sterically accessible, and then attempt to dock a given inhibitor structure into the model as a rigid body by matching its geometry to that of the active site. The most successful example of this approach is DOCK,^{4,5} which is efficient enough to screen entire chemical databases rapidly for lead compounds. The second class of docking techniques model the docking of a ligand to a target in greater detail: the ligand begins randomly outside the protein, and explores translations, orientations, and conformations until an ideal site is found. These techniques are typically slower than the matching techniques, but they allow flexibility within the ligand to be modeled and can utilize more detailed molecular mechanics to calculate the energy of the ligand in the context of the putative active site. They allow computational chemists to investigate modifications of lead molecules suggested by the chemical intuition and expertise of organic synthetic chemists.

AUTODOCK^{6,7} is an example of the latter, more physically detailed, flexible docking technique. Previous releases of AUTODOCK combine a rapid grid-based method for energy evaluation,^{8,9} precalculating ligand–protein pairwise interaction energies so that they may be used as a look-up table during simulation, with a Monte Carlo simulated annealing search^{10,11} for optimal conformations of ligands. AUTODOCK has been applied with great success in the prediction of bound conformations of enzyme–inhibitor complexes,^{12,13} peptide–antibody complexes,¹⁴ and even protein–protein interactions¹⁵; these and other applications have been reviewed elsewhere.¹⁶

We initiated the current work to remedy two limitations of AUTODOCK. (i) We have found that the simulated annealing search method performs well with ligands that have roughly eight rotatable bonds or less: problems with more degrees of freedom rapidly become intractable. This demanded a more efficient search method. (ii) AUTODOCK is often used to obtain unbiased dockings of flexible inhibitors in enzyme active sites: in computer-assisted drug-design, novel modifications of such lead molecules can be investigated computationally. Like many other computational approaches, AUTODOCK performs well in predicting relative quantities and rankings for series of similar molecules; however, it has not been possible to estimate in AUTODOCK whether a ligand will bind with a millimolar, micromolar, or nanomolar binding constant. Earlier versions of AUTODOCK used a set of traditional molecular mechanics force-field parameters that were not directly correlated with observed binding free energies; hence, we needed to develop a force field that could be used to predict such quantities.

Molecular docking is a difficult optimization problem, requiring efficient sampling across the entire range of positional, orientational, and conformational possibilities. Genetic algorithms (GA) fulfill the role of global search particularly well, and are increasingly being applied to problems that suffer from combinatorial explosions due to their many degrees of freedom. Both canonical genetic algorithms^{17–21} and evolutionary programming methods²² have been shown to be successful in both drug design and docking.

In this report, we describe two major advances that are included in the new release of AUTODOCK, version 3.0. The first is the addition of three new search methods: a genetic algorithm; a local search method; and a novel, adaptive global–local search method based on Lamarckian genetics, the Lamarckian genetic algorithm (LGA). The second advance is an empirical binding free energy force field that allows the prediction of binding free energies, and hence binding constants, for docked ligands.

Methods

GENETIC ALGORITHMS

*Genetic algorithms*²³ use ideas based on the language of natural genetics and biological evolution.²⁴ In the case of molecular docking, the particular arrangement of a ligand and a protein can be defined by a set of values describing the translation, orientation, and conformation of the ligand with respect to the protein: these are the ligand's *state variables* and, in the GA, each state variable corresponds to a gene. The ligand's state corresponds to the genotype, whereas its atomic coordinates correspond to the phenotype. In molecular docking, the *fitness* is the total interaction energy of the ligand with the protein, and is evaluated using the energy function. Random pairs of individuals are mated using a process of *crossover*, in which new individuals inherit genes from either parent. In addition, some offspring undergo random *mutation*, in which one gene changes by a random amount. *Selection* of the offspring of the current *generation* occurs based on the individual's fitness: thus, solutions better suited to their environment reproduce, whereas poorer suited ones die.

A variety of approaches have been adopted to improve the efficiency of the genetic algorithm. Classical genetic algorithms represent the genome as a fixed-length bit string, and employ *binary* crossover and *binary* mutation to generate new individuals in the population. Unfortunately, in many problems, such binary operators can generate values that are often outside the domain of interest, leading to gross inefficiencies in the search. The use of real encodings helps to limit the genetic algorithm to reasonable domains. Alternative genetic algorithms have been reported²⁵ that employ more complicated representations and more sophisticated operators besides crossover and mutation. Some of these retain the binary representation, but must employ decoders and repair algorithms to avoid building illegal individuals from the chromosome, and these are frequently computationally intensive. However, the search performance of the genetic algorithm can be improved by introducing a local search method.^{26,27}

HYBRID SEARCH METHODS IN AUTODOCK

Earlier versions of AUTODOCK used optimized variants of simulated annealing.^{6,7} Simulated annealing may be viewed as having both global and

local search aspects, performing a more global search early in the run, when higher temperatures allow transitions over energy barriers separating energetic valleys, and later on performing a more local search when lower temperatures place more focus on local optimization in the current valley. AUTODOCK 3.0 retains the functionality of earlier versions, but adds the options of using a genetic algorithm (GA) for global searching, a local search (LS) method to perform energy minimization, or a combination of both, and builds on the work of Belew and Hart.^{27,28} The local search method is based on that of Solis and Wets,²⁹ which has the advantage that it does not require gradient information about the local energy landscape, thus facilitating torsional space search. In addition, the local search method is *adaptive*, in that it adjusts the step size depending upon the recent history of energies: a user-defined number of consecutive failures, or increases in energy, cause the step size to be doubled; conversely, a user-defined number of consecutive successes, or decreases in energy, cause the step size to be halved. The hybrid of the GA method with the adaptive LS method together form the so-called *Lamarckian genetic algorithm* (LGA), which has enhanced performance relative to simulated annealing and GA alone,^{21,26} and is described in detail later. Thus, the addition of these new GA-based docking methods enhances AUTODOCK, and allows problems with more degrees of freedom to be tackled. Furthermore, it is now possible to use the same force field as is used in docking to perform energy minimization of ligands.

IMPLEMENTATION

In our implementation of the genetic algorithm, the chromosome is composed of a string of real-valued genes: three Cartesian coordinates for the ligand translation; four variables defining a quaternion specifying the ligand orientation; and one real-value for each ligand torsion, in that order. Quaternions are used to define the orientation³⁰ of the ligand, to avoid the gimbal lock problem experienced with Euler angles.³¹ The order of the genes that encode the torsion angles is defined by the torsion tree created by AUTOTORS, a preparatory program used to select rotatable bonds in the ligand. Thus, there is a one-to-one mapping from the ligand's state variables to the genes of the individual's chromosome.

The genetic algorithm begins by creating a random population of individuals, where the user

defines the number of individuals in the population. For each random individual in the initial population, each of the three translation genes for x , y , and z is given a uniformly distributed random value between the minimum and maximum x , y , and z extents of the grid maps, respectively; the four genes defining the orientation are given a random quaternion, consisting of a random unit vector and a random rotation angle between -180° and $+180^\circ$; and the torsion angle genes, if any, are given random values between -180° and $+180^\circ$. Furthermore, a new random number generator has been introduced that is hardware-independent.³² It is used in the LS, GA, and LGA search engines, and allows results to be reproduced on any hardware platform given the same seed values. The creation of the random initial population is followed by a loop over generations, repeating until the maximum number of generations or the maximum number of energy evaluations is reached, whichever comes first. A generation consists of five stages: mapping and fitness evaluation, selection, crossover, mutation, and elitist selection, in that order. In the Lamarckian GA, each generation is followed by local search, being performed on a user-defined proportion of the population. Each of these stages is discussed in more detail in what follows.

Mapping translates from each individual's genotype to its corresponding phenotype, and occurs over the entire population. This allows each individual's *fitness* to be evaluated. This is the sum of the intermolecular interaction energy between the ligand and the protein, and the intramolecular interaction energy of the ligand. The physicochemical nature of the energy evaluation function is described in detail later. Every time an individual's energy is calculated, either during global or local search, a count of the total number of energy evaluations is incremented.

This is followed, in our implementation, by proportional *selection* to decide which individuals will reproduce. Thus, individuals that have better-than-average fitness receive proportionally more offspring, in accordance with:

$$n_o = \frac{f_w - f_i}{f_w - \langle f \rangle} \quad f_w \neq \langle f \rangle$$

where n_o is the integer number of offspring to be allocated to the individual; f_i is the fitness of the individual (i.e., the energy of the ligand); f_w is the fitness of the worst individual, or highest energy,

in the last N generations (i.e., N is a user-definable parameter, typically 10); and $\langle f \rangle$ is the mean fitness of the population. Because the worst fitness, f_w , will always be larger than either f_i or $\langle f \rangle$, except when $f_i = f_w$, then for individuals that have a fitness lower than the mean, $f_i < \langle f \rangle$, the numerator in this equation, $f_w - f_i$, will always be greater than the denominator $f_w - \langle f \rangle$, and thus such individuals will be allocated at least one offspring, and thus will be able to reproduce. AUTODOCK checks for $f_w = \langle f \rangle$ beforehand, and if true, the population is assumed to have converged, and the docking is terminated.

Crossover and *mutation* are performed on random members of the population according to user-defined rates of crossover and mutation. First, crossover is performed. Two-point crossover is used, with breaks occurring only between genes, never within a gene—this prevents erratic changes in the real values of the genes. Thus, both parents' chromosomes would be broken into three pieces at the same gene positions, each piece containing one or more genes; for instance, ABC and abc . The chromosomes of the resulting offspring after two-point crossover would be AbC and aBc . These offspring replace the parents in the population, keeping the population size constant. Crossover is followed by mutation; because the translational, orientational, and torsional genes are represented by real variables, the classical bit-flip mutation would be inappropriate. Instead, mutation is performed by adding a random real number that has a Cauchy distribution to the variable, the distribution being given by:

$$C(\alpha, \beta, x) = \frac{\beta}{\pi(\beta^2 + (x - \alpha)^2)}$$

$$\alpha \geq 0, \beta > 0, -\infty < x < \infty$$

where α and β are parameters that affect the mean and spread of the distribution. The Cauchy distribution has a bias toward small deviates, but, unlike the Gaussian distribution, it has thick tails that enable it to generate large changes occasionally.²⁶

An optional user-defined integer parameter *elitism* determines how many of the top individuals automatically survive into the next generation. If the elitism parameter is non-zero, the new population that has resulted from the proportional selection, crossover, and mutation is sorted according to its fitness; the fitness of new individuals

having resulted from crossover and/or mutation is calculated as necessary. Because populations are implemented as heaps, selection of the best n individuals is efficient.

The genetic algorithm iterates over generations until one of the termination criteria is met. At the end of each docking, AUTODOCK reports the fitness (the docked energy), the state variables, and the coordinates of the docked conformation, and also the estimated free energy of binding. AUTODOCK performs the user-specified number of GA dockings, and then carries out conformational cluster analysis on the docked conformations to determine which are similar, reporting the clusters ranked by increasing energy.

LAMARCKIAN GENETIC ALGORITHM

The vast majority of genetic algorithms mimic the major characteristics of Darwinian evolution and apply Mendelian genetics. This is illustrated on the right-hand side of Figure 1 (note the one-way transfer of information from the genotype to the phenotype). However, in those cases where an inverse mapping function exists (i.e., one which yields a genotype from a given phenotype), it is possible to finish a local search by replacing the individual with the result of the local search; see the left-hand side of Figure 1. This is called the Lamarckian genetic algorithm (LGA), and is an allusion to Jean Batiste de Lamarck's (discredited) assertion that phenotypic characteristics acquired during an individual's lifetime can become heritable traits.³³

The most important issues arising in hybrids of local search (LS) techniques with the GA revolve around the *developmental mapping*, which transforms genotypic representations into phenotypic ones.²⁶ The genotypic space is defined in terms of the genetic operators—mutation and crossover in our experiments—by which parents of one generation are perturbed to form their children. The phenotypic space is defined directly by the problem, namely, the energy function being optimized. The local search operator is a useful extension of GA global optimization when there are local "smoothness" characteristics (continuity, correlation, etc.) of the fitness function that local search can exploit. In hybrid GA + LS optimizations, the result of the LS is always used to update the fitness associated with an individual in the GA selection algorithm. If, and only if, the developmental mapping function is *invertible*, will the Lamarckian option—

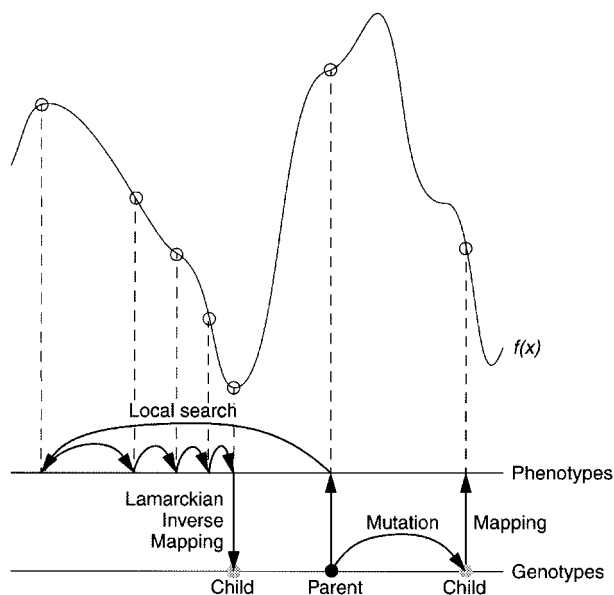


FIGURE 1. This figure illustrates genotypic and phenotypic search, and contrasts Darwinian and Lamarckian search.²⁷ The space of the genotypes is represented by the lower horizontal line, and the space of the phenotypes is represented by the upper horizontal line. Genotypes are mapped to phenotypes by a developmental mapping function. The fitness function is $f(x)$. The result of applying the genotypic mutation operator to the parent's genotype is shown on the right-hand side of the diagram, and has the corresponding phenotype shown. Local search is shown on the left-hand side. It is normally performed in phenotypic space and employs information about the fitness landscape. Sufficient iterations of the local search arrive at a local minimum, and an inverse mapping function is used to convert from its phenotype to its corresponding genotype. In the case of molecular docking, however, local search is performed by continuously converting from the genotype to the phenotype, so inverse mapping is not required. The genotype of the parent is replaced by the resulting genotype, however, in accordance with Lamarckian principles.

converting the phenotypic result of LS back into its corresponding genotype become possible.

In our case, the fitness or energy is calculated from the ligand's coordinates, which together form its phenotype. The genotypic representation of the ligand, and its mutation and crossover operators, have already been described. The developmental mapping simply transforms a molecule's genotypic state variables into the corresponding set of atomic coordinates. A novel feature of this application of hybrid global–local optimization is that the

Solis and Wets LS operator searches through the genotypic space rather than the more typical phenotypic space. This means that the developmental mapping does not need to be inverted. Nonetheless, this molecular variation of the genetic algorithm still qualifies as Lamarckian, because any "environmental adaptations" of the ligand acquired during the local search will be inherited by its offspring.

At each generation, it is possible to let a user-defined fraction of the population undergo such a local search. We have found improved efficiency of docking with local search frequencies of just 0.06, although a frequency of 1.00 is not significantly more efficient.²⁶ Both the canonical and a slightly modified version of the Solis and Wets method have been implemented. In canonical Solis and Wets, the same step size would be used for every gene, but we have improved the local search efficiency by allowing the step size to be different for each type of gene: a change of 1 Å (1 Å = 10⁻¹⁰ m) in a translation gene could be much more significant than a change of 1° in a rotational or torsional gene. In the docking experiments presented here, the translational step size was 0.2 Å, and the orientational and torsional step sizes were 5°.

In the Lamarckian genetic algorithm, genotypic mutation plays a somewhat different role than it does in traditional genetic algorithms. Traditionally, mutation plays the role of a local search operator, allowing small, refining moves that are not efficiently made by crossover and selection alone. With the explicit local search operator, however, this role becomes unnecessary, and is needed only for its role in replacing alleles that might have disappeared through selection. In LGA, mutation can take on a more exploratory role. The Cauchy deviates are a compromise between radical jumps to arbitrary sections of the conformation space and detailed exploration of the local topography.

DERIVATION OF THE EMPIRICAL BINDING FREE ENERGY FUNCTION

The study of molecular structure underpins much of computational molecular biology. There are several established methods for performing molecular mechanics and molecular dynamics, notably AMBER,^{34,35} CHARMM,³⁶ DISCOVER,³⁷ ECEPP,³⁸ and GROMOS.³⁹ Many of these traditional force fields model the interaction energy of a molecular system with terms for dispersion/repulsion,⁴⁰ hydrogen bonding,⁴¹ electrostatics,⁴²⁻⁴⁵ and devia-

tion from ideal bond lengths and bond angles. These methods are excellent for studying molecular processes over time, for optimizing bound conformations, and for performing free energy perturbation calculations between molecules with a single atom change,⁴⁶ but they often require considerable investments of computer time and, unfortunately, these approaches tend to perform less well in ranking the binding free energies of compounds that differ by more than a few atoms. What is needed is an empirical relationship between molecular structure and binding free energy.

The first thoroughly established linear free energy relationship was observed by Hammett as early as 1933, and reported in 1937.⁴⁷ It was used to relate structure and reactivity of small organic molecules on a quantitative basis. Hammett was able to derive substituent constants and reaction constants that could then be used to calculate rate constants and equilibrium constants for a specific reaction of a specific compound. It could be said that Hammett's work was the forerunner of modern-day quantitative structure-activity relationships (QSAR), pioneered by Hansch and coworkers in the 1960s. Here it is assumed that the sum of the steric, electronic, and hydrophobic effects of substituents in a compound determines its biological activity; see, for example, Fujita,⁴⁸ Hansch,⁴⁹ and more recently Selassie et al.⁵⁰

Current structure-based scoring functions seek to remedy some of the deficiencies of traditional force fields by developing empirical free energy functions that reproduce observed binding constants. Most of these approaches use an expanded "master equation" to model the free energy of binding, adding entropic terms to the molecular mechanics equations⁵¹:

$$\Delta G = \Delta G_{\text{vdw}} + \Delta G_{\text{hbond}} + \Delta G_{\text{elec}} + \Delta G_{\text{conform}} + \Delta G_{\text{tor}} + \Delta G_{\text{sol}}$$

where the first four terms are the typical molecular mechanics terms for dispersion/repulsion, hydrogen bonding, electrostatics, and deviations from covalent geometry, respectively; ΔG_{tor} models the restriction of internal rotors and global rotation and translation; and ΔG_{sol} models desolvation upon binding and the hydrophobic effect (solvent entropy changes at solute-solvent interfaces). This latter term is the most challenging. Most workers use variants of the method of Wesson and Eisen-

berg,⁵² calculating a desolvation energy based on the surface area buried upon complex formation, with the area of each buried atom being weighted by an atomic solvation parameter. Böhm built on earlier work with the *de novo* inhibitor design program LUDI,⁵³ and used linear regression to calibrate a similar function against a set of 45 diverse protein–ligand complexes with published binding constants.⁵⁴ The final function predicted binding constants for a set of test complexes with a standard deviation equivalent to about a factor of 25 in binding constant: more than sufficient to rank inhibitors with millimolar, micromolar, and nanomolar binding constants. Jain devised a continuous, differentiable scoring function,⁵⁵ which is, in essence, very similar to that of Böhm, but based on non-physical pairwise potentials using Gaussians and sigmoidal terms.

We have implemented a similar approach using the thermodynamic cycle of Wesson and Eisenberg.⁵² The function includes five terms:

$$\begin{aligned} \Delta G = & \Delta G_{\text{vdW}} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \\ & + \Delta G_{\text{hbond}} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) \\ & + \Delta G_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \\ & + \Delta G_{\text{tor}} N_{\text{tor}} \\ & + \Delta G_{\text{sol}} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)} \quad (1) \end{aligned}$$

where the five ΔG terms on the right-hand side are coefficients empirically determined using linear regression analysis from a set of protein–ligand complexes with known binding constants, shown in Table I. The summations are performed over all pairs of ligand atoms, i , and protein atoms, j , in addition to all pairs of atoms in the ligand that are separated by three or more bonds.

The *in vacuo* contributions include three interaction energy terms, used in previous versions of AUTODOCK: a Lennard–Jones 12-6 dispersion/repulsion term; a directional 12–10 hydrogen bonding term, where $E(t)$ is a directional weight based on the angle, t , between the probe and the target atom⁹; and a screened Coulombic electrostatic potential.⁵⁶ Each of these terms, including their parameterization, have already been described.⁷

A measure of the unfavorable entropy of ligand binding due to the restriction of conformational

degrees of freedom is added to the *in vacuo* function. This term is proportional to the number of sp^3 bonds in the ligand, N_{tor} .⁵⁴ We investigated variants that included and excluded methyl, hydroxyl, and amine rotors.

In the development of an empirical free energy function for AUTODOCK, the desolvation term was most challenging, because AUTODOCK uses a grid-based method for energy evaluation, and most published solvation methods are based on surface area calculations. We investigated two different methods of calculating the desolvation energy term. The first of these methods was based on estimating atom-by-atom contributions to the interfacial molecular surface area between the ligand and the protein using the difference in the surface areas of the complex and the unbound protein and unbound ligand. Both the solvent-accessible and solvent-excluded surface areas were considered, being calculated with MSMS,⁵⁷ a fast and reliable program that computes analytical molecular surfaces. Unfortunately, there can be significant errors in the value of the interfacial solvent-accessible surface areas, due to the “collar” of accessible surface that surrounds the ligand–protein interface in the complex. We also tested seven variants of the pairwise, volume-based method of Stouten et al.⁵⁸: this method has the advantage that it is consistent with the pre-calculated affinity grid formulation used by AUTODOCK. For each atom in the ligand, fragmental volumes of surrounding protein atoms are weighted by an exponential function and then summed, evaluating the percentage of volume around the ligand atom that is occupied by protein atoms. This percentage is then weighted by the atomic solvation parameter of the ligand atom to give the desolvation energy. The full method may be broken into four separate components: burial of apolar atoms in the ligand, burial of apolar protein atoms, burial of polar and charged atoms in the ligand, and burial of polar and charged protein atoms. Great success has also been reported in using simply the amount of hydrophobic surface area buried upon complexation as a measure of the “hydrophobic effect,”⁵⁴ so we tested several formulations that included only the volume lost around ligand carbon atoms. The burial of polar atoms caused particular problems, as discussed in what follows. Apart from the volume-based method, we tested a simpler formulation for the solvent transfer of polar atoms; that is, a constant term corresponding to the favorable free energy of interaction of a polar atom with solvent

TABLE I.
Protein–Ligand Complexes Used to Calibrate Empirical Free Energy Function, Along with Brookhaven Protein Data Bank (PDB) Accession Codes and Binding.

Protein–ligand complex	PDB code	Log(K_i) ^a
Concanavalin A / α -methyl-D-mannopyranoside	4cna	2.00
Carboxypeptidase A / glycyl-L-tyrosine	3cpa	3.88
Carboxypeptidase A / phosphonate ZAA=P=(O)F	6cpa	11.52
Cytochrome P-450 _{cam} / camphor	2cpp	6.07
Dihydrofolate reductase / methotrexate	4dfr	9.70
α -Thrombin / benzamidine	1dwb	2.92
Endothiapepsin / H-256	zer6	7.22
ε -Thrombin / MQPA	1etr	7.40
ε -Thrombin / NAPAP	1ets	8.52
ε -Thrombin / 4-TAPAP	1ett	6.19
FK506-binding protein (FKBP) / immunosuppressant FK506	1fkf	9.70
D-Galactose / D-glucose binding protein / galactose	2gbp	7.60
Hemagglutinin / sialic acid	4hmg	2.55
HIV-1 Protease / A78791	1hvj	10.46
HIV-1 Protease / MVT101	4hvp	6.15
HIV-1 Protease / acylpepstatine	5hvp	5.96
HIV-1 Protease / XK263	1hr	9.51
Fatty-acid-binding protein / C ₁₅ COOH	2ifb	5.43
Myoglobin (ferric) / imidazole	1mbi	1.88
McPC603 / phosphocholine	2mcp	5.23
β -Trypsin / benzamidine	3ptb	4.74
Retinol-binding protein / retinol	1rbp	6.72
Thermolysin / Leu-hydroxylamine	4tln	3.72
Thermolysin / phosphoramidon	1tlp	7.55
Thermolysin / <i>n</i> -(1-carboxy-3-phenylpropyl)-Leu-Trp	1tmn	7.30
Thermolysin / Cbz-Phe- <i>p</i> -Leu-Ala (ZFP ₁ LA)	4tmn	10.19
Thermolysin / Cbz-Gly- <i>p</i> -Leu-Leu (ZGpLL)	5tmn	8.04
Purine nucleoside phosphorylase (PNP) / guanine	1ulb	5.30
Xylose isomerase / CB3717	2xis	5.82
Triose phosphate isomerase (TIM) / 2-phosphoglycolic acid (PGA)	2ypi	4.82

^a Adapted from Böhm.⁵⁴

is estimated, and this is subtracted from the binding free energy.

Trilinear interpolation is used to evaluate rapidly the intermolecular dispersion/repulsion energy, the hydrogen bonding energy, the electrostatic potential, and the solvation energy of each atom in the ligand, using grid maps that have been pre-calculated over the protein for each atom type in the ligand. In AUTODOCK 3.0, we have implemented a faster method of trilinear interpolation⁵⁹ than was available in earlier versions of AUTODOCK. Both methods are mathematically equivalent. The original implementation used 24 multiplications to perform each three-dimensional trilinear interpolation, but, by cascading seven one-dimensional interpolations, the number of multiplications has been reduced to 7.

Thirty protein–ligand complexes with published binding constants were used in the calibration of AUTODOCK's free energy function (Table I), and were chosen from the set of 45 used by Böhm,⁵⁴ omitting all complexes that he modeled (i.e., using only complexes for which crystallographic structures were available). One of the limitations of these binding constant data is that the conditions under which they were determined vary, which intrinsically limits the accuracy of our best model. We converted between the inhibition constant, K_i , and the observed free energy change of binding, ΔG_{obs} , using the equation:

$$\Delta G_{\text{obs}} = RT \ln K_i$$

where R is the gas constant, 1.987 cal K⁻¹ mol⁻¹, and T is the absolute temperature, assumed to be

room temperature, 298.15 K.⁶⁰ Note that this equation lacks a minus sign because the inhibition constant is defined for the dissociation reaction, $EI \rightleftharpoons E + I$,⁶¹ whereas ΔG_{obs} refers to the opposite process of binding, $E + I \rightleftharpoons EI$; where E is the enzyme and I is the inhibitor.

To remove any steric clashes in the crystallographic complexes, each ligand was optimized using AUTODOCK's new Solis and Wets local minimization technique described earlier, but with the previously reported force field.⁷ The separate contributions from the hydrogen bonding, dispersion/repulsion, electrostatic, and solvation energies were evaluated. Empirical free energy coefficients for each of these terms were derived using linear regression in the S-PLUS software package,⁶² and cross-validation studies were performed. In total, 900 different binding free energy models were tested: each linear model consisted of a van der Waals term, a hydrogen bonding term (one of 6 variants), an electrostatic term, a torsional entropy term (one of 5 variants), and a desolvation term (one of 15 variants). We also investigated whether the inclusion of a constant term improved the model. Six of the seven test systems used to test the docking procedure, which were originally used to test AUTODOCK, version 2.4,⁷ were also in the training set of 30 protein–ligand complexes; therefore, to validate the chosen coefficients, linear regression was repeated for the set of 24 protein–ligand complexes, excluding the 6 overlapping test systems.

TESTING DOCKING METHODS

Seven protein–ligand complexes, with a range of complexity and chemical properties, were chosen from the Brookhaven Protein Data Bank^{63,64} to compare the performance of the docking techniques (see Fig. 2). To facilitate comparison with the previous force field, we chose the same set of six test systems investigated earlier,⁷ but added a harder docking problem to challenge all the search methods (see Table II). The simplest test cases were the β -trypsin/benzamidine and cytochrome P-450_{cam}/camphor complexes, which had small, rigid ligands. Interactions in the former are dominated by electrostatic interactions and hydrogen bonds to the substrate amidine, whereas the latter is dominated by hydrophobic interactions. McPC-603/phosphocholine and streptavidin/biotin were moderately flexible, and represented test systems having an intermediate level of difficulty. HIV-1

protease/XK263, hemagglutinin/sialic acid, and dihydrofolate reductase/methotrexate provided more difficult tests, with many rotatable bonds and diverse chemical characteristics.

We compared the performance of Monte Carlo simulated annealing (SA), the genetic algorithm (GA), and the Lamarckian genetic algorithm (LGA). The new empirical free energy function presented here was used for energy evaluation in all cases. Dockings were performed using approximately the same number of energy evaluations (~ 1.5 million), so each method could be judged given similar computational investments. The CPU time taken for a single docking varied from 4.5 to 41.3 minutes, on a 200-MHz Silicon Graphics MIPS 4400 with 128 MB of RAM, depending on the number of rotatable bonds and the number of atoms in the ligand.

At the end of a set of dockings, the docked conformations were exhaustively compared to one another to determine similarities, and were clustered accordingly. The user-defined root-mean-square positional deviation (rmsd) tolerance was used to determine if two docked conformations were similar enough to be included in the same cluster, and symmetrically related atoms in the ligand were considered. These clusters were ranked in order of increasing energy, by the lowest energy in each cluster. Ordinarily, the structure of the protein–ligand complex would not be known, so the criteria by which the dockings would be evaluated are the energies of the docked structures, and, in cases where there are several plausible, low-energy structures, the number of conformations in a conformationally similar cluster. Because one of our goals was to test the ability of the methods to reproduce known structures, we also compared the rmsd between the lowest energy docked structure and the crystallographic structure.

DOCKING- AND SEARCH-METHOD-SPECIFIC PARAMETERS

The proteins and ligands in the seven docking tests were treated using the united-atom approximation, and prepared using the molecular modeling program, SYBYL.⁶⁵ Only polar hydrogens were added to the protein, and Kollman united-atom partial charges were assigned. Unless stated otherwise, all waters were removed. Atomic solvation parameters and fragmental volumes were assigned to the protein atoms using a new AUTODOCK utility, ADDSOL. The grid maps were calculated using AUTOGRID, version 3.0. In all seven protein–ligand

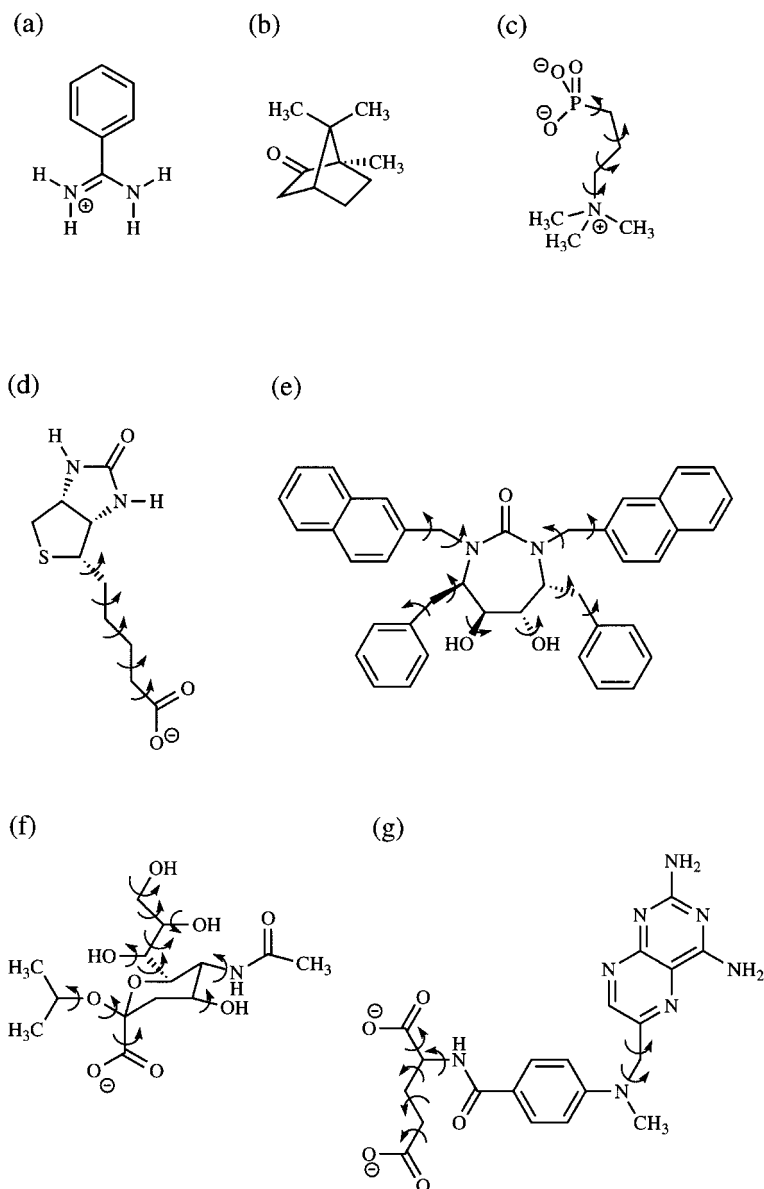


FIGURE 2. The seven ligands chosen for docking, showing the rotatable bonds as curly arrows: (a) benzamidine; (b) camphor; (c) phosphocholine; (d) biotin; (e) HIV-1 protease inhibitor XK-263; (f) isopropylated sialic acid; and (g) methotrexate. Note that two ligands, (e) and (f), contain hydroxyl rotors, which are not counted in the total number of torsional degrees of freedom; note also that cyclic rotatable bonds are excluded.

cases, we used grid maps with $61 \times 61 \times 61$ points, a grid-point spacing of 0.375 \AA , and, because the location of the ligand in the complex was known, the maps were centered on the ligand's binding site. The ligands were treated in SYBYL initially as all atom entities, that is, all hydrogens were added, then partial atomic charges were calculated using the Gasteiger–Marsili method.^{66,67} AUTOTORS, an AUTODOCK utility, was used to define the rotatable bonds in the ligand, if any, and also to unite the

nonpolar hydrogens added by SYBYL for the partial atomic charge calculation. The partial charges on the nonpolar hydrogens were added to that of the hydrogen-bearing carbon also in AUTOTORS.

In all three search methods, 10 dockings were performed; in the analysis of the docked conformations, the clustering tolerance for the root-mean-square positional deviation was 0.5 \AA , and the crystallographic coordinates of the ligand were used as the reference structure. For all three search

TABLE II. X-Ray Crystal Structure Coordinates Used in Docking Experiments, Their Brookhaven Protein Data Bank Accession Codes and Resolution, Number of Rotatable Bonds in the Ligand, Number of Torsional Degrees of Freedom, Total Number of Degrees of Freedom, and Energy of Crystal Structure Using the Empirical Force Field Presented Here.

Protein–ligand complex	PDB code	Resolution (Å)	Reference	Number of rotatable bonds	N_{tor}^a	Total number of degrees of freedom	Energy of crystal structure (kcal mol ⁻¹)
β -Trypsin / benzamidine	3ptb	1.7	69	0	0	7	-7.86
Cytochrome P-450 _{cam} / camphor	2cpp	1.63	70	0	0	7	-4.71
McPC-603 / Phosphocholine	2mcp	3.1	71	4	4	11	+5.48 ^b
Streptavidin / biotin	1stp	2.6	73	5	5	12	-8.86
HIV-1 protease / XK263	1hvr	1.8	75	10	8	17	-18.62
Influenza hemagglutinin / sialic acid	4hmg	3.0	76	11	7	18	-4.71
Dihydrofolate reductase / methotrexate	4dfr	1.7	79	7	7	14	-13.64

^a N_{tor} is the number of torsional degrees of freedom used in the calculation of the predicted free energy change of binding, ΔG_{pred} . Note that this excludes rotatable bonds that only move hydrogens, such as hydroxyl, amino, and methyl groups.

^b This energy is dominated by a large, positive contribution from C2 and O1 to the internal nonbonded energy, of +6.13 kcal mol⁻¹; these atoms are 2.26 Å apart.

methods, the step sizes were 0.2 Å for translations and 5° for orientations and torsions. These step sizes determined the amount by which a state variable could change when a move is made in simulated annealing and the relative size of mutation in the local search, whereas the α and β parameters determined the size of the mutation in the genetic algorithms, GA and LGA. The Cauchy distribution parameters were $\alpha = 0$ and $\beta = 1$. Note that in simulated annealing, random changes were generated by a uniformly distributed random number generator; in the Solis and Wets local search, by a normal distribution; and, in the genetic algorithm, by a Cauchy distribution. In the simulated annealing tests, the initial state of the ligand was chosen randomly by AUTODOCK. We used the optimal set of simulated annealing parameters that were determined from the schedule experiments described earlier.⁷ These included an initial annealing temperature of 616 cal mol⁻¹, a linear temperature reduction schedule, 10 runs, 50 cycles, and a cycle-termination criterion of a maximum of 25,000 accepted steps or 25,000 rejected steps, whichever came first. The minimum energy state was used to begin the next cycle; the only exception was for 1hvr, where the initial annealing temperature was increased to 61,600 cal mol⁻¹. The maximum initial energy allowed was 0.0 kcal mol⁻¹, and the maximum number of retries was 1000, used to generate a low energy random initial state to begin each simulated annealing docking.

In the GA and LGA dockings, we used an initial population of random individuals with a population size of 50 individuals; a maximum number of 1.5×10^6 energy evaluations; a maximum number of generations of 27,000; an elitism value of 1, which was the number of top individuals that automatically survived into the next generation; a mutation rate of 0.02, which was the probability that a gene would undergo a random change; and a crossover rate of 0.80, which was the probability that two individuals would undergo crossover. Proportional selection was used, where the average of the worst energy was calculated over a window of the previous 10 generations. In the LGA dockings, the pseudo-Solis and Wets local search method was used, having a maximum of 300 iterations per local search; the probability of performing local search on an individual in the population was 0.06; the maximum number of consecutive successes or failures before doubling or halving the local search step size, ρ , was 4, in both cases; and the lower bound on ρ , the termination criterion for the local search, was 0.01.

Results and Discussion

CALIBRATION OF EMPIRICAL FREE ENERGY FUNCTION

Several linear regression models were tested for their ability to reproduce the observed binding

TABLE III.
Calibration of Empirical Free Energy Function.

Model ^a	Residual standard error	Multiple R^2	ΔG_{vdw}^b	ΔG_{estat}	ΔG_{hbond}	ΔG_{tor}	ΔG_{solv}
A	2.324	0.9498	0.1795 (0.0263)	0.1133 (0.0324)	0.0166 (0.0625)	0.3100 (0.0873)	0.0101 (0.0585)
B	2.232	0.9537	0.1518 (0.0269)	0.1186 (0.0246)	0.0126 (0.0382)	0.3548 (0.0890)	0.1539 (0.1050)
C	2.177	0.9559	0.1485 (0.0237)	0.1146 (0.0238)	0.0656 (0.0558)	0.3113 (0.0910)	0.1711 (0.1035)

^a Models differ in the formulation of the solvation term and the hydrogen bonding term. Model A: full volume-based solvation term and standard 10–12 hydrogen bonding, as in Eq. (1). Model B: apolar ligand atoms only in the solvation term, and standard 10–12 hydrogen bonding. Model C: apolar ligand atoms only in the solvation term, and the standard 10–12 hydrogen less the estimated average, as in Eq. (2).

^b Values for the model coefficients, with standard deviations in parentheses.

constants of structurally characterized complexes. Table III shows the results for the three major candidates, and Figure 3 shows the correlation between the observed and the predicted binding free energies for the 30 protein–ligand complexes

in the calibration set, using the chosen model (model C). Model A adds the full volume-based solvation method and the torsional restriction term to the original molecular mechanics force field. Model B simplifies the solvation method by evalu-

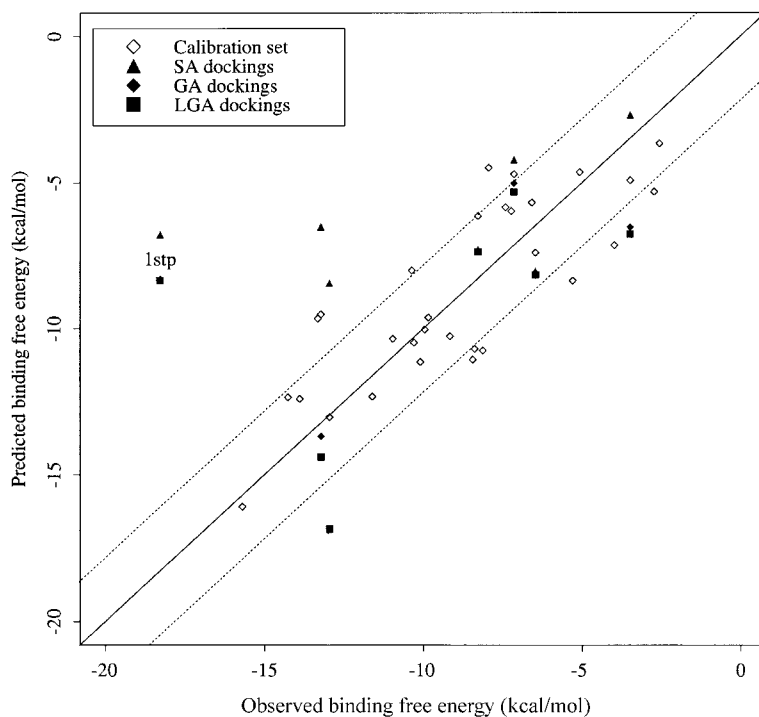


FIGURE 3. Predicted versus observed binding free energies for the calibration set and the docking tests. The solid line shows a perfect fit, and the dotted lines show one standard deviation above and below this. Hollow diamonds show the 30 protein–ligand complexes used in fitting the terms of the binding free energy function. Solid triangles show the results of the simulated annealing (SA) dockings, solid diamonds show the genetic algorithm (GA) dockings, and the solid squares show the Lamarckian genetic algorithm (LGA) dockings. Note the outlying biotin–streptavidin complex (1stp), where it is believed there are significant contributions to the binding free energy due to protein rearrangements.

ating the volume buried for only the carbon atoms in the ligand. Model C also uses only ligand carbon atoms in the desolvation calculation, and also adds a constant term to the hydrogen bonding function, modeling desolvation of polar atoms. Model C was chosen for incorporation into AUTODOCK 3.0, based on its better overall statistics, and on criteria discussed in what follows. The form of this free energy function is:

$$\begin{aligned} \Delta G = & \Delta G_{\text{vdW}} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \\ & + \Delta G_{\text{hbond}} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{\text{hbond}} \right) \\ & + \Delta G_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \\ & + \Delta G_{\text{tor}} N_{\text{tor}} \\ & + \Delta G_{\text{sol}} \sum_{i_c, j} S_i V_j e^{(-r_{ij}^2/2\sigma^2)} \end{aligned} \quad (2)$$

where E_{hbond} is the estimated average energy of hydrogen bonding of water with a polar atom, and the summation in the solvation term is performed over all pairs consisting of only carbon atoms in the ligand, i , and atoms of all types, j , in the protein. Note that the internal or intramolecular interaction energy of the ligand is *not* included in the calculation of binding free energy; during docking, however, internal energy *is* included in the total docked energy, because changes in ligand conformation can affect the outcome of the docking, so this must be taken into consideration. We looked at linear regression models that did include the internal energy, and found that adding this term did not improve the model. The assumption made is that the internal energy of the ligand in solution and in the complex are the same. The energies used and reported by AUTODOCK should be distinguished: there are *docked energies*, which include the intermolecular and intramolecular interaction energies, and which are used during dockings; and *predicted free energies*, which include the intermolecular energy and the torsional free energy, and are only reported at the end of a docking. Because the intermolecular energy grid maps include the desolvation term, dockings using the new, empirical force field in AUTODOCK version 3.0 may be qualitatively different from results found using earlier versions.

Three coefficients, for dispersion/repulsion, electrostatics, and loss of torsional freedom were

very stable in the linear regression analysis, with consistent coefficient values in different formulations and reasonable standard deviations. In our best model, dispersion/repulsion energies, with parameters taken from AMBER,³⁴ were weighted by a factor of 0.1485, yielding an energy of about $-0.2 \text{ kcal mol}^{-1}$ for the most favorable atom–atom contacts. Electrostatics, modeled with a screened Coulomb potential,⁵⁶ were weighted by a factor of 0.1146, yielding an energy of about $-1.0 \text{ kcal mol}^{-1}$ for an ideal salt bridge. In the torsional restriction term, each torsional degree of freedom requires $0.3113 \text{ kcal mol}^{-1}$.

The major differences between models occurred with the interaction of the hydrogen bonding term and the desolvation term. Hydrogen bonding is modeled with a directional 12–10 potential.⁶⁸ We encountered a major problem when calibrating this hydrogen bonding function. Because the test set included only natural enzyme–ligand complexes, optimized by millions of years of evolution, hydrogen bonding groups in the ligands are nearly always paired with the appropriate hydrogen bonding group in the protein. Thus, the number of hydrogen bonds that the ligand forms in the complex and the number it forms with solvent when free in solution are approximately the same; that is, there is little change in the free energy of hydrogen bonding, and ΔG_{hbond} was evaluated to be approximately zero. Unfortunately, this provides no information on the cost of burying a hydrogen bonding group without forming a bond with the protein, and our data set did not include cases to evaluate this. Of course, the volume-based solvation method should account for this—the unfavorable polar contribution to the solvation energy should compensate for the favorable 12–10 hydrogen bonding energy. The linear regression, however, consistently returned coefficients that set the hydrogen bonding energy and desolvation energy to nearly zero, and increased the dispersion/repulsion term to compensate (see Model A in Table III). We chose an alternative formulation to resolve this problem.

We obtained the best results by separating the desolvation of polar atoms from the volume-based calculation. We assumed that the extent of hydrogen bonding in the complexes was roughly the same as the extent of hydrogen bonding in solution. The calculated hydrogen bonding energy, using the directional 12–10 hydrogen bonding function, was divided by the maximal number of possible hydrogen bonds, counting two for each oxygen atom and one for each polar hydrogen. For

the 30 complexes used in calibration, it was found that 36% of the maximum possible hydrogen bonding sites were actually utilized. Values of E_{hbond} ranging from 36% to 100% of the maximal well depth of 5 kcal mol⁻¹ had little effect on the success of the formulation (data not shown), and a value of 36% was chosen. Optimized weights yielded an ideal hydrogen bonding energy in the complex of -0.328 kcal mol⁻¹, and the estimated average energy of each hydrogen bond in solution of -0.118 . Because hydrogen bonding was modeled by this difference, the typical hydrogen bonding free energy of a complex was approximately zero, but there was a penalty of about 0.2 kcal mol⁻¹ for oxygen and nitrogen atoms that did not form hydrogen bonds, driving the simulation toward docked conformations with maximal hydrogen bonding. We are currently exploring an appropriate data set for evaluating this formulation more rigorously.

The desolvation for carbon atoms in the ligand was evaluated using two different classes of atom type, aliphatic and aromatic, as in the original study.⁵⁸ The desolvation term was weighted by a factor of 0.1711 in our final empirical free energy force field, so a typical aliphatic carbon atom yields an energy of about -0.2 kcal mol⁻¹ upon binding.

We cross-validated the free energy model in two ways. First, we investigated the influence of each member of the training set on the final coefficients of the model, by removing each one from the training set and calculating the coefficients from the remaining 29 complexes. We found that none of them had a strong effect on the final values of the coefficients.

We also performed a second kind of cross-validation of the free energy model, by performing Solis and Wets local search using AUTODOCK and the new free energy function, starting from the x-ray crystallographic conformations of each inhibitor in 20 HIV-1 protease-inhibitor complexes, to compare the resulting optimized conformations' predicted free energy change of binding, $\Delta G_{\text{binding, pred}}$, with the experimentally determined values. These protease inhibitors were quite different, having from 7 to 28 torsions, and widely different side chains—charged, polar, and hydrophobic, and constituted a diverse test set. As can be seen from the results in Table IX, the correlation was very good, with an overall rmsd between the experimental and calculated values of $\Delta G_{\text{binding}}$ of 1.92 kcal mol⁻¹.

The final form of the free energy function may seem overparameterized, with additional weight-

ing parameters added to a previously optimized parameterization. We retained the molecular mechanics formulation, however, specifically for its ability to model the distance dependence of each energetic term. This distance dependence (and angular dependence in hydrogen bonding) is essential for finding valid docked conformations, but the amount and resolution of the available protein–ligand data do not support a full re-parameterization of the functions.

DOCKING EXPERIMENTS

Because we are comparing different search methods, it is important to ensure that the methods are treated equally. It is therefore important that each search method be allowed approximately the same number of energy evaluations in a docking. The number of energy evaluations in a docking depends on the termination criteria, and because it is not possible to predict how many accepted or rejected steps the stochastic SA method will make at a given temperature, the number of evaluations varies in SA. The range was from 1.19×10^6 to 2.33×10^6 , depending on the protein–ligand test system, even though the same parameters were used for the number of cycles, accepted steps and rejected steps. In the case of the GA dockings, the population was 50 and the number of generations was 27,000, which gave a total of 1.35×10^6 energy evaluations in a docking; thus, the GA dockings were terminated by reaching the maximum number of generations. In the case of the LGA dockings, 6% of the population underwent Lamarckian local search, each search consisting of 300 iterations and each iteration using an extra energy evaluation. Thus, the LGA dockings, even with the same population size and number of generations as the GA, were terminated by reaching the maximum number of energy evaluations, 1.50×10^6 .

The results of the simulated annealing, genetic algorithm, and the Lamarckian genetic algorithm docking experiments are summarized in Tables IV, V and VI, respectively. The lowest energy docked structure found by each method is compared with the crystal structure of the ligand in Figure 4. The predicted change in free energy upon binding, ΔG_{pred} , for the lowest energy found by LGA is shown in Table VII, along with the experimentally observed change in free energy upon binding, ΔG_{obs} . In addition, the breakdown of the energy of the lowest energy docked conformation is shown, in terms of the intermolecular interaction energy,

TABLE IV.
Results of Simulated Annealing Dockings.^a

PDB code	Number of clusters	Number in rank 1	Energy (kcal mol ⁻¹) and rmsd (Å)				Mean rmsd	Number of energy evaluations	
			Lowest energy	rmsd of lowest energy	Mean energy				
3ptb	5	6	-8.03	0.21	-7.84	(0.08)	0.50	(0.17)	2.01 × 10 ⁶
2cpp	4	6	-7.29	0.81	-7.22	(0.03)	0.91	(0.30)	2.33 × 10 ⁶
2mcp	10	1	-4.09	0.88	70.89	(2.10 × 10 ²)	5.40	(4.80)	1.85 × 10 ⁶
1stp	10	1	-8.48	1.27	-7.71	(0.66)	1.24	(0.35)	2.00 × 10 ⁶
1hvr	10	1	-11.77	1.15	1.12 × 10 ⁵	(3.36 × 10 ⁵)	6.13	(2.61)	1.19 × 10 ⁶
4hmg	10	1	-2.59	3.77	6.99 × 10 ⁴	(1.52 × 10 ⁵)	6.20	(2.94)	1.55 × 10 ⁶
4dfr	10	1	-8.73	4.83	6.13 × 10 ²	(1.96 × 10 ³)	5.04	(1.74)	1.30 × 10 ⁶

^a The parameters used were 10 runs, 50 cycles, and a cycle-termination criterion of 25,000 accepted steps or 25,000 rejected steps, whichever came first. The rmsd conformational clustering tolerance was 0.5 Å, calculated from the ligand's crystallographic coordinates. Standard deviations given in parentheses.

ΔG_{inter} the intramolecular energy, ΔG_{intra} , and the torsional free energy, ΔG_{tor} . These results are discussed case-by-case in what follows; "crystallographic rmsd" refers to the root-mean-square positional deviation of a given conformation from the crystallographic coordinates.

β -Trypsin/ Benzamidine (3ptb)

The recognition of benzamidine by β -trypsin, which binds tightly in the specificity pocket of trypsin, is chiefly due to the polar amidine moiety

and the hydrophobic benzyl ring.⁶⁹ The amidine moiety was treated as being protonated. It was assumed that delocalization of the π -electrons of the benzene ring extended to the π -system of the amidine, and thus the ligand was treated as a rigid body. All three search methods succeeded in finding lowest energy conformations that were also the ones with the lowest crystallographic rmsd. In this case, the method that found the docked structure with the lowest energy was GA (Table V), but that found by the LGA method (Table VI) was practically the same. The mean of the final docked

TABLE V.
Results of Genetic Algorithm Dockings.^a

PDB code	Number of clusters	Number in rank 1	Energy (kcal mol ⁻¹) and rmsd (Å)				Mean rmsd	Number of energy evaluations	
			Lowest energy	rmsd of lowest energy	Mean energy				
3ptb	2	9	-8.17	0.32	-7.72	(1.35)	1.50	(3.39)	1.35 × 10 ⁶
2cpp	4	7	-7.36	0.93	-6.65	(2.11)	2.18	(3.42)	1.35 × 10 ⁶
2mcp	10	1	-5.17	0.85	-3.61	(0.95)	5.26	(2.98)	1.35 × 10 ⁶
1stp	7	4	-10.09	0.75	-8.42	(1.82)	2.96	(3.04)	1.35 × 10 ⁶
1hvr	7	4	-21.41	0.82	-11.09	(9.79)	2.79	(1.97)	1.35 × 10 ⁶
4hmg	9	2	-7.60	1.11	-5.72	(1.77)	2.32	(1.43)	1.35 × 10 ⁶
4dfr	10	1	-16.10	0.95	-10.24	(3.95)	4.39	(2.37)	1.35 × 10 ⁶

^a The parameters used were 10 runs a population size of 50, and a run-termination criterion of a maximum of 27,000 generations or a maximum of 1.5×10^6 energy evaluations, whichever came first. Note that, in this case all runs terminated after the maximum number of generations was reached, which equals the product of the population size and the number of generations. The rmsd conformational clustering tolerance was 0.5 Å, calculated from the ligand's crystallographic coordinates. Standard deviations are given in parentheses.

TABLE VI.
Results of Lamarckian Genetic Algorithm Dockings.^a

PDB code	Number of clusters	Number in rank 1	Energy (kcal mol ⁻¹) and rmsd (Å)				Mean rmsd	Number of energy evaluations	
			Lowest energy	rmsd of lowest energy	Mean energy	Mean energy			
3ptb	1	10	-8.15	0.45	-8.15	(0.00)	0.46	(0.01)	1.50 × 10 ⁶
2cpp	1	10	-7.36	0.93	-7.36	(0.00)	0.93	(0.00)	1.50 × 10 ⁶
2mcp	6	2	-5.54	1.05	-4.15	(0.15)	1.10	(0.07)	1.50 × 10 ⁶
1stp	1	10	-10.14	0.69	-10.06	(0.05)	0.66	(0.06)	1.50 × 10 ⁶
1hvr	2	9	-21.38	0.76	-19.11	(6.92)	0.85	(0.35)	1.50 × 10 ⁶
4hmg	3	7	-7.72	1.14	-7.54	(0.19)	1.18	(0.12)	1.50 × 10 ⁶
4dfr	2	7	-16.98	1.03	-16.90	(0.07)	0.98	(0.07)	1.56 × 10 ⁶

^a The parameters used were 10 runs, a population size of 50, and a run-termination criterion of a maximum of 27,000 generations or a maximum of 1.5×10^6 energy evaluations, whichever came first. Because local search also uses energy evaluations, the total number of energy evaluations for the LGA method was greater than that for the GA method, using the same population size and maximum number of generations; in the LGA dockings, the runs terminated because the maximum number of energy evaluations was exceeded. The rmsd conformational clustering tolerance was 0.5 Å, calculated from the ligand's crystallographic coordinates. Standard deviations are given in parentheses.

energy across the ten dockings was lowest for LGA, followed by SA, and finally GA. This is reflected in a comparison of the mean rmsd of the docked conformation from the crystallographic structure for each of the methods: GA had the highest mean rmsd, followed by SA, and on average, the LGA produced conformations with the lowest crystallographic rmsd. Thus, considering their average performance, the best search method at finding the lowest energy and the lowest rmsd was the LGA. The predicted binding free energy, ΔG_{pred} , of the lowest docked energy structure obtained using the LGA method was -8.15 kcal mol⁻¹ (Table VII), whereas the observed value, ΔG_{obs} was -6.46 kcal mol⁻¹: this is within the estimated error of the model.

Cytochrome P-450_{cam} / Camphor (2cpp)

Camphor binds to the monooxygenase cytochrome P-450_{cam} such that the 5-*exo* C—H bond is hydroxylated stereospecifically. The active site is deeply sequestered within the enzyme, and the crystal structure of the complex does not possess an obvious substrate access channel.⁷⁰ This buried active site presents a more challenging docking problem than 3ptb. Once bound, however, the substrate is "tethered" by a hydrogen bond that is donated from the Tyr-96 hydroxyl to the carbonyl oxygen of camphor, while the subtle complementarity of the pocket and the hydrophobic skeleton of camphor help to position the rest of the sub-

strate. The lowest energy found was -7.36 kcal mol⁻¹, found by both the GA (Table V) and LGA methods (Table VI); SA's lowest energy was -7.29 kcal mol⁻¹ (Table IV), which is practically the same. All methods found the crystallographic structure, SA succeeding in 9 of 10 dockings, GA in 7 out of 10 dockings, and LGA in all of the dockings (with success, once again, being measured as having a crystallographic rmsd of less than 1 Å). In all three search method cases, the lowest energy cluster was the most populated, with 6, 9, and 10 members using SA, GA, and LGA, respectively. The predicted binding free energy, ΔG_{pred} , of the lowest docked energy structure, was -7.36 kcal mol⁻¹ using the LGA method (see Table VII), whereas the observed value, ΔG_{obs} , was -8.27 kcal mol⁻¹—once again, this was within the estimated error of the model.

McPC-603 / Phosphocholine (2mcp)

Antibody molecules bind their target antigens with exquisite specificity, having close complementarity between antigen and antibody surfaces, hydrogen bonding, van der Waals, and electrostatic interactions. Phosphocholine binds to Fab McPC-603,⁷¹ and is an example of recognition is predominantly electrostatic in character, primarily due to the influence of Arg H52.⁷² There is little conformational change in the side chains of Fab McPC-603 upon binding, as indicated from the unbound crystal structure. We allowed all four

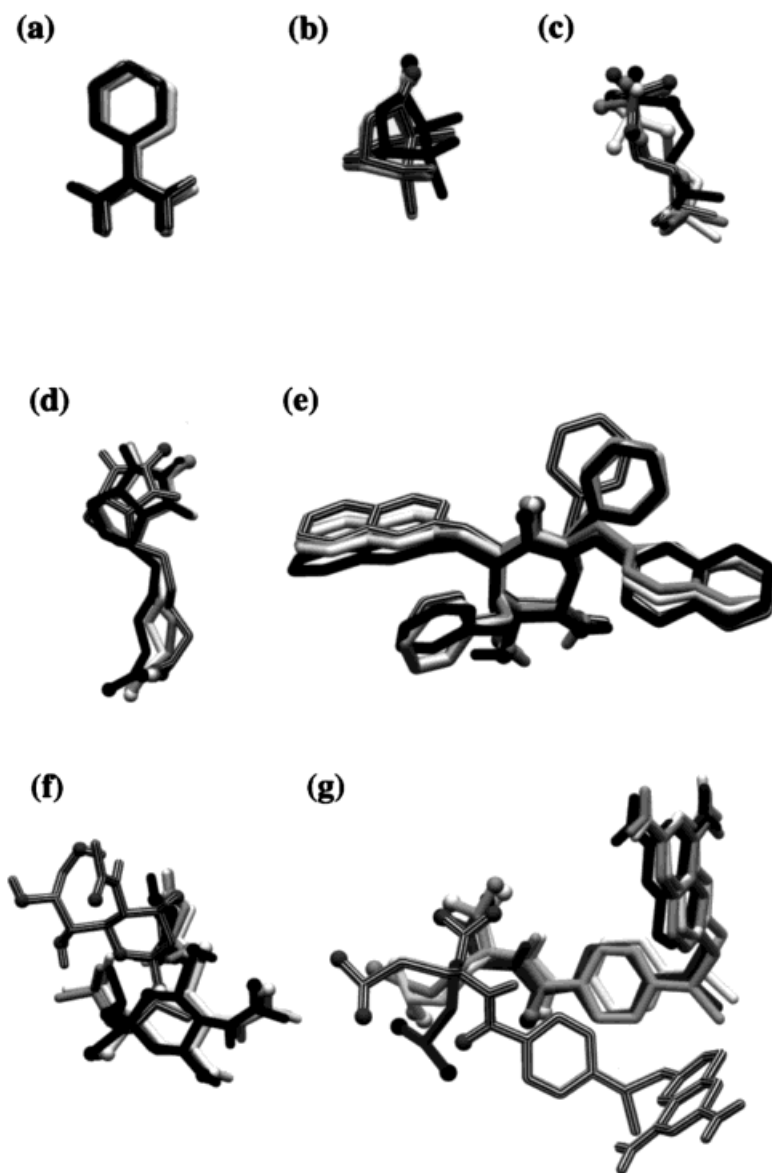


FIGURE 4. A comparison of the lowest energy structure found by each search method and the crystal structure. The latter is shown in black. The simulated annealing results are rendered with a striped texture, the genetic algorithm results are shaded gray, and the Lamarckian genetic algorithm results are white. Oxygen atoms are shown as spheres; other heteroatoms are not shown. Note that simulated annealing failed in the last two test cases, 4hmg and 4dfr, but both the genetic algorithm and the Lamarckian genetic algorithm succeeded.

bonds to rotate during docking. The energy of the crystal structure was positive, due in most part to a large, positive internal energy dominated by C2 and O1 being too close (2.26 Å); this could be improved if local minimization had been performed on the crystal structure before docking. The lowest energy found by each of the three search methods were -4.09 , -5.17 , and -5.54 kcal mol $^{-1}$, using SA, GA, and LGA, respectively. Unlike 3ptb and 2cpp, these differences in energy

were more significant. Both SA and GA found 10 different clusters, whereas LGA found 6 clusters. The mean energy of the 10 dockings was $+70.89$, -3.61 , and -4.15 kcal mol $^{-1}$ for SA, GA, and LGA, respectively. Thus, on average, the LGA performed best in finding the lowest energy docked structure. Furthermore, the mean rmsd from the crystallographic coordinates was 5.40, 5.26, and 1.10 Å for SA, GA, and LGA, respectively, indicating that LGA also reproduced the crystal structure

TABLE VII. Comparison of Predicted Free Energy of Binding, ΔG_{pred} , of Lowest Energy Docked Structure Obtained Using Lamarckian Genetic Algorithm, and Observed Free Energy of Binding, ΔG_{obs} .^a

PDB code	Energy (kcal mol ⁻¹) and rmsd (Å)							
	Lowest energy	rmsd of lowest energy	ΔG_{inter}	ΔG_{intra}	ΔG_{tor}	ΔG_{pred}	ΔG_{obs}	$(\Delta G_{\text{pred}} - \Delta G_{\text{obs}})$
3ptb	-8.15	0.45	-8.15	0.00	0.00	-8.15	-6.46	-1.69
2cpp	-7.36	0.93	-7.36	0.00	0.00	-7.36	-8.27	+0.91
2mcp	-5.54	1.05	-6.57	+1.03	+1.25	-5.32	-7.13	+1.81
1stp	-10.14	0.69	-9.90	-0.24	+1.56	-8.34	-18.27	+9.93 ^b
1hrv	-21.38	0.76	-19.34	-2.04	+2.49	-16.85	-12.96	-3.89
4hmg	-7.72	1.14	-8.93	+1.21	+2.18	-6.75	-3.48	-3.27
4dfr	-16.97	1.03	-16.57	-0.40	+2.18	-14.39	-13.22	-1.17

^a ΔG_{inter} is the intermolecular interaction energy between the ligand and the receptor, ΔG_{intra} is the intramolecular interaction energy of the ligand, and ΔG_{tor} is the torsional free energy change of the ligand upon binding.

^b This large discrepancy may be due to neglect of the conformational rearrangements of streptavidin upon binding biotin, which are neglected in the docking simulation and binding free energy calculation.

most often. The predicted binding free energy, ΔG_{pred} , of the lowest docked energy structure was -5.32 kcal mol⁻¹ using the LGA method (see Table VII), whereas the observed value, ΔG_{obs} , was -7.13 kcal mol⁻¹—this was also within the estimated error of the model.

Streptavidin/Biotin (1stp)

One of the most tightly binding noncovalent complexes is that of streptavidin/biotin, with an experimentally observed dissociation constant, K_d , of $10^{-15}M$. Comparison of the *apo* form and the complex⁷³ shows that the high affinity results from several factors, including formation of multiple hydrogen bonds and van der Waals interactions between the biotin and the protein, in addition to the ordering of surface polypeptide loops of streptavidin upon binding biotin. The method that found the lowest energy was LGA, at -10.14 kcal mol⁻¹, although GA was not significantly different, followed by SA with -8.48 kcal mol⁻¹. The method with the lowest mean energy was LGA at -10.06 kcal mol⁻¹, then GA with -8.42 kcal mol⁻¹, and finally SA with -7.76 kcal mol⁻¹. The method that found the crystallographic complex coordinates most often was LGA, having a mean rmsd of 0.66 Å, then SA at 1.24 Å, and finally GA with 2.96 Å. At the rmsd tolerance chosen for these experiments, 0.5 Å, SA found 10 different conformational clusters, GA found 7 clusters (the most populated was rank 1, with 4 members), and LGA found 1 cluster.

It was not possible to include the entropic effects of the flexible surface loops of streptavidin in the docking of biotin, although they make significant contributions to the binding free energy as revealed by a recent set of experiments involving an atomic force microscope.⁷⁴ It was found that the unbinding forces of discrete complexes of streptavidin with biotin analogs were proportional to the enthalpy change of the complex formation but independent of changes in the free energy, which indicates that the unbinding process is adiabatic and that entropic changes occur after unbinding. This may help to explain why the predicted binding free energy of the streptavidin/biotin complex (ΔG_{pred}) -10.14 kcal mol⁻¹, underestimated the magnitude of the observed value (ΔG_{obs}) -18.27 kcal mol⁻¹ (Table VII).

HIV-1 Protease/XK263 (1hrv)

HIV-1 protease inhibitors prevent the maturation of virions of HIV, and are a major target for computer-assisted drug design in the development of AIDS therapies. Substrates and inhibitors of HIV-1 protease are typically extended peptides or peptidomimetics, with a dozen or more freely rotatable bonds and, as such, they present a challenging target for automated docking techniques. In addition, considerable protein motion is expected in the flaps upon binding, to allow the continuous polypeptide to reach the active site. However, most docking methods use a rigid protein target, and explicit modeling of the opening

and closing of the flaps is not performed: thus, the ligand must “thread” its way into the active site. The cyclic urea HIV-protease inhibitor, XK-263, has 10 rotatable bonds, excluding the cyclic urea’s flexibility. All three search methods found solutions near to the crystal structure⁷⁵: interestingly, the lowest docking energy found by SA was $-11.77 \text{ kcal mol}^{-1}$ and had an rmsd of 1.15 \AA from the crystal structure, whereas GA and LGA found much lower energies but were still near to the active site, having crystallographic rmsd values of 0.82 \AA and 0.76 \AA , respectively. The lowest docking energy found overall was $-21.41 \text{ kcal mol}^{-1}$, and was found using GA, although that found by LGA was practically the same. The predicted binding free energy, ΔG_{pred} , of the lowest energy structure was $-16.85 \text{ kcal mol}^{-1}$ using LGA, whereas the observed value, ΔG_{obs} , was $-12.96 \text{ kcal mol}^{-1}$. The larger discrepancy between the predicted and observed values may be due to the entropic contributions of protein side chain and flap conformational rearrangements, or may be due to other low-energy conformational states of the cyclic urea moiety of XK-263, which are neglected in our calculations.

Influenza Hemagglutinin/Sialic Acid (4hmg)

The recognition of sialic acid by influenza hemagglutinin is chiefly mediated through hydrogen bonding: sialic acid has five hydroxyls, three in the glycerol group, one carboxylate, a cyclic ether oxygen, and an acetamido group, with a total of 11 rotatable acyclic bonds. We used the crystal structure of Weis et al.,⁷⁶ although the low resolution meant that the overall coordinate error was approximately $0.35\text{--}0.40 \text{ \AA}$, which, in itself, presents a potential challenge in the docking tests. We modeled an isopropylated derivative of sialic acid to mimic part of an adjacent six-membered ring that would normally be present in this complex, but was not seen due to disorder: this introduced an extra rotatable bond, giving a total of 11 torsions. Furthermore, in these tests, we used the crystal conformation of the six-membered ring, although normally we would use several of the lowest energy conformations of the ring system and dock these separately.

This was one of two cases where simulated annealing failed to find a docking that was near the crystal structure: the lowest energy structure found had an rmsd of 3.77 \AA from the crystallographic structure, and the docking with the lowest

crystallographic rmsd was 2.36 \AA . The mean energy of all 10 SA dockings was very high ($6.99 \times 10^4 \text{ kcal mol}^{-1}$); only 4 of the 10 SA dockings found negative energies. Both GA and LGA, however, succeeded in finding conformations near ($\leq 1.5 \text{ \AA}$ rmsd) the crystal conformation. The lowest energy found was by LGA, and was $-7.72 \text{ kcal mol}^{-1}$. This structure had a crystallographic rmsd of 1.14 \AA , and had a predicted binding free energy, ΔG_{pred} , of $-6.75 \text{ kcal mol}^{-1}$; the observed binding free energy, ΔG_{obs} , for the sialic acid-hemagglutinin complex was $-3.48 \text{ kcal mol}^{-1}$. The difference in predicted and observed binding free energies may be due to the structural differences between the isopropylated derivative that was docked and sialic acid itself.

Dihydrofolate Reductase/Methotrexate (4dgr)

Methotrexate is an antimetabolite that attacks proliferating tissue selectively induces remissions in certain acute leukemias⁷⁷; however, dangerous side effects of methotrexate in normal cells continue to make DHFR an important target in computer-assisted anticancer drug design.⁷⁸ We used the crystal structure of *E. coli* dihydrofolate reductase complexed with methotrexate⁷⁹ to investigate a more challenging docking problem. We assumed that waters 603, 604, and 639, which mediate hydrogen bonding between the inhibitor and the protein, were conserved biowaters, and included them in the protein structure in our grid calculations. Ideally, these should be predicted, and recently a method based on a *k*-nearest-neighbors classifier and a genetic algorithm called *Consolv* was reported to do just this.⁸⁰

This is one of the two test cases where simulated annealing failed: the lowest energy structure that it found had an rmsd of 4.83 \AA from the crystal structure. This could be because the final docked conformation in simulated annealing is arrived at after a series of continuous steps, and if the route to the active site is blocked, the docking will tend to fail before the ligand reaches the active site. Note that, in the case of the camphor-cytochrome P-450_{cam} docking, the random initialization loop was able to find initial states that were inside the binding pocket, but in this case the dockings failed to start near the active site.

The lowest energy found was $-16.98 \text{ kcal mol}^{-1}$, and was found using LGA: this structure had an rmsd from the crystal structure of 1.03 \AA .

The predicted binding free energy, ΔG_{pred} , of the lowest docked energy structure was -14.39 kcal mol $^{-1}$ using LGA, whereas the observed value, ΔG_{obs} , was -13.22 kcal mol $^{-1}$. This finding was within the estimated error of the model.

JUDGING SEARCH METHODS

To evaluate the new search methods, and to compare them with the earlier search method of simulated annealing, we addressed the following questions: Which search method is most *efficient*? That is, which finds the lowest energy in a given number of energy evaluations? Which search method is most *reliable*? That is, which method finds the most conformations similar to that of the lowest energy? Finally, which search method is most *successful*? That is, which finds the crystallographic conformation most often after a given number of dockings? Furthermore, because these comparisons were carried out using the new, empirical free energy force field, these tests also represent an evaluation of the force field itself, and, if the global minimum of the force field is unable to reproduce observed crystallographic structures, its usefulness will be limited. Because it is very difficult to determine the global minimum of such a complex function, we cannot answer this question definitively; however, we can report the lowest energy found by any of the methods and its structural similarity to that of the crystal structure.

If we calculate statistics across all seven protein–ligand test systems for each search method, we obtain a quantitative estimate of relative per-

formance of each search method (see Table VIII). If, in each of the seven test systems, we assume that the lowest docked energy found by any method is the effective global minimum energy, and then calculate the difference between this energy and all of the docked energies found by each search method, we can then calculate the mean and standard deviation of this difference energy for each search method. Ideally, the mean and standard deviation of this value would be zero. The mean of this difference energy was lowest for LGA (0.40 kcal mol $^{-1}$), followed by GA (3.41 kcal mol $^{-1}$), and finally SA (2.62×10^5 kcal mol $^{-1}$): the very high mean difference energy for SA is indicative of the cases in which this method failed to escape a local minimum, where the ligand was partially or wholly trapped within the protein. Hence, in answer to the first question, the Lamarckian genetic algorithm, LGA, is the most efficient search method.

In terms of how often the structure with lowest energy was found, LGA performed best: the mean of the number of docked structures in rank 1 was 78% for LGA, 40% for GA, and 24% for SA. The mean of the number of clusters found was lowest for LGA (2.29), followed by GA (7.00), and finally SA (8.43). Hence, the most reliable search method was LGA.

In comparing the relative success of each search method in reproducing the crystallographic structure, considering the crystallographic rmsd across all 10 dockings in each of the 7 test systems, the mean rmsd was lowest for LGA (0.88 Å, standard

TABLE VIII. Statistical Comparison of Three Search Methods in AutoDock 3.0 Across all Seven Test Systems.^a

Search method	Statistic	Number of clusters	Number in rank 1	Energy (kcal mol $^{-1}$) and rmsd (Å)				Number of energy evaluations
				Difference from effective global minimum energy	rmsd of lowest energy	Mean energy	Mean rmsd	
SA	Mean	8.43	2.43	2.62×10^5	1.85	2.61×10^5	3.63	1.75×10^6
	SD	2.70	2.44	1.40×10^5	1.74	4.60×10^4	2.61	4.15×10^5
GA	Mean	7.00	4.00	3.41	0.82	-7.64	3.06	1.35×10^6
	SD	3.06	3.06	5.31	0.25	2.59	1.32	0.00
LGA	Mean	2.29	7.86	0.40	0.86	-10.47	0.88	1.50×10^6
	SD	1.80	2.91	2.62	0.24	5.47	0.25	0.00

^a The search methods are simulated annealing (SA), genetic algorithm (GA), and Lamarckian genetic algorithm (LGA). The mean and standard deviation (SD) for each criterion is shown. The effective global minimum energy for each of the seven test systems is the lowest docked energy found by any method for that test system. For each of the 10 dockings, the difference between the final docked energy and this effective global minimum energy was calculated; the mean and standard deviation was calculated across all 7 test systems, which was repeated for each search method.

TABLE IX.
Results of Cross-Validation of Free Energy
Function Using Local Search on 20 HIV-1
Protease-Inhibitor Complexes.

PDB code	Experimental $\Delta G_{\text{binding}}$ (kcal/mol)	Calculated $\Delta G_{\text{binding}}$ (kcal/mol)
1hvs	-14.04	-10.95
1hvk	-13.79	-11.60
1hvi	-13.74	-12.39
7hvp	-13.11	-12.19
1hps	-12.57	-11.80
1hpv	-12.57	-8.24
4phv	-12.51	-14.36
1hef	-12.27	-9.52
1hiv	-12.27	-13.02
1hvl	-12.27	-10.35
8hvp	-12.27	-9.36
1aaq	-11.62	-9.68
1htg	-11.58	-13.13
9hvp	-11.38	-10.54
1hih	-10.97	-11.43
1heg	-10.56	-8.60
1sbg	-10.56	-10.35
1htf	-9.31	-8.21
1hbv	-8.68	-9.75
1hte	-7.69	-7.28

deviation 0.25 Å), followed by GA (3.06 Å, standard deviation 1.32 Å), and finally SA (3.63 Å, standard deviation 2.61 Å). These average results indicate that, of the three search methods, LGA will find the crystallographic structure most often. Thus, the answer to the last question, "Which method is most successful?" is LGA.

In two different cases, 4hmg and 4dfr, the simulated annealing method failed to reproduce the corresponding crystal structure, although it succeeded with 1hvr (see Fig. 4). This is important because methotrexate has 7 rotatable bonds, and would be expected to be solvable using our rule-of-thumb that SA succeeds in problems with 8 torsions or less; however, the HIV-1 protease inhibitor XK-263, has 10 rotatable bonds, and was successfully docked using SA. Thus, the degree of difficulty of a docking problem is not as simple as how many rotatable bonds there are; other factors, such as the nature of the energy landscape, clearly play an important role.

It could be said that the crystallographic rmsd of the lowest energy structure found by any of the

search methods is an estimate of the quality of the force field, although this is complicated by the fact that the search method itself must determine a docking near to the global minimum, an unknown state. We can calculate the energy of the ligand in the crystal structure using the new force field (see Table II), which we assume to be near the global minimum, but, unfortunately, the crystal structure may contain frustrations and bad contacts. This appears to be the case in 2mcp, where a close contact between C2 and O1 causes a positive total energy to be calculated for the crystal structure. In all cases, the lowest energy found, considering all the search methods, was lower than that of the corresponding crystal structure.

The crystallographic rmsd of the lowest energy (found by any search method) for each of the protein-ligand test systems were all within 1.14 Å, or less, of the crystal structure. This suggests that the force field's global minimum in each of the protein-ligand cases was near to the crystal structure, if we accept the assumption that the crystal structure was near to or at the global minimum, and that the lowest energy found was near to the global minimum. In some cases, dockings were found that had lower crystallographic rmsd values but slightly higher energies than the lowest energy found. All of the lowest crystallographic rmsd values were 0.89 Å or less, indicating that low-energy structures found by the force field were very similar to the corresponding crystal structure.

Conclusion

AUTODOCK is a software package of general applicability for automated docking of small molecules, such as peptides, enzyme inhibitors, and drugs, to macromolecules, such as proteins, enzymes, antibodies, DNA, and RNA. New search methods have been introduced and tested here, using a new, empirical binding free energy function for calculating ligand-receptor binding affinities.

We have shown that, of the three search methods tested in AUTODOCK (simulated annealing, genetic algorithm, and Lamarckian genetic algorithm), the most efficient, reliable, and successful is the Lamarckian genetic algorithm LGA. We defined efficiency of search in terms of lowest energy found in a given number of energy evaluations; reliability in terms of reproducibility of finding the

lowest energy structure in independent dockings, as measured by the number of conformations in the top ranked cluster; and success in terms of reproducing the known crystal structure. Simulated annealing failed to reproduce the crystal structures for the influenza hemagglutinin–sialic acid complex (4hmg) and the dihydrofolate reductase–methotrexate complex (4dfr). However, both the genetic algorithm and the Lamarckian genetic algorithm methods succeeded. Thus, the introduction of the LGA search method extends the power and applicability of AUTODOCK to docking problems with more degrees of freedom than could be handled by earlier versions.

The predicted binding affinities of the lowest energy docked conformations, using the LGA method and the new empirical free energy function, were within the standard residual error of the force field in four of the seven cases (3ptb, 2cpp, 2mcp, and 4dfr), and reasonably close in two other cases (1hvr and 4hmg). The large discrepancy between the predicted and the observed binding affinity of biotin for streptavidin (1stp), even though the crystal structure was successfully reproduced, may be due to the large free energy change that accompanies conformational changes in the protein upon binding, in particular the surface loops. This remains a limitation of the method, because protein motion is not modeled and successfully predicting such large-scale protein conformational changes is difficult. The AUTODOCK method works well when there is little change between the *apo* and ligand-bound forms of the protein, even if the protein undergoes significant conformational changes during binding.

AUTODOCK predicts the binding affinity using one conformation of the ligand–protein complex. A new class of models for predicting receptor–ligand and binding affinities has been reported recently that considers not just the *lowest* energy state of the complex, but the *predominant states* of the binding molecules.⁸¹ These approaches are grounded in statistical thermodynamics, and combine a modest set of degrees of freedom with aggressive conformational sampling to identify the low-energy conformations of the complex and the free molecules. AUTODOCK version 3.0 currently performs extensive conformational sampling, information that could be incorporated into the calculation of the binding affinity. We are studying how the search methods can be modified such that statistical thermodynamics calculations can be performed while

the docking proceeds, to improve the calculation of the binding affinity.

Availability

More information about AUTODOCK and how to obtain it can be found on the World Wide Web at: <http://www.scripps.edu/pub/olson-web/doc/autodock>.

Acknowledgments

The authors thank Dr. Bruce S. Duncan and Dr. Christopher Rosin for their helpful comments and suggestions. This work is publication 10887-MB from The Scripps Research Institute.

References

1. J. M. Blaney, and J. S. Dixon, *Perspect. Drug Discov. Design*, **1**, 301 (1993).
2. I. D. Kuntz, E. C. Meng, and B. K. Shoichet, *Acc. Chem. Res.*, **27**, 117 (1994).
3. R. Rosenfeld, S. Vajda, and C. DeLisi, *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 677 (1995).
4. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin, *J. Mol. Biol.*, **161**, 269 (1982).
5. B. K. Shoichet and I. D. Kuntz, *Prot. Eng.*, **6**, 723 (1993).
6. D. S. Goodsell and A. J. Olson, *Prot. Struct. Func. Genet.*, **8**, 195 (1990).
7. G. M. Morris, D. S. Goodsell, R. Huey, and A. J. Olson, *J. Comput.-Aided Mol. Des.*, **10**, 293 (1996).
8. N. Pattabiraman, M. Levitt, T. E. Ferrin, and R. Langridge, *J. Comput. Chem.*, **6**, 432 (1985).
9. P. J. Goodford, *J. Med. Chem.*, **28**, 849 (1985).
10. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
11. S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, *Science*, **220**, 671 (1983).
12. E. A. Lunney, S. E. Hagen, J. M. Domagala, C. Humblet, J. Kosinski, B. D. Tait, J. S. Warmus, M. Wilson, D. Ferguson, D. Hupe, P. J. Tummino, E. T. Baldwin, T. N. Bhat, B. Liu, and J. W. Erickson, *J. Med. Chem.*, **37**, 2664 (1994).
13. J. V. N. Vara Prasad, K. S. Para, D. F. Ortwine, J. B. Dunbar Jr., D. Ferguson, P. J. Tummino, D. Hupe, B. D. Tait, J. M. Domagala, C. Humblet, T. N. Bhat, B. Liu, D. M. A. Guerin, E. T. Baldwin, J. W. Erickson, and T. K. Sawyer, *J. Am. Chem. Soc.*, **116**, 6989 (1994).
14. A. R. Friedman, V. A. Roberts, and J. A. Tainer, *Prot. Struct. Func. Genet.*, **20**, 15 (1994).
15. B. L. Stoddard and D. E. Koshland, *Nature*, **358**, 774 (1992).
16. D. S. Goodsell, G. M. Morris, and A. J. Olson, *J. Mol. Recogn.*, **9**, 1 (1996).

17. C. M. Oshiro, I. D. Kuntz, and J. S. Dixon, *J. Comput.-Aided Mol. Design*, **9**, 113 (1995).
18. D. R. Westhead, D. E. Clark, D. Frenkel, J. Li, C. W. Murray, B. Robson, and B. Waszkowycz, *J. Comput.-Aided Mol. Design*, **9**, 139 (1995).
19. P. Willet, *TIBTECH*, **13**, 516 (1995).
20. D. E. Clark and D. R. Westhead, *J. Comput.-Aided Mol. Design*, **10**, 337 (1996).
21. C. D. Rosin, R. S. Halliday, W. E. Hart, and R. K. Belew, In *Proceedings of the Seventh International Conference on Genetic Algorithms (ICGA97)*, T. Baeck, Ed., Morgan Kaufman, San Francisco, CA, 1997.
22. D. K. Gehlhaar, G. M. Verkhivker, P. A. Rejto, C. J. Sherman, D. B. Fogel, L. J. Fogel, and S. T. Freer, *Chem. Biol.*, **2**, 317 (1995).
23. J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI, 1975.
24. S. S. Četverikov, *J. Exper. Biol.*, **2**, 3 (1926).
25. Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, New York, 1996.
26. W. E. Hart, *Adaptive Global Optimization with Local Search*, Ph.D. Thesis, Computer Science and Engineering Department, University of California, San Diego, 1994. See also: "ftp://ftp.cs.sandia.gov/pub/papers/wehart/thesis.ps.gz."
27. W. E. Hart, T. E. Kammeyer, and R. K. Belew, In *Foundations of Genetic Algorithms III*, D. Whitley and M. Vose, Eds., Morgan Kaufman, San Francisco, CA, 1994.
28. R. K. Belew and M. Mitchell, *Adaptive Individuals in Evolving Populations: Models and Algorithms. Santa Fe Institute Studies in the Science of Complexity*, XXVI, Addison-Wesley, Reading, MA, 1996.
29. F. J. Solis and R. J.-B. Wets, *Math. Oper. Res.*, **6**, 19 (1981).
30. P.-G. Mailliot, In *Graphics Gems*, A. S. Glassner, Ed., Academic Press, London, 1990, p. 498.
31. A. Watt and M. Watt, In *Advanced Animation and Rendering Techniques—Theory and Practice*, ACM Press, New York.
32. P. L'Ecuyer and S. Cote, *ACM Trans. Math. Software*, **17**, 98 (1991).
33. J. B. Lamarck, *Zoological Philosophy*, Macmillan, London, 1914.
34. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta Jr., and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984).
35. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 5179 (1995).
36. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187 (1983).
37. A. T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5319 (1977).
38. G. Némethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.*, **87**, 1883 (1983).
39. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. diNola, and J. R. Haak, *J. Chem. Phys.*, **81**, 3684 (1984).
40. J. H. van der Waals, *Lehrbuch der Thermodynamik, Part 1*, Mass and Van Suchtelen, Leipzig, 1908.
41. I. K. McDonald and J. M. Thornton, *J. Mol. Biol.*, **238**, 777 (1994).
42. M. K. Gilson and B. Honig, *Nature*, **330**, 84 (1987).
43. D. Bashford and M. Karplus, *Biochemistry*, **29**, 10219 (1990).
44. D. Bashford and K. Gerwert, *J. Mol. Biol.*, **224**, 473 (1992).
45. B. Honig and A. Nicholls, *Science*, **268**, 1144 (1995).
46. P. A. Bash, U. C. Singh, F. K. Brown, R. Langridge, and P. A. Kollman, *Science*, **235**, 574 (1987).
47. L. P. Hammett, *J. Am. Chem. Soc.*, **59**, 96 (1937).
48. T. Fujita, J. Iwasa, and C. Hansch, *J. Am. Chem. Soc.*, **86**, 5175 (1964).
49. C. Hansch, A. R. Steward, J. Iwasa, and E. W. Deutsch, *Mol. Pharmacol.*, **1**, 205 (1965).
50. C. D. Selassie, Z. X. Fang, R. L. Li, C. Hansch, G. Debnath, T. E. Klein, R. Langridge, and B. T. Kaufman, *J. Med. Chem.*, **32**, 1895 (1989).
51. D. H. Williams, J. P. L. Cox, A. J. Doig, M. Gardner, U. Gerhard, P. T. Kaye, A. R. Lal, I. A. Nicholls, C. J. Salter, and R. C. Mitchell, *J. Am. Chem. Soc.*, **113**, 7020 (1991).
52. L. Wesson and D. Eisenberg, *Prot. Sci.*, **1**, 227 (1992).
53. H.-J. Böhm, *J. Comput.-Aided Mol. Design*, **6**, 593 (1992).
54. H.-J. Böhm, *J. Comput.-Aided Mol. Design*, **8**, 243 (1994).
55. A. N. Jain, *J. Comput.-Aided Mol. Design*, **10**, 427 (1996).
56. E. L. Mehler and T. Solmajer, *Prot. Eng.*, **4**, 903 (1991).
57. M. F. Sanner, A. J. Olson, and J.-C. Spohner, *Biopolymers*, **38**, 305 (1996).
58. P. F. W. Stouten, C. Frömmel, H. Nakamura, and C. Sander, *Mol. Simul.*, **10**, 97 (1993).
59. S. Hill, In *Graphics Gems IV*, P. S. Heckbert, Ed., Academic Press, London, 1994, p. 521.
60. P. W. Atkins, *Physical Chemistry*, Oxford University Press, Oxford, 1982, p. 263.
61. H. R. Horton, L. A. Moran, R. S. Ochs, J. D. Rawn, and K. G. Scrimgeour, *Principles of Biochemistry*, Prentice-Hall, London, 1993.
62. S-PLUS, Statistical Sciences, Inc., Seattle, WA.
63. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977).
64. E. E. Abola, F. C. Bernstein, S. H. Byant, T. F. Koetzle, and J. Weng, In *Crystallographic Databases-Information Content, Software Systems, Scientific Applications*, F. H. Allen, G. Bergerhoff, and R. Sievers, Eds., Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, 1987, p. 107.
65. SYBYL, Tripos Associates, Inc., St. Louis, MO.
66. M. Marsili and J. Gasteiger, *Chim. Acta*, **52**, 601 (1980).
67. J. Gasteiger and M. Marsili, *Tetrahedron*, **36**, 3210 (1980).
68. D. N. A. Boobbyer, P. J. Goodford, P. M. McWhinnie, and R. C. Wade, *J. Med. Chem.*, **32**, 1083 (1989).
69. M. Marquart, J. Walter, J. Deisenhofer, W. Bode, and R. Huber, *Acta Crystallogr. (Sect. B)*, **39**, 480 (1983).
70. T. L. Poulos, B. C. Finzel, and A. J. Howard, *J. Mol. Biol.*, **195**, 687 (1987).
71. E. A. Padlan, G. H. Cohen, and D. R. Davies, *Ann. Immunol. (Paris) (Sect. C)*, **136**, 271 (1985).

72. J. Novotny, R. E. Brucoleri, and F. A. Saul, *Biochemistry*, **28**, 4735 (1989).
73. P. C. Weber, D. H. Ohlendorf, J. J. Wendolski, and F. R. Salemme, *Science*, **243**, 85 (1989).
74. V. T. Moy, E.-L. Florin, and H. E. Gaub, *Science*, **266**, 257 (1994).
75. P. Y. S. Lam, P. K. Jadhav, C. J. Eyerman, C. N. Hodge, Y. Ru, L. T. Bachelier, J. L. Meek, M. J. Otto, M. M. Rayner, Y. Wong, C.-H. Chang, P. C. Weber, D. A. Jackson, T. R. Sharpe, and S. Erickson-Viitanen, *Science*, **263**, 380 (1994).
76. W. I. Weis, A. T. Brünger, J. J. Skehel, and D. C. Wiley, *J. Mol. Biol.*, **212**, 737 (1990).
77. C. K. Matthews and K. E. van Holde, *Biochemistry*, Benjamin/Cummings, Redwood City, CA, 1990.
78. C. A. Reynolds, W. G. Richards, and P. J. Goodford, *Anti-Cancer Drug Des.*, **1**, 291 (1987).
79. J. T. Bolin, D. J. Filman, D. A. Matthews, R. C. Hamlin, and J. Kraut, *J. Biol. Chem.*, **257**, 13650 (1982).
80. M. L. Raymer, P. C. Sanschagrin, W. F. Punch, S. Venkataraman, E. D. Goodman, and L. A. Kuhn, *J. Mol. Biol.*, **265**, 445 (1997).
81. M. K. Gilson, J. A. Given, and M. S. Head, *Chem. Biol.*, **4**, 87 (1997).

Software News and Update

A Semiempirical Free Energy Force Field with Charge-Based Desolvation

RUTH HUEY, GARRETT M. MORRIS, ARTHUR J. OLSON, DAVID S. GOODSSELL
Department of Molecular Biology, Scripps Research Institute, La Jolla, California 92102

Received 26 January 2006; Revised 13 March 2006; Accepted 24 April 2006

DOI 10.1002/jcc.20634

Published online 1 February 2007 in Wiley InterScience (www.interscience.wiley.com).

Abstract: The authors describe the development and testing of a semiempirical free energy force field for use in AutoDock4 and similar grid-based docking methods. The force field is based on a comprehensive thermodynamic model that allows incorporation of intramolecular energies into the predicted free energy of binding. It also incorporates a charge-based method for evaluation of desolvation designed to use a typical set of atom types. The method has been calibrated on a set of 188 diverse protein–ligand complexes of known structure and binding energy, and tested on a set of 100 complexes of ligands with retroviral proteases. The force field shows improvement in redocking simulations over the previous AutoDock3 force field.

© 2007 Wiley Periodicals, Inc. J Comput Chem 28: 1145–1152, 2007

Key words: computational docking; prediction of free energy; force fields; desolvation models; computer-aided drug design; AutoDock4

Introduction

Computational docking currently plays an essential role in drug design and in the study of macromolecular structure and interaction.^{1–5} Docking simulations require two basic methods: a search method for exploring the conformational space available to the system and a force field to evaluate the energetics of each conformation. Given the desire to minimize computational effort, there is a trade-off between these two elements. We may choose to use a highly sophisticated force field while searching only a small portion of the conformational space. This is the approach taken by methods such as molecular dynamics and free energy perturbation, which use physically based energy functions combined with full atomic simulation to yield accurate estimates of the energetics of molecular processes.⁶ However, these methods are currently too computationally intensive to allow blind docking of a ligand to a protein. Computational docking is typically performed by employing a simpler force field and exploring a wider region of conformational space. This is the approach taken by AutoDock (<http://autodock.scripps.edu>) and other computational docking methods, which may predict bound conformation with no *a priori* knowledge of the binding site or its location on the macromolecule.^{7–9}

Empirical free energy force fields, which define simple functional forms for ligand–protein interactions, and semiempirical free energy force fields, which combine traditional molecular mechanics force fields with empirical weights and/or empirical functional forms, have been used by a wide variety of computa-

tional docking methods.^{7–9} These force fields provide a fast method to rank potential inhibitor candidates or bound states based on an empirical score. In some cases, this score may be calibrated to yield an estimate of the free energy of binding.

The AutoDock3 force field¹⁰ is an example. It uses a molecular mechanics approach to evaluate enthalpic contributions such as dispersion/repulsion and hydrogen bonding and an empirical approach to evaluate the entropic contribution of changes in solvation and conformational mobility. Empirical weights are applied to each of the components based on calibration against a set of known binding constants. The final semiempirical force field is designed to yield an estimate of the binding constant.

We describe here the development and testing of an improved semiempirical free energy force field for AutoDock4, designed to address several significant limitations of the AutoDock3 force field. The new force field, which has been calibrated using a large set of diverse protein–ligand complexes, includes two major advances. The first is the use of an improved thermodynamic model of the binding process, which now allows inclusion of intramolecular terms in the estimated free energy. Second, the force field includes a full desolvation model that includes terms for all atom types, including the favorable energetics of desolvating carbon atoms as well as the unfavorable

Correspondence to: D.S. Goodsell; e-mail: goodsell@scripps.edu

Contract/grant sponsor: National Institutes of Health; contract/grant numbers: R01 GM069832, P01 GM48870

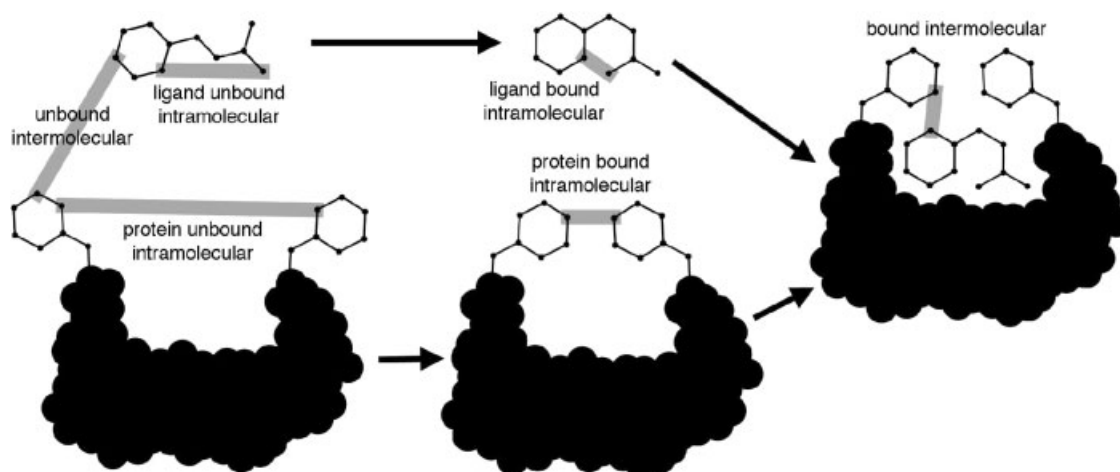


Figure 1. The force field evaluates binding in two steps. The ligand and protein start in an unbound conformation. The first step evaluates the intramolecular energetics of the transition from these unbound states to the conformation that the ligand or protein will adopt in the bound complex. The second step evaluates the intermolecular energetics of combining the ligand and protein in their bound conformations.

energetics of desolvating polar and charged atoms. The force field also incorporates an improved model of directionality in hydrogen bonds, now predicting the proper alignment of groups with multiple hydrogen bonds such as DNA bases.¹¹

Methods

Overview

The semiempirical free energy force field estimates the energetics of the process of binding of two (or more) molecules in a water environment using pair-wise terms to evaluate the interaction between the two molecules and an empirical method to estimate the contribution of the surrounding water. This differs from a traditional molecular mechanics force field, which also relies on pair-wise atomic terms, but typically uses explicit water molecules to evaluate solvation contributions. The goal of the empirical free energy force field is to capture the complex enthalpic and entropic contributions in a limited number of easily evaluated terms.

The approach taken in AutoDock is shown in Figure 1. The free energy of binding is estimated to be equal to the difference between (1) the energy of the ligand and the protein in a separated unbound state and (2) the energy of the ligand–protein complex. This is broken into two steps for the evaluation: we evaluate the intramolecular energetics of the transition from the unbound state to the bound conformation for each of the molecules separately, and then evaluate the intermolecular energetics of bringing the two molecules together into the bound complex.

The force field includes six pair-wise evaluations (V) and an estimate of the conformational entropy lost upon binding (ΔS_{conf}):

$$\Delta G = (V_{\text{bound}}^{\text{L-L}} - V_{\text{unbound}}^{\text{L-L}}) + (V_{\text{bound}}^{\text{P-P}} - V_{\text{unbound}}^{\text{P-P}}) + (V_{\text{bound}}^{\text{P-L}} - V_{\text{unbound}}^{\text{P-L}} + \Delta S_{\text{conf}}) \quad (1)$$

In this equation, L refers to the “ligand” and P refers to the “protein” in a protein–ligand complex; note, however, that the approach is equally valid for any types of molecules in a complex. The first two terms are intramolecular energies for the bound and unbound states of the ligand, and the following two terms are intramolecular energies for the bound and unbound states of the protein. The change in intermolecular energy between the bound and unbound states is in the third parentheses. It is assumed that the two molecules are sufficiently distant from one another in the unbound state that $V_{\text{unbound}}^{\text{P-L}}$ is zero. In the current study, we did not allow motion in the protein, so the bound state of the protein is identical with the protein unbound state, and the difference in their intramolecular energy is zero.

The pair-wise atomic terms include evaluations for dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation:

$$V = W_{\text{vdw}} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{\text{hbound}} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + W_{\text{sol}} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)} \quad (2)$$

The weighting constants W are the ones that are optimized to calibrate the empirical free energy based on a set of experimentally characterized complexes. The first term is a typical 6/12 potential for dispersion/repulsion interactions. Parameters A and B were taken from the Amber force field.¹² The second term is a directional H-bond term based on a 10/12 potential.¹³ The parameters C and D are assigned to give a maximal well depth of 5 kcal/mol at 1.9 Å for O–H and N–H, and a depth of 1 kcal/mol at 2.5 Å for S–H. Directionality of the hydrogen bond interaction $E(t)$ is dependent on the angle t away from ideal bonding geometry and is described fully in previous work.^{10,14} Directionality is further enhanced by limiting the number of hydrogen bonds available to each point in the grid to the actual

number of hydrogen bonds that could be formed.¹¹ Electrostatic interactions are evaluated with a screened Coulomb potential identical with that used in AutoDock3.¹⁵ The final term is a desolvation potential based on the volume (V) of the atoms surrounding a given atom, weighted by a solvation parameter (S) and an exponential term based on the distance.¹⁶ As in the original report, the distance weighting factor σ is set to 3.5 Å.

As in our previous work, this force field is calibrated for a united atom model, which explicitly includes heavy atoms and polar hydrogen atoms. Intramolecular energies are calculated for all pairs of atoms within the ligand (or protein, if it has free torsional degrees of freedom), excluding 1–2, 1–3, and 1–4 interactions.

The term for the loss of torsional entropy upon binding (ΔS_{conf}) is directly proportional to the number of rotatable bonds in the molecule (N_{tors}):

$$\Delta S_{\text{conf}} = W_{\text{conf}} N_{\text{tors}} \quad (3)$$

The number of rotatable bonds include all torsional degrees of freedom, including rotation of polar hydrogen atoms on hydroxyl groups and the like.

Desolvation

The desolvation term is evaluated using the general approach of Wesson and Eisenberg.¹⁷ Two pieces of information are needed (1) an atomic solvation parameter for each atom type, which is an estimate of the energy needed to transfer the atom between a fully hydrated state and a fully buried state [S_i in eq. (2)] and (2) an estimate of the amount of desolvation when the ligand is docked [V_i in eq. (2)]. The amount of desolvation is calculated using a volume-summing method similar to the Stouten et al. method.¹⁶ We have developed a modified approach for the atomic solvation parameters based on the chemical type and the atomic charge of the atom. This approach is designed to use the simple set atom types used in AutoDock and other docking methods. Incorporation of the atomic charge into the solvation parameter removes the need to use two discrete charged and uncharged atom types for oxygen and nitrogen. This is particularly useful with hybridized charged groups such as carboxylic acids, which require arbitrary assignment of a formal charge on one oxygen or the other when using previous approaches.

The solvation parameter for a given atom (S , used in the equation above) is calculated as:

$$S_i = (\text{ASP}_i + \text{QASP} \times |q_i|) \quad (4)$$

where q_i is the atomic charge and ASP and QASP are the atomic solvation parameters derived here. The ASP is calibrated using six atom types: aliphatic carbons (C), aromatic carbons (A), nitrogen, oxygen, sulfur, and hydrogen. A single QASP is calibrated over the set of charges on all atom types.

The estimation of the extent of desolvation is calibrated to use atomic volumes calculated as a sphere with radius equal to the contact radius (C/A, 2.00; N, 1.75; O, 1.60; S, 2.00). The amount of shielding in a typical protein was evaluated using the

Table 1. Calibration of the Desolvation Model.

Type	ASP (std error)
C	−0.00143 (0.00019)
A	−0.00052 (0.00012)
N	−0.00162 (0.00182)
O	−0.00251 (0.00189)
H	0.00051 (0.00052)
S	−0.00214 (0.00118)

QASP = 0.01097 (0.00263). Note that these values must be multiplied by the empirically determined weighting factor [W_{sol} in eq. (2)] for the estimate of the free energy.

188 proteins in the calibration set (described later). For each atom in the protein, the volume term in the free energy equation was evaluated:

$$\Delta V_i = \sum_{k \neq i} V_k \times e^{(-r_{ik}^2/2\sigma^2)} \quad (5)$$

where the sum is performed over all atoms k in the protein, excluding all atoms in the same amino acid residue as i . The maximal value of ΔV for each amino acid type over the entire set of proteins was then determined. These values were used to perform a least-squares fit of the model to a set of experimental vacuum-to-water transfer energies,¹⁸ to determine values for the atomic solvation parameters ASP and QASP. The program *R* (<http://www.r-project.org>) was used to perform the fit. Several formulations were tested. The best results were obtained using two carbon types, a nonaromatic (C) and an aromatic (A) type (Table 1), where the aromatic type is limited to carbon atoms within aromatic ring structures. The maximum error was 1.88 kcal/mol in that case. Using a single carbon type, the maximum error increased to 3.00 kcal/mol, and use of a single type for all atoms, along with the charge, gave a maximum error of 4.12 kcal/mol.

We used a simple approximation for incorporation of additional atom types in the desolvation model. The ASP is assigned to the average of the values from the six atom types used in the calibration and the same QASP is applied.

Unbound States

This method relies on assignment of an unbound state for the ligand and protein. In this work, we tested three approaches to the unbound state, as shown in Figure 2. These states are simple approximations to the ensemble of unbound conformations, making a few extreme assumptions about which conformations dominate the energetics of the ensemble.

The first approach (the “extended” state) is a fully extended conformation, which models a fully solvated conformation with few internal contacts. A short optimization was performed on the ligand in isolation using a uniform potential inversely proportional to the distance between each pair of atoms. This pushes all atoms as far away from one another as possible.

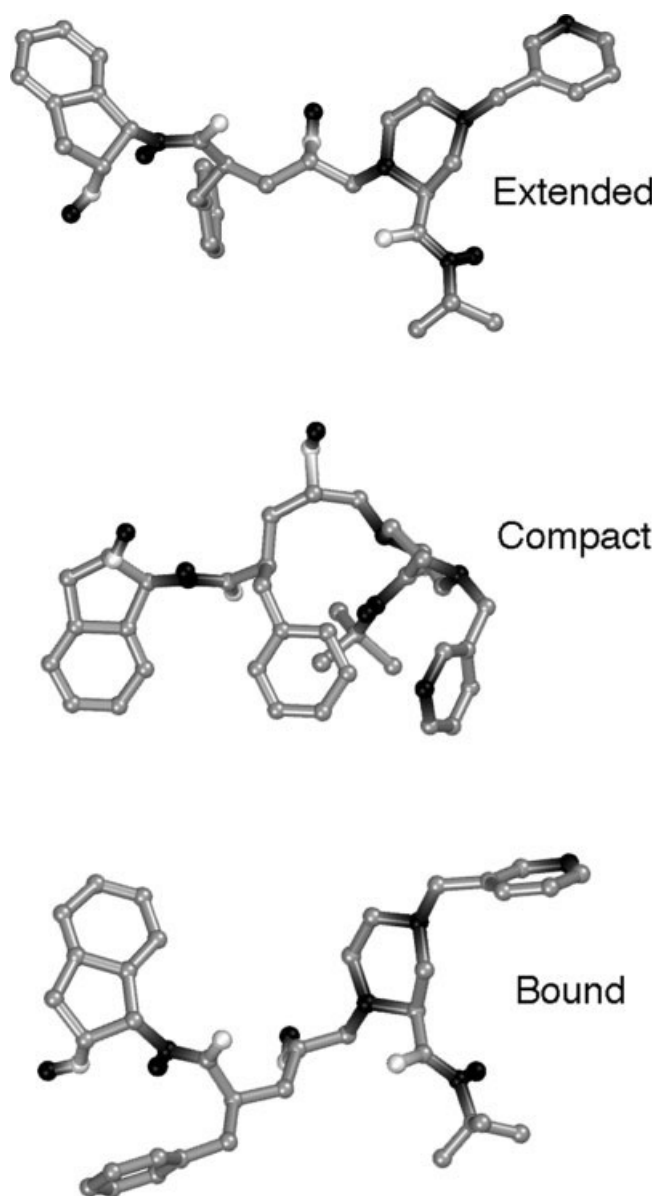


Figure 2. Comparison of the extended, compact, and bound conformations of the HIV protease inhibitor indinavir, taken from PDB entry 1hsg. Note that the extended and bound states are quite similar, and that the hydrophobic groups have formed a cluster in the compact state.

The second approach (the “compact” state) is a minimized conformation that has substantial internal contacts, modeling a folded state for the unbound ligand. We wished to include the new solvation model in the determination of this unbound state, so we used AutoDock4 with values of the energetic parameters taken from the calibration using the extended state. A short Lamarckian genetic algorithm conformational search was performed, using an empty affinity grid. As expected, these conformations tend to bury hydrophobic portions inside and form internal hydrogen bond interactions.

The final approach (the “bound” state) uses the assumption used in AutoDock3 and many other docking methods. In this, it is assumed that the conformation of the unbound state is identical to the conformation of the bound state.

Coordinate Sets

The force field was calibrated and tested using a large collection of protein complexes for which experimental information on binding strength is available. These complexes were taken from two sources.

The force field was calibrated on a set of 188 complexes. Binding data were obtained from the Ligand–Protein Database (<http://lpdb.scripps.edu>), and coordinates were obtained from the Protein Data Bank (<http://www.pdb.org>). These complexes were checked and corrected if necessary for the proper biological unit, and files were regularized to remove alternate locations and to have consistent naming of atoms. Hydrogen atoms were then added automatically using Babel,¹⁹ atomic charges were added using the Gasteiger PEOE method,²⁰ and then nonpolar hydrogen atoms were merged. The Gasteiger method was chosen for its fast and easy operation and ready availability as part of Babel. Formal charges were assigned to metal ions. Charges on terminal phosphate groups were assigned improperly, with a total charge of -0.5 , so the remaining -0.5 charge was split manually between the four surrounding oxygen atoms. Babel also treated a sulfonamide group in several ligands as neutral; a single negative charge, as reported in the original structure reports, was split manually between the oxygen and nitrogen atoms in these groups. Ligands were processed in ADT (the graphical interface to AutoDock, <http://autodock.scripps.edu/resources/adt>) to assign atom types and torsion degrees of freedom. Finally, a short optimization of the ligand was performed using the local search capability of AutoDock3, to relax any unacceptable contacts in the crystallographic conformation.

Binding data for a test set of 100 retroviral protease complexes was obtained from the PDBBind database (<http://www.pdbbind.org>), and coordinates were obtained from the Protein Data Bank. They were processed similarly to the calibration set.

Redocking

Redocking experiments were performed with AutoDock4 and the new empirical free energy force field. For each complex, 50 docking experiments were performed using the Lamarckian genetic algorithm conformational search with the default parameters from AutoDock3. A maximum of 25 million energy evaluations was applied for each experiment. The results were clustered using a tolerance of 2.0 Å. For comparison, docking experiments with the same parameters were performed with AutoDock3.

Results and Discussion

Performance

The force field was calibrated on a set of 188 complexes tabulated in the Ligand–Protein Database. Calibration with the three

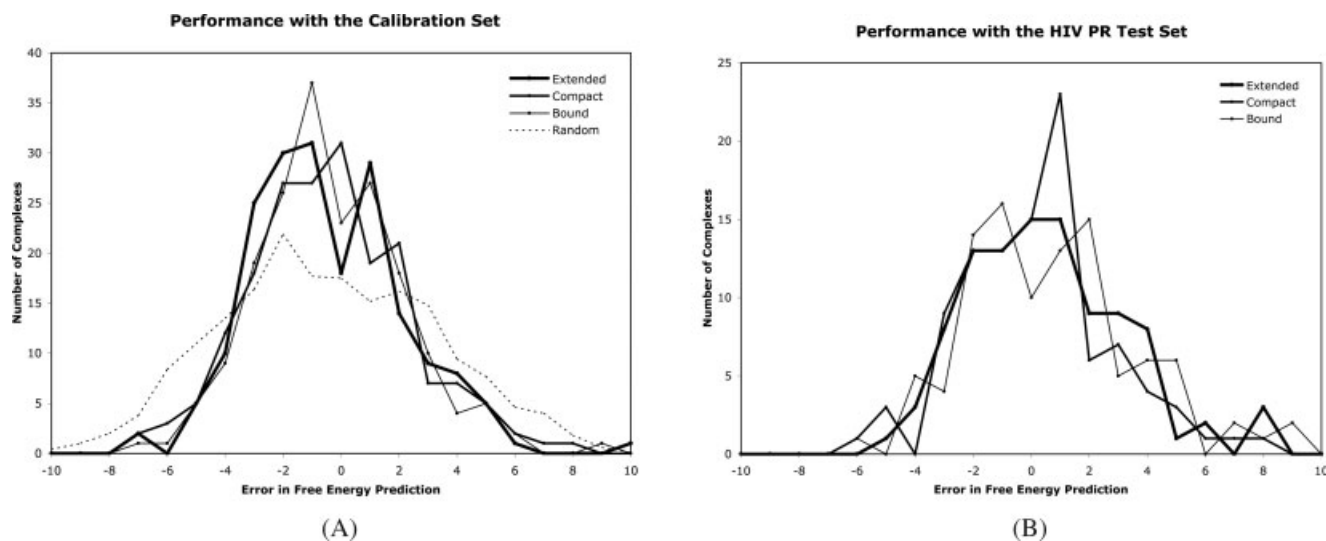


Figure 3. Performance of the new force field. These graphs are histograms showing the number of complexes with a given error in the predicted free energy of binding. Values of zero correspond to complexes that are perfectly predicted, positive values are cases where the predicted energy is too favorable (too negative). The random curve was generated by using a random number between 0 and 1 for values of the h-bond, dispersion/repulsion, electrostatic, and desolvation energies for complexes in the calibration set, and then deriving parameters based on these randomized energies. (A) Results for the calibration set. (B) Results for the HIV protease test set.

different unbound states gave similar results, as shown in Figure 3. Table 2 shows standard errors for the three unbound states for several sets. The first line is standard error over the entire set of 188 complexes used in the calibration. The second line is the standard error calculated for the set of 155 complexes that do not have close ligand–metal contacts, using parameters calibrated against the whole set. The performance is significantly better on this smaller set of proteins. The third line is the standard error when evaluating binding energies for a naive set of 100 HIV protease complexes, not used in the calibration. The final four lines are results from a cross-validation study, where 1/4 of the complexes were removed from the calibration set, and then evaluated with the parameters calibrated with the remaining 3/4 of the set.

Table 3 shows parameters for the three calibrations, including standard errors and *t* values for the parameters. The compact state is clearly the worst of the three. The extended and bound

states have similar performance, with the bound state showing slightly better statistics in the calibration and the cross-validation, but poorer performance with the HIV test set. Based on these results, we have chosen the extended conformation as the default unbound state in AutoDock4. However, it is possible within AutoDock4 to use any of the models for the unbound state, with the appropriate values for the force field parameters, if there is specific information for a given application about the nature of the unbound state.

The new force field requires slightly more computation time for the energy evaluation than the previous AutoDock3 force field. It requires precalculation of the energy of the unbound state, which is not performed in AutoDock3. During the docking simulation, it requires one additional operation to look-up the charge-dependent desolvation component from a precalculated AutoGrid map.

Magnitude of Terms

The total predicted free energy is surprisingly well distributed between the four pair-wise terms, as shown in Table 4. The dispersion/repulsion term provides an average of about -0.3 to -0.5 kcal/mol per nonhydrogen ligand atom. Hydrogen bonding provides approximately -0.6 kcal/mol for single hydrogen bonds and twice that with oxygen atoms that accept two hydrogen bonds. Electrostatic energies range widely, but give an overall average that is favorable by a few tenths of a kcal/mol. However, in the best cases of interactions with metal ions, electrostatics can provide significant stabilization of several kcal/mol. Desolvation of hydrophobic groups provides a favorable interaction of -0.13 kcal/mol in the best cases and the worst case of

Table 2. Calibration Results.

Coordinate set	Extended	Compact	Bound	AutoDock3
Std error (188 complexes)	2.62	2.72	2.52	2.63
Std error (155 complexes)	2.35	2.51	2.25	2.33
Std error (HIV test set)	2.80	2.52	2.99	3.48
Std error (cross validation)	2.41	2.57	2.36	
	2.44	2.47	2.36	
	2.83	2.92	2.73	
	3.24	3.38	3.07	

Values for standard errors are in kcal/mol.

Table 3. Parameters for the Force Field.

Parameter	Extended	Compact	Bound
1. h-bond	0.097 (0.020) 4.9	0.053 (0.021) 2.5	0.121 (0.022) 5.4
2. desolvation	0.116 (0.026) 4.5	0.060 (0.025) 2.4	0.132 (0.026) 5.0
3. vdW	0.156 (0.009) 17.1	0.164 (0.010) 15.7	0.166 (0.009) 18.0
4. estat	0.147 (0.019) 7.5	0.127 (0.021) 6.1	0.141 (0.019) 7.5
5. tors entropy	0.274 (0.040) 6.9	0.227 (0.040) 5.7	0.298 (0.039) 7.7

Values are given for weight (std error) *t* value.

desolvation of polar oxygen is unfavorable at about +0.6 kcal/mol. Since the desolvation potential is based largely on the atomic charge, note that there are also cases where burial of carbon is unfavorable, as with the carbon atoms in the middle of guanidinium or carboxyl groups.

For the polar atoms, the desolvation term has the desired effect of counterbalancing the hydrogen bonding energy. Hydrogen bonds in complexes do not add substantially to the binding free energy, since they simply replace hydrogen bonds that are formed with water in the free state. When polar groups are buried in proteins without forming hydrogen bonds, however, this can have an unfavorable effect on binding. The combination of the favorable hydrogen bond potential with the unfavorable desolvation potential models this effect in the force field. Polar atoms that form hydrogen bonds in the complex will have a slightly favorable contribution to the energy, but polar atoms that do not form hydrogen bonds will only have the unfavorable desolvation component.

Intramolecular Energies

One of the major advances incorporated into this force field is the inclusion of intramolecular energies in the estimated free energy of binding. Often, these are not included in empirical free energy force fields, causing a persistent problem: the function used for scoring during docking is not the same function used to predict the free energy of binding. In many cases, this is solved by using a simple yes-or-no scoring for intramolecular energies; conformations with unacceptable clashes are simply discarded, so the intramolecular energy is not included explicitly in the scoring function. In AutoDock3, a pair-wise intramolecular energy is used in scoring and optimization, but omitted when

the free energy is predicted for the best conformations. This has an undesirable effect: occasionally a complex will rank high in docking score, but will have a poor free energy, or the converse.

The current method incorporates the intramolecular energy through the use of an unbound structure, so that the difference in energy between the bound and unbound forms is included in the predicted free energy. In most cases, this is close to zero.

Desolvation

The desolvation method used here builds upon two approaches. The basic approach is taken from work from Wesson and Eisenberg.¹⁷ They postulate that the desolvation energy is proportional to the change in the surface area that is available to water. The amount of shielding upon binding is determined by comparing the solvent accessible surface area of the unbound and bound forms. Each type of atom makes a different contribution to this energy, depending on how polar or hydrophobic it is. They classified atoms into a few simple atom types, each with its own atomic solvation parameter. Typically, this set includes two carbon types, one type for polar O or N, and separate types for charged N and charged O. In cases where the charge is delocalized over several atoms (carboxylates, guanidinium), the charge is placed on the most solvent-exposed atom. Stouten et al.¹⁶ modify this basic approach, casting it into a form that is amenable to a pair-wise volume-summing method to evaluate the amount of shielding.

The approach used here was chosen to accommodate the needs of the grid-based evaluation used in AutoDock and many other methods. The number of atom types must be kept to a minimum, since each new atom type requires a new map of interaction potentials. For this reason, we calibrated the method

Table 4. Magnitude of Terms.

Type	Disp/rep	H-bond	Electrostatics	Desolvation
C	+0.14 (−0.43) −0.82	n/a	+1.00 (−0.04) −0.85	+0.48 (+0.01) −0.13
A	+0.00 (−0.41) −0.80	n/a	+0.31 (−0.02) −0.83	+0.75 (+0.02) −0.05
N	−0.02 (−0.34) −0.56	0.00 (−0.03) −0.59	+0.54 (−0.25) −3.67	+0.56 (+0.11) −0.08
O	+0.17 (−0.32) −0.68	0.00 (−0.23) −1.24	+1.52 (−0.09) −7.07	+0.60 (+0.10) −0.06
S	−0.15 (−0.54) −0.93	0.00 (−0.01) −0.11	+0.05 (−0.01) −0.10	+0.00 (−0.08) −0.15
H	+0.09 (−0.05) −0.11	0.00 (−0.32) −0.63	+1.14 (−0.21) −1.15	+0.35 (+0.18) +0.00

Minimum (average) maximum values in kcal/mol evaluated over all ligand atoms in the set of 188 complexes from the calibration set. A includes aromatic carbons, and C includes all other carbon atoms. N, O, and S include only those nitrogen, oxygen, and sulfur atoms that accept hydrogen bonds. H includes polar hydrogen atoms.

Table 5. Parameters Calibrated with the Reduced set of 155 Complexes.

Parameter	Weight (std error)	<i>t</i> value
1. h-bond	0.172 (0.020)	8.416
2. desolvation	0.065 (0.024)	2.621
3. vdw	0.090 (0.008)	17.184
4. estat	-0.014 (0.034)	-0.424
5. tors entropy	0.336 (0.036)	9.292

using simple volumes calculated from contact radius, instead of more complex atomic volumes used in the Stouten et al. formulation. To account for the differences in polar atoms with a formal charge and those without, we incorporated a term based on the charge, instead of breaking oxygen and nitrogen atoms into two or more types.

Electrostatics

A slightly improved performance may be obtained by calibrating the force field using only the 155 complexes with no close ligand–metal contacts, with a standard error of 2.25 kcal/mol (compared with 2.35 kcal/mol using parameters calibrated from the entire set). The disadvantage, however, is that this formulation leads to a parameterization that does not include an electrostatic term, as shown in Table 5. These results are presented for

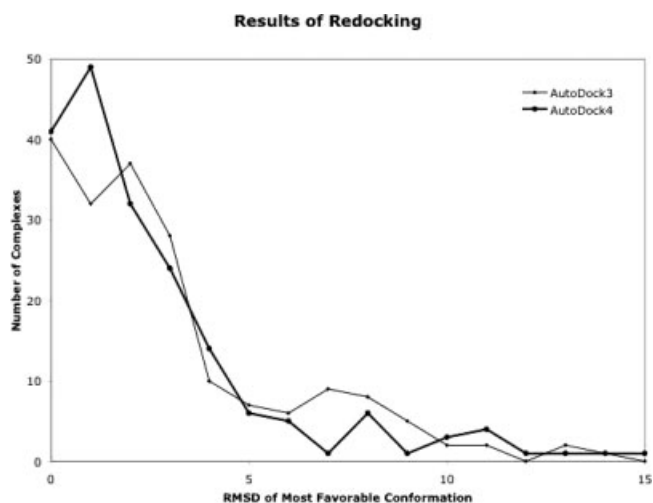


Figure 4. Performance of the new force field in docking experiments, compared with AutoDock3. The graph is a histogram showing the number of complexes within a given RMSD of the crystallographic structure. In each case, the conformation with most favorable estimated energy is used as the predicted conformation.

the extended state of the unbound ligand; similar results were obtained using the other two approximates for the unbound state. Note that the statistics for the electrostatic term are poor; in

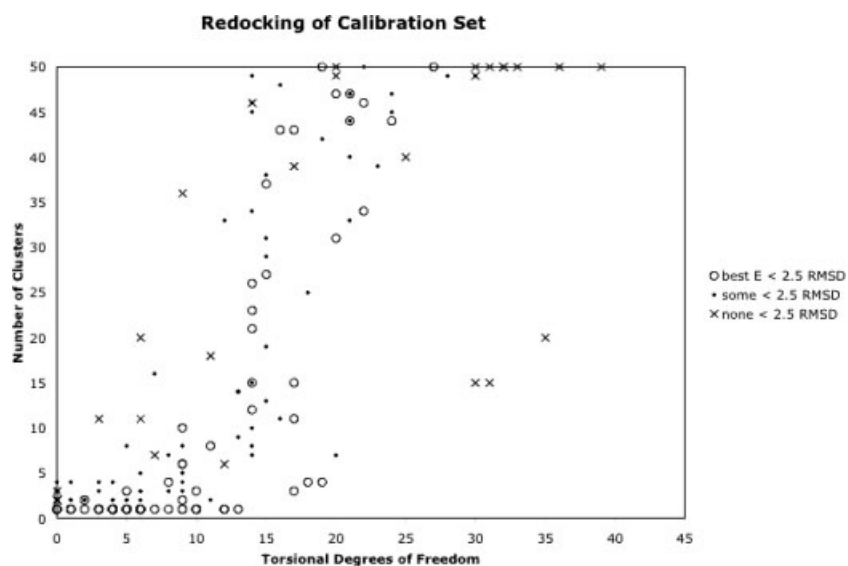


Figure 5. Performance of the new force field in redocking experiments. Each point in the graph represents one protein–ligand complex in the calibration set. Open circles are cases where the conformation of best predicted energy is within 2.5 Å of the crystallographic conformation. Dots are cases where AutoDock finds a conformation within 2.5 Å of the crystallographic conformation, but it is not the best energy. Complexes that were not successfully redocked are shown with an X. In each case, 50 docking simulations were performed and results that were within 2.0 Å of each other were clustered. The vertical axis shows the number of clusters found for each complex (ideally, if AutoDock was able to find the global minimum structure, we would see one cluster). The horizontal axis shows the number of torsional degrees of freedom in the ligand. Note that AutoDock fails with ligands with greater than 15–20 rotatable bonds.

fact, the model is not significantly affected if the electrostatic term is removed. Instead, the contribution of electrostatics appears to be included in the hydrogen-bonding term, which increases by a factor of two over the full force field described above. This is caused by the fact that there are very few electrostatic interactions within this dataset that do not involve hydrogen bonds. In the reduced set, there are not enough examples to separate the effect of hydrogen bonding from ionic interactions. In the full force field calibration, the metal ion contacts controlled the weighting of the electrostatic term.

Conformational Entropy

We tested two models for evaluation of conformational entropy, both based on a simple sum of the torsional degrees of freedom. The first includes all degrees of freedom and the second excludes terminal rotors that only move hydrogen atoms, such as rotation of hydroxyl groups or amines. Use of all torsional degrees of freedom gave slightly better results: standard error of 2.62 kcal/mol compared with standard error of 2.70 when hydrogen rotors are excluded.

Redocking

We tested the new semiempirical free energy force field in a redocking experiment with the calibration set of 188 protein–ligand complexes. The new force field when used in AutoDock4 performs better than similar redocking experiments using AutoDock3, as shown in Figure 4. In 85 cases, AutoDock4 with the new force field found a conformation within 2.5 Å of the known conformation, and scored this conformation with the best estimated free energy. In 75 cases, the correct conformation was found, but an alternative conformation was scored with better energy. However, in 67 of these cases, the proper conformation and the noncrystallographic conformation were scored within 1 kcal/mol of each other, which is not a significant difference given the precision of the force field. The remaining 28 complexes were not predicted correctly by AutoDock 4, most cases due to the fact that they were very large ligands with greater than 15 degrees of torsional freedom (see Fig. 5).

Conclusions

The performance of the new force field is similar to the existing AutoDock3 force field, which has shown proven effectiveness in hundreds of laboratories.²¹ The advantage of the new force field is that it is based on a comprehensive thermodynamic model, which is extensible for use in protein–protein docking and for incorporation of protein flexibility. The new force field has the strong advantage of incorporating the intramolecular energy into

the calculation, and has a conceptually satisfying formulation for the desolvation that includes terms for all atoms and does not require any assumptions for placement of charges. The force field will be made available with the release of AutoDock version 4.

Acknowledgment

This is manuscript 18017-MB from the Scripps Research Institute.

References

1. Schneider, G.; Bohm, H. J. *Drug Discovery Today* 2002, 7, 64.
2. Alvarez, J. C. *Curr Opin Chem Biol* 2004, 8, 365.
3. Oprea, T. I.; Matter, H. *Curr Opin Chem Biol* 2004, 8, 349.
4. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat Rev Drug Discov* 2004, 3, 935.
5. Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L. *Curr Pharm Des* 2005, 11, 323.
6. Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. *Annu Rev Biophys Biomol Struct* 2001, 30, 211.
7. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Struct Funct Genet* 2002, 47, 409.
8. Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J Comput Aided Mol Des* 2002, 16, 151.
9. Brooijmans, N.; Kuntz, I. D. *Annu Rev Biophys Biomol Struct* 2003, 32, 335.
10. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J Comput Chem* 1998, 19, 1639.
11. Huey, R.; Goodsell, D. S.; Morris, G. M.; Olson, A. J. *Lett Drug Des Discov* 2004, 1, 178.
12. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J Am Chem Soc* 1984, 106, 765.
13. Goodford, P. J. *J Med Chem* 1985, 28, 849.
14. Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. *J Med Chem* 1989, 32, 1083.
15. Mehler, E. L.; Solmajer, T. *Protein Eng* 1991, 4, 903.
16. Stouten, P. F. W.; Frommel, C.; Nakamura, H.; Sander, C. *Mol Simul* 1993, 10, 97.
17. Wesson, L.; Eisenberg, D. *Protein Sci* 1992, 1, 227.
18. Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. *Biochemistry* 1991, 30, 9686.
19. Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D. J.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. *J Comput Aided Mol Des* 2005, 19, 453.
20. Gasteiger, J.; Marsili, M. *Tetrahedron* 1980, 36, 3219.
21. Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins* 2006, 65, 15.

Software News and Updates

AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility

GARRETT M. MORRIS,¹ RUTH HUEY,¹ WILLIAM LINDSTROM,¹ MICHEL F. SANNER,¹
RICHARD K. BELEW,² DAVID S. GOODSSELL,¹ ARTHUR J. OLSON¹

¹*Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037*

²*Department of Cognitive Science, University of California, San Diego, La Jolla, California 92093*

Received 10 December 2008; Revised 27 January 2009; Accepted 3 February 2009

DOI 10.1002/jcc.21256

Published online 27 April 2009 in Wiley InterScience (www.interscience.wiley.com).

Abstract: We describe the testing and release of AutoDock4 and the accompanying graphical user interface AutoDockTools. AutoDock4 incorporates limited flexibility in the receptor. Several tests are reported here, including a redocking experiment with 188 diverse ligand-protein complexes and a cross-docking experiment using flexible side-chains in 87 HIV protease complexes. We also report its utility in analysis of covalently bound ligands, using both a grid-based docking method and a modification of the flexible sidechain technique.

© 2009 Wiley Periodicals, Inc. J Comput Chem 30: 2785–2791, 2009

Key words: AutoDock; computational docking; protein flexibility; covalent ligands; computer-aided drug design

Introduction

Automated docking is widely used for the prediction of biomolecular complexes in structure/function analysis and in molecular design. Dozens of effective methods are available, incorporating different trade-offs in molecular representation, energy evaluation, and conformational sampling to provide predictions with a reasonable computational effort.^{1–8} AutoDock combines an empirical free energy force field with a Lamarckian Genetic Algorithm, providing fast prediction of bound conformations with predicted free energies of association.⁹

In our hands, AutoDock3 has proven to be effective in roughly half of the complexes that we have studied. The remaining half show significant motion of the receptor upon binding, and thus have required a more sophisticated model of motion in the receptor, typically performed outside of AutoDock3. The new version of AutoDock described here—AutoDock4—incorporates explicit conformational modeling of specified sidechains in the receptor to address this problem. This capability also provides an effective method for analysis of covalently attached ligands.

Methods

Overview of AutoDock4

Since its release in 1990,¹⁰ AutoDock has proven to be an effective tool capable of quickly and accurately predicting bound conformations and binding energies of ligands with macromolec-

ular targets.^{9,11–14} To allow searching of the large conformational space available to a ligand around a protein, AutoDock uses a grid-based method to allow rapid evaluation of the binding energy of trial conformations. In this method, the target protein is embedded in a grid. Then, a probe atom is sequentially placed at each grid point, the interaction energy between the probe and the target is computed, and the value is stored in the grid. This grid of energies may then be used as a lookup table during the docking simulation.

The primary method for conformational searching is a Lamarckian genetic algorithm, described fully in Morris et al.⁹ A population of trial conformations is created, and then in successive generations these individuals mutate, exchange conformational parameters, and compete in a manner analogous to biological evolution, ultimately selecting individuals with lowest binding energy. The “Lamarckian” aspect is an added feature that allows individual conformations to search their local conformational space, finding local minima, and then pass this information to later generations. A simulated annealing search method and a traditional genetic algorithm search method are also available in AutoDock4.

AutoDock4 uses a semiempirical free energy force field to predict binding free energies of small molecules to macromolec-

Correspondence to: A. J. Olson; e-mail: olson@scripps.edu

Contract/grant sponsor: National Institutes of Health; contract/grant number: RO1 GM069832

ular targets. Development and testing of the force field has been described elsewhere.¹¹ The force field is based on a comprehensive thermodynamic model that allows incorporation of intramolecular energies into the predicted free energy of binding. This is performed by evaluating energies for both the bound and unbound states. It also incorporates a new charge-based desolvation method that uses a typical set of atom types and charges. The method has been calibrated on a set of 188 diverse protein–ligand complexes of known structure and binding energy, showing a standard error of about 2–3 kcal/mol in prediction of binding free energy in cross-validation studies.

Receptor Flexibility

AutoDock4 allows fully flexible modeling of specific portions of the protein, in a similar manner as the ligand. The user selects specific sidechains that are separated from the protein. During the simulation, these are treated explicitly, allowing rotation around torsional degrees of freedom using the same methods used to explore the conformational space of the flexible ligand. The remaining portion of the protein is represented using the affinity grids described earlier.

Methods for Covalent Docking

We have developed and tested two methods for docking of covalently-attached complexes: a grid-based approach and a modification of the flexible sidechain method. The grid-based approach calculates a special map for the site of attachment of the covalent ligand. A Gaussian function is constructed with zero energy at the site of attachment and steep energetic penalties at surrounding areas. The docking analysis is then performed by assigning a special atom type in the ligand for the atom that forms the covalent linkage. The docking simulation places this atom within the Gaussian well. One caveat is that this approach does not constrain the geometry of the covalent attachment to reasonable bond angles. To overcome this limitation, we tested the method using two Gaussian grids to define the bond that is formed during covalent linkage. Note, however, that the conformational freedom allowed with a single Gaussian grid may be an advantage if the method is used, for instance, to target ligands to metal coordination sites.

We also tested use of the flexible sidechain method for docking of covalent ligands. In this case, a coordinate file is created with the ligand attached to the proper sidechain in the protein, by overlapping ideal coordinates of the ligand onto the proper bond in the protein. This sidechain–ligand structure is then treated as flexible during the docking simulation, searching torsional degrees of freedom to optimize the interaction with the rest of the protein.

Overview of AutoDockTools

With the release of AutoDock3, it became apparent that the tasks of coordinate preparation, experiment design, and analysis required an effective graphical user interface to make AutoDock a widely accessible tool. AutoDockTools was created to fill this need. AutoDockTools facilitates formatting input molecule files, with a set of methods that guide the user through protonation,

calculating charges, and specifying rotatable bonds in the ligand and the protein (described later). To simplify the design and preparation of docking experiments, it allows the user to identify the active site and determine visually the volume of space searched in the docking simulation. Other methods assist the user in specifying search parameters and launching docking calculations. Finally, AutoDockTools includes a variety of novel methods for clustering, displaying, and analyzing the results of docking experiments.

AutoDockTools is implemented in the object-oriented programming language Python and is built from reusable software components.^{15,16} The easy-to-use graphical user interface has a gentle learning curve and an effective self-taught tutorial is available online. Reusable software components are used to represent the flexible ligand, the sets of parameters and the docking calculation, enabling a range of uses from a single use to thousands of docking experiments involving many different sets of molecules, facilitating automated high-throughput applications. For example, converting the NCI diversity database of small molecules into AutoDock-formatted ligand files was possible with a short Python script of less than 20 lines by leveraging the existing software components underlying AutoDockTools.

AutoDockTools exists in the context of a rich set of tools for molecular modeling, the Python Molecular Viewer (PMV).^{16,17} PMV is a freely distributed Python-based molecular viewer. It is built with a component-based architecture with the following software components: ViewerFramework, a platform for interactive visualization; Deja Vu, a generic OpenGL-based three-dimensional viewing component; and MolKit, a hierarchical data representation of molecules. AutoDockTools consist of a set of commands dynamically extending PMV with commands specific to the preparation, launching and analysis of AutoDock calculations. Hence, all PMV commands (such as reading/writing files, calculating and displaying secondary structure, adding or deleting hydrogens, calculating charges and molecular surfaces, and many others) are also naturally available in AutoDockTools. PMV also provides access to the Python-interpreter so that commands or scripts can be called interactively. PMV commands log themselves, producing a session file that can be rerun. In summary, AutoDockTools is an example of a specialization of the generic molecular viewer PMV for the specific application of AutoDock.

Specification of Ligand and Receptor Flexibility in AutoDockTools

AutoDockTools provides an interactive method for defining the torsional tree for a given ligand and receptor. Each step in this process has been automated to allow automatic assignment, for use in batch processes such as virtual screening. Ligand flexibility is assigned in several steps. First, a root atom is chosen, which will act as the fixed position during coordinate transformation in the docking simulation. To find the optimal atom, we evaluate the number of atoms in each branch, and choose the root atom that minimizes the size of the largest branch. In some cases, the user may wish to limit the flexibility of the ligand. AutoDockTools provides two choices to do this assignment automatically. One choice selects the set of torsional degrees of freedom that will move the largest number of atoms (torsions

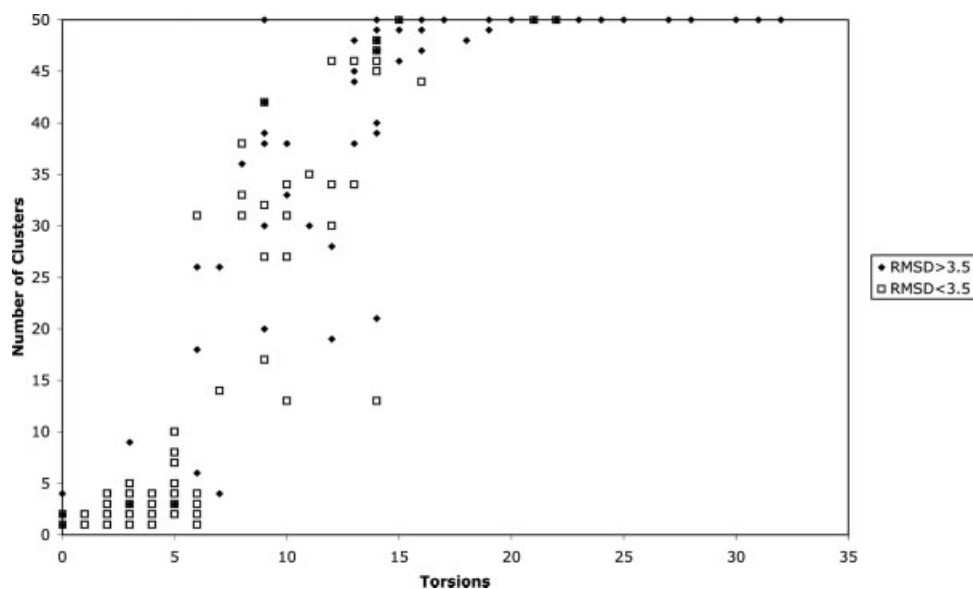


Figure 1. Redocking results. Results of redocking of 188 diverse ligand-protein complexes. Open squares represent complexes where the docked conformation with best predicted energy was less than 3.5 Å RMSD relative to the experimentally observed conformation. Dots are complexes where the best docked conformation was greater than 3.5 Å RMSD from the experimental structure.

near the root), the other adds torsions progressively from the leaves, moving the fewest number of atoms and leaving the core of the molecule rigid.

Validation Data Sets

Two sets of complexes were used for the validation of AutoDock4, both of which have been described previously.¹¹ A set of 188 diverse protein-ligand complexes were taken from the Ligand-Protein Database (<http://lpdb.scripps.edu>) and a set of 87 HIV protease complexes were taken from the PDDBind database (<http://www.pdbbind.org>). All coordinates were checked manually for the proper biological unit and consistency in naming schemes. Hydrogen atoms and charges were added in AutoDockTools, using Babel for hydrogens and the Gasteiger PEOE¹⁸ method for charges. Several misassigned charges were modified manually, as described in the previous report.

The 87 HIV protease complexes were aligned to allow easy comparison of docked conformations during cross dockings. An analysis of steric clashes was performed by swapping ligands within the set of 87 aligned complexes. ARG8 showed the largest number of bad contacts, with 877 cases where a ligand atom contacted a sidechain atom with less than 2 Å distance. Other contacts occurred with ILE50 (140 cases), PRO81 (95 cases), and PHE82 (88 cases, in two V82F mutants).

Docking Experiments

Docking experiments were performed with AutoDock4 and compared with docking experiments with AutoDock3. For each complex, 50 docking experiments were performed using the Lamarckian genetic algorithm with the default parameters from AutoDock3. A maximum of 25 million energy evaluations was

applied for each experiment. The results were clustered using a tolerance of 2.0 Å.

In the HIV cross dockings, ligand flexibility was limited to 10 torsional degrees of freedom, selecting torsions that allowed the fewest number of atoms to move (freezing the core of the molecule). Flexible docking was performed allowing three torsions to rotate in residue ARG8, in both the A and B chains. The structural water (water 301) was included in complexes that included this water in the crystallographic structure, and hydrogen atoms were added in geometry that allowed hydrogen bonding to the flaps.

We have also changed the default model for the unbound system in the current version of AutoDock. Our previous method calculated internal energies for an extended form of the molecule, mimicking a conformation that might be expected when fully solvated.¹¹ Results from beta testers, however, showed that this protocol has severe limitations when used for virtual screening. In cases where the ligand is sterically crowded, the artificial force field used to drive the ligand into an extended conformation tends to lead to conformations with sub-optimal energy. When the difference is calculated between this unbound conformation and the bound conformation, it leads to artificially favorable predictions of the free energy of binding. In response to this problem, we have returned to the default model of assuming that the unbound conformation of the ligand is the same as the bound conformation. Other options in AutoDock allow the user to use an energy-minimized conformation of the ligand as the unbound model.

Implementation and Availability

AutoDock 4.2 is currently being distributed free of charge as open source under a GPL license at the WWW site: <http://>

autodock.scripps.edu. ADT is being distributed free of charge as part of the MGLTools package, at the WWW site: <http://mgltools.scripps.edu/downloads>.

Results and Discussion

Redocking Analysis

Our first test of AutoDock4 is a redocking experiment using a set of 188 diverse protein-ligand complexes. The results, presented in Figure 1, are similar to the redocking study performed during the energy function calibration.¹¹ AutoDock4 successfully redocks most complexes with about 10 or fewer torsional degrees of freedom, but fails for most complexes with higher conformational flexibility. In 100 of 188 complexes, the docked conformation with lowest energy was within 3.5 Å RMSD of the crystallographic conformation. This is slightly better than a similar redocking experiment with AutoDock3. AutoDock3 suc-

cessfully docked complexes with about 10 or fewer torsional degrees of freedom, and in 97 of 188 complexes, the lowest energy conformation showed RMSD less than 3.5 Å (data not shown).

Cross Docking Analysis with Protein Flexibility

To validate the use of flexible sidechains in docking, we have used a set of 87 retroviral proteases with inhibitors. These proteins have a tunnel-shaped active site that wraps around a peptidomimetic inhibitor. An arginine-aspartate salt bridge forms at each end of the tunnel, bridging the two subunits. In complexes with large inhibitors, these arginines move to make space, whereas they adopt a more closed position in complexes with small inhibitors. In previous work, we have demonstrated that steric clash with these arginines prevents the docking of large inhibitors to proteins in the more closed conformation.¹²

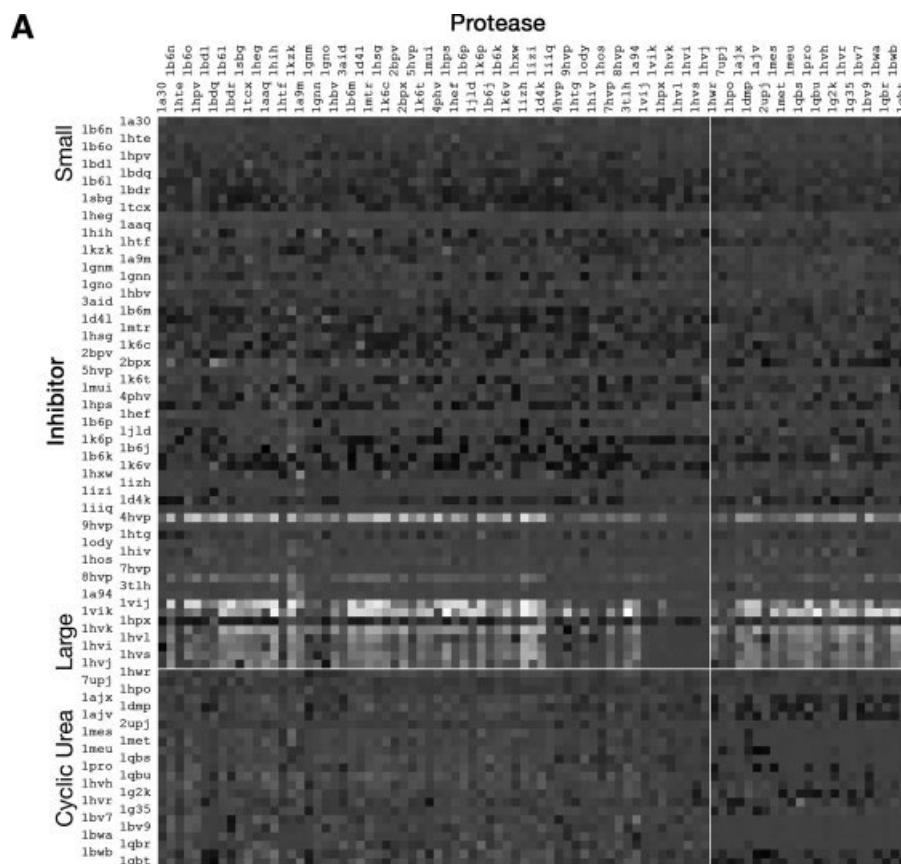


Figure 2. Cross docking results. (A) Difference in energy between rigid docking and flexible docking. White points are docking experiments where the flexible docking showed 20 kcal/mol more favorable predicted free energy of docking and black points are docking experiments where the rigid docking showed similar or worse free energy of docking. Inhibitors are ordered from small to large, with the cyclic urea inhibitors separated at the bottom. The protease structures are ordered similarly. (B) Cross docking with rigid protease structures. Each point is colored by the RMSD of ligand atoms from the crystallographic structure, with RMSD = 0 Å in white and RMSD >5 Å in black. (C) Cross docking with ARG8 treated as flexible in the protease. Each point is colored with the same scale as in (B).

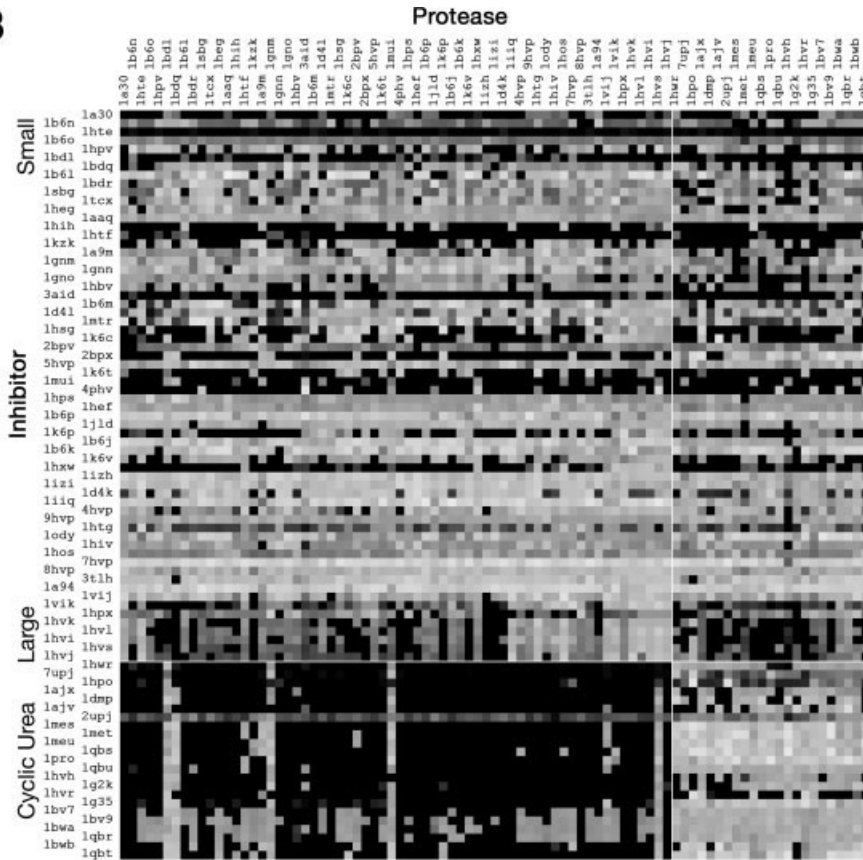
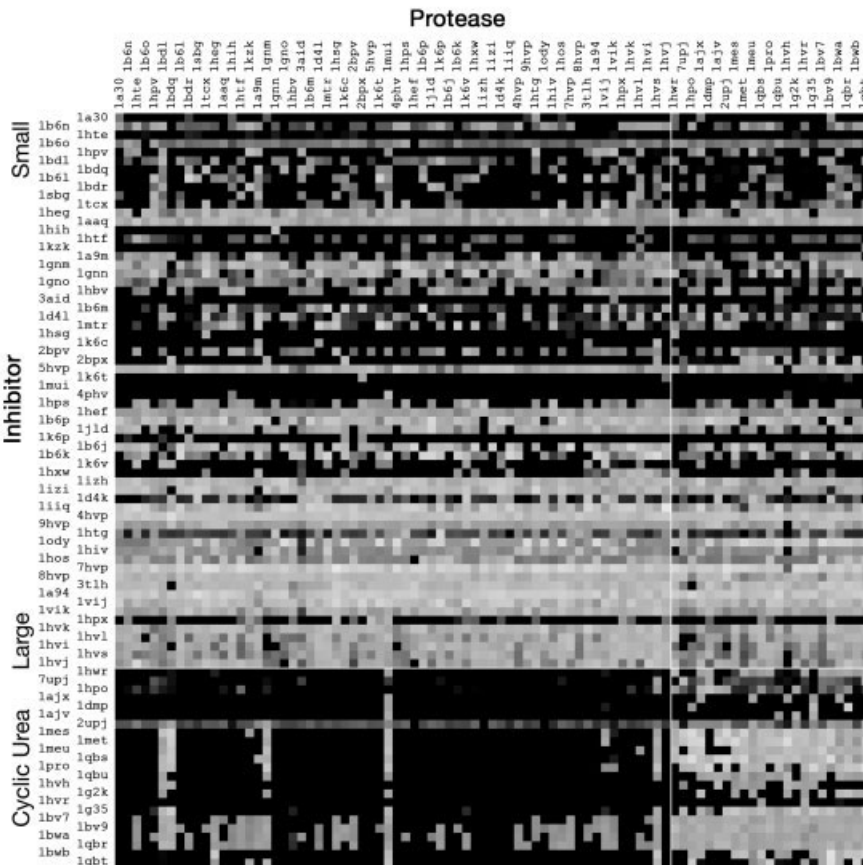
B**C**

Figure 2. (Continued)

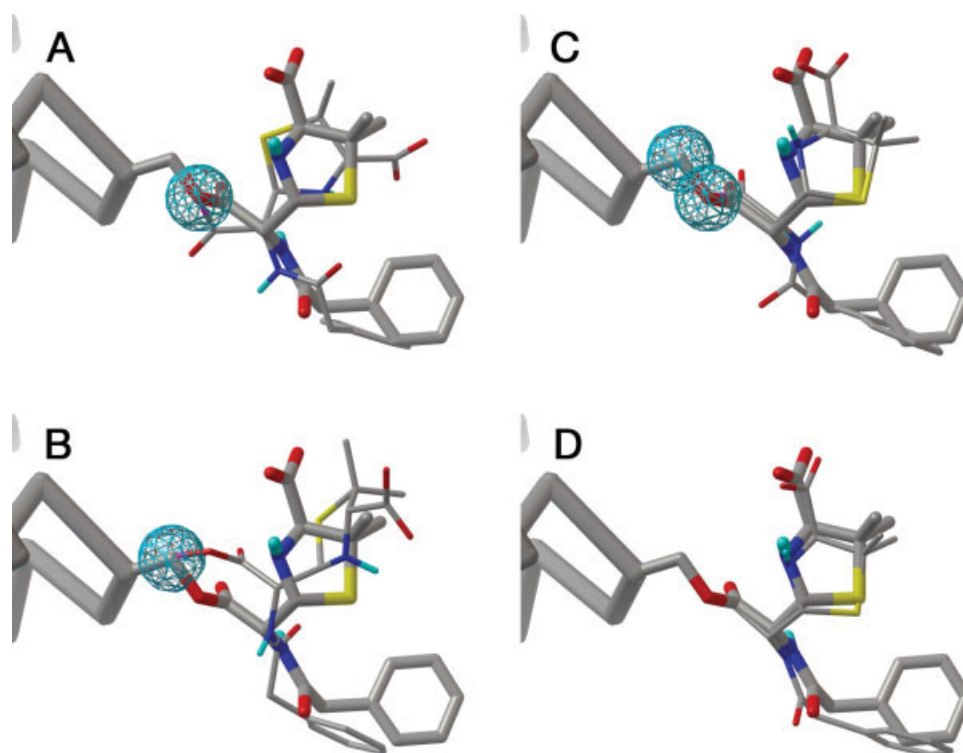


Figure 3. Results of covalent docking. (A) Using a Gaussian map centered on serine OG. The crystallographic structure is shown in large bonds and the best docked conformation is shown in thinner bonds. The blue sphere surrounds the region of most favorable energy in the Gaussian map. (B) Using a Gaussian map centered on serine CB. (C) Using two Gaussian maps. (D) Using a flexible sidechain to model the covalent ligand.

Using a distributed computing environment, we performed a large cross docking experiment, taking inhibitors from 87 crystallographic structures and docking them to the protein conformations from these structures. We performed two parallel experiments, one with a rigid protein target, and one modeling the two ARG8 amino acids as flexible. Results are shown in Figure 2. Parameters were chosen that correspond to a typical computational effort for an individual docking experiment, with 25,000,000 evaluations and 50 docked conformations. The torsional degrees of freedom in the ligand were limited to 10, with an added six degrees of freedom for the two arginine sidechains in the flexible sidechain tests.

The complexes in Figure 2 are arranged to highlight the structural features of the complexes. They are arranged from top to bottom based on the size of the inhibitor, with small inhibitors at the top and large inhibitors at the bottom. The cyclic urea inhibitors and similar inhibitors that are designed to displace the structural water 301 are separated at the bottom, also ordered from small to large. The protease structures are arranged in identical order from left to right, so the diagonal corresponds to redocking experiments. Note that the large block at lower left corresponds to cyclic urea inhibitors docked to protease structures that include water 301, so we don't expect that they will dock correctly. The large block at upper right includes peptidomimetic inhibitors docked to protease structure that do not

include water 301, so we might expect some loss of specificity in docking, although this is not observed.

As seen in Figure 2A, flexible docking improves docking in the complexes that are most expected to benefit. The white band below center corresponds to large inhibitors binding to protease structures that were solved with small inhibitors. These complexes show the most favorable change in docking energy with flexible docking, as unfavorable contacts with the ARG8 residues are relaxed.

Looking to the RMSD results in Figure 2B, we see that rigid docking fails with several types of complexes. The large block of cyclic urea inhibitors at lower left show poor RMSD values, as expected. Roughly 2/3 of the small inhibitors were docked successfully, and the mid-size ones were very successful. The horizontal block just below center (inhibitors 1hvk through 1hvj) are cases where large inhibitors are docked to active sites with ARG8 in a nonpermissive position, so these are the complexes where we expect improvement with a flexible protease model.

These results underscore the fact that in this experiment, the larger inhibitors are actually an easier problem, for two reasons. First, the space searched in each experiment includes the active site and a small region outside the protein. In some cases, small ligands find incorrect conformations entirely outside the active site, but the larger ligands are forced to be at least partially inside the active site, due to their size. Second, several torsional

degrees of freedom are frozen in the proper orientation in the large inhibitors, and this tends to favor the crystallographic conformation.

The results for the cyclic urea inhibitors follow a similar trend. The block at lower left are results for docking cyclic urea inhibitors into protease structures that include the structural water, so it is not surprising that the experiments do not find the proper answer. The block at lower right shows docking of cyclic urea inhibitors with protease structures without the structural water. It shows a similar trend of large inhibitors docking more consistently than the smaller inhibitors. The block at upper right shows docking of other inhibitors to proteases without the structural water. These experiments perform similarly to the experiments with the structural water.

When we add flexibility to the two ARG8 at the ends of the active site tunnel, we achieve the desired result with the large inhibitors that contact this residue. AutoDock4 successfully docks the large inhibitors in cases where the docking failed with rigid receptors (Fig. 2C). Unfortunately, the experiments with flexible sidechains showed significantly worse results for the small inhibitors, failing in over half of the cases. This underscores another limitation of use of flexible sidechains. They increase the size of the conformational space to be searched-in these cases, increasing it from a problem that is accessible to AutoDock to a problem with too many degrees of freedom.

Covalent Docking

We tested two methods for docking covalently attached ligands: a grid-based approach and an approach using flexible sidechains. The results are shown in Figure 3 for docking of penicillin in a covalent complex with D-alanyl-D-alanine carboxypeptidase. In the first two experiments, we created a Gaussian map centered on the CB or OG atoms of SER62 (Figs. 3A and 3B). We then performed docking experiments with the appropriate atoms in the ligand targeted to that map. Poor results were obtained: a wide range of conformations was found and the conformation of best energy showed only slight resemblance to the crystallographic conformation. Better results were obtained when two Gaussian maps were used, corresponding to CB and OG of the serine (Fig. 3C). This forced the docked ligand to adopt the appropriate geometry of bonding with the serine, and resulted in docked conformations that were much more similar to the crystallographic conformation.

Use of a flexible sidechain to model the covalent ligand gave excellent results. All 10 docking runs gave similar conformations, all of which were very similar to the crystallographic conformation (Fig. 3D).

Conclusions

Dependence on grid-based energy evaluation is a major limitation of AutoDock4. It is required to allow rapid evaluation of binding energies during the docking simulation, but it places a

severe restriction on the representation of the target macromolecule: all of the atoms included in the grid must be treated as rigid. The off-grid modeling of specific sidechains is a method for incorporating limited flexibility within this paradigm, and the results presented here show that it will be effective in some cases. However, adding flexibility presents several problems: (1) the calculation of the receptor energy is more computationally intensive since flexible regions must be evaluated by a full pairwise energy evaluation, and (2) the conformational space is larger, and hence, there is more potential for false positives. Future solutions to these problems will require advances in energetic functions, simplifying and/or speeding up pairwise approaches, and use of hierarchical approaches^{19,20} that allow different levels of sophistication in a docking simulation.

Acknowledgments

This is manuscript 19885 from the Scripps Research Institute.

References

1. Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. *J Med Chem* 2006, 49, 5851.
2. Coupez, B.; Lewis, R. A. *Curr Med Chem* 2006, 13, 2995.
3. Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins* 2006, 65, 15.
4. Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L. *Curr Pharm Des* 2005, 11, 323.
5. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat Rev Drug Discov* 2004, 3, 935.
6. Brooijmans, N.; Kuntz, I. D. *Annu Rev Biophys Biomol Struct* 2003, 32, 335.
7. Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J Comput Aided Mol Design* 2002, 16, 151.
8. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Struct Funct Genet* 2002, 47, 409.
9. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J Comput Chem* 1998, 19, 1639.
10. Goodsell, D. S.; Olson, A. J. *Proteins: Struct Funct Genet* 1990, 8, 195.
11. Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. *J Comput Chem* 2006, 28, 1145.
12. Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. *Proteins: Struct Funct Genet* 2002, 46, 34.
13. Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. *J Comput Aided Mol Design* 1996, 10, 293.
14. Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J Mol Recogn* 1996, 9, 1.
15. Lutz, M.; Asher, D. *Learning Python*; O'Reilly & Associates: Sebastopol, CA, 1999.
16. Sanner, M. F. *J Mol Graphics Mod* 1999, 17, 57.
17. Sanner, M. F. *Structure* 2005, 13, 447.
18. Gasteiger, J.; Marsili, M. *Tetrahedron* 1980, 36, 3219.
19. Zhao, Y.; Stoffler, D.; Sanner, M. F. *Bioinformatics* 2006, 22, 2768.
20. Zhao, Y.; Sanner, M. F. *J Comput Aided Mol Des* 2007, 22, 673.