PROTEINS: Structure, Function, and Genetics 23:472–490 (1995)

RESEARCH ARTICLES

An Automated Method for Dynamic Ligand Design

Andrew Miranker and Martin Karplus
Committee on Higher Degrees in Biophysics and Department of Chemistry, Harvard University, Cambridge Massachusetts 02138

ABSTRACT **An automated method for the dynamic ligand design (DLD) for a binding site of known structure is described. The method can be used for the creation of de novo ligands and for the modification of existing ligands. The binding site is saturated with atoms (sp^3 carbon atoms in the present implementation) that form molecules under the influence of a potential function that joins atoms to each other with the correct stereochemistry. The resulting molecules are linked to precomputed functional group minimum energy positions in the binding site. The generalized potential function allows atoms to sample a continuous parameter space that includes the Cartesian coordinates and their occupancy and type, e.g., the method allows change of an sp^3 carbon into an sp^2 carbon or oxygen. A parameter space formulated in this way can then be sampled and optimized by a variety of methods. In this work, molecules are generated by use of a Monte Carlo simulated annealing algorithm. The DLD method is illustrated by its application to the binding site of FK506 binding protein (FKBP), an immunophilin. De novo ligands are designed and modification of the immunosuppressant drug FK506 are suggested. The results demonstrate that the dynamic ligand design approach can automatically construct ligands which complement both the shape and charge distribution of the binding site.** © 1995 Wiley-Liss, Inc.

Key words: drug design, FKBP, FK506, immunophilin, MCSS, DLD

INTRODUCTION

Rational ligand design endeavors to suggest new compounds which bind strongly to biologically active regions of proteins of known three-dimensional structure. Ideally, synthetic feasibility and clinical effectiveness would be taken into account in such a design, but most of the present research in the field concentrates on the binding alone. Our approach to ligand design is to divide the task into parts, which is much more efficient that doing the design process in a single step. Such a strategy permits one to perform combinatorial chemistry on the computer.[31] First, we developed a method, the multiple copy simultaneous search (MCSS) method, for the determination of energetically favorable positions and orientations for small organic functional groups (e.g., methanol, acetonitrile) on a protein surface. MCSS has already proved to be a useful tool for the analysis of binding regions[1] and as a starting point for ligand design.[2] Given the MCSS minima, the next step is to connect them to form molecules. One fruitful approach is to search a database for compounds ("skeletons") that can link several MCSS functional groups and fill into the binding site.[3] The method presented here is more general because it can generate new molecules not represented in databases. Also, it is well suited for modifying known ligands by linking them to MCSS minima.

The essential idea of the present approach is to saturate a site with atoms and/or small molecules (sp^3 carbon in the present implementation) and to form molecules in an automatic fashion by using a generalized potential function that joins them to each other and to functional groups, whose positions have been determined by MCSS. The potential function is designed so that bonded systems with proper geometry (i.e., molecules) have lower energies than the separated species. A potential function with a continuous first derivative is used to permit existing and novel techniques for sampling and optimization to be employed. Thus, molecular dynamics or Monte Carlo methods can serve to sample the space,[4] while minimization,[5] genetic algorithms,[6] simulated annealing,[7] energy embedding,[8] etc., can be used for optimization. In this study, the potential function is sampled with a Metropolis scheme and optimized with a simulated annealing protocol.

Received February 3, 1995; revision accepted May 24, 1995.
Address reprint requests to Martin Karplus, Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138.
Current address of A. Miranker: Oxford Centre for Molecular Sciences and New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, U.K.

© 1995 WILEY-LISS, INC.

When we first began to develop methods for computer aided ligand design, there were very few approaches available. Now there are many. One widely used approach to the ligand design problem is based on the searching of structural databases. The methods used for this vary considerably depending on whether the database is compiled from existing molecules or constructed using geometric rules, whether the binding site is represented explicitly, with complimentary surfaces or functionality maps, and whether the result of the search represents the final molecule or a linker joining different pharmacophoric positions.[3,9–13] Also, algorithms to combine multiple hits in a database search into a single molecule have been develped.[14] Database searching has several desirable properties. These include computational speed with some methods fast enough to provide an interactive tool. In the case of databases made up of existing molecules, the results are immediately testable or readily synthesized in most cases. The main limitations are the relatively small number ($\sim 10^6$) of members of a database and the representation of flexible molecules in databases. The latter limitation has been addressed, in part, by duplicating database members in alternative conformations.[15] The effect of the former limitation can be overcome by increasing the size of the database. Since this is usually not based on the specific site of interest, the number of suggested ligands resulting from a search will, at best, increase linearly with the size of the database. It is also possible to make the use of a database secondary to the de novo design process. Toward this end, ligand design techniques have been developed in which the database consists of relatively few members which are used as building blocks. One such algorithm begins with a user specified nucleation site and constructs a polypeptide chain one residue at a time.[16] All specified side chains in a variety of conformations are sampled for the new residue and a scoring function is used to evaluate the interactions with the binding site. A related approach,[2] which does not depend on an initial seed position, uses a pseudo-energy function to evaluate all possible ways of connecting favorable sites for *N*-methylacetamide groups obtained with the MCSS approach.[1] The best scoring main chains are then given side chains by application of the same pseudo-energy function using functionality maps of protein side chains (e.g., methanol, ethyl guanidinium). Both of these methods have shown predictive success. Other algorithms focus on the problem of connecting disparate pharmacophoric positions in the binding site. One example builds connections iteratively from a library of small aliphatic templates.[17] Another first determines families of acyclic chains which can connect pairs of pharmacophoric sites and then introduces additional connections to form rigid cyclic structures.[18]

The goal of the present DLD approach is to suggest many plausible ligands or fragments of ligands for a protein binding site while using a minimum of computational resources. This frees us from some of the many restrictions that lead to the computationally intensive nature of standard molecular simulations. We make use of this freedom to make important simplifications, e.g., rigid body transformations to move groups of atoms, reduction of the protein surface to its solvent exposed atoms, and precomputed functional group minima. Most importantly, we are able to introduce a novel potential function that includes nonphysical terms and degrees of freedom so as to provide a mechanism for transforming one molecule into another and generate new molecules within a continuous function space. Within this space, atoms can form and break bonds, appear and disappear, and change atomic number or hybridization. By applying the pseudo-energy function to the system, we are able to generate molecules which have shape complementarity and favorable functional group interactions with the binding site of interest. These molecules can serve to introduce modifications to known ligands or to develop new lead compounds. The method is described in the next section with emphasis on the nature of the pseudo-energy function and its use for generating molecules. Consideration is also given to the optimization procedure. To illustrate the method, it is applied to ligand generation in the binding pocket of the immunophilin FK506 binding protein, FKBP.

METHODS

Potential Function

Overview

The DLD approach can be described in terms of the pseudo-energy function and the algorithms which govern sampling and optimization. The pseudo-energy function as currently implemented consists of a sum of bond, bond–angle, nonbonded, and protein–ligand terms. The function, $f_{\text{total}}[\vec{x}, \vec{o}]$ is given in Eq. (1) and its contributing terms are considered in detail in the sections that follow.

$$\begin{aligned} f_{\text{total}}[\vec{x},\vec{o}] = & \sum_{i,j} o_i o_j f_{\text{nonbond}}[|\vec{x_i} - \vec{x_j}|] + \\ & \sum_{i,j} o_i\, o_j\, f_{\text{nonbond}}[|\vec{x_i} - \vec{x_j}|] f_{\text{bond}}[|\vec{x_i} - \vec{x_j}|] \\ & + \sum_{i,j,k} o_i o_j o_k f_{\text{angle}}[|\vec{x_i} - \vec{x_k}|] \qquad (1) \\ & g\{f_{\text{bond}}[|\vec{x_i} - \vec{x_j}|]\} g\{f_{\text{bond}}[|\vec{x_k} - \vec{x_j}|]\} \\ & + \sum_{i,j} f_{\text{protein–ligand}}[|\vec{x_i} - \vec{x_j}|] \end{aligned}$$

The symbol $\vec{x}$ represents the $3n$ Cartesian coordinates for the n atoms in the system, $\vec{o}$ are the occupancy degrees of freedom of which there can be as many as n, and $\vec{x_i}$ and o_i are the 3 Cartesian and occupancy degrees of freedom for atom i, respectively. The first summation, f_{nonbond}, describes the

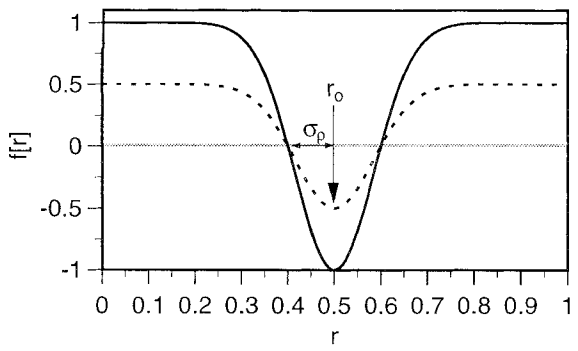


Fig. 1. Plot of Eq. (2) illustrating the effect of occupancy on the value of f. $r_0 = 0.5$, $\sigma_\rho = 0.1$. The solid line represents unit occupancy while the dashed line is the result when the product of the component occupancies equals 0.5.

nonbonded term which is applied to all atom pairs in which neither atom is designated as bond forming and whose occupancies in the binding site are $o_i o_j$. The second summation introduces nonbonded and bonded potentials for overlapping atoms in which both are designated as capable of bond formation. The third term is for bond angles and is applied to all bond pairs with a common vertex. The final term represents protein–ligand interactions.

Equation (1) indicates certain properties of the potential function that play a role in their use. First, all terms have pairwise distances as their arguments. This makes implementation more efficient, especially on machines with a vector architecture. Second, the bond angle components are attenuated by a function $g\{f_{\text{bond}}[|\vec{x}_i - \vec{x}_j|]\}$ which varies between zero and one, and has the two bond lengths involved in the particular bond angle as its arguments. This permits bonds to form and break without discontinuous changes in the bond angle energy. Future implementations could use a pairwise distance potential for 1–4 interactions, which could be attenuated similarly by $g\{f_{\text{angle}}[|\vec{x}_i - \vec{x}_k|]\}$. Lastly, all purely ligand terms are scaled by a fourth degree of freedom, the occupancy which varies between zero and one. Protein–ligand terms are not scaled so as to ensure that ligand atoms do not penetrate the protein surface. The potential functions are designed so that they are negative for acceptable configurations (e.g., 1.5 ± .1 Å for sp^3–sp^3 bonds) and positive otherwise. Optimization of the system forces unrealistic configurations toward zero occupancy and realistic configurations to unit occupancy. Occupancy is fundamental to DLD as it provides a mechanism for enhanced sampling of putative ligands by dynamically scaling interactions between ligand atoms by factors between 0 and 1. Similar to MCSS methodology, this enables more atoms or groups of atoms to be placed in a binding site than would be sterically possible at unit occupancy.[1] Furthermore, occupancy provides a mechanism for eliminating unrealistic configurations and, as will be shown, could allow atoms to change atom type.

The representation of the binding site is designed to maximize the specificity of DLD generated ligands while using a minimum of computational resources. Only atoms of the protein which are solvent accessible are included. Specificity is achieved by using MCSS calculated positions for polar and charged functional groups whose subsequent translation during DLD is restricted and whose rotation is limited to a single axis so as to maintain the electrostatic contacts made with the protein. Only linker atoms (sp^3 carbon atoms) are permitted to freely diffuse in the space of the binding site with their motions ultimately restricted by the protein surface and a cubic reflecting boundary.

Functional form

The functions f in Eq. (1) are inverted Gaussians and accomplish several tasks, three of which recur in the bond, bond–angle, and nonbonded terms [Eq. (2)] (Fig. 1).

$$f_\rho[r] = \left\{1 - 2\exp\left(\frac{(r-r_0)^2\ln(1/2)}{\sigma_\rho^2}\right)\right\} \qquad (2)$$

First, for a given property (ρ), such as a bond length, f has only one minimum at the ideal value: e.g., the minimum, r_0, for an sp^3-sp^3 C–C bond is 1.5Å; the rounded value 1.5Å, rather than 1.54Å, is used for simplicity. Second, within a user specified tolerance (σ_ρ) of r_0,f is negative so that occupancy of the component atoms ($o_i o_j$) tends to increase, while outside of this tolerance, f is positive and the occupancy tends to decrease (Fig. 1). Since a scaled Gaussian is used, f has a well depth and barrier height independent of r_0 and σ_ρ. Because f rises asymptotically to a limiting value (usually one), little distinction is made between different nonphysical configurations. This ensures that optimization is not inadvertently skewed to encourage particular structures; e.g., addition of a carbon to methanol should not be more easily optimized than addition of a carbon to acetonirile, unless there is a specific reason for searching for special ligands. If it is desired to bias the sampling of particular structures, this can be done deliberately by applying weights to terms in the potential function.

Nonbonded term

A nonbonded potential (Eq. 3) is applied to all pairs of putative ligand atoms in the site, i.e.,

$$o_i o_j f_{\text{nonbond}}[r] = \begin{cases} o_i o_j\left\{1 - \exp\left(\dfrac{(r-r_0)^2\ln(1/2)}{\sigma_{\text{nonbond}}^2}\right)\right\} & r < r_0 \\ = 0 & r \geq r_0 \end{cases} \qquad (3)$$

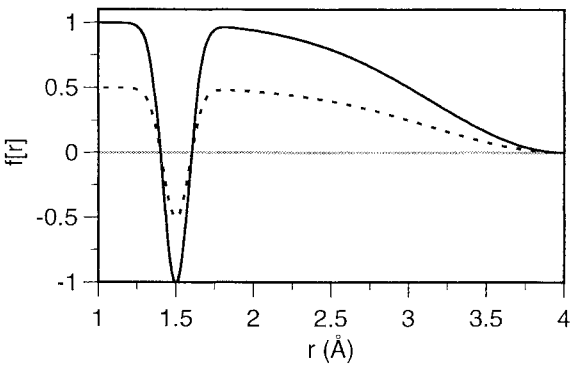


Fig. 2. Bonding term [Eq.(4)] for a pair of sp^3 carbons. r_{ob} = 1.5 Å, r_{onb} = 4.0 Å, σ_b = 0.1 Å, σ_{nb} = 1.0 Å. The solid line is the full potential while the dashed line shows the effect of diminished occupancy.

Protein–ligand interactions are treated separately. There are no attractive or electrostatic contributions to the nonbonded potential; only a repulsive term is included to restrict atomic overlap. Implementation of a nonbonded potential which includes only short-range repulsive terms allows much smaller distances to be considered when building lists of interacting atom pairs. Eq. 3 differs from Eq. 2 in that its range is from zero to one rather than minus one to one. The potential must go to zero at some specified contact distance or else all atom pairs in the system have to be evaluated to avoid discontinuities in the potential. Since the nonbonded potential is positive everywhere, occupancy is never increased due to nonbonded interactions.

The zero point (r_0) for the potential is set at a distance where the Lennard–Jones potential is between one and 2 kT (T = 298 K) using CHARMM parameters.[19] This allows unrestricted movement of ligand atoms until their overlap exceeds what is reasonable at room temperature. The deviation $\sigma_{nonbond}$ is typically assigned a value greater than 1.0 Å. This ensures that the gradient at r = 0 is greater than the square root of the computer precision. Although this consideration is unnecessary for Monte Carlo sampling, it is important for optimization by first derivative methods.[5]

Bonding term

A bonding potential is applied between specified atoms (see below) when they are closer than their nonbonded contact distance. The potential used is shown in Eq. (4).

$$o_i o_j f_{nonbond}[r] f_{bond}[r] = o_i o_j \left\{1 - \exp\left(\frac{(r-r_{onb})^2 \ln(1/2)}{\sigma^2_{nonbond}}\right)\right\} \left\{1 - 2\exp\left(\frac{(r-r_{ob})^2 \ln(1/2)}{\sigma^2_{bond}}\right)\right\} \quad r < r_{onb}$$
$$= 0 \quad r \geq r_{onb} \tag{4}$$

It contains the nonbonded potential as a factor so that atom pairs flagged as bond forming experience a potential with both bond and nonbonded character. The bonding factor dominates in the region close to the optimum bond length, r_{ob}; the parameter σ_b is the allowed deviation from the optimum bond length. The bond minimum, r_{ob}, is taken from CHARMM parameters with the zero points for the energy function occurring at $r = r_{ob} \pm \sigma_b$. The parameters r_{onb} and σ_{nb} correspond to the van der Waals contact distance and deviation, respectively (Fig. 2). A narrow σ_b ($\sim 10^{-2}$) would reflect the vibrational fluctuations accessible to a bond at room temperature. However, this would result in flat regions of the pseudo-energy function leading to inefficient optimization. Instead, we choose a σ_b (σ_b = 0.1 Å) which is large enough for effective optimization, but small enough so that resultant candidate ligands are not excessively altered by subsequent minimization under a traditional molecular mechanics potential.

In the present implementation, carbon atoms are the only atom type permitted to dynamically form bonds. These carbon atoms can have dynamic hybridization and they can be freely moving linker atoms, part of an MCSS functional group or a rigid collection of atoms (e.g., benzene). This restriction has the advantage that it allows valence to be dealt with implicitly since once the valence of a carbon is satisfied, additional bonds cannot be made to it without forming unfavorable bond angles. Other atom types could be allowed to dynamically form bonds with the same implicit valence criterion by introducing dummy atoms (e.g., the lone pairs of an ether oxygen). After optimization, any unsatisfied valency is taken care of by adding hydrogen atoms.

Established molecular mechanics potentials use a Coulomb potential and partial charges to approximate nonbonded electrostatic interactions. For atoms which are close in space as a result of one or two covalent bonds, the Coulomb potential is either not calculated, or is scaled by an empirically determined factor. In a system which allows dynamic bond formation, this results in significant added complexity. In DLD, dynamic bonding occurs only between carbon atoms with the MCSS probes and multiatom linkers including one or more additional carbon atoms in their representation of basic functionalities, e.g., methanol is used instead of hydroxyl and the carbon atoms can only bond to the methanol CH_3 group. This means that the polar atoms (i.e., those atoms with significant partial charge) of MCSS functionalities and multiatom linkers will be separated by at least 3 bonds involving two or more carbon atoms. In the resulting structures, possible complexities associated with through bond attenuation of partial charges are eliminated.

Bond-angle terms

All pairs of bonds with a common vertex atom, $\vec{x}_2$, introduce a bond angle potential term [Eq.(5)].

$$o_1o_2o_3 f_{\text{angle}}[r_{13}]g\{f_{\text{bond}}[r_{12}]\}g\{f_{\text{bond}}[r_{23}]\} = o_1o_2o_3\left\{1 - 2\exp\left(\frac{(r_{13}-r_{o13})^2 \ln(1/2)}{\sigma^2_{\text{angle}}}\right)\right\} \times \frac{1 - \left\{1 - 2\exp\left(\frac{(r_{12}-r_{o12})^2 \ln(1/2)}{\sigma^2_{\text{bond}}}\right)\right\}}{2} \frac{1 - \left\{1 - 2\exp\left(\frac{(r_{23}-r_{o23})^2 \ln(1/2)}{\sigma^2_{\text{bond}}}\right)\right\}}{2} \quad (5)$$

It is a product of functions, of which the first, involving atoms $\vec{x}_1$ and $\vec{x}_3$, is an angle term and the second and third are attenuation terms based on the bonding potential for atoms $\vec{x}_1 - \vec{x}_2$ and $\vec{x}_3 - \vec{x}_2$. For simplicity and speed, the distance, r_{13}, between the nonvertex atoms, $\vec{x}_1$ and $\vec{x}_3$, is used to represent the angle. The bond–angle energy [Eq.(5)] passes through zero at $r_{o13} \pm \sigma_{13}$, where σ_{13} is typically set to 0.2 Å. Values of r_{o13} are chosen so that the angle is tetrahedral (for sp^3 carbon) for bond lengths equal 1.5 Å; this leads to $r_{o13} = 2.45$ Å.

The bond angle function is attenuated by terms corresponding to the bonds which form the angle for three reasons. First, there is little point in large bond angle contributions to the total pseudo-energy if the bonds comprising it are highly strained. Second, it is desirable to couple the bond and bond angle terms so that a realistic bond geometry is favored if the resultant angle is near its optimum and to discourage bond formation otherwise. Finally, to maintain a continuous function, it is important that the bond angle term goes to zero at the point where the bond is broken (i.e., at the van der Waals contact distance). These three considerations are met by the attenuation terms of the bond angle function. The bond attentuation terms, $g\{\}$, tend to one when the bond is near its minimum position and tend to zero as the bonds are stretched.

Protein ligand interaction term

Calculation of interactions between a ligand and the protein are CPU intensive, especially if one includes long-range interactions and flexibility in the protein. Since MCSS minima are precomputed in the presence of the full protein potential,[1] the steric and electrostatic properties of the binding site that are needed for constructing complete molecules can be introduced by using a simplified representation of the protein surface, and by constraining functional groups to remain in the vicinity of their precomputed minima. Since we are considering freely diffusing carbon atoms as linkers, only a repulsive potential is applied when the ligand and protein overlap [Eq.(6)].

$$f_{\text{protein}}[r] = (r - r_0)^2 \quad r < r_0 \qquad = 0 \quad r \geq r_0 \quad (6)$$

The zero point (r_0) for the potential is the same as that of the nonbond term (see Nonbonded term). The form adopted here is that now used in the X-PLOR program for structure determinations. This provides a simple approximation to the steric interactions between protein and ligand. Many descriptions such as Voronoi polyhedra[20] or complimentary surface generation[21] could have been employed for a simplified description of a protein surface. We used a solvent accessible surface calculation[22] performed on the protein using a 1.4 Å radius probe. All atoms of the protein binding site with exposed surface are included in the protein ligand potentials; this yields on the order of 100 atoms for a typical binding site.

To constrain atoms to the vicinity of the binding site, a cubic reflective boundary is present. A harmonic boundary potential was tested first; however, the reflective boundary is more suitable because the available coordinate space is finite so that objective optimization of the simulated annealing schedule is facilitated.[23,24]

Occupancy

Variable occupancy is introduced in Eq. (1) to make sampling more efficient. By reducing the pseudo-energy cost when the occupancy is less than unity, it enhances sampling of the DLD potential in regions of internal strain and atomic overlap and allows more atoms to be present in a binding site than its volume would otherwise permit. In addition, it provides a mechanism for eliminating atoms forming unphysical structures. Variable occupancy is implemented by using a fourth dimension, θ, for each atom or group of atoms (e.g., a benzene or cyclopropane ring) which have the freedom to appear or disappear. The occupancy [Eq.(7)] varies smoothly between zero and one; in converged simulations, all atoms have limiting occupancies. The occupancy factor scales all ligand–ligand terms in the potential for atoms which have been designated by the user as having variable occupancy. Ligand–protein interactions are not scaled to ensure that the ligand does not penetrate the protein surface.

$$o = 1 - \cos^2(\theta) \quad (7)$$

The occupancy function is generalizable to the sampling of two or more mutually exclusive properties. Consider that a physical property ρ_{total}, is the weighted sum of **n** possibilities [Eq.(8)]. The cosines of this equation are simply the elements of a unit vector of length **n.** By perturbing the orientation of this unit vector,[25] weighted average properties are sampled. Orientations of the unit vector which do

not correspond to a single cosine equal to unity represent a nonphysical system. Correspondingly, the product of the cosines are used to generate an occupancy [Eq.(9)] which is applied to all energy terms which depend on the property.

$$\rho_{\text{total}} = \sum_{i=1}^{n} \rho_i \cos^2(\theta_i) \tag{8}$$

$$o = 1 - n^n \prod_{i=1}^{n} \cos^2(\theta_i) \tag{9}$$

An application of the use of the occupancy variable is the optimization of the hybridization state of a carbon atom. In this case, the optimum bond distance, r_0, between an sp^3 carbon and a carbon with variable hybridization would be given by Eq. (10).

$$r_0 = r_0^{sp^3-sp^3} \cos^2(\theta_{sp^3}) + r_0^{sp^3-sp^2} \cos^2(\theta_{sp^2}) + r_0^{sp^3-sp} \cos^2(\theta_{sp}) \tag{10}$$

As different weighted averages for the bond length are sampled, the occupancy of the variable carbon (or of the MCSS group containing the variable carbon) is given by Eq. (11).

$$o = 1 - 27\cos^2(\theta_{sp^3})\cos^2(\theta_{sp^2})\cos^2(\theta_{sp}) \tag{11}$$

In this way, intermediate, nonphysical hybridization states are sampled, but with low occupancy. In the event that several properties of an atom are being sampled, the total occupancy of the atom would be the product of all the occupancies associated with each property.

The occupancy degree of freedom resembles the use of the parameter, λ, in free energy perturbation calculations;[26] in fact, such a smooth interpolation function may be useful in achieving improved convergence in free energy simulations. Although DLD and free energy difference calculations have different goals, both λ and DLD occupancy vary between zero and one and sample nonphysical configurations so as to be able to go smoothly from one physical system to another. Whereas free energy difference calculations converge best when the change between perturbed (λ = 1) and unperturbed (λ = 0) systems is small and often (but not always) use a single λ,[27] many DLD occupancies may be introduced to sample large ranges of physical properties. Also, in the DLD application, there is no need for changes to go reversibly from one state to another.

There are many possible applications of the occupancy variable. For example, the two planar conformations of a tyrosine hydroxyl could be sampled without the need to sample conformations where the hydrogen is out of the plane. In addition, a DLD optimization could be interrupted periodically by database searches to connect certain regions of the developing ligand with elements from a library of small rigid linkers. By introducing candidates from the search with zero occupancy, the pseudo-energy would not change abruptly.

Sampling and Optimization

Functional group flexibility

Functional group minima computed for a rigid protein make contacts which suggest the nature of the motional freedom accessible to them. Methanol minima which both donate and accept hydrogen bonds can rotate freely about the hydroxyl O–H bond axis without affecting the magnitude of electrostatic interactions with the protein. Similarly, if each of the oxygen atoms of an acetate group accepts a hydrogen bond, the group can rotate about the axis formed by the line connecting the two carboxyl oxygen atoms. Rotational freedom for polar and charged MCSS minima are always limited to such a specified axis, while translational freedom is restricted to ±0.25 Å. These choices reflect a balance between maximizing the ability of a functional group to form links to other atoms and minimizing the displacement from precomputed MCSS positions. More generally, these restrictions represent a reduced description of the potential energy surface in the neighborhood of a functional group minimum. This is an aspect of DLD whose importance will grow as consideration is given to protein flexibility and to the effects of one minimum energy position on another.

Monte Carlo sampling

The phase space of DLD is sampled by choosing a random atom or rigid collection of atoms (e.g., an MCSS functional group) and applying a random displacement in its translation, rotation, and occupancy degrees of freedom; the step is accepted according to a Metropolis criterion. One Metropolis step for this system requires 3 energy evaluations. The first evaluation determines the contribution of a particular atom or group to the total system energy, the second determines the change in energy upon a small perturbation in position, and the third determines the change in system energy after a small perturbation in occupancy. Changes in the pseudo-energy due to the positional and occupancy perturbations are computed separately so that independent step sizes can be used for these degrees of freedom. For efficiency, the maximum step size is increased or diminished by 5% every 100 steps of Monte Carlo so that the acceptance ratio is approximately 0.5.[4]

For multiatom rigid groups, the rotation step size is coupled to the translation step size. This ensures that atomic displacement as a result of translation or rotation is of a similar magnitude regardless of the size of the rigid body being transformed. Translation steps are randomly selected in the range $\pm s_x$, and rotation steps in the range $\pm s_\theta$ about an axis.

The quantities s_x and s_θ are related by Eq. (12) where l is the distance from the center of mass of the rigid body to its most distant point. For freely rotating rigid bodies, the rotation axis is chosen at random and passes through the center of mass. In this work, the only rigid bodies are MCSS functional group minima whose rotation has been limited to a single axis, as described above. In this case, l is the distance from the axis of rotation to the atom of the group most distant from the axis. Use of Eq. (12) introduces an exact correspondence between translation step size and the maximum distance that the atom furthest from the rotation axis can move.

$$s_\theta = 2\sin^{-1}\left\{\min\left(1,\frac{s_x}{2l}\right)\right\} \tag{12}$$

Optimization

Optimization was performed with simulated annealing of the Metropolis algorithm. Formal properties of the simulated annealing algorithm have been determined for stochastic sampling methods.[23,24] One important property is that simulated annealing converges asymptotically to the global minimum as the number of iterations approaches infinity. A corollary of this is that the algorithm samples the equilibrium distribution of states throughout the annealing process. Since sampling is limited to a finite length of time, a succession of temperature decrements must be made that are large so as to minimize CPU usage, yet small enough so that a near equilibrium distribution of states is sampled at each temperature. At high temperatures in systems with finite size, near equilibrium sampling is quickly attained since the density of states approaches uniformity. A system of finite size was achieved in this work by using a reflective boundary. Also, since the number of available states diminishes with temperature, the number of steps required to achieve near-equilibrium sampling after a temperature decrement becomes smaller. The variance of the pseudo-energy at a given temperature was used as a crude measure of the number of available states at that temperature in the implementation of the cooling schedule [Eq. (13)].[23]

$$\Delta T = \alpha\frac{T^2}{\sigma(T)} \tag{13}$$

Here, α is a constant (see Annealing schedule), T is the DLD temperature, and $\sigma(T)$ is the variance of the pseudo-energy. Since DLD samples nonphysical configurations of molecules, no attempt was made to rationalize the temperature units to a physical scale (e.g., kcal/mol). A starting temperature 0.1–1.0 (in arbitrary units) was determined as high enough for most energy barriers to be crossed, but as low as possible to conserve CPU time. The global minimum can neither be guaranteed using this optimization method, nor is it necessary. Rather, a series of optimizations using different random number seeds are used to produce a variety of possible ligands with low DLD pseudo-energies.

RESULTS

A Simple Example: Propane

To illustrate the methodology, we consider the simplest example that contains all of the elements of the potential used in DLD. It consists of three sp^3 carbon atoms, each capable of bond formation and with dynamic occupancy. A complete molecule would be formed by adding hydrogen atoms as appropriate for sp^3 carbon atoms. It is unnecessary to specifically introduce these in the present procedure. For configurations in which all atoms are within 4 Å of each other, this system has 3 bond terms and 3 bond angle terms since each atom is considered a vertex of a bond angle. Several pertinent structures along with their energy contributions rounded to their limiting values (± 1) for simplicity are shown in Figure 3a. If only one optimal bond ($r = r_o$) is present and the two other distances are strained ($|r - r_o| >> \sigma_{bond}$), all three bond-angle terms contain at least one bond factor that is zero [Eq. (5)] and therefore make no contributions to the total energy (Fig. $3a_i$,$3a_{ii}$), e.g., $3a_{ii}$ has a favorable angle, but the pseudo-energy contribution is zero because one bond factor is zero. The two carbon atoms comprising the optimal bond in Figure $3a_i$ and $3a_{ii}$ have a favorable bond term (-1) and an unfavorable bond term ($+1$); therefore the occupancy gradient is zero and neither carbon is encouraged nor discouraged from existence. The third carbon generates two unfavorable bond terms ($+2$) and tends toward zero occupancy. Finally, there remains a two carbon system with a valid bond corresponding to an ethane molecule.

For a configuration with two optimum bonds but no favorable bond angles (Fig. $3a_{iii}$), the vertex carbon of the two optimal bonds has a positive occupancy gradient since it is associated with two favorable bonds (-2) and one unfavorable bond angle ($+1$). The remaining carbon atoms have negative occupancy gradients which will drive this configuration to a one body system (methane) with a pseudo-energy equal to zero.

Although it would be desirable, the present DLD energy function does not admit cyclopropane as an acceptable configuration (Fig. $3a_v$). The cyclopropane configuration has three favorable bonds and three strained bond angles measured relative to sp^3 carbon atoms. Since the occupancy of any given atom affects three bond angle terms ($+3$), but only two bond terms (-2), and since all atoms are equivalent, the entire system tends toward zero occupancy. Unusual geometries, such as cyclopropane or epoxide, could be introduced into DLD most easily

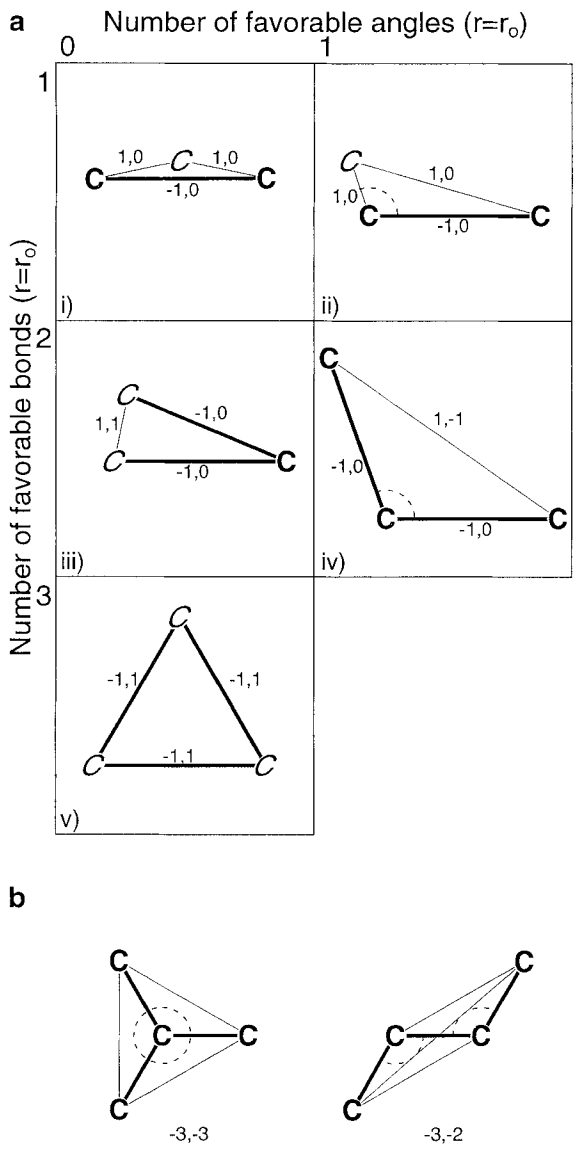


Fig. 3. Several simple states for three and four body systems of bond forming sp^3 carbons. Italicized carbon labels indicate that the occupancy gradient is negative, encouraging the atom to disappear. Optimal bond length (indicated by a thick line) is 1.5 Å, contact distance is 4 Å, and optimal bond angle (indicated by a dotted arc) is 109.5°. Deviations are 0.1 and 0.2 Å for bond and bond angles, respectively. For each pairwise interaction in **a**, the bond and bond angle contributions are equally weighted and are shown adjacent to the bond, e.g., (1,0) indicates that the bond term is 1 and the bond angle term is zero. In **b**, the total pseudo-energies are shown as a pair of values for each configuration. The first element of the pair is the total pseudo-energy given equal weighting of all terms of the potential, while the second element is the pseudo-energy with bond contributions scaled by 0.1. All pseudo-energies are given rounded to the nearest whole number.

by adding such molecules in addition to sp^3 carbon atoms as linking groups.

In the ideal configuration (Fig. $3a_{iv}$), two optimal bonds form an optimal bond angle of 109.5° at their vertex. The two other bond angle terms contain bond factors of zero and make no contribution to the pseudo-energy. This configuration contains two optimum bonds (−2), one unfavorable bond (+1) and one optimum bond angle (−1) giving a total pseudo-energy of −2 and positive occupancy gradients for all the atoms yielding a stable three body system (propane). This optimal configuration has two counterbalancing contributions, one bond angle (−1) and one bond (+1), for the carbon atoms forming the bond opposite the 109.5° angle (Fig. $3a_{iv}$). The effect of this is that the pseudo-energy is not weighted towards satisfying the valence of bond forming carbon atoms with nonhydrogen atoms. For example, 2-methylpropane has the same pseudo-energy as butane (Fig. 3b). Rigid structures, such as substituted rings, are desirable as inhibitors to minimize the conformational entropy loss on binding. Since these are more likely to have sp^3 carbon centers with 1 or 0 hydrogen atom substituents, the bond and nonbonded terms were weighted by 0.1. The features of the three body potential surface are unchanged, i.e., propane remains the only viable three atom configuration (Fig. $3a_{iv}$). The four body potential energy surface is, however, changed and favors the formation of structures with satisfied valence in preference to extended chains; in the example of Figure 3b, 2-methylpropane has a pseudo-energy of −3 while butane has a pseudo-energy of −2.

Annealing Schedule

To determine the appropriate annealing schedule, a test system consisting of a random distribution of 50 sp^3 carbon atoms each having dynamic occupancy (virtual carbon atoms) was placed in a cubic region of length 9.0 Å without any protein (Fig. 4a).

As the DLD potential function is not physically realistic, the selection of a starting temperature had to be determined de novo. In the presence of a harmonic boundary potential, which was tried first, increasingly higher temperatures simply allowed atoms to penetrate further into the boundary resulting in ever higher pseudo-energies. From calculations on our test system in which phase space is rendered finite by a reflective boundary, a clear phase transition can be seen (Fig. 5b). This indicates a nominal starting temperature in the range of 0.1–1.0.

Implementation of a cooling schedule according to Eq. (13) required empirical determination of the parameter, α. Annealing trials were performed on this test system using different values of α and a value of 0.25 was chosen as this represented a balance between achieving low energy minima and increasingly long optimization times. In simulations with similar CPU requirements, a linear cooling scheme showed final pseudo-energies which were at least 25% higher than this variance mediated cooling schedule with observed connectivity also considerably improved.

The most dramatic improvement to optimization

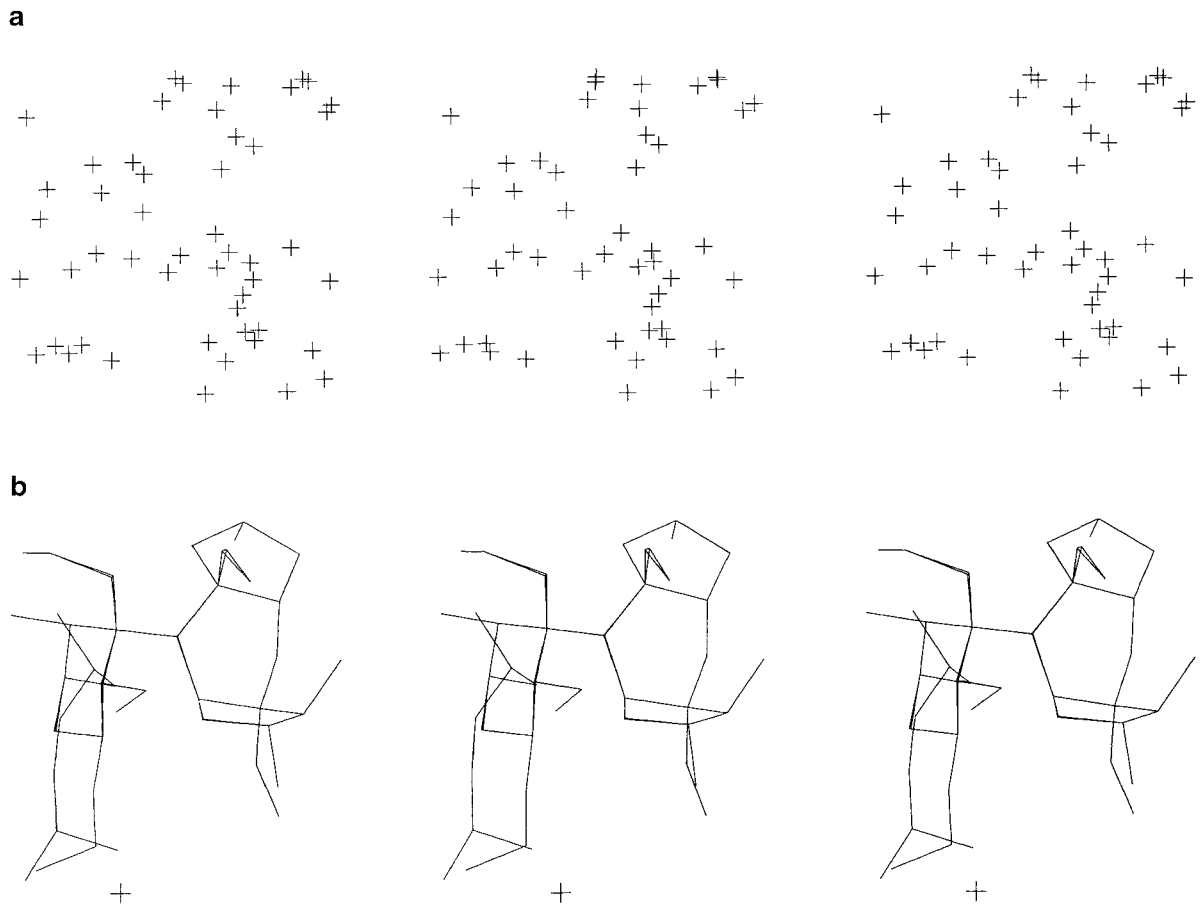


Fig. 4. Test system of 50 sp^3 carbon atoms before, **a**, and after, **b**, optimization. Carbon atoms having no bonds formed with another carbon atom are represented with a cross.

resulted from examination of the per atom contributions (Fig. 5a) to the total pseudo-energy (Fig. 5b). It can be seen that on annealing, the energy of most atoms make a transition from zero to a negative contribution to the total energy while the remaining atoms remain at zero. Furthermore, prior to the neighborhood of the transition temperature (~0.05), all atoms make a contribution of near zero to the total energy. This indicates that atoms can be partitioned into those that are forming structure and those that are not. Two independently adjustable step sizes for both occupancy and positional degrees of freedom were established, with the assignment of an atom to a particular step size chosen at every Monte Carlo iteration according to whether the sum of its bond, nonbonded, and bond–angle terms was above or below a threshold; based on Figure 5a, this value was set to −0.5. This gives a total of 4 dynamically adjusted step sizes, and resulted in a 2-fold decrease in the energies after annealing, as well as a considerable increase in the observed level of branched and ringed structures. Annealing in all subsequent work was performed using α equal to 0.25, 4 dynamically adjusted step sizes and bond and nonbonded pseudo-energy terms scaled by 0.1 (Fig. 5b).

Application to FKBP

The DLD method was tested on the site at which the immunosuppressant drug, FK506, binds to the immunophilin, FKBP. We first determined MCSS functional group minimum energy positions. These were then combined with freely diffusing sp^3 carbon atoms to generate new compounds that might bind to FKBP and to modify FK506 so as to introduce interactions with additional functional group sites.

Determination of MCSS minima

Methyl ammonium, acetate, acetonitrile, methanol, and *N*-methyl acetamide were used as functional groups and 183 MCSS minima were determined, as previously described,[1] in a 24 Å diameter sphere (Fig. 6b) positioned on the center of geometry of the FK506 coordinates of the cocrystal struc-

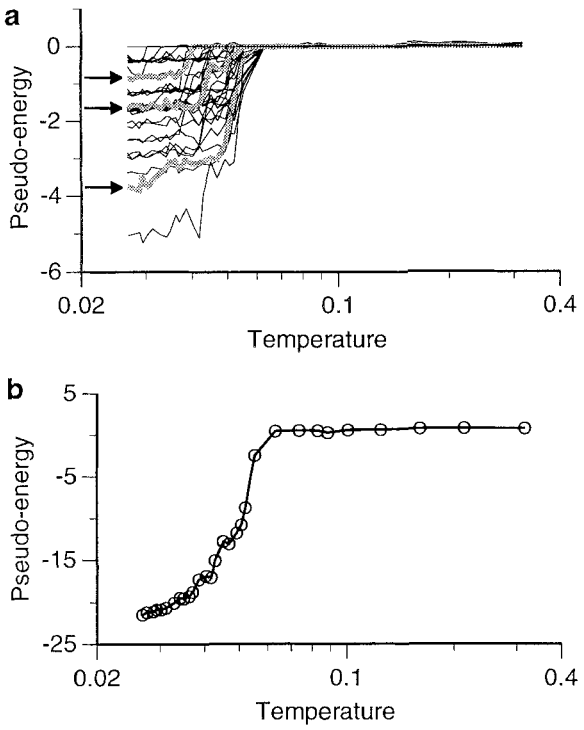


Fig. 5. The sum of terms in the pseudo-energy function associated with each of the 50 carbon atoms of the test system (Fig. 4a) is shown in **a**, while the total cost is shown in **b**. In **a**, 3 of the 50 groups are singled out with arrows and drawn with bold, gray lines to show the underlying annealing behavior.

ture.[28] Only polar and charged probes were examined since DLD uses freely diffusing carbon atoms to connect the functional groups. Although all 183 minima could have been used in DLD calculations, this would have been computationally intensive and unnecessary as many of the minima for a given group are part of a single cluster; clusters are defined as consisting of copies of the functional group with the same hydrogen bonds to the protein. To make an objective selection of a subset of the MCSS minima, two rules were applied. First, groups were selected which formed two or more polar contacts with solvent exposed groups of the protein, provided that these elements were separated by at least one dihedral degree of freedom. This criterion was used as a screen for minimum energy positions which are more likely to result from geometry specific interactions with the FKBP binding site than from nonspecific protein contacts. A positive example of this rule would be an acetate probe which accepts hydrogen bonds from the NH_3^+ of a lysine side chain and its main chain amide. A counter example, which would not be selected, is an acetate probe forming a double salt bridge with a lysine side chain. Second, only the lowest energy member of every cluster was selected. Exclusion of other members of a given cluster is compensated for, in part, by the translation and rotation allowed each functional group (see Functional group flexibility). The two criteria were applied manually using interactive molecular graphics in the present application.

The two criteria were applied to the 183 MCSS minima for FKBP. Furthermore, all minima outside a 20.6 Å box roughly centered on the geometric center of FK506 (Fig. 6b) were discarded. This resulted in a subset of 54 minimum energy positions. For the DLD studies presented here, MCSS minima were given a fixed occupancy of one while linking carbon atoms were given dynamic occupancy. This was done to ensure that resulting molecules retained electrostatic as well as steric complementarity with the binding site. A selection of a nonoverlapping subset of the 54 minima was therefore required. To do this objectively, a DLD optimization was run in the presence of FKBP, with the 54 MCSS minima assigned dynamic occupancies in the absence of linker carbon atoms. The minima remaining with occupancy greater than 1/2 formed a nonoverlapping set and were kept for subsequent DLD calculations. There were 7 methyl ammonium, 1 acetate, 2 acetonitrile, 8 methanol, and 5 *N*-methyl acetamide minima.

Overview of DLD application to FKBP

All DLD studies of FKBP reported here were performed in the same 20.6 Å cube centered near the center of geometry of FK506. The protein surface was represented by 182 protein atoms (as calculated by their accessibility to a 1.4 Å radius probe[22]) whose solvent exposed surface extends into the sampling region.

Two sets of results are presented. First, MCSS groups and randomly distributed linker sp^3 carbon atoms are annealed in the absence of FK506 to demonstrate de novo design. Second, MCSS groups and linker carbon atoms are annealed in the presence of FK506 to demonstrate the possibility of modifying a known ligand. The sampling and annealing parameters were adjusted empirically to provide low energy minima within several hours of CPU. At each temperature, the number of Monte Carlo steps taken was set to between 50 and 200 times the number of groups in the simulation. Annealing was performed with 45 temperature decrements, starting from 0.1 and ending at temperatures near ~0.01. On average, 0.15 CPU seconds (on a Silicon Graphics 4D/340) are taken per Monte Carlo step; a typical annealing run requires 200,000 steps, or about 10 h.

De novo design

The MCSS minima were given a fixed occupancy of one, and annealed in the presence of 50 linker carbon atoms with dynamic occupancy (virtual sp^3 carbon atoms) and the fixed solvent exposed atoms of the FKBP binding site. The CPU requirements for one annealing run of this system were between 2

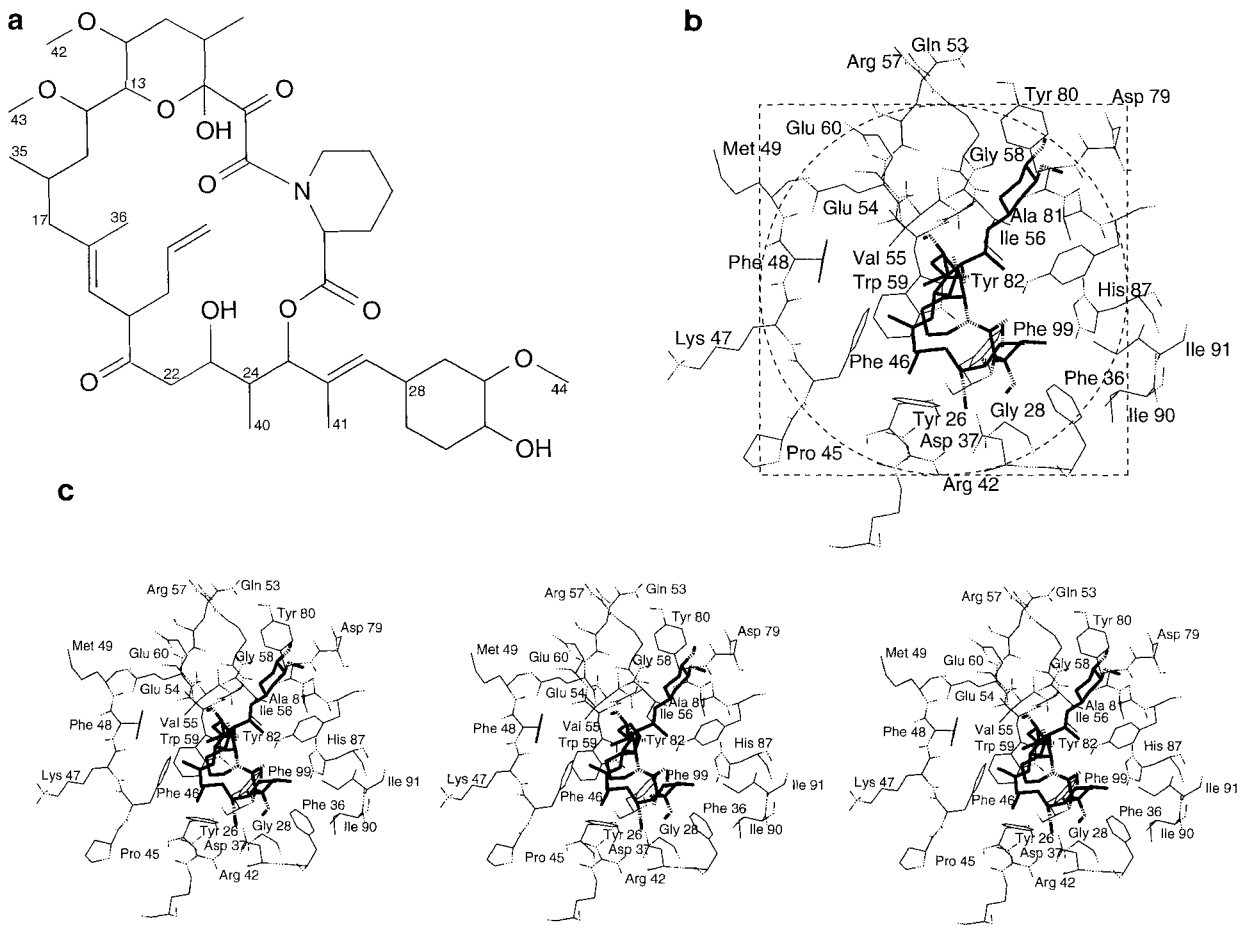


Fig. 6. Schematic of FK506, **a**, and FK506 in the binding pocket of FKBP, **b, c.** In **b**, the cubic boundary region was used for subset selection of MCSS minima and all subsequent DLD calculations, while the spherical region was used for the initial distribution of linker carbons. Atom numbers for FK506 shown in **a** are discussed in the text and are the same as found in the Brookhaven Protein Data Bank. In **b** and **c**, nitrogens and oxygens are shown as dotted lines.

and 5 h on one processor of a Silicon Graphics 4D/340. A total of 7 molecules linking 3 or more MCSS minima together were generated in 5 optimization runs, each run differing only in the starting random number seed. Four of the molecules are shown in Figure 7; the two molecules shown in Figure 7e,f were generated in the same annealing run.

The variety of configurations is striking. In Figure 7a,b, 8 functional groups are joined into a single molecule forming hydrogen bonds with polar side chains (Asp-37, Tyr-26, and Arg-42) below the hydrophobic pocket. Similar contacts are made by the 10 MCSS minima joined in Figure 7c,d and the 9 minima joined in the molecule on the right in Figure 7e,f. In addition, the molecule in Figure 7c,d makes three further contacts with the main and side chain of Lys-47. The right most molecule in Figure 7e,f forms additional contacts to His-87 on the right side of the pocket and extends an aliphatic chain through the binding pocket to a methanol group forming hydrogen bonds with the main chain of Ile-56 and the side chain of Tyr 82.

A substituted octane is present on the left side of Figures 7e,f and 8e. Cyano, methyl, hydroxyl, methyl, and ammonium substituents are at the 1, 2, 3, 6, and 7 positions, respectively, with average dynamically formed bond lengths of 1.50 ± 0.01 Å, angles of 109.8° ± 3.0°, and dihedrals at ±27° of staggered conformations. This chain is in a minor groove adjacent to the hydrophobic pocket of FKBP and makes a number of polar contacts with the site as a result of the restricted motion of the MCSS minima during DLD optimization. The acetonitrile

Fig. 7. MCSS minima annealed with fixed unit occupancy in the binding site of FKBP. Fifty virtual carbons with dynamic occupancy were present during the optimization. Annealing was performed with 90 steps/group/temperature decrement and 45 temperature decrements using a starting temperature of 0.1. Only molecules containing 3 or more MCSS functionalities and virtual carbons with occupancy greater than 0.5 are shown. Each set of figures (**a** + **b, c** + **d, e** + **f**) is from simulations differing only in starting random number seed and are shown in the presence (**b, d, f**) and absence (**a, c, e**) of the FKBP binding site. Nitrogens and oxygens are drawn as dotted lines.

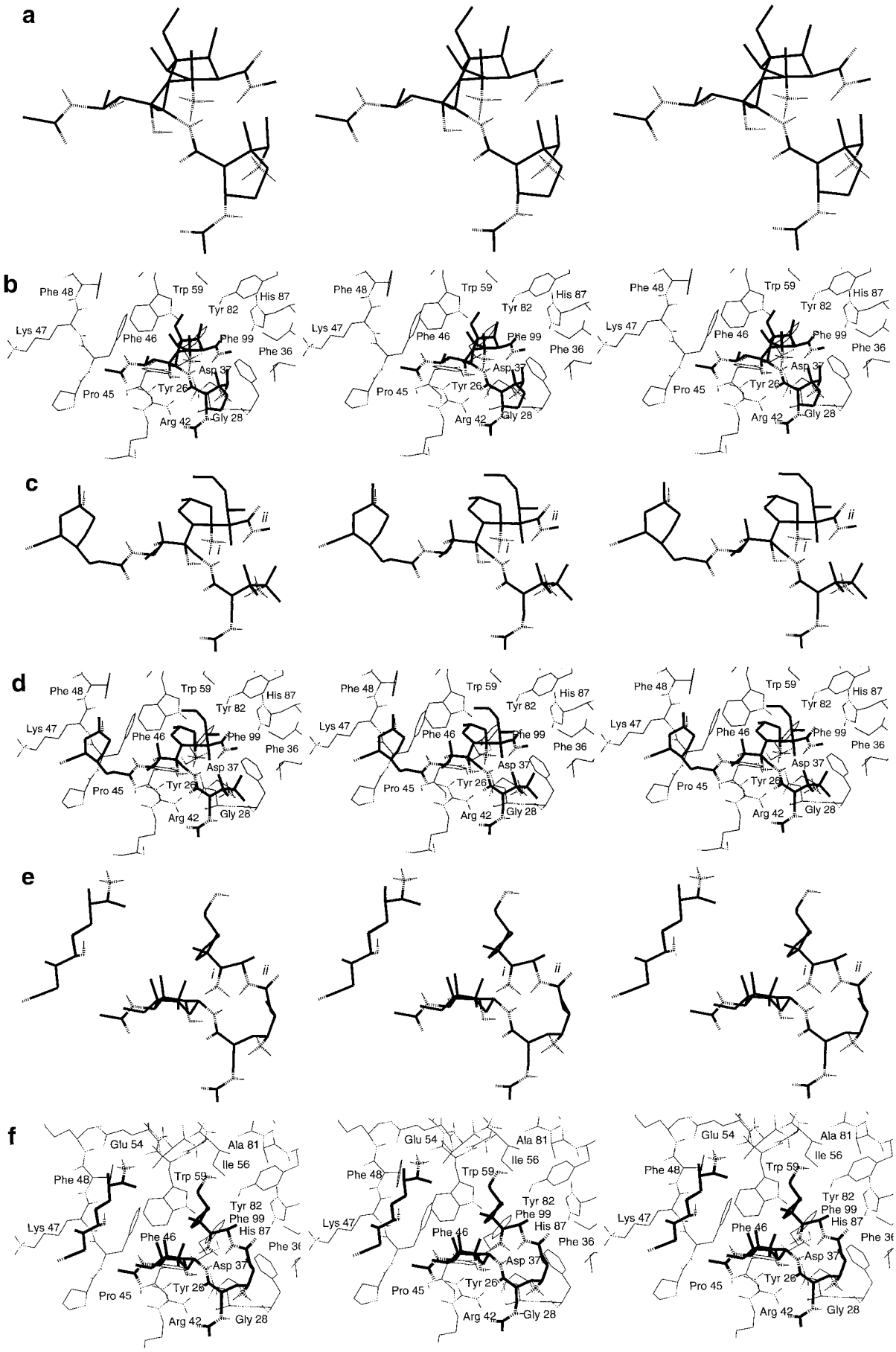

a
b
Phe 48
Trp 59
His 87
Tyr 82
Lys 47
Phe 46
Phe 99
Phe 36
Asp 37
Pro 45
Tyr 26
Arg 42
Gly 28
c
i
ii
d
e
f
Glu 54
Ala 81
Ile 56

Fig. 7.

group moved only 0.33 Å RMS from its minimum in contact with the side chain of Lys-47, a methanol moves 0.45 Å RMS but still donates and accepts hydrogen bonds with the main chain of Lys-47 and the methyl ammonium still donates to the side and main chain of Glu-54 after moving 0.30 Å RMS. Nonpolar contacts are made with the aliphatic portion of the Lys-47 side chain and to the side chain of Phe-46.

Each optimization results in alternative ways of connecting the same functional group minima. An example of this can be seen in the linking of an *N*-methyl acetamide probe (i) to a methyl ammonium group (ii) in Figure 7c,e. This methyl ammonium group donates two hydrogen bonds to the side chains of Tyr-26 and Asp-37, respectively, so that it has a single degree of rotational freedom about the line joining the donating hydrogen atoms. Since the *N*-methyl acetamide group only donates a hydrogen bond to the side chain of Asp-37, it is free to rotate about its NH bond during DLD optimization. In Figure 7c,d, the sp^3 carbon atoms of the *N*-methyl acetamide probe are 3.1 Å RMS distant from the same atoms in Figure 7e,f while still maintaining its contact with Asp-37. This results in the two observed linkages, the former attaching the methyl ammonium to the carbonyl carbon via a one carbon linker, and the latter attachment is directly to the *N*-methyl carbon of the *N*-methyl acetamide.

Modification of FK506

A simple way to explore the possibility of modification of FK506 is to overlay results of calculations done in the absence of FK506 with the coordinates of the FK506 molecule. We present three DLD molecules which are in regions of the FKBP binding site that are not occupied by FK506. We do not give precise suggestions for linking these fragments to FK506: the lack of atomic overlap suggests that attachment is plausible. The first example (Fig. 8a,b) would introduce polar contacts to the side and main chains of Lys-47, as well as accepting a hydrogen bond from the side chain of Arg-42. This conformation of the fragment is stabilized by an internal polar contact from its amide hydrogen to the oxygen of the hydroxyl substituent of the cyclohexane. The strained conformation of cyclohexane could be stabilized by forming [2,2,2]-bicyclooctane. It may be possible to attach this fragment to C35 or C17 of FK506 (Fig. 6a). To do this, the hydroxyl group of the fragment that donates a hydrogen bond to the side chain of Tyr-26 and accepts hydrogen bonds from the main chain of Arg-42 would have to be eliminated. A second fragment (Fig. 8c,d) is composed of *N*-methyl acetamide, methanol, and methyl ammonium groups. The amide accepts a hydrogen bond from the side chain of Arg-57 and donates to the main chain oxygen of Val-55. The hydroxyl group of methanol donates a hydrogen bond to the main chain of Ala-81 and accepts a hydrogen bond from the main chain of Arg-57. Lastly, the methyl ammonium donates to the main chain of Asp-79 and Ala-81. A total of six additional hydrogen bonds would be gained by linking this fragment to the cyclohexyl ring of FK506. The third example (Fig. 8e,f) is the substituted octane, already discussed (Fig. 7e,f), which makes several polar and nonpolar contacts with the binding pocket. The terminal carbon of this fragment is approximately two carbon–carbon bond lengths from C22, C40, and C36 of the FK506 molecule.

A more direct method for finding modifications to FK506 is to do simulated annealing runs with it in the binding site. The carbon atoms of the extended atom representation of FK506 from the cocrystal structure (Fig. 6a) were catalogued according to their hybridization state and intramolecular bonds. Forty-one out of 44 carbon atoms of FK506 have one or more attached hydrogen atoms that can be replaced by carbon. All of these carbon atoms were designated as bond forming regardless of their position in relation to the FKBP surface. In actual applications, only the carbon atoms close to the part of the FKBP surface that are not involved in the complex formation of FK506, FKBP would be designated as bond forming. Also, no consideration was given to the ease of introducing substituents at the selected positions. Since some of the MCSS minima overlap FK506 in the bound crystal structure, a subset of nonoverlapping minima was prepared. Annealing studies were performed with the 60 FK506 atoms in the presence of 6 methyl ammonium, 1 acetate, 2 acetonitrile, 7 methanol, 3 *N*-methyl acetamide minima, 50 virtual sp^3, and 182 fixed protein atoms representing the binding site, as before. Five runs were made with different starting random number seeds. The modifications of FK506 that resulted included one or two MCSS minima. As can be seen in Figure 9a,b, a single *N*-methyl acetamide group that accepts hydrogen bonds from the Arg-42 side chain is linked directly to C35 of FK506. Additional modifications could be generated by allowing the atoms of FK506 to move according to the internal degrees of freedom of FK506. The latter was not done in this work, but would be simple to implement, especially for the ether groups terminated by C42, C43, and C44.

A number of internal linkages were formed that would make FK506 more rigid in its bound conformation, e.g., three virtual carbon atoms join C41, C42, and C24. Also, two carbon atoms join C41 to C28 to form a structure which fixes the relative positions of the cyclohexyl and hexose rings of FK506. This is a very useful property of DLD as a more rigid ligand will have less entropy to lose on binding. Indeed, the flexibility of the immunosuppressant cyclosporin A (CsA) and the observation that the bound and unbound conformations are different

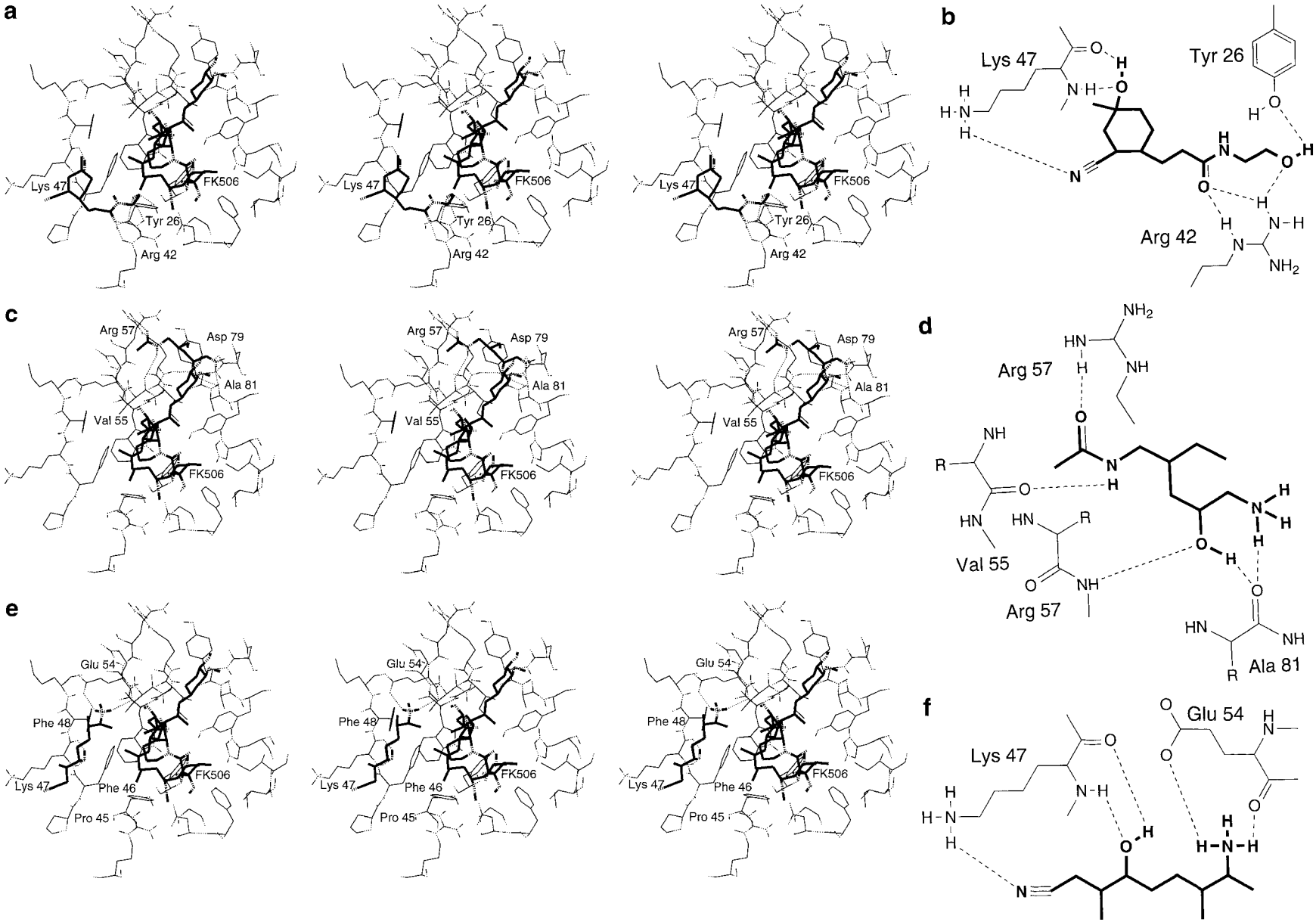


Fig. 8. Overlay of FK506 with nonoverlapping fragments from three separate DLD optimizations of MCSS functionalities in the presence of linker carbons (**a, c, e**). In the corresponding schematic diagrams (**b, d, f**), hydrogen bonding contacts of the DLD fragments with the protein are shown.

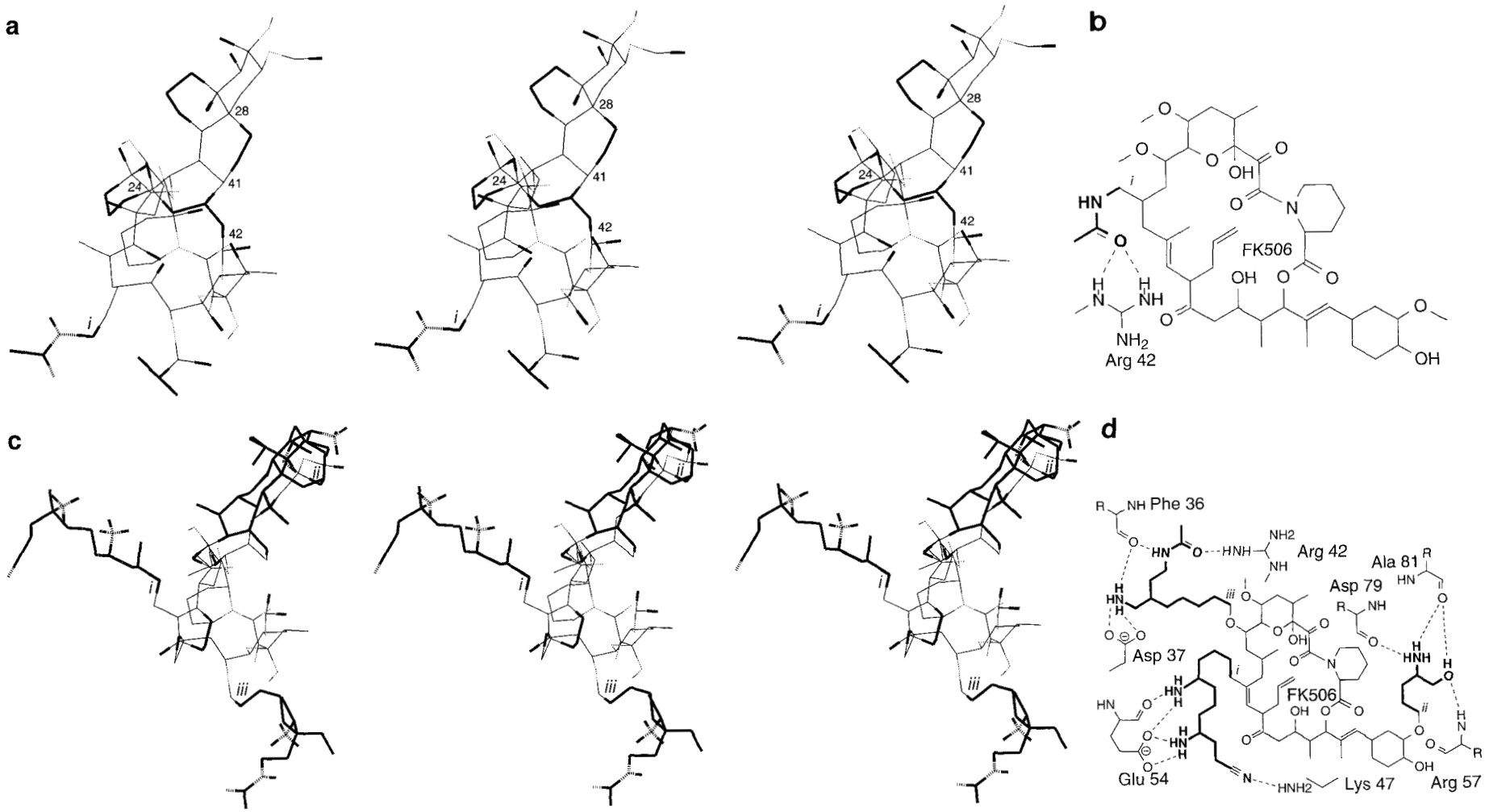


Fig. 9. Two examples DLD optimization of MCSS minima in the presence of FK506. MCSS minima had a fixed unit occupancy. FK506 was fixed in space with 41 FK506 carbon atoms permitted to form bonds. Fifty (**a, b**) and 100 (**c, d**) linker carbons were included in the simulation. Annealing steps per group per temperature decrement were kept constant for both optimizations and only MCSS minima and virtual carbons linked to FK506 are drawn. Stereo triplets are representations in the bound conformation (**a, c**) while schematic diagrams show resultant suggestion of linkages of MCSS functional groups to FK506 (**b, d**). Thick lines are drawn for linker carbon atoms and MCSS functional groups while thin lines represent FK506. Roman numerals relate the attachment sites of these linkages shown in the schematics to their positions in their respective triplets. Arabic numbers in **a** correspond to those in Figure 6c and are discussed in the text.

prompted the design a tricyclic variant of CsA. The designed molecule was synthesized and found to bind cyclophilin with three times greater affinity.[29]

A problem with this simulation is that in the presence of the 41 bond forming carbon atoms of FK506, relatively few virtual carbon atoms are available to attach functional groups. In Figure 9a for example, 26 of the virtual carbon atoms have attached to 23 separate FK506 carbon atoms. This reduces by half the number of virtual carbon atoms which can form links to MCSS minima in the binding site.

The number of MCSS groups attached to FK506, as well as the number of internal crosslinks, was increased by using more virtual carbon atoms. In a simulation with 100 virtual carbon atoms, fragments containing two or more MCSS minima are attached to the C36(i), C44(ii), and C43(iii) carbon atoms of FK506 (Fig. 9c,d). In particular, the fragment attached to C36 uses 7 virtual carbon atoms to link FK506 to two methyl ammonium minima, which donate hydrogen bonds to the main and side chain of Glu-54, and to an acetonitrile that makes a polar contact with Lys-47. The fragment attached to C43 includes an amide which accepts a hydrogen bond from the side chain of Arg-42 and donates to the main chain carbonyl of Phe-36. It also contains an ammonium group which donates two hydrogen bonds to the side chain of Asp-37 and donates to the main chain carbonyl of Phe-36. Finally, the fragment attached to C44 contains a methanol group which donates a hydrogen bond to the main chain of Ala-81 and accepts a hydrogen bond from the main chain of Arg-57. Its methyl ammonium group donates to the main chain of Asp-79 and Ala-81. The results are clearly improved through the use of increased numbers of virtual carbon atoms. However, as a result of bond angle terms, the CPU time scales with the cube of the number of particles. For this system in which 125 virtual carbon atoms were used (not shown) several days of CPU were required for optimization.

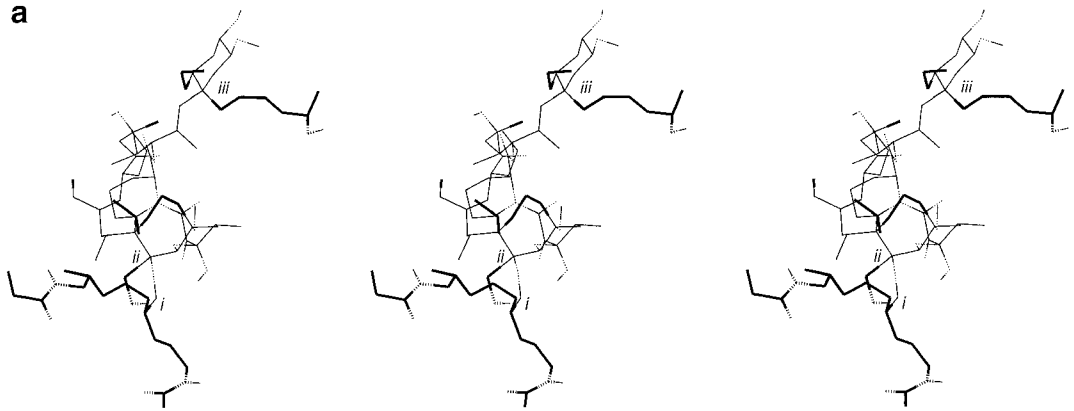
As an alternative to the approach with more virtual carbon atoms, a series of optimizations were run in which different sets of 10 carbon atoms of FK506 were flagged as bond forming and annealing was performed in the presence of the same MCSS functionalities and only 50 virtual carbon atoms. While we flagged random sets of carbon atoms, a modeler could use a particular collection of carbon atoms based on the ligand structure and synthetic ease. Nine optimizations were performed; the most promising modifications to FK506 are shown in Figure 10. Some of these modifications to FK506 are similar to those obtained when all of the FK506 carbon atoms could form new bonds. For example, an amide accepts a hydrogen bond from Arg-42 and is linked by 3 linker carbon atoms to C43 (*i* in Fig. 10a,b); this amide and C43 of FK506 were also connected in Figure 9c,d(*iii*) by a 6 carbon linker. The identical attachment of ammonium and hydroxyl groups to the ether substituent of the cyclohexyl ring of FK506 can be seen in Figure 10e,f(*ii*) as well as Figure 9c,d(*ii*). There are also a number of modifications which differ from Figure 9. These include a four carbon link from the C28 to a methanol group which accepts a hydrogen bond from His-87 (*iii* in Fig. 10a,b). A second fragment attaches an ammonium group and an amide to C13 of FK506 (*i* in Fig. 10c,d). The ammonium group donates three hydrogen bonds to the side chain and main chain of Asp-37 while the amide accepts a hydrogen bond from Arg-42. Finally, we show the attachment of an amide, a nitrile, and a hydroxyl group as a single fragment to C35 (*i* in Fig. 10e,f). The amide accepts hydrogen bonds from the side chain of Arg-42, the nitrile makes a polar contact with Lys-47, and the hydroxyl donates and accepts hydrogen bonds from the main chain of Lys-47.

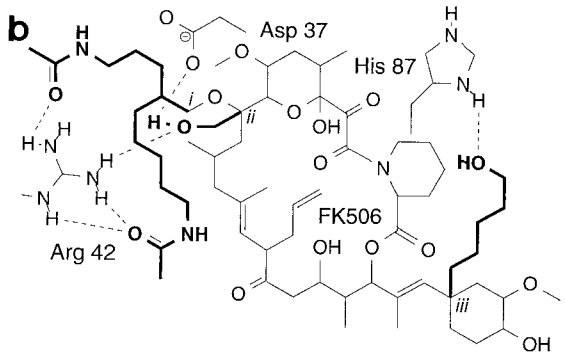
DISCUSSION

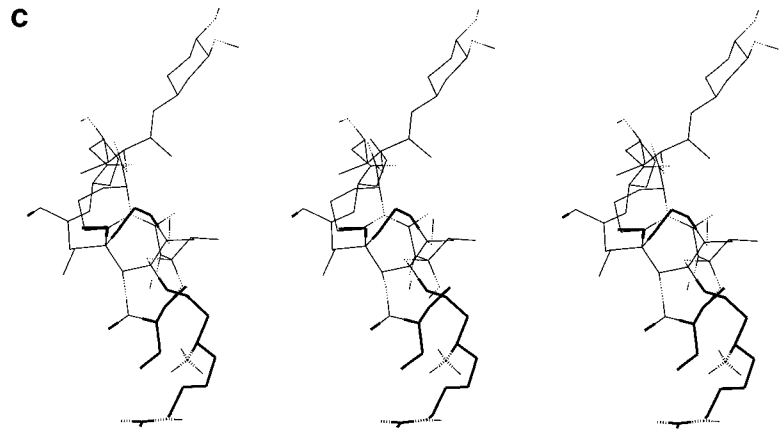
Dynamic ligand design is a novel method for the automatic generation of molecules which are complementary to protein surfaces. In the illustrations that focus on the FKBP binding pocket for FK506, we have demonstrated that DLD can form new compounds from precomputed MCSS minima, and can attach MCSS functional groups to FK506. The modeled compounds are complimentary to the binding pocket as a result of the polar contacts supplied by MCSS and the steric interaction provided by the DLD pseudo-potential function. The examples demonstrate the ability of DLD to form cyclic and rigid structures which reduce the loss of conformational entropy of a ligand upon binding to its target. In the current implementation of DLD, the best way to proceed in designing ligands is to first determine the MCSS minima for the binding site of interest. Since many minima are usually found, a subset of these must be selected. We kept only minima with different hydrogen bonding geometries and selected from these a set of nonoverlapping minima. The minima are given fixed occupancy and allowed to combine with linker carbon atoms, which have dynamic occupancy. The entire system is annealed to optimize the resulting molecules.

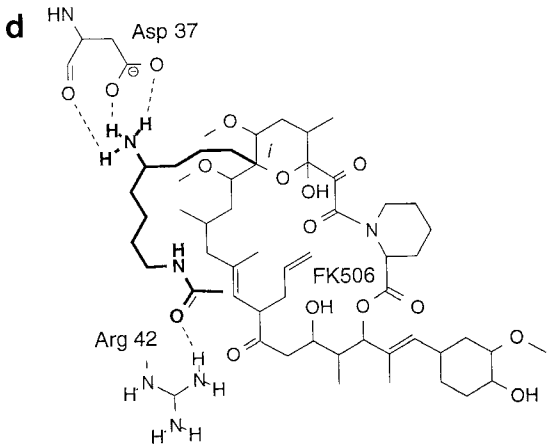
Future implementations of DLD require that we address its shortcomings as well as expand its capabilities. The principal shortcoming of DLD is the extent to which the molecular constructs have strained geometries. While it is trivial to postprocess the results to eliminate such molecules, their existence indicates that computational resources are being wasted. An approach that optimizes dihedral angles, as well as bond lengths and bond angles, has been implemented.[30]

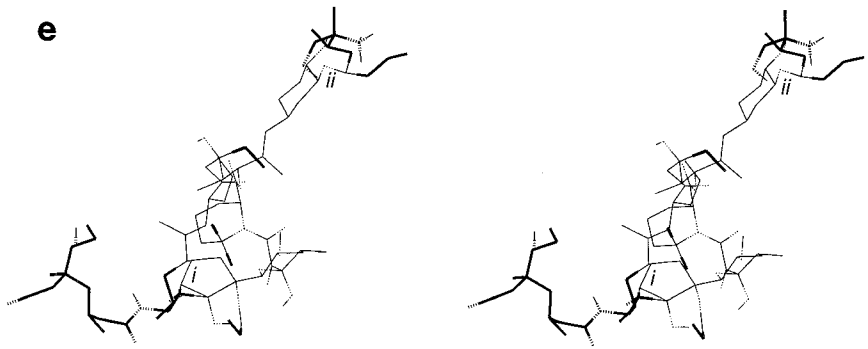
Additional capabilities can be added to DLD to increase the variety of molecules that can be constructed. This includes the addition of dynamic atom

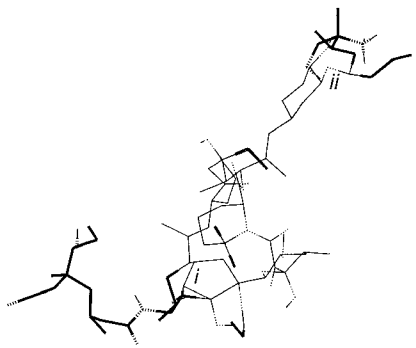

a
iii
ii
i

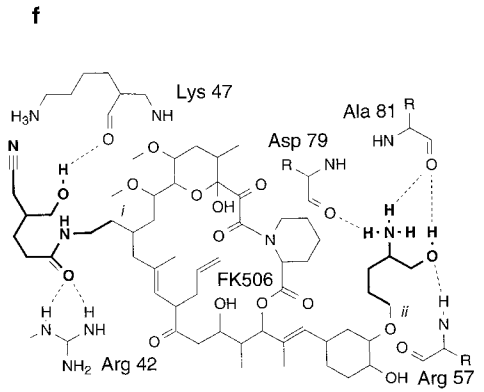

b
Asp 37
His 87
FK506
Arg 42


c
i


d
Asp 37
FK506
Arg 42


e
ii
i


ii
i


f
Lys 47
Ala 81
Asp 79
FK506
Arg 42
Arg 57

type which will allow, for example, carbon atoms to smoothly change their hybridization states from sp^3 to sp^2. Also, existing techniques for small molecule database searching can be introduced to connect regions which cannot be connected by sp^2 and sp^3 hybridized carbon atoms.[3] Elements of the database could include epoxide, cyclopentane, alkynes, etc. Such small molecule fragments could be smoothly introduced into the ongoing DLD optimization by giving them an initial occupancy of zero.[30] An attractive term could be added to the protein–ligand interaction potential. This would bias the location of structure formation to be on the protein surface so as to yield better shape complementarity. Other drug design methodologies which have incorporated attractive terms in the protein ligand potential have displayed this attribute.[3] Because the primary interactions of the molecules being constructed in DLD to connect the MCSS groups are hydrophobic in character, it is expected that solvent effects will scale approximately with the van der Waals term.[31]

DLD will also benefit from more diverse types of MCSS functional groups and from functional group positions that come from sources other than MCSS, e.g., groups from compounds that interact strongly with a binding site with no known three-dimensional structure can be used to generate pharmacophore positions for functional groups.[30] In the inhibition of macromolecular complexes of known structure, it would be possible to use the functional groups forming one side of the macromolecular interface. Drugs designed with this information could interfere with protein–protein, protein–nucleic acid, and dimer formation, for example. The interfaces within a protein due to its tertiary structure (e.g., subdomains of lysozyme) could be used to generate ligands which yield insights into protein folding. The linking groups used in DLD could also evolve to include rigid aliphatic compounds such as benzene and cyclopropane. Such compounds can be included as freely diffusing linkers, or as an MCSS minimum energy positions for regions of the binding site which contain few polar groups. Lastly, as MCSS functionality maps on a flexible protein are developed,[30] DLD will be able to include these representations, as well as any rigid body motions of the protein necessary to more readily accommodate the designed ligands (e.g., Asn and Gln side chain rotations).

This first implementation of DLD includes only a reduced set of features necessary to effectively sample the space of molecules which will complement a protein surface. The DLD approach clearly has much promise since even in the present implementation, the interesting suggestions for candidate ligands have been generated. Introduction of some of the possible improvements will lead to an even more useful ligand design methodology.

Fig. 10. DLD optimization of MCSS functionalities in the presence of FK506 and 50 virtual carbons. MCSS minima had a fixed unit occupancy. FK506 was fixed in space with 10 random carbons permitted to form bonds. A different set of FK506 carbons was randomly designated as bond forming for each of three optimizations (**a** + **b, c** + **d, e** + **f**). Linker carbon atoms and MCSS minima which were bonded to FK506 after optimization are shown (**a, c, e**). Several elements with strained conformation were deleted from **a.** Stereo triplets are representations in the bound conformation (**a, c, e**) while schematic diagrams show resultant suggestion of linkages of MCSS functional groups to FK506 (**b, d, f**). Roman numerals relate the attachment sites of these linkages shown in the schematics to their positions in their respective triplets. Thick lines are drawn for linker carbon atoms and MCSS functional groups while thin lines represent FK506 and FKBP.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the National Institutes of Health. We thank Dr. L. Caves, Dr. S. Fischer, and Dr. R. Stote for helpful discussions during program development. We thank Dr. S. Michnick and Dr. T. Wandless for discussions on the FKBP results. We thank C. Stultz for comments on the manuscript. A version of CHARMM developed at Harvard by Dr. R. Elber for use with multiple copies of a molecule was employed in the MCSS calculations; CHARNM 23 and all subsequent versions of the program contain a REPLICA code, developed by L. Laves for use with MCSS scripts.

REFERENCES

1. Miranker, A., Karplus, M. Functionality maps of binding sites: A multiple copy simultaneous search method. Proteins 11:29–34, 1991.
2. Caflisch, A., Miranker, A., Karplus, M. Multiple copy simultaneous search and construction of ligands in binding sites: Application to inhibitors of HIV-1 aspartic proteinase. J. Med. Chem. 36:2142–2167, 1993.
3. Eisen, M.B., Wiley, D.C., Karplus, M., Hubbard, R.E. Hook—a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding-site. Proteins 19:199–221, 1994.
4. Allen, M.P., Tildesley, D.J. "Computer Simulations of Liquids." Oxford: Clarendon Press, 1987.
5. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. "Numerical Recipes: The Art of Scientific Computing." Cambridge: Cambridge University Press, 1987.
6. Davis, L. "Handbook of Genetic Algorithms." New York: Van Nostrand Reinhold, 1991.
7. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P. Optimization by simulated annealing. Science 220:671–680, 1983.
8. Purisima, E.O., Scheraga, H.A. An approach to the multiple-minima problem in protein folding by relaxing dimensionality—tests on enkephalin. J. Mol. Biol. 196:697–709, 1987.
9. Desjarlais, R.L., Seibel, G.L., Kuntz, I.D., Furth, P.S., Alvarez, J.C., Demontellano, P., Decamp, D.L., Babe, L.M., Craik, C.S. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus-1 protease. Proc. Natl. Acad. Sci. U.S.A. 87:6644–6648, 1990.
10. Chau, P.L., Dean, P.M. Automated site-directed drug design—the generation of a basic set of fragments to be used for automated structure assembly. J. Comput.-Aided Mol. Design 6:385–396, 1992.
11. Lauri, G., Bartlett, P.A. Caveat—a program to facilitate the design of organic-molecules. J. Comput.-Aided Mol. Design 8:51–66, 1994.
12. Lawrence, M.C., Davis, P.C. Clix—a search algorithm for finding novel ligands capable of binding-proteins of known 3-dimensional structure. Proteins 12:31–41, 1992.
13. Böhm, H.-J. The computer-program ludi—a new method

for the denovo design of enzyme-inhibitors. J. Comput.-Aided Mol. Design 6:61–78, 1992.
14. Ho, C., Marshall, G.R. Splice—a program to assemble partial query solutions from 3-dimensional database searches into novel ligands. J. Comput.-Aided Mol. Design 7:623–647, 1993.
15. Miller, M.D., Kearsley, S.K., Underwood, D.J., Sheridan, R.P. Flog—a system to select quasi-flexible ligands complementary to a receptor of known 3-dimensional structure. J. Comput.-Aided Mol. Design 8:153–174, 1994.
16. Moon, J.B., Howe, W.J. Computer design of bioactive molecules—a method for receptor-based denovo ligand design. Proteins 11:314–328, 1991.
17. Gillet, V., Johnson, A.P., Mata, P., Sike, S., Williams, P. Sprout—a program for structure generation. J. Comput.-Aided Mol. Design 7:127–153, 1993.
18. Leach, A.R., Kilvington, S.R. Automated molecular design: A new fragment-joining algorithm. J. Comput.-Aided Mol. Design 8:283–298, 1994.
19. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comp. Chem. 4:187–217, 1983.
20. Crippen, G.M. Voronoi binding site models. J. Comp. Chem. 8:943–955, 1987.
21. Shoichet, B.K., Kuntz, I.D. Protein docking and complementarity. J. Mol. Biol. 221:327–346, 1991.
22. Connolly, M.L. Analytical molecular surface calculation. J. Appl. Cryst. 16:548–558, 1983.
23. Otten, R.H.J.M., van Ginneken, L.P.P.P. "The Annealing Algorithm." Boston: Kluwer Academic Publishers, 1989.
24. Aarts, E., Korst, J. "Simulated annealing and Boltzmann machines: A stochastic approach to combinatorial optimization and neural computing." Chichester: John Wiley, 1989.
25. Marsaglia, G. Choosing a point from the surface of a sphere. Ann. Math. Stat. 43:645–646, 1972.
26. Brooks, C.L.I., Karplus, M., Pettitt, B.M. "Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics." Chichester: John Wiley, 1988.
27. Boresch, S., Archontis, G., Karplus, M. Free-energy simulations—the meaning of the individual contributions from a component analysis. Proteins 20:25–33, 1994.
28. Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., Clardy, J. Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. Science 252:839–842, 1991.
29. Alberg, D.G., Schreiber, S.L. structure-based design of a cyclophilin-calcineurin bridging ligand. Science 262:248–250, 1993.
30. Stultz, C., Karplus, M. Computational combinatorial chemistry for de novo ligand design: review. (submitted).
31. Caflisch, A., Karplus, M. Perspectives in drug discovery and design. (in press).