Annu. Rev. Biophys. Biomol. Struct. 1995. 24:677–700

FLEXIBLE DOCKING AND DESIGN

R. Rosenfeld, S. Vajda, and C. DeLisi

Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

KEY WORDS: automated molecular docking, computer-aided drug design, empirical free-energy evaluation, peptide antigen–MHC interaction, protease inhibitor design

CONTENTS

ABSTRACT

Docking and design are the major computational steps toward understanding and affecting receptor-ligand interactions. The flexibility of many ligands makes these calculations difficult and requires the development and use of special methods. The need for such tools is illustrated by two examples: the design of protease inhibitors and the analysis and design of peptide antigens binding to specific MHC receptors. We review the computational concepts that have been extended from rigid-body to flexible docking, as well as the following important strategies for flexible docking and design: (*a*) Monte Carlo/molecular dynam-

1056-8700/95/0610-0677$05.00

ics docking, (*b*) in-site combinatorial search, (*c*) ligand build-up, and (*d*) site mapping and fragment assembly. The use of empirical free energy as a target function is discussed. Due to the rapid development of the methodology, most new methods have been tested on only a limited number of applications and are likely to improve results obtained by more traditional computational or graphic tools.

INTRODUCTION

A critical factor in virtually all biological processes is the specificity of ligands for larger proteins, such as membrane-bound receptors, or enzymes. Ligands may be flexible; for example, neurotransmitters, inhibitors, and cofactors are often small flexible organic molecules. Many hormones and antigens are small or intermediate-size polypeptides. A predictive understanding of the molecular basis of recognition, which would encompass the ability to design ligands that bind specific receptors and to predict the change in stability caused by site-specific substitutions, is central to developing rational drug- and vaccine-design strategies as well as to elucidating normal and pathological cellular behavior.

Here, we analyze aspects of computational approaches to the docking of flexible molecules to proteins, as well as to the design of oligopeptides and other flexible ligands that would bind with high affinity to a specified receptor site. Design means either determining the chemical composition of a ligand with desired binding properties (e.g. the amino acid sequence of an oligopeptide) or, in a somewhat more restrictive sense, ranking a given set of ligands according to their binding affinities. Docking means determining the geometry of a receptor-ligand complex using the structure of the free receptor.

In general, accurate approaches to docking and design require progress on two major problems: finding a target function (in the most general case, a computationally viable free energy–evaluation model) that can accurately weight any specified conformation of the system and developing an algorithm that efficiently searches a complex energy landscape for the target function's minimum value.

In spite of the complexity of these problems, several experimental systems allow simplifying approximations in the computations. In the simplest systems, the geometries of the reactants remain, to a good first approximation, unchanged by complex formation (2, 16, 34, 40, 51, 52). In such rigid-body systems, an algorithm need only search the six-dimensional rotational/transitional space and can generally make

effective predictions with simplified target functions. Because execution time is relatively rapid, these methods can be used to search large structural databases for small molecules that fit the binding site and that can serve as lead compounds for drug design (1, 39, 63, 71).

The rigid-body approach has limited uses, except in the search for small compounds. In almost all test cases, docking accuracy is much greater when one uses the geometries that the receptor and ligand have in the complex than when one uses the geometries of the uncomplexed molecules (70). The difference in accuracy implies that the assumption of rigidity is not fully valid even for protein-protein complexes in which the observed structural changes between the bound and free forms are small (77). In addition, simple scoring functions, such as measures of surface complementarity (34, 84), surface-area burial (85), solvation free energy, electrostatic-interaction energy, or the total molecular-mechanics energy (70), often fail to distinguish between near-native structures and others far from native. Hence, simple methods, though useful, are incomplete.

In the past few years, several groups have tried to improve docking procedures by allowing for flexibility. Introducing local minimization of a molecular-mechanics energy function such as CHARMm (7) yields only limited improvement. A general approach to flexible docking requires an unbiased sampling of the allowed conformational space of the ligand and a concurrent exploration of its six rigid-body degrees of freedom in the anisotropic environment of the receptor. This goal is difficult to achieve using traditional approaches, except perhaps for very small ligands (68). We review several flexible-docking and -design strategies for reducing computational efforts by fully exploiting the available structural and chemical information. Although most of the methods reviewed apply to any flexible ligand, we concentrate on peptides and peptide-like compounds that play a major role in many biological systems (49).

The success of docking depends heavily on the choice of a target function that accounts for the enthalpic, entropic, and solvation interactions that govern ligand-receptor association. The scoring is critical in design, in which the entropy loss of a flexible ligand, and the change in its internal energy upon binding, can greatly affect the binding affinity. In principle, free-energy perturbation (FEP) or thermodynamic-integration (TI) methods should permit accurate determination of the effects of mutagenesis on a receptor-ligand complex (35, 37, 54). In practice, however, such relatively rigorous procedures are computationally intensive and are not likely to be useful in docking or design

for at least another five to ten years. The alternative approach to determining free energy is to use empirical functions (31, 58, 60), several of which are reviewed here.

PROBLEMS REQUIRING THE USE OF FLEXIBLE METHODS

Design of Protease Inhibitors

OVERVIEW Protein cleavage is a ubiquitous aspect of biological processes: for example, the digestion of food, the control of blood pressure, and the activation of many proteins from precursors all require cutting, often at specific locations, of covalent bonds. The enzymes that carry out these diverse cleavage processes are collectively referred to as proteases. Errors in the production or regulation of any number of these enzymes cause serious pathologies. Protease inhibitors can be potential drugs (such as the HIV aspartic protease inhibitor), or investigative tools to elucidate the function of specific proteases.

INHIBITOR DESIGN The design of high-affinity inhibitors for many proteases has been accomplished already because the enzymatic mechanisms of these families of enzymes are well understood and many derivatives of natural inhibitors have proven to be potent inhibitors. For serine proteases, the existence of a structural binding motif (Figure 1), found in all the small inhibitors of this family (5), may simplify the search for possible inhibitor conformations. The major obstacle to utilizing these inhibitors for therapeutic purposes is specificity. The major-

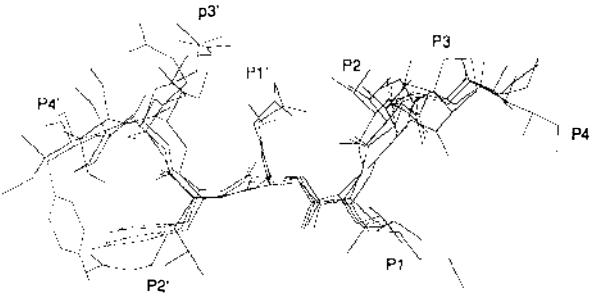

Figure 1 Superposition of four different inhibitors bound to different proteases: bovine pancreatic trypsin inhibitor (BPTI) complexed with kallikrein, chymotripsin inhibitor 2 complexed with subtilisin, eglin C complexed with thermitase, turkey ovomucoid inhibitor complexed with proteinase B from *Streptomyces griseus*. P4′ through P4 denote inhibitor side chains at given positions of the specificity pocket.

ity of known protease inhibitors are not specific for a single protease but rather inhibit, to varying levels, several proteases from a given family.

Because the naturally occurring protease inhibitors are proteins, one obvious approach to inhibitor design is to modify a segment of the inhibitor that is implicated in binding. Peptide inhibitors can occasionally be used directly for therapeutic purposes—for example, topical application of peptide-based inhibitors of proteases blocks parasite penetration through the skin (12). However, they more often serve as first-generation compounds that can be transformed into peptidomimetic inhibitors resistant to proteolysis.

Peptide Binding to Major Histocompatibility Complex Receptors

OVERVIEW The ability to discriminate foreign from self, the major role of the immune system, is regulated mainly by the major histocompatibility complex (MHC) receptors. These comprise a highly polymorphic set of glycoproteins that bind and display fragments of intracellular proteins on the cell surface, thereby informing other components of the immune system of the protein composition within the cell. The two major classes of MHC receptors, I and II, have highly similar structures. In both, the antigenic binding site—a cleft approximately 20 Å long, 8 Å deep, and 12 Å wide—is formed by a β-sheet floor bounded by α-helical walls. Variations in the amino acids that line the cleft modulate both its shape and chemical nature, thus determining haplotype specificity.

Class I molecules are found on all nucleated cells, displaying 8- to 10-residue-long fragments of cytosolic proteins. In infected cells, a certain fraction of these fragments will have originated in proteins that are not native to the host, thus presenting a foreign molecular arrangement recognized by T-cell receptors (TCR) on $CD8^+$ cytotoxic T lymphocytes. This recognition stimulates, among other things, the release of cytolytic agents that mediate lysis of the infected cell. Class II products are generally confined to cells of the immune system, such as B cells, macrophages, and other dendritic cells. The pathway leading to the presentation of class II receptors on B cells consists of protein recognition by immunoglobulin receptors, followed by endocytosis of the antigen-receptor complex, proteolysis of the antigen, and recycling of 13- to 15-residue-long fragments complexed with class II molecules to the surface. These peptide-MHC complexes, when recognized by the TCR of $CD4^+$ helper T cells, stimulate production of the various lymphokines that regulate a humoral (antibody) immune response.

In contrast to its numerous immunoglobulin receptors and TCRs, the immune system has few MHC haplotypes. To elicit an immune response against the many possible pathogens, each haplotype must be able to bind many different peptides.

VACCINE DESIGN The central role of MHC-peptide complexes in communicating the health of the cell to TCRs suggests that a predictive understanding of their structure could contribute substantially to progress in applied, as well as in basic, immunological research. Of particular therapeutic and prophylactic importance is the ability to design peptides with desired MHC and TCR specificities. One of the central complications in designing a peptide vaccine is MHC gene polymorphism. Different individuals, even when infected with an identical virus, will generally present different peptide fragments of the antigen. Any attempt to design a peptide vaccine must account for the various MHC haplotypes in the population of interest.

Several studies on HIV have reported a correlation between specific MHC class I alleles and either fast (33) or slow (76) progression of AIDS, which suggests an important role for a cell-mediated immune response. The induction of a strong antibody response may weaken a cell-mediated response caused by the effects of cross-regulatory cytokines, and therefore an effective vaccine strategy may need to specifically target cell-mediated immunity without eliciting neutralizing antibodies.

SEQUENCE AND STRUCTURAL MOTIFS In peptides that bind a particular MHC haplotype, one or two positions (so-called anchors) tend to be occupied by one or a few specific residues, whereas the other positions tend to permit a relatively high degree of substitution (22, 23). However, the anchor residues are neither sufficient nor always necessary for binding.

Crystallographic studies (24, 47, 48, 72) have shown that class I–bound peptides are relatively extended, with a kink at or near position 3. Although the various peptide structures can be juxtaposed so that their terminal residues are closely aligned at both ends, their backbone conformations and side-chain orientations can differ substantially (Figure 2). Nevertheless, docking peptides to class I receptors is easier than flexible docking in general, because the known structural motif substantially constrains the translational and rotational search space.

The only class II complex that has been crystallized displays a fully extended, twisted peptide conformation (74). Several biochemical stud-

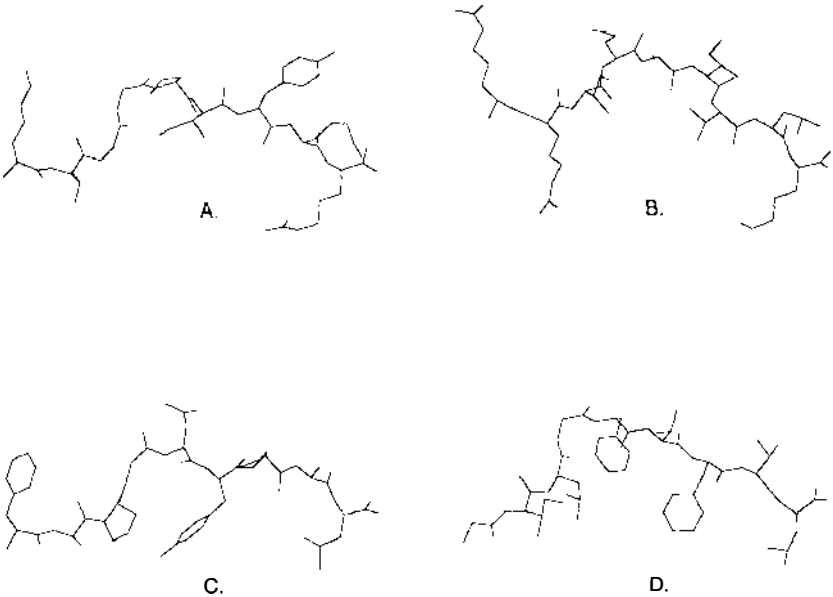

Figure 2 Bound conformation of four peptides bound to different MHC haplotypes. (*A*) HLA-Aw68 bound peptide. (*B*) HLA-B27 bound peptide. (*C*) H2-K^b bound peptide. (*D*) HLA-A2 bound peptide.

ies have suggested that class II–bound peptides share a common three-dimensional structure (50, 61).

METHODS OF DOCKING AND DESIGN

Docking encompasses a variety of different problems. The prediction of the optimum location of a ligand and its free energy of binding would ideally emerge from the crystal structure of the unliganded macromolecule without any a priori restrictions on the structure of the ligand. In practice, computations have been restricted to systems that allow various simplifications.

We assume that the binding sites, structural motifs of bound ligands, and structural variations within a given class of complexes are known, as in the proteases and MHC molecules as targets. Although this information significantly reduces the translational space that must be explored, ligand flexibility introduces substantial complications. Although the structure of the class I MHC has been known since 1987 (3), not even an approximate structure was predicted for the bound

peptide, despite several attempts (11, 64), before a high resolution X-ray structure of an MHC-peptide complex was obtained (47).

From Rigid to Flexible Algorithms

Rigid-body docking contributed to the development of the following computational concepts that play important roles in flexible docking and design: (*a*) shape descriptors, (*b*) grid-based energy evaluation, (*c*) soft potentials, (*d*) local minimization, and (*e*) multiple-copy techniques. We discuss these concepts in turn.

SHAPE DESCRIPTORS Shape descriptors consist of a relatively small number of characteristic points on the surfaces of the two molecules in a complex (69). One can search for the docked conformations by optimally superimposing or matching these points, which is computationally more efficient than matching two molecular surfaces, each defined by hundreds of atoms. In the DOCK program (16, 40), a set of spheres represents the negative image of the receptor, and a fast sphere-matching algorithm docks candidate ligands from a shape database. Katchalski-Katzir and coworkers (34) use three-dimensional discrete functions that distinguish between the surface and the interior of proteins. A correlation function is calculated to assess the extent of geometric match between surfaces of the molecules.

Shape complementarity alone is a poor target function that frequently fails to distinguish near-native complex conformations from ones that have a different structure but relatively good complementarity (70). Therefore, the concept of shape descriptors has been extended by requiring pairwise interactions. In the directed version of DOCK, a hydrogen-bonding parameter is introduced (43). Jiang & Kim (32) match opposing surface normal vectors. In addition to surface overlap, Bacon & Moult (2) include Coulombic interactions in the scoring function. Pellegrini & Doniach (62) use a full molecular-mechanics interaction potential, but only among relatively few fiducial atoms. Shape descriptors are used for docking molecular fragments in flexible docking programs such as GRID (26), LUDI (6), and CLIX (41) (see below).

GRID-BASED ENERGY EVALUATION A scoring scheme that accounts for electrostatic and van der Waals interactions as well as solvation effects generally improves accuracy of the docking and reduces the number of false positive solutions, i.e. complex structures that differ from the native (70). However, the use of more realistic potential function increases computational costs. These can be reduced by precomputing terms in the potential of the rigid receptor for each point on a three-

dimensional grid (26, 27). Any target function in which the ligand and receptor terms are separable can be treated in this manner. The approach can be combined with the use of shape descriptors (32, 52). Although grid-based energy evaluation applies to rigid receptors only, ligand energy values are not precomputed and hence the ligand can be flexible (27).

SOFT POTENTIALS Soft potentials are target functions that allow for some penetration of the protein surfaces being matched. Even the most rigid proteins undergo some changes when interacting with another molecule. Thus, rigid docking procedures frequently rank near-native conformations after structures that are far from the native, and a more realistic potential function does not necessarily improve performance. One way to address this difficulty is to use an artificial soft potential (2, 32, 62, 84, 88). Soft potentials can and have been used in conjunction with both shape descriptors and grid-based energy evaluations. Softening of the potential is also beneficial in flexible docking procedures.

LOCAL MINIMIZATION Another way of introducing limited flexibility involves local minimization. Rigid-body docking with six degrees of freedom is augmented by local energy minimization in the space of N additional continuous variables that describe the conformation of either the ligand only, or both the ligand and receptor (70). The score function in these procedures is usually the potential energy of the complex, based on a molecular-mechanics energy function such as CHARMm (7). Local minimization can be combined both with shape descriptors and soft potentials and also with grid-based energy evaluation if the receptor structure is kept rigid. Although minimization leads to some conformational changes, it yields results that depend heavily on the starting ligand conformation, and hence the method does not qualify as truly flexible docking (68).

MULTIPLE-COPY TECHNIQUES Several methods use numerous ligand copies, each transparent to the others but subject to the full force of the receptor, to determine energetically favorable positions and orientations of small molecules or functional groups in the binding site of a protein (21, 55, 65). Although the multiple-copy, simultaneous-search (MCSS) method was originally proposed in the context of a rigid receptor (55), it was extended recently to allow for ligand and receptor flexibility (65). In this latter case, several ligand copies and one copy of the receptor are minimized simultaneously, which substantially increases

computational efficiency because the internal energy of the receptor is calculated only once for all ligand copies.

Strategies for Flexible Docking and Design

Flexible docking strategies require no a priori knowledge of the ligand conformation and hence consist of searches in the space of $6+N$ translational, rotational, and conformational variables (11, 27, 28, 42, 88). Such an approach becomes a necessity if there is no useful information on the conformation of the ligand, as in the case of docking short linear peptides. Because flexible design relies heavily on flexible docking procedures, the two problems can be discussed simultaneously.

Four different strategies are currently in use for docking flexible ligands: (*a*) Monte Carlo or molecular-dynamics docking of complete molecules, (*b*) in-site combinatorial search, (*c*) ligand buildup, and (*d*) site mapping and fragment assembly. Some of the recent docking and design procedures combine several of these strategies. Although the same ideas are also important in pharmacophore-matching programs such as ALLADIN (81), CATALIST (73), and FOUNDATION (30), we restrict our discussion to docking and design for receptors with known three-dimensional structures.

DOCKING BY MONTE CARLO OR MOLECULAR-DYNAMICS METHODS The number of possible conformations increases exponentially with the conformational degrees of freedom, N. One way to avoid this combinatorial explosion is to sample the $6+N$ dimensional space by using Monte Carlo or molecular-dynamics simulations. Monte Carlo methods explore the thermodynamically accessible states of a system by generating small random changes that are either accepted or rejected according to the Metropolis algorithm (53). The probability of accepting steps that increase the energy depends on a temperature-like parameter T, which can be decreased according to a schedule known as simulated annealing (36). This popular method was first used for docking by Goodsell & Olson (27), who allowed rotations about a single bond in each step in addition to rigid body rotations and translations. Because a grid-based energy evaluation increased the efficiency of the procedure, the receptor was assumed to be rigid. Yue (88) used simulated annealing with a soft potential given in terms of distance constraints.

Simulated annealing should, in theory, always converge to the global minimum. Guaranteeing this convergence, however, would require an infinite number of temperature decrements, at each of which the system should reach equilibrium. In practice, one can only explore a few temperatures, and the number of equilibration steps at each temperature

is also limited. Furthermore, in macromolecular applications the numerous degrees of freedom in a Monte Carlo search limits it to only small regions of the conformational space, and the results depend on the starting conformation. The investigator can reduce this deficiency somewhat by using the method in tandem with a coarse-grain rigid search (11, 28) or by simply repeating the analysis at various starting conformations. One can increase the relatively small acceptance ratio yielded by Monte Carlo calculations by using Monte Carlo minimization, i.e. by combining each Monte Carlo step with local minimization (77).

Molecular-dynamics techniques generally work as well as Monte Carlo methods in the simulation of macromolecules, particularly for calculating thermodynamic properties. Although molecular dynamics is efficient in exploring low-energy states, conformational changes that require the crossing of even relatively low energy barriers are extremely rare, and because of limits on simulation times the analysis is strictly local. Investigators have explored larger regions of the conformational space of free peptides (14) and protein loops (9) simply by running the simulation at artificially high temperatures. High temperatures tend to unfold the receptor, but recent modifications sidestep these unwanted effects (18) by separating the center-of-mass motion of the ligand from its internal and rotational motions and by using different thermal baths to study each type of ligand motion and to assess the receptor motion. Thus, the temperatures and the time constants of coupling to the baths can be arbitrarily varied for these three types of motion, allowing either a frozen or a flexible receptor and allowing control of search rate without disturbing the internal structure of the receptor (18). The new method has been applied to only a single problem, but the idea is certainly appealing.

With increasing degrees of freedom, both Monte Carlo and molecular-dynamics strategies become more unreliable, because of the restricted sampling of the conformational space (11). Even such limited exploration is very time-consuming, and these methods are not very useful for screening large databases for suitable ligands (82).

IN-SITE COMBINATORIAL SEARCH The conformational space of small molecules, particularly short peptides, can be explored in discrete steps. Thus, in the conformational analysis of an oligopeptide, we assign a conformational state to each residue from a collection of discrete states, which are usually defined as low-energy conformations of isolated amino acids represented by points on the (ϕ,ψ) map (79, 80). Assigning side chains from a table of rotamers is a similar problem (17).

The problem is finding the combination of discrete states that minimizes a target function such as the potential energy. In principle, the optimal assignment can be found by a systematic search, but the number of conformations to be explored increases exponentially with the size of the problem (e.g. the number of residues). Combinatorial optimization algorithms such as branch-and-bound (80) or dead-end elimination (17) can efficiently limit the number of conformations whose energies need to be evaluated during the search for the global minimum.

When we consider a ligand in a given orientation relative to a protein, we can perform a conformational analysis of the ligand directly in the binding site (42) and can extend the analysis to receptor side chains that interact with the ligand (42). Because many of the conformations do not fit into the site, the search is computationally more efficient for a given ligand position than for an isolated molecule. Coupled to a search in the rotational and translational space of the ligand, the method can perform flexible docking that takes a significant portion of the receptor's conformational freedom into account.

LIGAND BUILDUP An efficient combinatorial strategy for solving stagewise optimization problems, known as ligand buildup (25, 79), involves constructing conformations by sequentially adding building blocks (i.e. fragments or atoms) in various conformations to a seed (i.e. starting) molecular fragment. Building ligands directly in the binding site is a remarkably efficient flexible-docking and -design strategy. With the GROW procedure (56), one can dock and/or design peptide inhibitors using a library of low-energy conformations of isolated amino acid residues as building blocks. The recent addition of a general template library extended this method to nonpeptide ligand design. The building blocks used in GroupBuild (66) are small functional groups such as hydroxy, carbonyl, or benzene rings. The buildup starts from a given ligand fragment or from a given position in the binding site. GenStar (67) and LEGEND (59) are buildup-type design programs that use single atoms rather than functional groups as building blocks.

Both combinatorial minimization and its special, buildup-type implementations are prone to combinatorial explosion, unless a large fraction of all conformations is discarded at every step according to some heuristic considerations. However, if the heuristics are based on valid assumptions, these approaches guarantee finding the lowest-energy structures available to the system, unlike random approaches such as the Monte Carlo technique.

SITE MAPPING AND FRAGMENT ASSEMBLY Goodford (26) introduced the idea of using molecular probes to map the binding site of a macromole-

cule to search for energetically favorable positions. Generally, the probes are functional groups (or, in the case of peptide ligands, possibly entire amino acid residues) that are docked to a target site. In particular, GRID (26) places small fragment probes at many regularly spaced grid points within the active site and determines the most favorable scores. Thus, the different groups (water, methyl group, amine nitrogen, carboxy oxygen, and hydroxyl) are used as specialized shape descriptors in a grid-based energy evaluation, and the energy contour surfaces for the various probes delineate regions of attraction between probe and protein. The procedure is well suited to multiple-copy techniques (55).

The output of GRID (and, more generally, the mapping of the receptor) provides information on favorable positions for several functional groups. Using local minimization, one can identify a few positions for each group at which it is both sterically acceptable and is likely to interact favorably with the surrounding side chains of the receptor. The goal of fragment-assembly approaches, pioneered by Lewis & Dean (44, 45), is to connect the individual molecular fragments into a single, viable molecule. The CLIX program (41) performs this task by screening a small molecular database. For each molecule, the program attempts to make a pair of the substituent's chemical groups spatially coincide with a pair of favorable interaction sites proposed by GRID in the binding site of the protein. Some site-mapping and fragment-assembly programs also make use of generalized shape descriptors. LUDI (6) positions molecular fragments into the binding site of an enzyme in such a way that hydrogen bonds with the enzyme are formed and hydrophobic pockets are filled. These fragments are then linked together with suitable spacers. The linked-fragment approach of Verlinde and coworkers (83) is somewhat more general, exploiting several properties (shape, hydrophobicity, hydrogen bonding, etc) as shape descriptors. Caflisch and coworkers (10) used MCSS to map a binding site and constructed possible ligands by building bonds to connect the various minima they found. To build the bonds, they used a simplified energy function that allowed an exhaustive search over all possible connections. In HOOK (19), MCSS is also used in the mapping stage, but the minima are connected by means of a search through a database of molecular scaffolds for possible connectors. In docking or designing peptide ligands, the use of entire residues instead of small functional groups generally increases the specificity of mapping and restricts the necessary considerations to a relatively small number of positions in the assembly stage (65, 68).

COMBINED STRATEGIES Several factors promote the development of combined strategies. First, in-site ligand buildup can be made more

efficient if the building blocks are favorable positions derived from site mapping. Conversely, site mapping results in numerous favorable positions and, in a design problem, in a variety of more or less equally favorable chemical groups at each position. Efforts to link such fragments clearly can profit from the methods of combinatorial optimization. Second, a hierarchical approach, the elements of which appear in almost every procedure, can increase efficiency. A hierarchical approach reduces the region of interest; for example, an initial rigid-body method might be used to search for a manageable number of conformations, which are then subjected to detailed exploration with sophisticated strategies and more realistic target functions. Finally, combining various methods can be very efficient in particular problems. For example, the end groups of peptides bound to class I MHC receptors are confined to small regions at the two ends of the site. As shown below, one can exploit this property by mapping these regions to find favorable positions for the end residues and then determining the conformation of the intervening peptide fragment by chain-closure algorithms (65, 68) based either on combinatorial search (8) or on a multiple copy scaling-relaxation algorithm (90, 91).

Empirical Evaluation of Binding Free Energy

The most natural score function, both in design and docking, is the free energy of binding given by $\Delta G = G^b - G^f$, where G^f and G^b, respectively, denote the free energies of the system before (free) and after (bound) the complex is formed. In design, the objective is to identify the ligand with the lowest attainable value of ΔG, whereas in docking, G^f is a constant, and hence the minimum of the binding free energy is attained at the minimum of G^b.

The established methods of free-energy evaluation by thermodynamic integration or perturbation approaches through molecular-dynamics or Monte Carlo simulations are far too demanding computationally to be used in docking or design (86). Several empirical free-energy evaluation models have been published to overcome this limitation (31, 58, 60, 86). In addition to a molecular-mechanics interaction energy, empirical models estimate the contributions of solvation and entropy loss. If these terms are simple enough, the model enables fast free-energy evaluation and hence can serve as a target function.

Novotny and coworkers (60) calculated the free-energy change accompanying the association of two rigid molecules according to the equation

$$\Delta G = E_{el} + \Delta G_h - T\Delta S_{sc} + \text{const}, \qquad 1.$$

where E_{el} is the electrostatic interaction energy between the ligand and the receptor, ΔG_h is the solvation free energy, and ΔS_{sc} is the loss of side-chain conformational entropy. The constant term includes further entropy loss resulting from changes in rotational and translational degrees of freedom and the cratic contribution to entropy. Equation 1 does not include a van der Waals component because the protein-protein and protein-water van der Waals interactions are assumed to cancel (60). The solvation free energy ΔG_h is assumed to be proportional to the change in solvent-accessible surface area upon binding. The side-chain entropy loss ΔS_{sc} is also calculated by taking into account the change in accessible atomic surface areas. The ΔG values calculated for nine endopeptidase–protein inhibitor complexes showed good agreement with experiments (38).

Horton & Lewis (31) estimated the free energy of rigid binding by

$$\Delta G = \alpha \Delta G_{apolar} + \beta \Delta G_{polar} + \text{const}, \qquad 2.$$

where both solvation free-energy terms, ΔG_{apolar} and ΔG_{polar}, are calculated by sums of the form used by Eisenberg & McLachlan (20). The polar term encompasses only atoms that are involved in hydrogen bonds and salt bridges; all other atoms contribute to the apolar term. The coefficients α and β are dimensionless parameters determined by fitting Equation 2 to binding free energies measured in 15 rigid-body complexes (31). On the basis of general relationships in the thermodynamics of protein folding (57), Murphy and coworkers described the structural energetics of angiotensin binding (58) with a similar expression that, in addition to a constant, consists of terms that depend on polar and apolar surface areas.

Whereas Equation 1 can be derived from the general expression $\Delta G = \Delta E - T\Delta S$ by considering the energy and entropy differences between bound and free states and using several simplifying assumptions, the model of Horton & Lewis (Equation 2) is completely empirical. Furthermore, the two expressions differ strongly, because of the Coulombic term E_{el} in Equation 1. However, in analyzing high-resolution X-ray structures of protein complexes, we found that the electrostatic interaction energy E_{el} can be accurately described as a linear combination of polar and apolar surface areas, probably because surface charges dominate the interaction energy (80a). Thus, the two free-energy expressions are, in fact, very similar.

Equations 1 and 2 assume that both the receptor and the ligand are rigid, but this is clearly a very poor approximation for many ligands, such as short peptides. The flexibility of the ligand has two consequences: Neither (*a*) the conformational (internal) energy change of the

ligand, ΔE, nor (*b*) the loss of conformational entropy upon binding are negligible. Because the free ligand adopts many different conformations, one needs to calculate the average energy change $\langle \Delta E \rangle$. The solvation free-energy change ΔG_h also depends on the conformation of the free ligand, and hence in Equation 1 it should be replaced by the ensemble average $\langle \Delta G_h \rangle$ (80a). For peptides, the equation must account for the backbone entropy loss ΔS_{bb}, which can be estimated by an empirical method such as conformational filtering (78). Thus, the free energy of binding of a flexible ligand is calculated by

$$\Delta G = \langle E_{el} \rangle + \langle \Delta G_h \rangle - T\Delta S_{sc} - T\Delta S_{bb} + \text{const.} \qquad 3.$$

Because $\langle E_{el} \rangle$ and $\langle \Delta G_h \rangle$ are averages over the ensemble of free ligand conformations, the evaluation of Equation 3 is far from simple.

Verlinde & Hol (82) recently formulated the most important qualitative rules for the design of high-affinity ligands as follows: (*a*) The steric and electrostatic complementarity should be excellent; (*b*) a fair amount of hydrophobic surface should be buried in the complex; and (*c*) substantial conformational rigidity of the ligand is required to reduce the entropy loss upon binding. The terms in Equation 1 and 3 are in good agreement with these rules. The only difference is that, because of the lack of the van der Waals term, the equations do not measure steric complementarity. In fact, the simplified model assumes good packing of both protein-ligand and protein-solvent interfaces, and hence all steric clashes have to be removed before the free-energy evaluation, e.g. by subjecting the complex to local energy minimization using a traditional molecular-mechanics potential function.

APPLICATIONS OF FLEXIBLE METHODS: CURRENT STATUS

Design of Protease Inhibitors

In each of the major families of proteases, the active site exhibits a high degree of structural conservation. Because the three-dimensional structure of at least one member of each of these families has been determined, attempts to design inhibitors for a protease whose structure is unknown can start by predicting the structure of the protease by means of homologous extension. The first successful study of this kind was the design of antihypertensive drugs—inhibitors of renin and the angiotensin-converting enzyme.

The first inhibitors designed for renin, an aspartic protease, simply incorporated the naturally occurring, small-molecule inhibitor of this

enzyme, statin, into an octapeptide from its natural substrate, angiotensinogen. The resulting inhibitor was improved by building a model of the protease through homologous extension, by using molecular graphics to dock the inhibitor, and by creating more potent derivatives by means of chemical intuition (4). A similar approach (13) yielded a potent inhibitor for angiotensin-converting enzyme, a zinc protease whose structure is unknown. This inhibitor is widely used as the antihypertensive drug Captopril. A recent study showed that this compound is highly flexible in its free state (46), and its success as a drug demonstrates that protease inhibitors can be designed using traditional tools. However, Hassall and coworkers (29) later addressed the issue of flexibility and successfully increased Captopril's potency. They fit a model of the molecule, which they considered to have two freely rotatable bonds, to the known structure of the thermolysin active site, and determined its most likely conformation using molecular graphics. They designed a bicyclic derivative based on this conformation, thereby abolishing freedom of rotation about the two rotatable bonds and imposing a conformation that they predicted is the most suitable for binding.

Many other systems have been developed with similar approaches, i.e. by homologous extension from a known protease-inhibitor complex. Cohen and coworkers (12) designed a peptide inhibitor for a serine protease of the blood fluke *Schistosoma mansoni*. A model for the protease was built based on the known structure of porcine pancreatic elastase, and a substrate was designed based on the structure and binding mode of boronic acid inhibitor bound to α-lytic protease. Sumiya and coworkers (75) designed a specific inhibitor for cathepsin B, a lysosomal thiol protease. The structure of the protease was modeled based on the crystal structure of a papain-inhibitor complex (also a thiol protease). The authors then used molecular graphics to dock a known inhibitor to the model and, on the basis of chemical intuition, synthesized 17 derivatives, one of which proved to be both potent and specific.

Similar studies have examined proteases of known structure, of which the best studied is the aspartic protease of HIV (for a review, see 87). The DOCK program (15) was applied to this enzyme to search through approximately 10^4 small molecules for potential inhibitors. An additional visual search among the 200 top-ranked compounds for proximity to the catalytic site, hydrogen-bonding capacity, and chemical synthesizability resulted in a candidate inhibitor. Although the affinity of the final compound is several orders of magnitude too low for clinical use, this molecule can be used as a lead compound for the design of more potent inhibitors.

Rini and coworkers (63) used the same techniques to design inhibi-

tors against proteases crucial to the pathogenicity of the schistosome and malaria parasites, for which the structures are unknown. The two enzymes, a serine and a cysteine protease, were modeled based on the known structure of homologous proteins, and a database of approximately 55,000 compounds was searched for suitable lead compounds. The compounds were ranked according to two target functions: the shape-complementarity function (15) and a simplified molecular-mechanics potential approximating the interaction energy between the protease and ligand. Both functions proved inadequate for choosing high-affinity compounds, and potential lead compounds were chosen based on visual inspection of the 2200 compounds with the best shape complementary and the 2200 with the best interaction energy. This study suggests that the efficacy of determining lead compounds may be sensitive to the accuracy of the proposed structure of the receptor when traditional methods and target functions are used.

Flexible-docking methods have been used only recently. Caflisch and coworkers (10) applied a fragment assembly method to dock a known peptide inhibitor to HIV protease. MCSS was used to find possible minima for functional groups corresponding to backbone and side-chain pieces and for the peptide constructed by connecting these fragments. With GroupBuild, Rotstein & Murcko (66) reconstructed a known inhibitor from a fragment used as a seed. Both studies obtained results in accord with crystallographic data, but the applicability of the methods to the design of novel inhibitors has yet to be demonstrated.

Peptide Binding to Major Histocompatibility Complex Receptors

As noted previously, attempts to dock peptides to class I MHC receptors have clearly demonstrated the limitations of the currently available docking methods. Energy minimization (68), molecular-dynamics (64), and Monte Carlo (11) docking strategies failed dramatically to predict the bound conformation of antigenic peptides: In all cases the peptide, initially put in an α-helical conformation, remained helical throughout the calculations despite the fact that the native conformation is extended. Of the several crystal structures of class I–peptide complexes now available—of the same MHC haplotype with varying peptides and of varying haplotypes—all show a highly conserved hydrogen-bond network between the backbone of the terminal peptide residues and conserved side chains of the receptors. If the peptide termini are in highly favorable energy minima, one can assume that their optimal position and orientation when isolated from the rest of the peptide will significantly overlap their crystallographic position.

This assumption forms the basis for two recent studies (65, 68) in which each of the terminal residues was initially docked in isolation and the conformation of the intervening chain was then calculated using a loop-closure algorithm. In the first study, the terminal residues were docked using a rigid-body grid search augmented by local energy minimization at energetically favorable grid positions; the second study utilized a multiple-copy (21, 55) method to simultaneously minimize the energy of 20 residue copies that had random initial conformations and positions. Because the electrostatic interaction between the ligand (single-residue) backbone and the receptor is strong, both of these methods succeed in finding the ligand's optimal position. The side-chain conformations, on the other hand, are harder to predict. In both methods, the predicted side-chain conformation depends strongly on the initial side-chain conformation and can be found only by testing a large number of starting conformations. The multiple-copy methods, which enhance sampling, address this problem well. In fact, we (R Rosenfeld, Q Zheng, S Vajda & C DeLisi, submitted) recently found the correct side-chain conformations using this method, having chosen as the target function the potential energy of the system augmented by a qualitative entropy-related term derived from the clustering of copies in a multiple-copy simulation (65, 89, 91).

Although the use of loop closure algorithms in determining the conformation of the intervening residues has met with only limited success, largely because of the omission of water from the simulations, it nevertheless is of practical value. These calculations correctly predict the function of each of the peptide side chains (anchor vs T-cell epitope), as well as the overall orientation of the peptide, as measured by the hydrogen bonds between the peptide backbone and the receptor (R Rosenfeld, Q Zheng, S Vajda & C DeLisi, submitted). This information is important for the development of vaccine-design strategies. In contrast, the correct prediction of the atomic coordinates awaits further methodological development.

CONCLUSIONS

The number of available docking and design methods is exploding. Some of the fundamental ideas, such as shape descriptors (40), grid-based energy evaluation, and mapping of the binding site (26), have been known for almost a decade. Nevertheless, essentially all general-purpose docking and structure-based design programs, particularly those explicitly addressing ligand flexibility, have been published in the past four years—most in the past two years (82). Clearly, the impor-

tance of designing inhibitors for proteases, particularly the HIV protease, has substantially promoted the development of these programs and has focused efforts on the problem of ligand flexibility.

The recent and explosive developments in methodology have led to an interesting situation. On one hand, some of the most powerful methods available for flexible drug design are so new that they have been tested in only a few relatively simple problems. On the other hand, other research groups have not waited for such state-of-the-art technologies and have solved difficult drug-design problems using traditional tools such as rigid-body docking and visual inspection on the computer screen. As discussed above, these traditional tools have been successful in the design of protease inhibitors, but the problem of docking peptide antigens in MHC receptors requires a genuine flexible methodology. The new programs are likely to result in improved inhibitors and ligands in both types of problems.

In spite of the spectacular rate at which computational tools are being developed, additional work will be required to solve two major difficulties. First, both docking and design as defined here lead to nondeterministic polynomial (NP)–complete combinatorial problems. Solving such problems by any exact method presumably requires computing time that grows faster than any polynomial in N, the number of degrees of freedom. In practice, investigators avoid the combinatorial explosion by using criteria that retain only some top-scoring solutions at various steps of the algorithms. Thus, the final solutions are likely to be suboptimal rather than optimal. Second, a related and even more important problem is selecting a target function that estimates the binding free energy efficiently and with adequate accuracy. Recently developed empirical models seem to satisfy these criteria when tested on complexes of rigid molecules with known structure, but these models have not yet been coupled with search algorithms. In particular, the success of such a procedure depends on the sensitivity of the calculated free energies to inevitable errors in the atomic coordinates. Hence, this sensitivity requires careful analysis. As we have discussed, the use of free energy as a target function can be easily extended to flexible ligands in docking problems, but design requires calculating the free energy of the unbound ligand, a problem that is substantially more difficult.

ACKNOWLEDGMENTS

We thank Zhiping Weng for reading the manuscript. This work was supported by NIH/NIAID grant AI30535 (CD and RR) and DOE grant FG02-93ER61656 (SV).

Literature Cited

1. Alberg DG, Schreiber SL. 1993. Structure-based design of a cyclophilin-calcineurin bridging ligand. *Science* 262: 248–50
2. Bacon DJ, Moult J. 1992. Docking by least-squares fitting of molecular surface patterns. *J. Mol. Biol.* 225:849–58
3. Bjorkman PJ, Saper MA, Samraoui B, Bennet WS, Strominger JL, Wiley DC. 1987. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329:506–12
4. Blundell TL, Cooper J, Foundling SI, Jones DM, Atrash B, Szelke M. 1989. On the rational design of renin inhibitors: X-ray studies of aspartic proteinases complexed with transition-state analogues. In *Perspectives in Biochemistry*, ed. H Neurath, pp. 83–89. Washington, DC: Am. Chem. Soc.
5. Bode W, Huber R. 1991. Ligand binding: proteinase-protein inhibitor interactions. *Curr. Opin. Struct. Biol.* 1: 45–52
6. Bohm HJ. 1992. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* 6:131–47
7. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMm: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217
8. Bruccoleri RE, Karplus M. 1987. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* 26: 137–68
9. Bruccoleri RE, Karplus M. 1990. Conformational sampling using high-temperature molecular dynamics. *Biopolymers* 29: 1847–62
10. Caflisch A, Miranker A, Karplus M. 1993. Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *J. Med. Chem.* 36:2142–64
11. Caflisch A, Niederer P, Anliker M. 1992. Monte Carlo docking of oligopeptides to proteins. *Proteins* 13: 223–30
12. Cohen FE, Gregoret LM, Amiri P, Aldape K, Railey J, McKerrow JH. 1991. Arresting tissue invasion of a parasite by protease inhibitors chosen with the aid of computer modeling. *Biochemistry* 30:11221–29
13. Cushman DW, Cheung HS, Sabo EF, Ondetti MA. 1977. Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptalkanoyl amino acids. *Biochemistry* 16:5484–91
14. Daggett V, Levitt M. 1992. Molecular dynamics simulations of helix denaturation. *J. Mol. Biol.* 233: 1121–38
15. DesJarlais RL, Seibel GL, Kuntz ID, Furth PS, Alvarez JC, et al. 1990. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. USA* 87: 6644–48
16. DesJarlais RL, Sheridan RP, Seibel, GL, Dixon JS, Kuntz ID. 1988. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* 31: 722–29
17. Desmet J, DeMaeyer M, Hazes B, Lastern I. 1992. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* 356: 539–42
18. DiNola A, Roccatano D, Berendsen HJC. 1994. Molecular dynamics simulation of the docking of substrates to proteins. *Proteins* 19:174–82
19. Eisen MB, Wiley DC, Karplus M, Hubbard RE. 1994. HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecular binding site. *Proteins* 19:199–221
20. Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature* 319:199–203
21. Elber R, Karplus M. 1990. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* 112:9161–75

22. Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee H-G. 1991. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 351:290–96
23. Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee H-G. 1994. Pool sequencing of natural HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules. *Immunogenetics* 39: 230–42
24. Fremont DH, Matsumura M, Stura EA, Peterson PA, Wilson IA. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-2 K^b. *Science* 257:919–27
25. Gibson KD, Scheraga HA. 1987. Revised algorithms for the build-up procedure for predicting protein conformations by energy minimization. *J. Comput. Chem.* 8: 826–34
26. Goodford PJ. 1985. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28:849–75
27. Goodsell DS, Olson AJ. 1990. Automated docking of substrates to proteins by simulated annealing. *Proteins* 8:195–202
28. Hart TN, Read NJ. 1992. A multiple-start Monte-Carlo docking method. *Proteins* 13:206–22
29. Hassall CH, Krohn A, Moody CJ, Thomas WA. 1982. The design of a new group of angiotensin-converting enzyme inhibitors. *FEBS Lett.* 147: 175–79
30. Ho CMW, Marshall G. 1993. FOUNDATION: a program to retrieve all possible structures containing a user-defined minimum number of matching query elements from three-dimensional databases. *J. Comput. Aided Mol. Des.* 7:3–22
31. Horton N, Lewis M. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci.* 1: 169–81
32. Jiang F, Kim SH. 1992. "Soft docking": matching the molecular surface cubes. *J. Mol. Biol.* 219:79–102
33. Kaslow RA, Duquesnoy R, Van Raden M. 1990. A1, Cw7, B8, DR3 HLA antigen combination associated with rapid decline of T-helper lymphocytes in HIV-1 infection. *Lancet* 335: 927–30
34. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem A, Aflalo C, Vakser IA. 1992. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA* 89:2195–99
35. King PM, Richards WG. 1992. Free energy calculations in molecular biophysics. *Mol. Phys.* 76:251–75
36. Kirpatrick S, Gelatt CD Jr, Vecchi MP. 1983. Optimization by simulated annealing. *Science* 220:671–80
37. Kollman PA. 1994. Theory of macromolecule-ligand interactions. *Curr. Opin. Struct. Biol.* 4:240–45
38. Krystek S, Stouch T, Novotny J. 1993. Affinity and specificity of serine endopeptidase-protein inhibitor interactions: empirical free energy calculation based on X-ray crystallographic structures. *J. Mol. Biol.* 234:661–79
39. Kuntz ID. 1992. Structure-based strategies for drug design and discovery. *Science* 257:1078–82
40. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. 1982. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* 161:269–88
41. Lawrence MC, Davis PC. 1992. CLIX: a search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. *Proteins* 12:31–41
42. Leach AR. 1994. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* 235:345–56
43. Leach AR, Kuntz ID. 1992. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comp. Chem.* 13:733–48
44. Lewis RA, Dean PM. 1989. Automated site-directed drug design: the concept of spacer skeletons for primary structure generation. *Proc. R. Soc. London Ser. B* 236:125–40
45. Lewis RA, Dean PM. 1989. Automated site-directed drug design: the formation of molecular templates in primary structure generation. *Proc. R. Soc. London Ser. B* 236:141- 62
46. Luke BT. 1994. A quantum mechanical conformational search of Captopril, a potent inhibitor of the angiotensin-converting enzyme. *J. Mol. Struct.* 309:1–11
47. Madden DR, Garboczi DN, Wiley DC. 1994. The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. *Cell* 75: 693–708
48. Madden DR, Gorga, JC, Strominger JL, Wiley DC. 1992. The three-dimensional structure of HLA-B27 at 2.1 Å

resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* 70:1035–48
49. Marshall G. 1992. Three-dimensional structure of peptide-protein complexes: implications for recognition. *Curr. Opin. Struct. Biol.* 2:904–19
50. Marshall KW, Liu AF, Canales J, Perahia B, Jorgensen B, et al. 1994. Role of the polymorphic residues in the HLA-DR molecules in allele-specific binding of peptide ligands. *J. Immunol.* 152:4946–57
51. Meng EC, Gschwend DA, Blaney JM, Kuntz ID. 1993. Orientational sampling and rigid-body minimization in molecular docking. *Proteins* 17: 266–78
52. Meng EC, Schoichet BK, Kuntz ID. 1992. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* 13:505–24
53. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–92
54. Mezei M, Beveridge DL. 1986. Free energy simulations. *Ann. N. Y. Acad. Sci.* 494:1–23
55. Miranker A, Karplus M. 1991. Functionality maps of binding sites: a multicopy simultaneous search method. *Proteins* 1:29–34
56. Moon JB, Howe J. 1991. Computer design of bioactive molecules: a method for receptor-based de novo ligand design. *Proteins* 11:314–28
57. Murphy KP, Freire E. 1992. Thermodynamics of structural stability and cooperative folding behaviours in proteins. *Adv. Protein. Chem.* 43:313–61
58. Murphy KP, Xie D, Garcia KC, Amzel LM, Freire E. 1993. Structural energetics of peptide recognition: angiotensin II/antibody binding. *Proteins* 15: 1120
59. Nishibata Y, Itai A. 1993. Confirmation of usefulness of a structure construction program based on three-dimensional receptor structure for rational lead generation. *J. Med. Chem.* 36: 2921–28
60. Novotny J, Bruccoleri RE, Saul FA. 1989. On the attribution of binding energy in antigen-antibody complexes MCPC603, D1.3, and HyHEL-5. *Biochemistry* 28:4735–49
61. O'Sullivan D, Arrhenius T, Sidney J, Del Guercio M-F, Albertson M, et al. 1991. On the interaction of promiscuous antigenic peptides with different DR alleles: identification of common structural motifs. *J. Immunol.* 147: 2663–69
62. Pellegrini M, Doniach S. 1993. Computer simulation of antibody binding specificity. *Proteins* 15:436–44
63. Ring CS, Sun E, McKerrow HJ, Lee GK, Rosenthal PH, et al. 1993. Structure-based inhibitor design by using protein models for the development of antiparasitic agents. *Proc. Natl. Acad. Sci. USA* 90:3583–87
64. Rognan D, Reddehase MJ, Koszinowski UH, Folkers G. 1992. Molecular modeling of an antigenic complex between a viral peptide and a class I major histocompatibility glycoprotein. *Proteins* 13:70–85
65. Rosenfeld R, Zheng Q, Vajda S, DeLisi C. 1993. Computing the structure of bound peptides: application to antigen recognition by class I MHCs. *J. Mol. Biol.* 234:515–21
66. Rotstein SH, Murcko MA. 1993. GroupBuild: a fragment-based method for de novo drug design. *J. Med. Chem.* 36:1700–10
67. Rotstein SH, Murcko MA. 1993. GenStar: a program for de novo drug design. *J. Comput. Aided Mol. Des.* 7: 23–43
68. Sezerman U, Cornette J, Vajda S, DeLisi C. 1993. Toward computational determination of peptide-receptor structure. *Protein Sci.* 2:1827–46
69. Shoichet BK, Bodian DL, Kuntz ID. 1992. Molecular docking using shape descriptors. *J. Comput. Chem.* 3: 380–97
70. Shoichet BK, Kuntz ID. 1991. Protein docking and complementarity. *J. Mol. Biol.* 221:327–46
71. Shoichet BK, Stroud RM, Santi D, Kuntz ID, Perry KM. 1993. Structure-based discovery of inhibitors of thymidilate synthase. *Science* 259:1145–50
72. Silver ML, Guo HC, Strominger JL, Wiley D. 1992. Atomic structure of a human MHC molecule presenting an influenza virus peptide. *Nature* 360: 367–68
73. Sprague PW. 1991. CATALIST: a computer aided drug design system specifically designed for medicinal chemists. In *Recent Advances in Chemical Information. Proc. 1991 Chemical Information Conf.*, ed. H Coller, pp. 107–11. London: R. Soc. Chem.
74. Stern LG, Brown JH, Jardetsky TS, Gorga JC, Urban RG, et al. 1994. Crystal structure of the human class

II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 368:215–21
75. Sumiya S, Yoneda T, Kitamura K, Murata M, Yokoo C, et al. 1992. Molecular design of potent inhibitor specific for cathepsin B based on the tertiary structure prediction. *Chem. Pharm. Bull.* 40:299–303
76. Taylor R. 1994. Histocompatibility antigens, protective immunity, and HIV-1. *J. NIH Res.* 6:68–71
77. Totrov M, Abagyan R. 1994. Detailed ab initio prediction of lysozyme-antibody complex with 1.6Å accuracy. *Nat. Struct. Biol.* 1:259–63
78. Vajda S. 1993. Conformational filtering in polypeptides and proteins. *J. Mol. Biol.* 229:125–45
79. Vajda S, DeLisi C. 1990. Determining minimum energy conformations of polypeptides by dynamic programming. *Biopolymers* 29:1755–72
80. Vajda S, DeLisi C. 1994. An adaptive branch-and-bound minimization method based on dynamic programming. In *The Protein Folding Problem and Tertiary Structure Determination,* ed. KM Mertz, SM LeGrand, pp. 409–32. Boston: Birkhauser
80a. Vajda S, Weng Z, Rosenfeld R, DeLisi C. 1994. Effect of conformational flexibility and solvation on receptor-ligand binding free energies. *Biochemistry* 33:13977–88
81. VanDrie JH, Weininger D, Martin YC. 1989. ALLADIN: an integrated tool for computer-assisted molecular design and pharmacophore recognition from geometric, steric, and substructure searching of three-dimensional molecular structures. *J. Comput. Aided Mol. Des.* 3:225–54
82. Verlinde CLMJ, Hol WGJ. 1994. Structure-based drug design: progress, results and challenges. *Structure* 2: 577–87
83. Verlinde CLMJ, Rudenko G, Hol WGJ. 1992. In search of new lead compounds for trypanosomiasis drug design: a protein structure-based linked-fragment approach. *J. Comput. Aided Mol. Des.* 6:131–47
84. Walls PH, Sternberg MJE. 1992. New algorithm to model protein-protein recognition based on surface complementarity: applications to antibody-antigen docking. *J. Mol. Biol.* 228: 277–97
85. Wang H. 1991. Grid-search molecular accessible surface algorithm for solving the protein docking problem. *J. Comp. Chem.* 12:746–50
86. Wilson C, Mace JE, Agard DA. 1991. Computational method for the design of enzymes with altered substrate specificity. *J. Mol. Biol.* 220:495–506
87. Wlodawer A, Erickson JW. 1993. Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* 62: 543–85
88. Yue S. 1990. Distance-constrained molecular docking by simulated annealing. *Protein Eng.* 4:177–84
89. Zheng Q, Rosenfeld R, DeLisi C, Kyle DJ. 1994. Multiple copy sampling in protein loop modeling: computational efficiency and sensitivity to dihedral angle perturbations. *Protein Sci.* 3: 493–506
90. Zheng Q, Rosenfeld R, Vajda S, DeLisi C. 1993. Loop closure via bond scaling and relaxation. *J. Comput. Chem.* 14:556–65
91. Zheng Q, Rosenfeld R, Vajda, S, DeLisi C. 1993. Determining protein loop conformation using scaling-relaxation techniques. *Protein Sci.* 2:1242–48