PROTEINS: Structure, Function, and Genetics 32:159–174 (1998)

Flexible Docking Allowing Induced Fit in Proteins: Insights From an Open to Closed Conformational Isomers

Bilha Sandak,[1]* **Haim J. Wolfson,**[2] **and Ruth Nussinov**[3,4]*
[1]*Department of Applied Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel*
[2]*Computer Science Department, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel*
[3]*Laboratory of Experimental and Computational Biology, SAIC, NCI-FCRDC, Frederick, Maryland*
[4]*Sackler Institute of Molecular Medicine, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel*

ABSTRACT **Here we dock a ligand onto a receptor surface allowing hinge-bending domain/substructural movements. Our approach mimics and manifests induced fit in molecular recognition. All angular rotations are allowed on the one hand, while a conformational space search is avoided on the other. Rather than dock each of the molecular parts separately with subsequent reconstruction of the consistently docked molecules, all parts are docked simultaneously while still utilizing the position of the hinge from the start. Like pliers closing on a screw, the receptor automatically closes on its ligand in the best surface-matching way. Movements are allowed either in the ligand or in the larger receptor, hence reproducing induced molecular fit. Hinge bending movements are frequently observed when molecules associate. There are numerous examples of open versus closed conformations taking place upon binding. Such movements are observed when the substrate binds to its respective enzyme. In particular, such movements are of interest in allosteric enzymes. The movements can involve entire domains, subdomains, loops, (other) secondary structure elements, or between any groups of atoms connected by flexible joints. We have implemented the hinges at points and at bonds. By allowing 3-dimensional (3-D) rotation at the hinge, several rotations about (consecutive or nearby) bonds are implicitly taken into account. Alternatively, if required, the point rotation can be restricted to bond rotation. Here we illustrate this hinge-bending docking approach and the insight into flexibility it provides on a complex of the calmodulin with its M13 ligand, positioning the hinges either in the ligand or in the larger receptor. This automated and efficient method is adapted from computer vision and robotics. It enables utilizing entire molecular surfaces rather than focusing a priori on active sites. Hence, allows attaining the overall optimally matching surfaces, the extent and type of motions which are involved. Here we do not treat the conformational flexibility of side-chains or of very small pieces of the molecules. Therefore, currently available methods addressing these issues and the method presented here, are complementary to each other, expanding the repertoire of computational docking tools foreseen to aid in studies of recognition, conformational flexibility and drug design. Proteins 32:159–174, 1998.**
© 1998 Wiley-Liss, Inc.

Key words: **molecular recognition; flexible docking; protein-ligand interaction; induced fit; structure-based drug design**

INTRODUCTION

Here we present and use a docking method which allows hinge induced conformational flexibility in either of the associating molecules. Hinge-bending transitions may occur both during molecular recognition and binding, and independently of such events. In either case, movements of whole domains or of rather small parts may take place at flexible joints, between or within secondary structure elements.[1]

Independently folding units may exhibit hinge-bending in the absence of a ligand, as has been shown in the case of the T4 lysozyme[2] and in the

---

Grant sponsor: National Cancer Institute, DHHS; Grant number: I-CO-74102; Grant sponsor: U.S.-Israel Binational Science Foundation (BSF), Jerusalem, Israel; Grant number: 95-0028; Grant sponsor: Recanati Fund; Grant sponsor: Tel Aviv University; Grant sponsor: Ministry of Industry and Technology (MAGNET).

*Correspondence to: Ruth Nussinov, Laboratory of Experimental and Computational Biology, SAIC, NCI-FCRDC, Building 469, Room 151, Fredrick, MD 21702, E-mail:ruthn@ncifcrf.gov or Bilha Sandak, Department of Applied Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: billie@wisdom.weizmann.ac.il

Received 17 October 1997; Accepted 27 January 1998

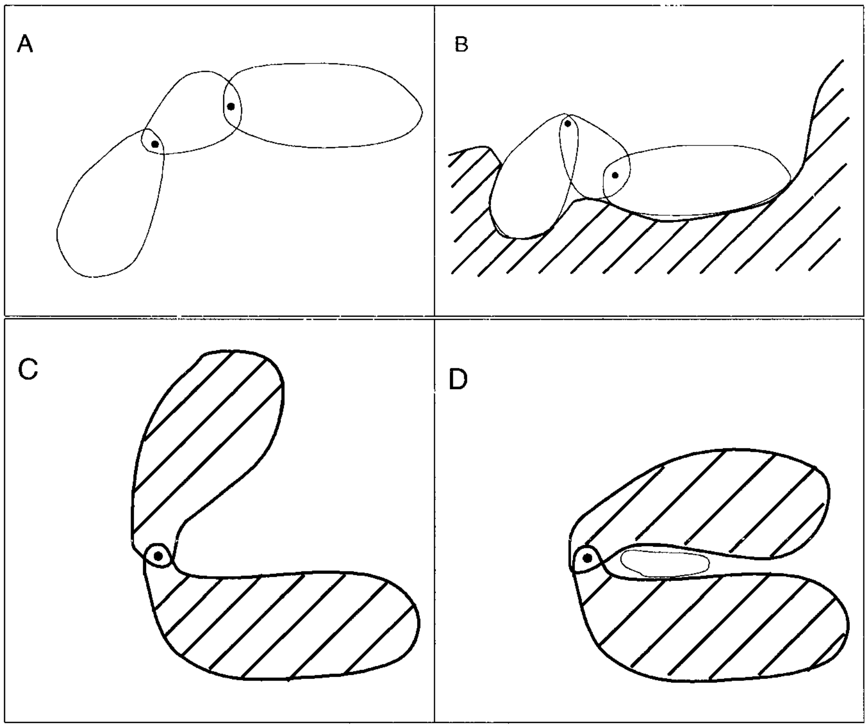

Fig. 1. A schematic illustration of hinge-bending movements accounted for by our method. (**A**) A flexible ligand consisting of three rigid parts connected by two hinges. Two consecutive parts share a common hinge. (**B**) The flexible ligand displayed in (A), is docked to a receptor (bold-face line type). (**C**) A flexible receptor consisting of two rigid domain parts connected by a hinge. (**D**) The flexible receptor displayed in (C), binds a ligand (light line type).

catabolite gene activator protein.[3] The movement of a small unit during molecular recognition has been observed in the triose phosphate isomerase, where the binding of substrates has induced a conformational transition of an 11-residue loop, apparently as a relatively rigid unit.[4] The binding of cyclosporin-A to cyclophilin is one example among many others of a flexible ligand.[5] Conformational changes in antigen-antibody binding have been reported for example in Rini et al.,[6] Stanfield et al.,[7] and more recently in Wedemayer et al.[8]

Large protein molecules are frequently built from domains which move with respect to each other. Domain rearrangements are critical for protein function. Early investigations of these movements have been carried out by Janin and Wodak[9] and by Bennet and Huber.[10] These studies have provided first insight into the mechanism underlying domain motions. Since then, much information has been accumulated leading to several comprehensive analyses of the types of motions which are involved (reviewed by Huber[11] and Schultz[12]). Recently, Gerstein et al.[13] have compiled and classified these domain rearrangements, dividing them into two main types, shear and hinge-bending. Whereas in a shear-type motion, a relatively large number of interfacing residues are involved, in a hinge-bending type of motion segments of the chain change significantly their position with respect to each other (Fig. 1). Domain motions are also manifested in domain-swapping (Bennett et al., 1995[14]), where domains may pair either with sister-domains from the same monomer, or swap to pair with counterparts from different chains within the oligomer. A small energetic cost is essential to enable a fast movement, which is necessary for protein function.

The binding of a receptor and a ligand generally elicits movements of segments of the molecules involved. The switch from an open to a closed conformation may both push out the water molecules, as well as aid in positioning of the catalytic groups in a favorable orientation with respect to the substrate. Furthermore, closure also traps the substrate, preventing escape of the reaction intermediates (Gerstein et al., 1994[13] and references therein). A particularly interesting case has been observed recently by Bernstein and colleagues,[15] showing a dramatic closing of a large cleft between two domains in phosphoglycerate kinase. This substrate induced effect brings the two ligands, 3-phosphoglycerate and ADP, in close proximity. Flap domain motions of the receptor in the binding of HIV-1 protease and a peptide inhibitor have been observed.[16] In the calmodulin-synthetic peptide binding, the two domains of the calmodulin receptor engulf and clamp the peptide ligand in a domain motion which is predominantly hinge.[13,17–19] In variants of adenylate kinase, the AMP and ATP substrates are buried by a domain rotating by 90°.[20,21] Lactoferrin,[22–24] LAO-binding protein,[25] and maltodextrin binding protein,[26,27] are examples of protein receptors that undergo ligand-induced conformational changes, all bearing similar morphology to the binding site. Domain movements are crucial for a variety of protein functions: cellular locomotion, transport of metabolites and, of particular interest, regulation, such as in allostery (recently reviewed by Mattevi and colleagues[28]). Allostery has been observed in e.g., haemoglobin,[29] in the phosphorylase,[30] and recently in chorismate mutase.[31] The induced movements of the domains, closing the open-form upon receptor-ligand association, has long been considered to reflect an induced fit.[1,32,33] Figure 1C and D depicts this concept.

Studies of domain movements often involve comparisons of crystal structures, with the domains in the open and closed ligand-bound configurations.[34] While the number of cases where both types of configurations have been crystallized is growing rapidly, they represent only a limited repertoire of the domain movements. Hence, approaches to predicting and modeling ensembles of likely domain motions are immensely valuable. Nevertheless, the task of modeling complex concerted rearrangements of domains at the atomic level is formidable (reviewed recently by Maiorov and Abagyan[35]). Despite this difficulty, progress has been made and some small-to-medium scale movements[36] and large scale rearrangements of domains,[35] have been sampled and reproduced. Such techniques are useful for automated identification of interdomain linkers.

The frequent occurrence of domain movements suggests that in seeking to predict docked bound configurations of a ligand with a receptor, they should be taken into account. Approaches carrying out docking between two molecules where the molecules are assumed to be rigid will be successful only in cases where the movements are relatively small, i.e., within the thresholds allowed in the docking procedures. Here we present a novel, computer-vision and robotics based method for automatically docking two molecules, allowing hinge-bending movements. In particular, we show the insight into the different conformers arising by the induced fit that this approach provides. Movements are allowed in either the ligand or the receptor molecules. If both open and closed conformations are available, then the hinge location can be specified via their comparisons. Alternatively, a wide range of potential hinge locations can be examined, comparing the conformational isomers, seeking the more likely chain linkers for the positioning of the hinge.

Previous docking techniques have been able to allow induced hinge flexibility within relatively small ligands (e.g., drug molecules).[37–47] Partial flexibility is enabled in protein receptors only by a few workers. Leach 1994[48] enables side-chain flexibility and Jones et al., 1995[49] allow partial flexibility of hydrogen bonding groups. None of the currently available approaches enables full scale domain movements, although theoretically it seems possible. Our approach allows hinge induced domain/subdomain/group of atoms motions to exist in either variable sized ligands[50,51] or in diverse sized receptors (see Fig. 1). Furthermore, we are able to incorporate several molecular hinges simultaneously. The ability to incorporate hinges in large molecules, allows us to model hinge induced allosteric effects in protein receptors. Putting hinges in ligands, reflects some of the conformational flexibility in the smaller molecules, between their more rigid (e.g., rings) subparts. We model full 3-D rotation at the hinge(s) or at any point in space and restrict the point rotations to covalent bond rotations if these are required. By enabling full point rotations we can model rotations taking place on several nearby bonds such as in the case of inter-domain movements. Such capabilities stem from the fact that the method is derived from computer vision and robotics.[52] Addressing local conformational changes is essential if we are to achieve correct docked solutions which would be of practical value. Nevertheless, while the aforementioned techniques sample the configurations locally, treat side-chain rotations, or allow flexibility only in the smaller ligands, they do not allow full scale domain movements. In that sense, the method presented here nicely complements existing methodologies. In particular, integrating the hinge-bending induced fit docking method presented here with a subsequent finer sampling of the local surface changes, may prove to be an efficient and practical tool kit for the researcher. Alternatively, it may also be combined with hinge bending location predicting algorithms, limiting the scope of the search and a priori focusing on the more flexible regions.

This work addresses the geometrical aspect of the docking. Because association between two molecules requires (partial, though acceptable) fit of their molecular surfaces, we seek optimally matched patches of surfaces. On the computational side, the problem of docking by allowing several parts of the molecule to move simultaneously with respect to each other is exceedingly complex. In general, there are two approaches to handling deterministic, geometric based docking. The first approach (e.g., Desjarlais et al.[37]) is to dock each of the structural parts separately. Subsequently, the separately docked conformations are screened, seeking consistent ones where the ligand parts do not inter-penetrate each other and concomittently are correctly joined at the hinge. In the second approach (e.g., Leach and Kuntz[38] and Mizutani et al.[41]), one first docks the most informative rigid part of the ligand. Next, a conformational space search is performed for the second and consecutively additional parts. Both approaches essentially view hinge-bending docking as a rigid docking. A major drawback of the first is that a priori it does not utilize an essential piece of information, namely, the location of the hinge. The second approach suffers from an inherent very time-consuming limitation, always associated with grid-based conformational sampling. We incorporate both the simple rigid subpart matching and the global consistency checks as an integral part of the recognition process. Matching evidence is collected simultaneously from all parts of the molecule, regardless of the parts sizes. We exploit the fact that the different parts belong to the same molecule and share common hinges. Furthermore, there is no dependency on the order at which the parts are matched. In addition, hinge induced conformational flexibility is allowed in either the ligand or the receptor molecules.

We have successfully applied our algorithm to a number of bound and unbound molecular configurations achieving fast matching (recognition) times of their surfaces, for both rigid and flexible docking.[53] The atom coordinates considered as input to our algorithm have been determined by X-ray crystallography and by nuclear magnetic resonance (NMR) (Brookhaven Protein Data Bank (PDB)).[54] The location of the hinge has been specified employing flexibility considerations, e.g., via a comparison of similar structures in different, i.e., open and closed conformations picked from the PDB. In order to verify our algorithm, we have first investigated bound, i.e., complexed molecular structures. We have applied our method to five complexes, the HIV-1 protease complexed with the U-75875 inhibitor; the dihydrofolate reductase complexed with methotrexate, and separately with NADPH; lactate dehydrogenase complexed with NAD-lactate; and a Fab fragment of an IgG antibody complexed with a peptide antigen (crystallized as residues 69–87 of myohemerythin). For each case, flexible docking was carried out by allowing hinge-bending in the ligand molecules. Applying our method to molecules extracted from bound configurations, we reproduce the binding mode which is in agreement with the experimental observation. Because the ligand and the receptor have been extracted from these types of complexes, the 'correct' geometrical solutions are those with rotations and translations close to zero. Indeed, the binding modes we have obtained have small root mean square (RMS) distance, as compared with the native crystal structures. The average RMS of a correct solution is 1.4Å and the average run-time for each complex is around 1 min (the execution of these cases have been conducted on a SGI-Challenge R8000 machine). The correct bound configurations typically rank high within our obtained binding modes (usually within the top three). In addition, geometrically well-fitting alternate binding modes have been generated as well for all cases tested. For further details see Sandak et al.[50,51] and Sandak.[53]

The unique capabilities of this methodology enable its application to obtain the extent and type of hinge-bending rotations. If the locations of the potential hinges, or the more flexible regions are a priori unknown, a large number of potential hinge locations need to be examined. The speed of our algorithm straightforwardly permits such repeated applications. By automatically allowing the molecules to attain their optimally matching configurations, the swiveling motion(s) introduced at the hinge-points is obtained. Hence, by applying this method to a pair of molecules, and systematically positioning a hinge either in the receptor, or in the ligand; by applying it to a single hinge versus two (or more) hinges; and by applying it to successive hinge locations, we obtain the angular rotations for each substructural part, for that hinge-choice. Because overall the final docked configurations may be similar in outline, we compile an ensemble of different angular rotations and translations, culminating in the different ways in which a final complexed structure can be obtained from a free unbound one. This allows an examination of the different substructural-part motions which may be involved. In particular, since our algorithm allows multiple hinges, we can uniquely simulate the effect of simultaneous rotations, as compared to the motions necessary at a single hinge to obtain the same docked configuration.

Here we apply this approach to the calmodulin (CaM) receptor and its M13 ligand. In all cases acceptable docked configurations have been achieved. As expected, depending on the location and number of hinges, different movements of the substructural parts are required to attain the optimal surface matching. We analyze the different rotational and translational movements, bending, rotating, and distorting the backbone. Hence, by using such an approach, we see the many ways in which a given open conformation can close on its respective ligand, and vice versa, to achieve the optimal matching of the molecular surfaces. By modeling the observed substructural-part rotations, the feasibility and favorability of such conformations can be enhanced, via the choice of particular side-chains.

MATERIALS AND METHODS

To simplify the exposition of the method, we describe it for the single hinge case where the flexibility is allowed in the ligand molecule. Although the method is described for flexible ligands, the roles of the ligands and the receptors can be interchanged, since the mathematical problem is symmetrical. At the end of this section we describe the enhancements carried out for multiple hinges.

A full 3-D rotation is allowed at the hinge. This geometric model is more general than the one with a single rotatable bond, since such a rotation has only one degree of freedom around a known axis. It allows a reasonable approximation of the hinge region of large molecular domains, and of the case of few consecutive (or, nearby) rotatable bonds. Chemical verification is optionally carried out obtaining solutions restricted to bond rotations at the hinge(s).

Our approach is based on a voting scheme for finding the most suitable ligands (out of a library of ligands), and for the respective transformations for docking each of their parts onto a target receptor. Initially, both the ligands and the receptor are represented as 3-D sets of interest points, describing their molecular surfaces. Below we briefly outline the molecular representation, followed by an overview of the method. We next detail its two phases, the preprocessing and the recognition phases. Its application to the different types of cases is given in the Results section.

Molecular Surface Representation

As our method is general and is not restricted to a particular molecular surface representation, our choices of surface descriptions further facilitate handling flexibility in either of the molecules. We have experimented with two types of surface representations, that of Kuntz et al.,[55] and that of Lin et al.[56,57] Both surface descriptions are derived from the solvent accessible surface generated by Connolly[58,59] as defined by Richards.[60] Both surface descriptions have been employed, and results from both are given below.

The attractive surface description devised by Kuntz is carried out to detect and construct the receptor's invaginations and cavities, into which the ligand molecule is docked. First, the molecular surface is calculated. This is done by employing Connolly's MS program, which generates the solvent accessible surface. The surface of the molecule is represented as sets of points, the reentrant (concave) and contact (convex) points, obtained by rolling a water probe ball having a radius of 1.4Å on the van der Waals spheres of the molecule. From the sets of reentrant and contact points, a set of spheres representing the complementary negative image of the receptor, is generated. Each sphere touches the molecular surface at two points and has its center on the surface normal of the first point. The number of spheres generated for each point is reduced so that only one sphere is retained per surface atom. The spheres are clustered such that overlapping spheres reside in the same cluster. The distinct clusters represent the various surface invaginations of different sizes. The largest cluster of sphere centers in the receptor, is the one on which the studies concentrate. For the ligand, the atom centers are employed.

The surface representation of Lin describes the surface in terms of accurately and sparsely spaced points, placed at key locations of the molecule. This representation is independent of the initial density of dot points used to describe the surface by Connolly. The density of the points is modest, while still adequately describing the shape and important characteristics of the surface, such as surface normals and area associated with each point (not employed by our algorithm). Critical points nick-named *cap*, *pit* and *belt* are computed for every convex, concave, and saddle regions of the surface, respectively. This is achieved by projecting the gravity center of points in the region onto the surface.

Lin's surface description differs from Kuntz's by the notion that its points cover the whole surface and are not constrained to the surface invaginations and cavities. Secondly, the points are positioned on the solvent accessible surface, rather than placed in the invaginations of the receptor. This facilitates the flexible docking of molecules enabling hinge induced domain movements in receptors. Kuntz assumes that the representation of the negative image of the receptor is a fixed 3-D spatial structure. This is not the case when accounting for a flexible receptor, as the structure of its binding site(s) may change upon docking. On the other hand, however, for docking flexible ligands, Kuntz's representation has been found preferable, as it facilitates this goal by considering only the relevant parts of the receptor, onto which the hinge bent ligand may be docked.

When using Kuntz's description, the clustered sphere representation, i.e., the coordinates of the sphere centers are considered as the receptor interest point set. The atom coordinates are considered as the interest point sets of the ligands. Lin's cap point representation is used for defining the receptor's interest point set and the pits as the ligand's interest point set. These sets of interest points describe the surfaces of the molecules and are handled by our geometric matching algorithm. That is, the matching is done between these point-sets.

Overview of the Algorithm

As the ligands may undergo translations and rotations of their parts in order to dock to a receptor, the ligands' information is stored in a database (a look-up table) invariant to this type of transformations. This procedure is carried out in the *preprocessing* phase of the algorithm. The position of the hinge in each stored ligand is determined in this phase by manually employing flexibility considerations. This phase can be executed off-line as it is independent of the receptor molecule. A receptor structure is presented to the system in the *recognition* phase of the algorithm. If a ligand, previously stored in the database, has a matching interest point configuration (i.e., a surface patch) which matches the receptor interest point configuration, a match is scored. That is, a vote is cast for this ligand, together with the computed location of its hinge within the docking site (matching stage). This location is derived from the transformation between the corresponding recep-

Preprocessing

(1) Construct the ligand's surface (interest point set).
(2) Specify the location of the hinge(s) in the ligand.
(3) Encode and store the ligand in the database (look-up table).

Recognition

(1) Construct the receptor's surface (interest point set).
(2) Recover candidate ligands from the database which match the receptor's surface patches.
(3) Choose significant scoring (voted for) candidate ligands and their hinge(s) locations (transformations).
(4) Verify the docked conformations obtaining geometrically acceptable solutions.
(5) Sort the docked solutions according to their 'goodness'.

Fig. 2. A general outline of the algorithm consisting of two major phases, the preprocessing and recognition phases.

tor and ligand interest point surface patches. We are interested in the highest scoring (voted for) hinge locations of candidate ligands. Note that we assume no knowledge of the hinge locations relative to the receptor. The high scoring transformations are verified obtaining geometrically acceptable solutions (verification stage). The verification is done by discarding the transformations which cause the ligands parts to collide with the receptor (collision check) and with other (self collision check). See Figure 2 for the general outline of the algorithm.

Detailed Description of the Algorithm

A detailed description of our two phase method is given below. These are divided into steps to clarify the exposition. A 2-D illustration of the two-basic phases of the algorithm is depicted in Figure 3.

Preprocessing

1. The ligand molecule (model) is represented as a set of interest points.[55–57]
2. The (known) hinge position is picked as the origin of a 3-D Cartesian coordinate frame, which will be called the ligand frame. The orientation of this frame is set arbitrarily.
3. For each non-ordered non-collinear triplet of interest points in each ligand part, we define a unique triplet based Cartesian frame, one for each triplet point. An internal cyclic order is defined, so that the frames are positioned as follows: The origin of each frame is defined at the respective triplet point, the x-axis as the line from the point to the neighboring point, the z-axis as the normal to the triangle plane obtained by the cross product of the x-axis with the second triangle side emanating from the origin, and the y-axis as the cross product of the x and z axes. These are the triplet frames. The non-ordered triplet of the triangle side lengths serves as an address to a look-up table. The information stored at this entry is the ligand identification, part number, and the transformations between the triplet frames and the ligand frame. Figure 4 illustrates this concept.

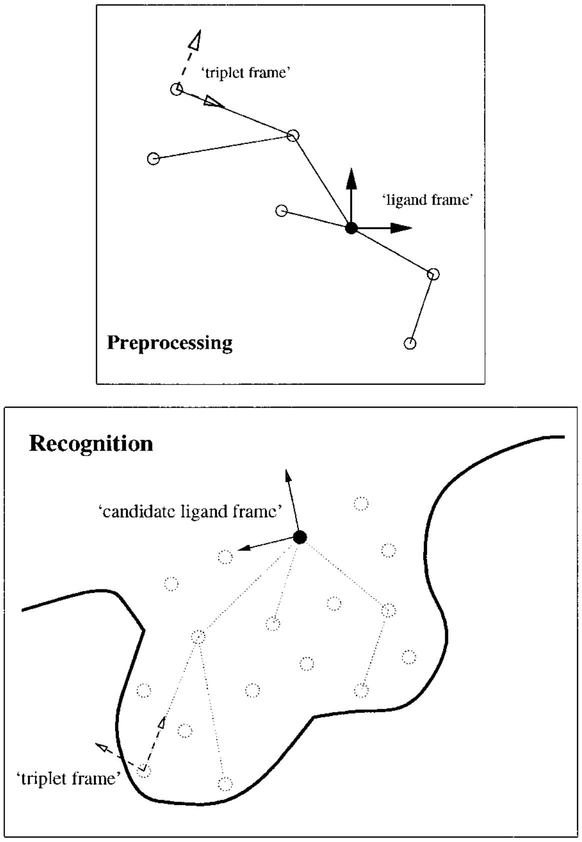

Fig. 3. A 2-D illustration of the two-basic phases of the algorithm. First phase—*preprocessing:* A flexible ligand consisting of two rigid sub-parts connected by a hinge. The small circles represent the ligands atoms; the lines represent the covalent bonds, which are disregarded by our algorithm. A Cartesian coordinate frame is defined and centered at the hinge (dark circle), referred to as the ligand frame. For each (non-ordered, non-collinear) triplet of atoms, we define three Cartesian triplet frames (see also Fig. 4). Here, we depict only a single triplet frame. We compute and store the transformations between the triplet frames and the ligand frame in the look-up table. If there are several ligands which will be docked to the receptor, the ligand's identification is stored in the table as well. Second phase—*recognition:* The flexible ligand presented above is docked onto the receptor surface. In this illustration, the small circles represent Kuntz's sphere centers of the negative image of the receptor wherein the ligands, previously stored in the look-up table. The lines represent the covalent bonds of the docked ligand, which are disregarded by the algorithm. The two rigid parts having matching sections with the receptor structure incorporate a candidate ligand frame at the same location (dark circle), although at two different orientations. The candidate ligand frames are calculated by applying the prerecorded transformations of the triplet frames of the ligands to all the triplet frames of the receptor.

Recognition

1. The molecular structure of the receptor is represented by its set of interest points which describe its surface.[55–57]

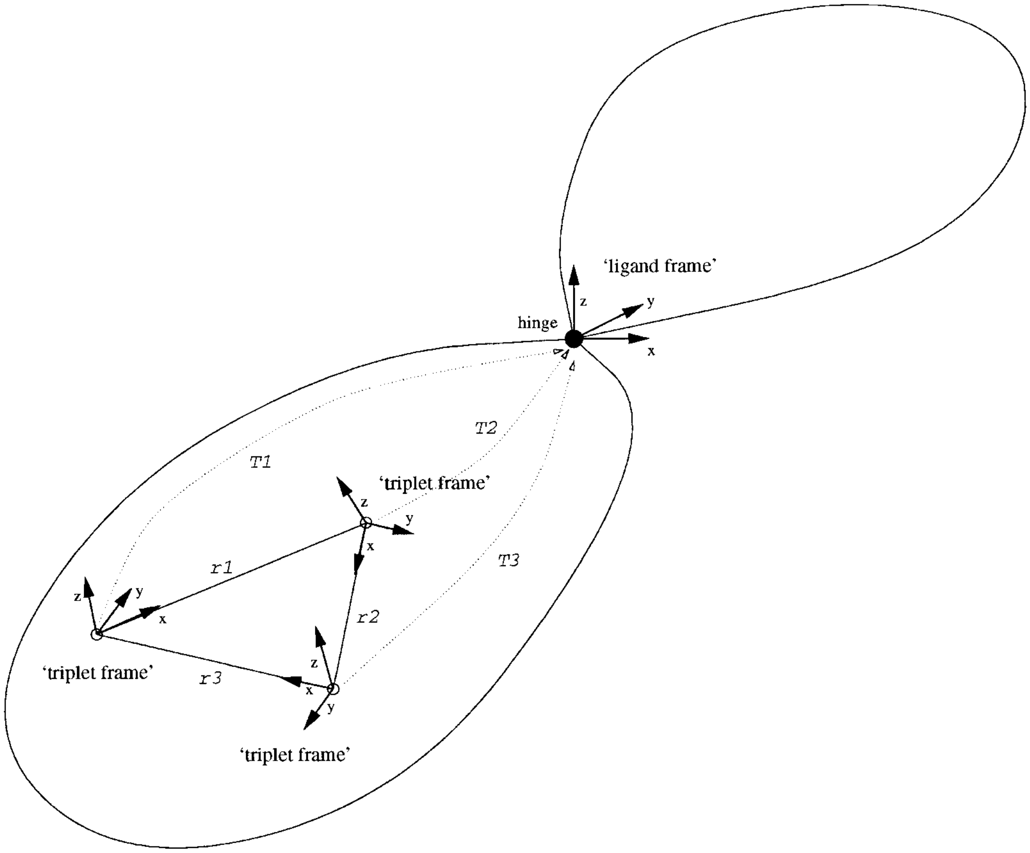

Fig. 4. A 3-D illustration of handling a triplet of points in respect to steps 2 and 3 of the preprocessing phase. A ligand frame is positioned at the hinge. Three triplet frames are positioned at triplet of interest points (this is carried out for every non-ordered non-collinear triplets of interest points of every ligand part). The ligand identification, part number and the transformations between the triplet frames and the ligand frame (marked as *T*1, *T*2 and *T*3) are stored in the look-up table. The address for this record of information is defined as the lengths of the triangle sides (marked as *r*1, *r*2 and *r*3).

2. All non-collinear triplets of the interest points are considered in the matching stage. For each such triplet, the triplet based Cartesian frames are calculated, as explained above (see also Fig. 4). The lengths of the triangle sides are computed. Note that this calculation is invariant under rotation and translation, thus (almost) congruent ligand triangles should have similar values. The look-up table is addressed according to the computed triplet of triangle side distances. For each ligand-record present at that entry in the table, the candidate ligand frame is computed by applying the prerecorded transformation at that entry to the receptor triplet frame. The origin of the candidate ligand frame is the candidate hinge location. A vote is cast for the identity of the ligand molecule together with the location and orientation of the candidate ligand frame.
3. The accumulator of votes is searched for high scoring pairs of (ligand, hinge location). Hinges receiving a large number of votes are picked. The high scoring hinge locations are determined according to the minimal percentage value of the number of votes received by the highest scoring hinge location (referred to as the voting threshold).
4. A potential match implies existence of complementarity between receptor-ligand surface patches. However, other regions of the two molecules may inter-penetrate (collide with) each other. The transformations for each part are optionally pruned by clustering them according to their rotations, and are applied to all atoms of each part. The two molecules are assumed to inter-penetrate if the distance between a transformed ligand atom and a receptor atom is smaller than the sum of their respective van der Waals radii subtracted by a proximity threshold (1.75Å). The same criterion is applied to the two ligand parts. To speed up the collision check we reduce the size of the space checked. The receptor molecule is divided into eight segments (octants) sharing the geometric center of the molecule. The collision check between a ligand atom and a receptor atom is conducted only in the appropriate receptor's segment (octant) which consists of the respective

receptor's atoms. Transformations yielding interpenetration (self-collision) between the two parts of the ligand or which result in the penetration of a ligand part into the receptor, are discarded. Optionally, a chemical verification is carried out for filtering configurations which are restricted to bond rotations.

5. The goodness of a solution is evaluated by employing a score which is based on the percentage of the ligand's van der Waals spheres which are in contact with the receptor spheres. We refer to this score as the contact percentage. A ligand sphere is assumed to be in contact with a receptor sphere if the distance between the ligand atom and a receptor atom is smaller than the sum of their respective van der Walls radii plus a proximity threshold (1Å). The acceptable docked solutions are ranked according to the contact percentage.

In this algorithm we have exploited the fact that both parts of the molecule share the same hinge. This has been done by locating the origin of the reference frame of the ligand at the hinge. In such a way, both parts contribute votes to a reference frame at the same location, although at different orientations (see Fig. 3). In particular, it is important to note that by picking up votes from both of its molecular parts, a ligand, which might otherwise have a small matching surface area with the receptor structure in each of its parts, can still score high. Hence, although each of the individual parts of the ligand can receive an insignificant score, by combining the votes from both parts an overall acceptable match can still be automatically detected.

Obviously, this flexible docking algorithm can handle the rigid docking as a particular case. For the rigid-body docking, the ligand reference frame is located arbitrarily.

Although here our algorithm is described for molecules with a single hinge, it is extended to handle multiple hinges by the following enhancement. Instead of having one ligand frame we define *multiple* ligand frames, each centered at a different hinge. In the preprocessing stage, for each ligand triplet in a single part, we encode the transformations of its triplet frame to all ligand frames located on that part. Thus, e.g., on a part with two hinges, two sets of transformations will be stored for each triplet, while on a part with one hinge only, one set of transformation will be stored. The recognition phase remains unchanged, except that each receptor triplet will vote for as many frames as the number of different transformations stored in its table entry.

For programming details on the various stages of the algorithm as well as analysis of the complexity and run times on typical examples see Sandak.[53]

RESULTS

Previously, we have allowed hinge-bending motions in ligands of diverse molecular complexes.[50,51,53] Here we study the potential binding modes of a peptide ligand and the calmodulin receptor (CaM), allowing flexibility in either the ligand or the receptor. Furthermore, we investigate both the bound (complexed) and the unbound structures of the calmodulin molecule. CaM is a 148-residue protein (2259 atoms), present in eukaryotic cells. The CaM molecule is responsible for regulating the activity of a large number of proteins, activating more than 20 different enzymes. Serving as a primary receptor for intra-cellular calcium (Ca^{2+}), it is involved in cellular Ca^{2+} dependent signaling pathways. That is, the CaM activates numerous proteins (enzymes) in response to changes in intra-cellular Ca^{2+} concentrations. The subsequent step in the signal transduction mechanism, requires the binding of calcium-bound calmodulin (Ca^{2+}-CaM) to an acceptor protein. The binding of synthetic peptides corresponding to the CaM-binding domains of relevant target proteins, enzymes, peptide drugs, and toxins has been investigated. The deployment of these peptides contributes to the knowledge of the physical interactions involved in the Ca^{2+}-CaM-protein induced signal mechanism. The extraordinary flexibility of CaM is manifested in the binding of Ca^{2+}-CaM to peptides, resulting in a considerably more compact conformation than that of the unbound form.[17–19] The two CaM domains, the amino- and carboxyl-termini, engulf and clamp the bound peptide, in a domain-motion which is predominantly hinge.[13] This hinge motion, upon the binding to the ligand, involves the disruption of the long central helix which connects the two domains. The helix is split to two helices which are connected by a long flexible loop (residues 74 to 82[17]). The two domains remain essentially unchanged. The ligand we use in our docking investigations, is a 26-residue synthetic peptide (441 atoms), referred to as M13.

The solutions of the bound and unbound cases are ranked according to the size of the contact area(s) of the docked ligand and receptor molecules, which is referred to as the contact percentage scoring (see definition in step number 5 of the Recognition-phase of the Materials and Methods section). We define the term *best* or *correct* solution, as the solution having the lowest RMS distance values of the molecule-parts extracted from bound complexes. This definition serves as a measurement of the performance of the algorithm when verifying it. An additional quantification of the RMS distance is employed. It is defined as the RMS distance of only the interface atoms of the relevant molecules. These atoms are the atoms of one molecule which are in contact with the

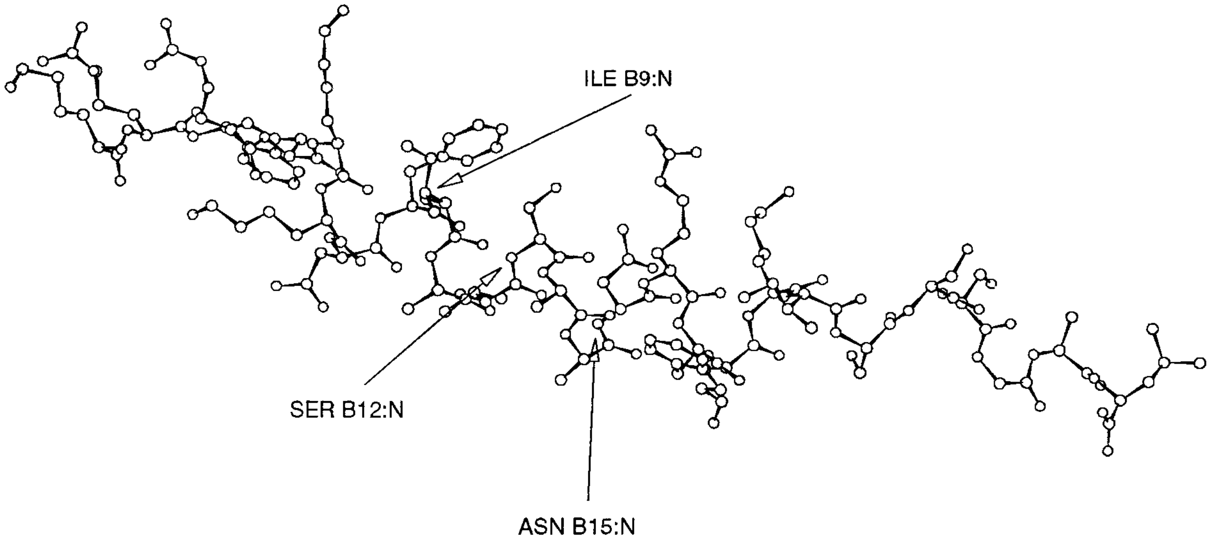

Fig. 5. The spatial structure of the M13 peptide, showing the positioning of the hinges. The hinge pointed at by the middle arrow (positioned at the N atom of Ser[12]), divides the peptide to two parts. The molecule is divided to three parts by the hinges pointed at by the external arrows (placed at the backbone N atoms of Ile[9] and Asn[15]).

atoms of the second. The executions of the program have been conducted on the SGI-R10000 machine.

Hinge-Bending Flexibility in the Ligand

Our first investigation focuses on binding-modes generated by allowing hinge-bending motions in the ligand, the M13 peptide. NMR, as well as computational studies, have indicated that all peptides, unless highly constrained by covalent cyclization, demonstrate an ensemble of conformations.[61] Furthermore, the protein receptor strongly influences the binding conformation of the flexible ligand. We have partitioned the M13 ligand both to two and to three submolecular parts by positioning either one hinge or two hinges, respectively, along the molecule (see Figure 5).

Single hinge case

For the single hinge case, we position the hinge on the N (Nitrogen) atom of the Ser[12]. The molecule is divided to two parts, where residues 1 to 11 compose the first part, and residues 12 to 26 compose the second. Since in the bound study the M13 ligand and the Ca^{2+}-CaM receptor have been extracted from a complex,[17] the correct conformations for the complexed case are those whose translations and rotations are near zero, having small RMS distances from the native structure. The M13 peptide was successfully docked to the CaM receptor, yielding the correct solution with RMS distances of the interface atoms of 1.91Å for the first part, and 0.97Å for the second. The RMS distances are 2.53Å and 1.17Å, respectively. The translation distance of the first part is 1.32Å, and that of the second is 1.0Å. The respective angular distances are 17.2 and 4.77 degrees.† Prior to the verification stage, i.e., prior to filtering docked solutions having inter-molecular and inter-part (intramolecular) penetration, the location of the hinge in the correct solution is in the first 0.64% highest scoring hinge locations among 77,424 voted for candidate hinge-locations (the default voting threshold parameter is 20%). At the end of the run, the correct solution is the second top scoring solution with respect to the contact percentage. The contact percent calculated for the ligand in the native crystal structure is 30%. The contact percent calculated for the ligand conformation obtained as the best solution is also 30%. That of the top scoring solution is 35%. The binding mode of the crystallographically correct conformation generated by our program is depicted in Figure 6. The CaM receptor surface is described by 440 points (constituting its negative image[55]). The M13 peptide is represented by 210 points which are its atomic coordinates. The breakdown of the execution times is as follows: The core of the algorithm, i.e., the surfaces-matching stage between the two molecules, takes 0.43 min. The collision (i.e., the inter-penetration) check between the peptide and the receptor, and the self-collision (self-penetration) check between the pep-

---

†The translation distance is the l_2-norm of the transformation's translation vector ($\sqrt{x^2 + y^2 + z^2}$), and the angular distance is the evaluation of the term which corresponds to the rotation angle around an equivalent axis $arccos \frac{tr(R)-1}{2}$ where R is the transformation's rotation matrix and tr is its trace).

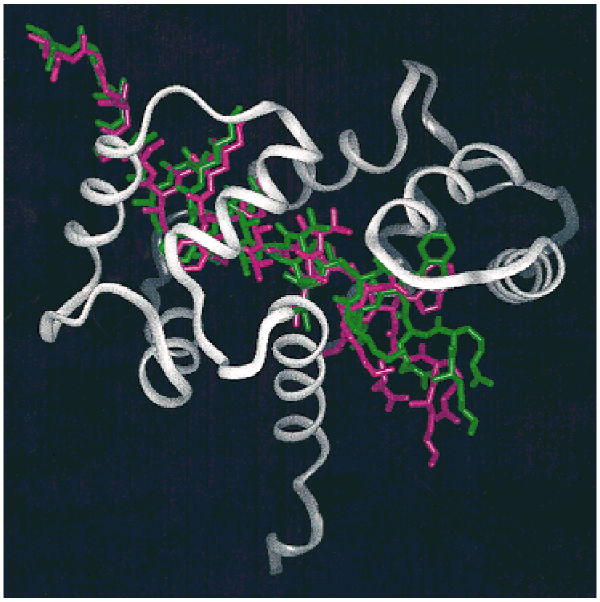

Fig. 6. The complexed (bound) case where the hinge is in the M13 ligand, and the calmodulin (CaM) receptor is rigid (2BBM file of the PDB). The hinge is positioned at the N atom of Ser[12] (see Fig. 5). The calcium-bound CaM receptor molecule of the same NMR complex, is drawn as a white ribbon. The CaM domain engulfing the M13 (stick models) face the viewer. The correct docked conformation of the M13 peptide predicted by our program, is represented as a green stick model. The correct conformation is plotted against the magenta stick model of the original M13 NMR structure.

tide's parts take 1.42 min, and under a second, respectively. The geometric feasibility checks handle 1164 atoms of the CaM receptor and 210 atoms of the M13 ligand (103 atoms in the first part and 108 atoms in the second; the hinge is joined by the two parts).

Double hinge case

We also obtain good results for the double hinge case. Here, we position the hinges on the backbone N atoms of the Ile[9] and Asn[15], dividing the molecule into three parts (see Fig. 5). The first part consists of residues 1 to 8, the second part of residues 9 to 14, and the third part comprises residues 15 to 26. The best solution obtained for this case has an RMS distance of 2.03Å for the first part, 0.98Å for the second part, and 1.03Å for the third. The RMS distances of interface atoms are 2.12Å, 1.0Å and 0.83Å for the three parts. The best solution has a translation distance of 1.81Å and an angular distance of 10.7 degrees for the first part, a translation distance of 1.15Å and an angular distance of 19.61 degrees for the second, and a translation distance of 1.02Å and an angular distance of 4.77 degrees for the third part. The best solution is generated from two hinge locations which are the first 1.51% and 2.97% highest scoring hinge locations, among 19,310 and 19,103 voted for candidates respectively, prior to the verification (collisions-checks) stage. At the end of the run, the best solution is ranked as the third top scoring solution with respect to the contact percentage. The contact percent calculated for the ligand conformation of the best solution is 27%. The increase in the number of hinges does not result in the growth of the execution times. On the contrary, there is even a significant decrease in the running times we have measured. This is explained by the fact that there are less triplets of interest points which are handled per part. The measured running times are: 0.37 min for the surfaces-matching stage, 1.55 min for the inter-penetration check, and under one second for the self-penetration check (87, 37 and 92 atoms in the ligand's subparts, respectively). This demonstrates the efficiency of our method, and its potential in handling multiple hinges cases in tractable time. In addition to the best solutions described above, we have obtained good-fitting predictive binding modes (results not shown).

Hinge-Bending Flexibility in the Receptor

We have explored the binding modes of the bound and unbound CaM receptor to the M13 peptide, enabling hinge movements of its two domains.

Bound molecular structures

For verifying the algorithm, in separate runs, we have positioned the hinge on a few backbone N atoms along part of the central helix of the CaM molecule. This region is known to be distorted upon binding. The distance between the atoms found at the ends of the flexible loop in the CaM is almost equal to that of the same atoms when they are in the helical form in the unbound structure (6% difference). Hence, the 3-D rotational movement at the hinge allowed by our algorithm, gives a close approximation to the total movements observed. These movements are a composite of both the distortion of a portion of the helix to a loop and the hinge rotation. Regardless of the positioning of the hinge consistent results are obtained, verifying our program (results not shown). Figure 7 depicts the CaM macromolecule of the complexed structure, and the positioning of a hinge at the backbone N atom of Lys[75]. The CaM is divided to two parts. The first part consists of residues 1 to 74, and the second part is composed of residues 75 to 148.

Figure 8 displays the binding mode of the correct conformation obtained for this partitioning of the complexed CaM. The translation distance of the first part obtained by the best solution is 3.14Å, and the angular distance is 27.89 degrees. A translation distance of 2.79Å and an angular distance of 15.23 degrees are obtained for the second part. Owing to their large sizes, the extent of the hinge-induced movement of the domains is not uniform. When analyzing the angular motions involved, we observe relatively small movements of the atoms in the

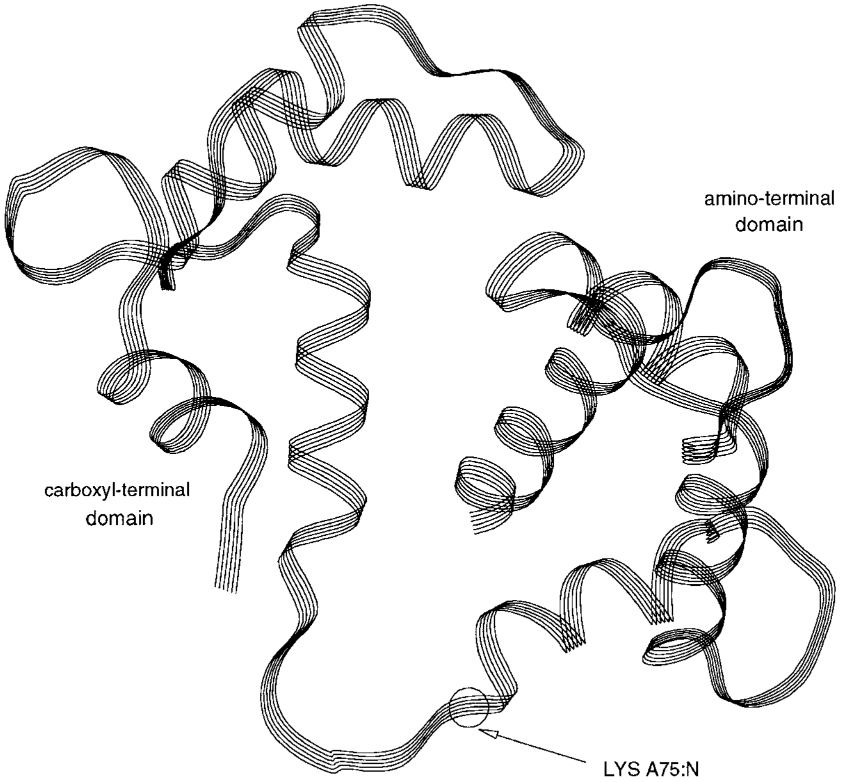

Fig. 7. The ribbon structure of the complexed calcium-bound CaM receptor, showing the hinge positioned at the backbone N atom of Lys^{75}. The hinge pointed at by an arrow divides the molecule into two large-parts, comprising the amino- and carboxyl-terminal domains, respectively.

vicinity of the hinge, and a larger movement farther away, thus obtaining RMS distance values of 8.91Å and 7.78Å for the domain parts. Utilizing the RMS distances of the atoms in the intermolecular interfaces (the ligand atoms which are in contact with the receptor's), we obtain the values of 7.27Å and 3.51Å for the first and second part, respectively. As the first domain part has a computed angular distance of 28 degrees, its atoms (and interface atoms in particular) have rotated, affecting the RMS distance calculations. The hinge location of the best solution is in the first 9.07% highest scoring hinge locations among 124,035 voted for candidate hinge locations. The best solution is ranked in the 450th position at the end of the run among 10,091 acceptable solutions, which is calculated in the first 4.5% highest scoring solutions with contact percentage of one percent. The contact percentage of the native CaM molecule with its complexed ligand is 5%. Here we employ the surface representation of Lin et al.[56,57] describing the surface in terms of points representing the convex, concave and saddle regions of the surface. In this investigation, we use convex points (526 interest points–265 points in the first part and 261 in the second) to represent the calcium-bound CaM molecule, and concave points (284 interest points) to represent the M13 peptide. The execution times recorded for this run are 0.35 min for the surfaces-matching stage between the CaM and M13, 12.13 min for the inter-molecular penetration check between the receptor and the ligand, and under one second for the self-penetration check between the two domain parts. This geometric verification stage (the inter-molecular penetration and self-penetration checks) handles 1,164 atoms of the CaM receptor, (572 atoms in the first part and 593 atoms in the second), and 210 atoms of the M13 ligand. Here the penetration checks take much longer than those of the two previous cases where the docked molecule has been the ligand (210 atoms), as the receptor is significantly larger (1,164 atoms).

Unbound molecular structures

In the unbound case of the CaM receptor, good-fitting binding modes have been obtained, some with close proximity to the binding mode of the native NMR structure. As done in the previous cases, their goodness is quantified by calculating the percent of the CaM van der Waals spheres which are in contact with those of the M13 peptide. In this case, we have obtained the best solution within the top 0.5% solutions (position 1,562 among 307,982 geometrically acceptable binding modes). The RMS values of the atoms of the interface is 4.32Å and 5.72Å for the first

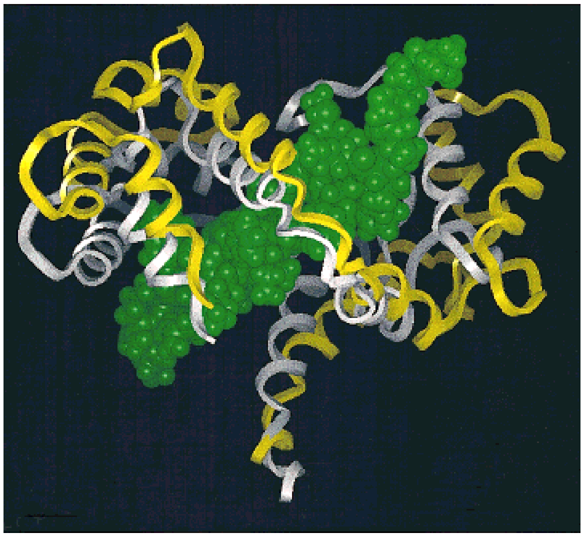

Fig. 8. The complexed (bound) case where the hinge is in the CaM receptor, and the ligand is rigid (2BBM file of the PDB). The hinge is positioned at the backbone N atom of Lys^{75} (see Fig. 7). The correct docked conformation of the complexed CaM receptor, predicted by our program, is represented as a yellow ribbon. The correct solution is plotted against the white ribbon representation of the CaM molecule, extracted from the NMR structure. The M13 peptide of the same NMR complex is drawn in its van der Waals green solid sphere representation. The disrupted central helix of the CaM molecule face the viewer.

and second domain, respectively. The native NMR CaM has a contact percentage of 5% with its complexed M13 ligand. The one calculated for the best solution is about one percent. Figure 9 depicts one of the top scoring potential binding modes of the unbound CaM having a 9% contact with the M13 peptide. The hinge is positioned on the backbone N atom of Asp^{78}. In its docked configuration, the unbound CaM clamps the M13 peptide in a reverse direction as compared to the complexed case discussed above. This is feasible owing to the highly pseudo-two fold symmetry of the CaM domains[18]. Figure 9 also illustrates how the flexibility we allow at the hinge, enables the two parts (composed of residues 1–77 and residues 78–148), to appreciably change the angle between them, resulting in a compact binding to the M13 ligand. Here, the hinge is positioned at an atom belonging to the helical region in the unbound structure. In the complexed conformation, this atom belongs to the flexible loop. This unwinding and expansion of a portion of the central helix to form the flexible loop, causes the CaM structure to tend to an ellipsoid, which seems to result in higher binding affinity. The model of a single hinge is not sufficient, since if this portion of the central helix is left intact, which is the case in our docking investigations, the obtained CaM bound conformation is less ellipsoidal. In spite of this difficulty, we have still been able to bind the M13 ligand to the CaM receptor which reflects our ability to detect imperfectly matched molecular conformations. Employing the molecular representation described for the previous bound case, the number of interest points representing the unbound CaM receptor is 465 (227 in the first part, and 238 in the second part), and 284 interest points representing the M13 ligand. These point sets are handled at the surface-matching stage of our algorithm. The execution times recorded for this run are: 0.28 min for the surfaces-matching stage between the two molecules; the geometric feasibility check, verifying the candidate transformations obtained from the former stage takes 26.02 min for the inter-penetration check between the receptor and the ligand, and under a second for the self-penetration check between the receptor's large parts. The latter two checks handle 1,132 atoms of the CaM receptor (572 atoms in the first part and 561 atoms in the second part), and 210 atoms of the M13 ligand.

DISCUSSION

Here we have presented a general approach to docking a ligand molecule (peptide) onto a protein receptor, allowing hinge bending motions of domains/subdomains/group of atoms, in either the ligand or the receptor, for both bound and unbound cases. This ability allows reproducing induced fit in molecular associations. Since molecules are inherently flexible entities, and movements of structural parts frequently recur during the binding of the ligand to the receptor, being able to automatically allow this type of motion is an important attribute of a docking (matching) tool. This is clearly of practical relevance in database searches and screening, docking a variety of potential drugs to a receptor of interest. Such a capability widens the scope of the search significantly. It further enables an insight into the ensemble of conformational isomers which may be found in the complexed form.

Using this method, we have successfully reproduced a number of known hinge-bent (and rigid) docked configurations of the associating molecules, obtaining binding modes which are consistent with experimental results (detailed in Sandak et al.[50,51] and Sandak[53]). The alternate docked configurations we predict, i.e., the additional good fitting binding modes between the two associating molecules, provide predictions for plausible molecular interactions and conformations. The docked configurations may also provide initial guesses for subsequent detailed conformational space sampling around the hinge-bent configurations or for molecular dynamics simulations.[62] Refinement energy minimization steps can follow to optimize the configurations.

There are some shortcomings to our method: (*i*) Here, the goodness of the docking is evaluated by quantifying the percent of the ligand van der Waals spheres which are in contact with those of the receptor, and vice versa. The scoring function serves

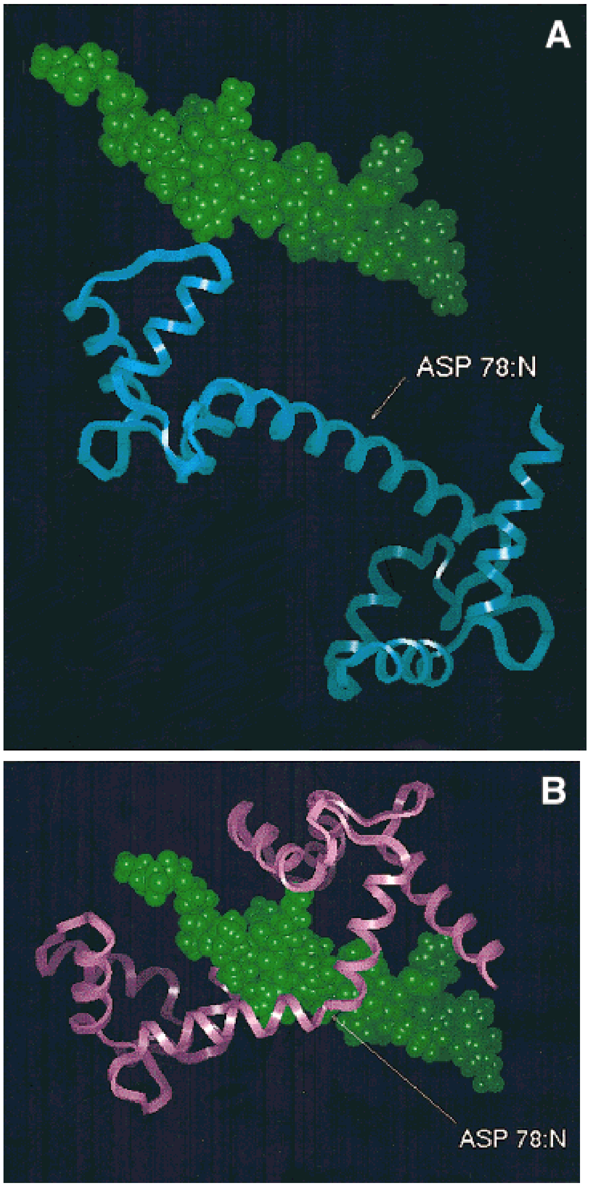

Fig. 9. The unbound case where the hinge is in the unbound CaM receptor, and the M13 ligand is rigid. (**A**) We position the hinge in the N atom of residue Asp^{78} of CaM, pointed at by an arrow. The native unbound structure of the CaM molecule is represented as a ribbon. The M13 peptide is represented in van der Waals solid spheres. The atom coordinates of the crystallized CaM are taken from the 1CLL file of the PDB. The M13 ligand is extracted from an NMR-determined complex (atom coordinates taken from the 2BBM file of the PDB). (**B**) A predictive binding mode of the CaM generated by our method, is plotted as a ribbon engulfing and clamping the M13 peptide.

as the filter for obtaining the correct solutions and other potential docking candidates, usually ranking the correct solution among the high scoring ones. Ours, is a simplified scoring function which is based on geometric attributes for discarding bad steric contacts (van der Waals overlaps) and considering good ones. Variations of this scoring technique have been employed by many workers (e.g., DesJarlais et al.,[37] Leach and Kuntz,[38] and Shoichet et al.[63]) all trying to tackle the problem of scoring validity and robustness (see Shoichet and Kuntz[64] for comparisons of scoring techniques). Hence, the problem of determining the optimal scoring function that will be rigorous, accurate and valid on the one hand and computationally efficient and robust on the other, is still an open problem. This problem is especially manifested in the cases where the flexibility is allowed in the receptor molecule for its binding to the ligand. The number of docked conformations is high since whole molecular surfaces are considered here. Thus, chemical matching in addition to geometrical matching and/or better scoring functions may be constructed to discriminate between favorable and unfavorable docked configurations. Due to the modularity of the algorithm and its implementation, newly devised scoring function(s) can be easily plugged-in and tested. Currently, the algorithm generates geometrically acceptable solutions sorted according to the scoring function described above. These have been shown by our trial runs to reproduce the configurations of the experimental complexes which usually rank high. Thus, the researcher employing the method can test the high scoring solutions for their biochemical feasibility. These rank within top scoring solutions (the 3 highest scoring) when docking flexible ligands, or within the top 0.5%–4.5% solutions for flexible receptor docking. (*ii*) Using Lin's[56,57] molecular surface representation when docking hinge-bent receptor to ligands has shown that the solutions obtained are less impressive than those obtained when docking flexible ligands onto rigid receptors (Kuntz's representation[55]). The former describes the whole molecular surface and the latter represents the receptor's invaginations, thus restricting the matching to potential binding sites. A denser point description of the molecular surface can be constructed using Lin's method for flexible receptor docking. It may improve the quality of the solutions for this type of docking, however it can effect the execution times. An automatic procedure that predicts potential docking regions can be employed here to restrict the matching. (*iii*) The inter-penetration checks are somewhat time consuming when considering relatively large molecules (thousands of atoms). This is manifested in the docking tests where the receptor is flexible, as compared to the cases where the flexibility is allowed in the ligand. One plausible remedy is to devise a many-body interaction collision check for reducing the complexity of this step by an order of magnitude. (*iv*) Considering the size of the molecular parts, if they are too small (e.g., a part consisting of a single bond), they may not be efficiently recognized by our method, as these subparts are described by only a small number of interest points. In such cases, entire exploration of the conformational space may be unavoidable. (*v*) We are still faced with the problem of surface flexibility, just as in the rigid body case. Molecules whose surfaces are highly variable, with extended flexible side-chains, may well interfere with our docking. Currently, this type of flexibility is handled through larger freedom in the inter-penetration. (*vi*) The data structures require significant memory space when dealing with molecular databases. Hence, alternative storage areas should be considered when extending the method for database screening.

On the other hand, there are several advantages to our method. (*i*) Our flexible docking method can handle efficiently molecules of diverse sizes.[53] The molecules we have docked, range in size from tens to thousands of atoms (23 to 2,862 atoms). In all cases, fast matching times are achieved. (*ii*) We allow flexibility is either the ligand or receptor molecules. (*iii*) The method has been implemented for a single, or for multiple hinges. (*iv*) Our experiments have shown that the execution time (i.e., matching) decreases as the number of hinges increases. (*v*) It is robust, being able to handle both high resolution crystal structures and NMR ones, as illustrated here. (*vi*) Full rotations are allowed on the hinge, which implicitly may take into account several rotatable bonds. Alternatively, one can choose to restrict the hinge motion to bond movement. (*vii*) Since all parts are docked simultaneously, voting for a joint hinge location, even if each of the parts is not well matched, the joined parts may still score high. Hence, a match which would otherwise be discarded, is kept for chemical scrutiny and further examination.

The method described here can be further enhanced to improve its robustness, efficiency and execution times. The algorithm may be extended to account for chemical filtering of the casted votes in the matching process. This may be done by labeling the 'interest points' according to chemical properties, and casting votes only for geometrically and chemically consistent matches (e.g., see Rarey et al. FlexX program[45] and Shoichet and Kuntz[65]). Such an implementation would filter out many biologically unacceptable solutions (false positives) also speeding the inter-penetration verification. Furthermore, since a rotational bond movement is more restrictive than a full 3-D hinge rotation, such a restriction may be incorporated (see, e.g., Rigoutsos et al.[66]). Such an implementation of the algorithm would enhance its performance for those special cases where rotations around single bonds is desired. This, however, would not be able to reproduce the full 3-D induced fit, inter-domain movements presented here. Parallelization of the algorithm to concurrently run on appropriate machines, is another direction for speeding up of the implementation.

This hinge bending method is appealing for domain motions in proteins and in ligands. The method can integrate into it complementary tools. These include local, smaller-scale search methods, which allow flexibility of small groups of atoms, either in small molecule drugs, or in receptor side-chains. If integrated with an automated, flexible hinge-region predictive tool, the search for likely energetically more plausible hinge locations can be restricted. Moreover, the generality of this approach allows its application toward additional structural/pattern matching problems in structural biology, such as searches for hinge-bent motifs. It further allows investigations of backbone flexibility, leading to conformational isomers. Obtaining similarly docked closed conformations by hinge-bending at different sites, demonstrates the variable routes that similar complexed configurations can arise. This illustrates the many ways evolution could have incorporated different mutations to attain similar results. Bending, rotating and distorting the backbone, and inevitably the side-chains as well, is energetically costly. Depending on the energy of the transition state, and of the closed docked conformation with the hinge at that location, the likelihood of the actual occurrence of the hinge may be estimated. An examination of the ensemble of the obtained conformers, with a gross overall similar complexed structure, yet differing in important details, may suggest which is a more favorable route for the molecule(s) to adopt. It may further serve as an alternate to a direct prediction of flexible hinge locations. Rather than examine a large number of allowed angular rotations, and analyze their energetic attractiveness, a substantially reduced, specific set would need to be scrutinized. This set comprises the angles a priori observed to bring about molecular surface complementarity. By repeatedly applying an approach such as the one presented here, either to one hinge at a time, or simultaneously to many hinge locations, one can simulate the evolution, and model the backbone to achieve the desired conformation.

ACKNOWLEDGMENTS

We thank Drs. S.L. Lin, R. Jernigan, and, in particular J. Maizel for helpful discussions and interest. The research of R. Nussinov has been sponsored by the National Cancer Institute, DHHS, under Contract N0. 1-CO-74102 with SAIC. The research of R. Nussinov and H.J. Wolfson in Israel has been supported in part by grant No. 95-0028 from the U.S.-Israel Binational Science Foundation (BSF), Jerusalem, Israel, by a grant from the Recanati Fund, by a Basic Research grant from Tel Aviv University and by the MAGNET grant, administered by the ministry of Industry and Technology. This work forms part of the Ph.D. Thesis of B. Sandak, Tel-Aviv University, and is supported in part by the Carl F. Gauss Minerva Center for Scientific Computation at the Weizmann Institute of Science.

REFERENCES

1. Dobson, C.M. Hinge-bending and folding. Nature 348:198–199, 1990.
2. Faber, H.R., Matthews, B.W. A mutant T_4 lysozyme displays five different crystal conformations. Nature 348:263–266, 1990.
3. Weber, I.T., Steitz, T.A. Structure of a complex of catabolic gene activator protein and cyclic AMP refined at 2.5Å resolution. J. Mol. Biol. 198:311–326, 1987.
4. Joseph, D., Petsko, G.A., Karplus, M. Anatomy of conformational change: Hinged "lid" motion of the trisephosphate isomerase loop. Science, 249:1425–1428, 1990.
5. Weber, C., Wilder, G., von Freyberg, B., Traber, R., Braun, W., Widmer, H., Wuthrich, K. The NMR structure of cyclosporin A bound to cyclophilin in aqueous solution. Biochemistry, 30:6563–6574, 1991.
6. Rini, J.M., Schulze-Gahmen, U., Wilson, I.A. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. Science, 255:959–965, 1992.
7. Stanfield, R.L., Fieser, T.M., Lerner, R.A., Wilson, I.A. Crystal structure of an antibody to a peptide and its complex with peptide antigen at 2.8Å. Science, 248:712–719, 1990.
8. Wedemayer, G.E., Patten, P.A., Wang, L.E., Schultz, P.G., Stevens, R.C. Structural insights into the evolution of an antibody combining site. Science, 276:1665–1669, 1997.
9. Janin, J., Wodak, S.J. Structural domains in proteins and their role in the dynamics of protein function. Prog. Biophys. Mol. Biol. 42:21–78, 1983.
10. Bennett, W.S., Huber, R. Structural and functional aspects of domain motion in proteins. CRC Crit. Rev. Biochem, 15:291, 1984.
11. Huber, R. Flexibility and rigidity, requirements of the function of proteins and protein pigment complexes. Biochem. Soc. Trans., 15:1009–1020, 1987.
12. Schultz, G.E. Domain motions in proteins. Curr. Opin. Struct. Biol., 1:883–888, 1991.
13. Gerstein, M., Lesk, A.M., Chothia, C. Structural mechanisms for domain movements in proteins. Biochemistry, 33(22):6739–6749, 1994.
14. Bennett, M.J., Schlunegger, M.P., Eisenberg, D. 3D domain swapping: a mechanism for oligomer assembly. Prot. Sci., 4:2455–2468, 1995.
15. Bernstein, B.E., Michels, P.A.M., Hol, W.G. Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation. Nature 385:275–278, 1997.
16. Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H., Wlodawer, A. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3Å resolution. Science, 246:1149–1152, 1989.
17. Ikura, M., Clore, G.M., Gronenborn, A.M., Zhu, G., Klee, C.B., Bax, A. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. Science, 256: 632–637, 1992.
18. Meador, W.E., Means, A.R., Quiocho, F.A. Target enzyme recognition by calmodulin: 2.4Å structure of a calmodulin-peptide complex. Science, 257:1251–1255, 1992.
19. Meador, W.E., Means, A.R., Quiocho, F.A. Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures. Science, 262:1718–1721, 1993.
20. Gerstein, M., Schulz, G., Chothia, C. Domain closure in adenylate kinase. Joints on either side of two helices close like neighboring fingers. J. Mol. Biol., 229:494–501, 1993.
21. Schulz, G.E., Muller, C.W., Diederichs, K. Induced-fit movements in adenylate kinases. J. Mol. Biol., 213:627–630, 1990.
22. Anderson, B.F., Baker, H.M., Norris, G.E., Rumball, S.V., Baker, E.N. Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins. Nature, 344:784–787, 1990.
23. Baker, E.N., Rumball, S.V., Anderson, B.F. Transferrins: insights into structures and function from studies of lactoferrin. Trends in Biochem. Sci., 12:350–353, 1987.
24. Gerstein, M., Anderson, B.F., Norris, G.E., Baker, E.N., Lesk, A.M., Chothia, C. Domain closure in lactoferrin. Two hinges produce a see-saw motion between alternative close-packed interfaces. J. Mol. Biol., 234:357–372, 1993.
25. Oh, B.H., Pandit, J., Kang, C.H., Nikaido, K., Gokcen, S., Ames, G.F.L., Kim, S.H. Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding proteins with and without a ligand. J. Bio. Chem., 268:11348–11355, 1993.
26. Sharff, A.J., Rodseth, L.E., Spurlino, J.C., Quiocho, F.A. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. Biochemistry, 31:10657–10663, 1992.
27. Spurlino, J.C., Lu, G.Y., Quiocho, F.A. The 2.3Å resolution structure of the maltose or maltodextrin binding protein, a primary receptor of bacterial transport and chemotaxis. J. Bio. Chem., 266:5202–5219, 1991.
28. Mattevi, A., Rizzi, M., Bolognesi, M. New structures of allosteric proteins revealing remarkable conformational changes. Curr. Opin. Struct. Biol., 6:824–829, 1996.
29. Perutz, M.F. Stereochemistry of cooperative effects in hemoglobin. Nature, 228:727–734, 1970.
30. Barford, D., Johnson, L.N. The allosteric transition of glycogen phosphorylase. Nature, 340:609–616, 1989.
31. Stråter, N., Håkansson, K., Schnappauf, G., Braus, G., Lipscomb, W.N. Crystal structure of the t state of allosteric yeast chorismate mutase and comparison with the r state. Proc. Natl. Acad. Sci. USA, 93:3330–3334, 1996.
32. Koshland, D.E. Application of a theory of enzyme specificity to protein-synthesis. Proc. Natl. Acad. Sci. USA, 44:98, 1958.
33. Jorgensen, W.L. Rusting the lock and key model for protein-ligand binding. Science, 254:954–955, 1991.
34. Wriggers, W., Schulten, K. Protein domain movements: Detection of rigid domains and visualization of hinges in comparisons of atomic coordinates. Proteins 29:1–14, 1997.
35. Maiorov, V., Abagyan, R. A new method for modeling large-scale rearrangements of protein domains. Proteins 27:410–424, 1997.
36. Paulsen, M.D., Orenstein, R.L. Dramatic differences in the motions of the mouth of open and closed cytochrome p450bm-3 by molecular dynamic simulations. Proteins 21:237–243, 1995.
37. DesJarlais, R.L., Sheridan, R.P., Dixon, J.S., Kuntz, I.D., Venkataraghavan, R. Docking flexible ligands to macromolecular receptors by molecular shape. J. Med. Chem., 29:2149–2153, 1986.
38. Leach, A.R., Kuntz, I.D. Conformational analysis of flexible ligands in macromolecular receptor sites. J. Comp. Chem., 13(6):730–748, 1992.
39. Ghose, A.K., Crippen, G.M. Geometrically feasible binding modes of a flexible ligand molecule at the receptor site. J. Comp. Chem., 6:350–359, 1985.
40. Goodsell, D.S., Olson, A.J. Automated docking of substrates to proteins by simulated annealing. Proteins, 8:195, 1990.
41. Mizutani, M.Y., Tomioka, N., Itai, A. Rational automatic search method for stable docking models of protein and ligand. J. Mol. Biol., 243:310–326, 1994.
42. Clark, K.P., Ajay. Flexible ligand docking without parameter adjustment across four ligand receptor complexes. J. Comp. Chem., 16:1210–1226, 1995.
43. Judson, R.S., Tan, Y.T., Mori, E., Melius, C., Jaeger, E.P., Treasurywala, A.M., Mathiowetz, A. Docking flexible molecules: a case study of three proteins. J. Comp. Chem., 16:1405–1419, 1995.
44. Welch, W., Ruppert, J., Jain, A.N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. Chem. Biol., 3:449–462, 1996.
45. Rarey, M., Kramer, B., Lengauer, T., Klebe, G. A fast flexible docking method using incremental construction algorithm. J. Mol. Biol., 261:470–489, 1996.
46. Knegtel, R.M.A., Kuntz, I.D., Oshiro, C.M. Molecular docking to ensembles of protein structures. J. Mol. Biol., 266:424–440, 1997.

47. Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R. Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol., 267:727–748, 1997.
48. Leach, A.R. Ligand docking to proteins with discrete side-chain flexibility. J. Mol. Biol., 235:345–356, 1994.
49. Jones, G., Willet, P., Glen, R.C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J. Mol. Biol., 245:43–53, 1995.
50. Sandak, B., Nussinov, R., Wolfson, H.J. An automated Computer-Vision & Robotics based technique for 3-D flexible biomolecular docking and matching. Computer Applications in the Biosciences, 11:87–99, 1995.
51. Sandak, B., Wolfson, H.J., Nussinov, R. Hinge-bending at molecular interfaces: Automated docking of a dihydroxyethylene-containing inhibitor of the HIV-1 protease. In: "Journal of Biomolecular Structure & Dynamics, Vol 1, Proceedings of the Ninth Conversation." Sarma, R.H., Sarma, M.H. (eds). New York: Adenine Press, 1996:233–252.
52. Wolfson, H.J. Generalizing the generalized Hough transform. Pattern Recognition Letters, 12(9):565–573, 1991.
53. Sandak, B. A Method for Biomolecular Structural Recognition Allowing Conformational Flexibility. PhD thesis, Computer Science Department, Tel Aviv University, 1997.
54. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanonchi, T., Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. J. Mol. Biol., 112:535–542, 1977.
55. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. J. Mol. Biol., 161:269–288, 1982.
56. Lin, S.L., Nussinov, R., Fischer, D., Wolfson, H.J. Molecular surface representation by sparse critical points. Proteins 18:94–101, 1994.
57. Lin, S.L., Nussinov, R. Molecular recognition via the face center representation of molecular surface. J. Mol. Graphics, 14:78–97, 1996.
58. Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. Science, 221:709–713, 1983.
59. Connolly, M.L. Analytical molecular surface calculation. J. Appl. Cryst., 16:548–558, 1983.
60. Richards, F.M. Areas, volumes, packing and protein structure. Ann. Rev. Biophys. Bioeng., 6:151–176, 1977.
61. Theriault, Y., Logan, T.M., Meadows, R., Yu, L., Olejniczak, E.T., Holzman, T.F., Simmer, R.L., Fesik, S.W. Solution structure of the cyclosporin A/cyclophilin complex by NMR. Nature, 361:88–94, 1993.
62. Yasri, A., Chiche, L., Haiech, J., Grassy, G. Rational choice of molecular dynamics simulation parameters through the use of the three-dimensional autocorrelation method: application to calmodulin flexibility study. Prot. Eng., 9:959–976, 1996.
63. Shoichet, B.K., Bodian, D.L., Kuntz, I.D. Molecular docking using shape descriptors. J. Comp. Chem., 13:380–397, 1992.
64. Shoichet, B.K., Kuntz, I.D. Protein Docking and Complementarity. J. Mol. Biol., 221:327–346, 1991.
65. Shoichet, B.K., Kuntz, I.D. Matching chemistry and shape in molecular docking. Protein Engineering, 6:723–732, 1993.
66. Rigoutsos, D., Platt, I., Califano, A. Flexible 3D-substructure matching & novel conformer derivation in very large databases of 3D-molecular information. Research Report RC 20497 (90854), IBM Research Division, 1996.