F^3Dock: A Fast, Flexible and Fourier Based Approach to Protein-Protein Docking

Chandrajit Bajaj * Rezaul Alam Chowdhury † Vinay Siddavanahalli ‡

ICES Report 08-01

January, 2008

(Updated April, 2009)

Abstract

Protein interactions, key to many biological processes, involves induced fit between flexible proteins which typically undergo conformational changes. Modeling this flexible protein-protein docking is an important step in drug discovery, structure determination and understanding structure-function relationships. In this paper, we present F^3Dock, a Fast Flexible and Fourier based docking algorithm which utilizes adaptive sampling of orientation and conformational spaces, and a hierarchical molecular flexibility and structure representation. Different conformations are adaptively sampled and docked using a Non-equispaced Fast Fourier based algorithm.

1 Introduction

Structural interactions between proteins is responsible for their functions as building blocks in our cells and their conformational changes is often critical during the induced fit. Hence accounting for different conformational states of proteins is important for accurate computational protein docking[1]. Flexibility often involves movements between large rigid parts of the protein, called domains, flexible loops on the molecular surface and large side chain at active sites (see [32] for an example of the HIV-1 protease flexibility simulation). In a previous paper, on F^2Dock [11], we presented a non-equispaced fast Fourier based algorithm for efficiently computing rigid protein-protein docking (based on shape complementarity and the electric potential). In this paper, we provide a data structure and file format for users to represent hierarchical flexibility meta data for the given proteins and extend a Normal Mode Analysis based approach for automatic domain identification to represent flexibility. Using a hierarchical representation of structure and multiresolution docking, an adaptive sampling of orientation space is performed to compute docking conformations. A multi-stage hierarchical rigid docking is used to drive the sampling and a final side chain fit at potential interfaces is performed to optimize the docking score.

2 Related work

In our F^2Dock paper [11], we provided a summary of previous approaches for rigid docking, and in this section we summarize previous work on flexible docking. There are several good reviews (e.g., [45, 46, 57, 60, 79, 130, 135, 139, 141]) that discuss protein-protein docking techniques at various resolutions, while [20] reviews research on flexible protein-protein docking in particular.

*Computational Visualization Center, Department of Computer Sciences and The Institute of Computational Engineering and Sciences, The University of Texas at Austin, Texas 78712, USA. Email:bajaj@cs.utexas.edu.

†Computational Visualization Center, Department of Computer Sciences and The Institute of Computational Engineering and Sciences, The University of Texas at Austin, Texas 78712, USA. Email:shaikat@cs.utexas.edu.

‡Google Inc., Mountain View, California, USA.

[1]according to the Abagyan Lab, TSRI, 'Only about one third of the protein complexes can be docked without serious considerations for the induced conformational changes upon docking.'

2.1 Flexibility in Proteins

Flexibility analysis of proteins can be performed through a wide variety of algorithms. Algorithms based on molecular dynamics (MD) are given in [71, 82, 98, 137, 142]. However, use of this method is limited since it usually simulates proteins at pico- or nano-second scale, while large conformational changes occur over micro- or milli-seconds. Though various methods [13, 18, 118, 129, 170] have been proposed to speed-up MD simulation, simulating such large conformational changes is still beyond the capabilities of today's state-of-the-art simulators. Xray Crystallography is used to obtain atomic resolution models of proteins in crystalline forms. NMR (Nuclear Magnetic Resonance) techniques are also used for smaller molecules.

Protein dynamics gives rise to a large number of conformations making its analysis computationally infeasible, and furthermore, many of the motions used in the simulation do not affect docking results. Many methods have been developed to reduce these conformations to a new basis, where the principal basis gives the large fluctuations efficiently. It has been shown that conformational changes of a protein can be captured by only a few bases and projection vectors [144, 147, 148]). Normal Mode Analysis (NMA), Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) are used to reduce the dimensionality of the problem. Low frequency normal modes can usually represent many observed conformational changes in proteins [23, 63, 67, 154]. Successful modeling of the Chaperonin GroEL was performed using NMA [105]. To avoid the computations on a large matrix, a blocked version of NMA was computed in [145]. Graph theory is also used for NMA based protein flexibility prediction [74]. In [165] deformations along principal components are treated as additional degrees of freedom to facilitate binding process which is shown to improve results of protein-protein docking [109]. Normal modes are also used to optimize complexes against density maps [69, 146] which has the potential of improving docking results. Gaussian Network Models (GNM) [51] model protein structures as elastic networks [149, 169]. GNM based on Kirchoff matrices were introduced for proteins in [84]. PCA is used in Essential Dynamics (ED) to generate protein conformaions [31, 117, 131]. Various tools and web services based on collective motions analysis are now available for identifying flexible and hinge regions [4, 14], and for generating starting conformations for docking [15, 99, 143].

Protein flexibility can also be modeled by decomposing it into rigid domains. Various techniques have been developed for identifying these domains from one or more conformation of a protein. Rigid domains are identified in [167] under the assumption that groups of atoms folded by hydrophobic effect behave as *compact units* during conformational changes. Static core or the backbone of molecules and their associated rigid domains were computed in [21] using two different conformations of a given protein. The algorithm in [136] assumes that a domain must have more internal contact than that with the rest of the protein. This idea is extended in [107] to perform a Monte Carlo sampling in internal coordinates.

HingeFinder [159] uses two conformations of a protein in order to identify rigid domains. It selects subsets of C^{α} atoms around randomly chosen points from both conformations. Atoms that show large movements in both subsets are removed from the sets and neighbors are screened and added if required. The remaining atoms are assumed to form a rigid domain, and are deleted from both conformations. The procedure is then repeated for the next set of randomly chosen atoms. The 6D transformation is approximated as a rotation around a hinge axis. DynDom [62, 64] also uses two conformations. FIRST/ProFlex [74] obtains flexible and rigid domains from a single conformation using a graph theoretical approach. DomainFinder [67, 68] computes NMA on the input protein and allows the user to select a set of modes and a deformation threshold to compute rigid domains. Cubes of approximately 6 residues are given 6 transformations from the selected modes. Clusters of cubes with similar transformations are grouped to form a rigid domain. Other software packages for domain decmposition include DomainParser [163] and PDP [3].

Side chains play an important role in protein flexibility. It has been shown that only a few conformational states (i.e., rotamers) of any side chain residue are predominantly present in nature (e.g., see [124]). Both backbone independent [40, 151] and backbone dependent [40–43] libraries for side chain rotamers have been developed. The Dead End Elimination theorem [34] is often used to prune the number of rotamers to consider for docking.

2.2 Flexible Protein-protein Docking

Protein-protein and protein-ligand (i.e., small protein) docking problems are mathematically equivalent, but the latter is computationally more feasible and hence has been tackled often. Unless specified otherwise, each approach mentioned below has been used for both problems.

Some docking algorithms treat side chain and small backbone flexibility implicitly by means of surface softening that allows slight interpenetration of surfaces to be matched [121], or by trimming long side chains [65]. Though this approach does not always provide good results due to frequent steric clashes at the atomic level [140], recently some improved docking results were obtained by utilizing snapshots of MD simulation for constructing the 3D grid onto which the molecules were mapped [104]. Many docking algorithms incorporate flexibility by performing cross docking, i.e., by rigid-body docking of ensembles of conformations obtained by applying some conformational sampling method (e.g., MD simulation [25,36,59,101,125,138,140], NMR structures [38], etc.). NMA [166] and ED [7, 117] have also been used to generate conformations for cross docking.

Many recent algorithms include flexibility explicitly in the docking process. These algorithms use explicit representations of molecules instead of mapping them onto a grid, and many of them apply flexibility during a refinement stage after rigid-body docking. HADDOCK [37] explicitly allows both backbone and side chain flexibility during its MD simulated annealing refinement stages which results in improved docking results [36]. Guided docking [50, 138], too, allows limited backbone flexibility. In [44] torsion angle dynamics was used to simulate the binding of barnase and barstar.

FlexDock [134] is able to model very large conformaional changes of proteins during docking by decomposing them into rigid domains, and then applying rigid-body docking on different conformations of the proteins obtained by applying hinge-bending motions between connected domains. Similar strategies are used by some other algorithms, too [19, 158]. Global search algorithms based on energy minimizations, heuristics based search methods, and geometric identifications of possible active sites are used for flexible docking. In DOCK [87] and later [33], receptor binding sites were identified as cavities and the complementary space as spheres. Ligand fragments were separately bound to the active site and then incrementally selected to form the entire ligand. Incremental approaches based on shape (e.g., [92], HammerHead [157]) and properties of the molecules (e.g., FLEXX [126, 127]) are used to dock fragments, pruning the exponential search by retaining only a fixed set of possible conformers at each step. Other global search techniques include fast Fourier Transform (FFT) [30, 52, 108, 152], hydrogen bond pattern based search [113], genetic algorithms [78](GOLD), [77, 80, 120], Monte-Carlo/simulated annealing [55](AUTODOCK), [8, 24, 86, 111], conformational space annealing [95–97], molecular dynamics [119] and evolutionary programming [53]. See [116] for the performance of different algorithms in AUTODOCK. Steered molecular dynamics, using a visualization and feedback toolkit has also been studied in SMD [98].

An algorithm based on geometric hashing that uses hinge bending for domain movements in proteins/ligands is given in [133]. In [156] conformations are sampled using a coarse set of values for torsion angles of rotating bonds, and conformations that do not form severe steric overlaps are used in rigid-body docking. Angles are sampled and matched using α-shapes [10].

Flexible side chains are more commonly modeled in protein-protein docking than movements in the backbone. Most side chain packing algorithms [29, 40–43, 58, 76, 103, 106, 110, 124, 151, 155, 160–162, 164, 165] use rotamer libraries. The side chain packing problem can be modeled as a combinatorial search problem that optimizes an energy function. This problem is known to be NP-Hard [2, 123] which cannot even be reasonably approximated in polynomial time [29]. Using rotamer libraries, and a greedy heuristic or branch and cut algorithm, [6] performs docking of proteins with flexible side chains as a second step to rigid protein-protein docking. Similar discrete side chain conformations were searched in [91]. By classifying residues as 'active' and 'inactive', and clustering them into connected graphs of interacting residues, SCWRL [28] is able to identify low energy conformation rotamers efficiently. A recent algorithm named TreePack [161, 162] claim to run up to 90 times faster than SCWRL3 [22, 28, 39]. A combination of pseudo-brownian Monte Carlo minimization followed by flexible side chain docking with ICM [1] (see also [49]) was tested on a variety of bound and unbound complexes in [48]. Many algorithms based on the Dead End Elimination (DEE) theorem [34] can reach the global minima provided they are able to converge [34,35,54,56,83,88,89,102,106,122,153], while some algorithms based on other techniques such as Monte Carlo simulation [70], cyclical search [42, 160], spatial restraint satisfaction [132], and approximation algorithms [29, 85, 162] run reasonably fast without any guarantee of global optima. Among other methods A* search [93], simulated annealing [94], mean-field optimization [73, 90], maximum edge-weighted clique [9], and integer linear programming [5, 47] are also used. All-atom representations are used in [1,26,27,37,48,49,100,140,150], all of which allow at least side chain flexibility.

Apart from backbone and side chain movements, loop flexibility can also affect docking. Flexible loops at known active sites is handled using a Monte Carlo, simulated annealing based docking approach in [16,17]. In [128] flexible loops are ignored in the docking step, and later rebuilt in a loop modeling step.

The *connexions project* at `http://cnx.rice.edu/content/m11464/latest/` maintains a summary of flexible docking algorithms. The CAPRI (Critical Assessment of Predicted Interaction) project [75,114,115] monitors the performance of state-of-the-art docking algorithms by arranging community-wide blind docking experiments which often include flexible proteins.

3 Flexible Chain Complex

Proteins have a naturally occurring backbone, forming chains which flex through their torsion angles as shown in figure 1. Structural (shape) and functional properties are described as a labeled *sheath* around the central *nerve*. This combined labeled representation of a *nerve* and a *sheath* is used to model a flexible protein's structure and properties and is referred to as a Flexible Chain Complex (FCC). Movement of loops and domains leading to large conformational changes occur due to backbone torsional angle changes and hinge type bending and shearing movements. Rearrangement of side chains in binding regions occur due to torsional angle changes in the side chains of various residues. Figure 1 shows three residues along a typical backbone with different relevant torsional angles marked. The backbones motion is mainly controlled by the pair of ϕ, ψ angles given for each residue. The side chains move through torsional changes in the χ angles. Depending on the amino acid type, there can be up to 5 such successive angles. In section §2.1, we have already summarized various algorithms to compute flexibility in proteins. Here we compute a hierarchical decomposition using Normal Mode Analysis. But since our data structure is general, we observe that users can augment their own descriptions to it. The FlexTree is a data structure introduced by Zhao, Stoffler and Sanner [168], and provides a method for storing a hierarchical description of flexibility. Their data structure should be easily parsed into our Flexible Chain Complex and used in the flexible docking algorithm. Our data structure provides similar capabilities, but with shear, bending and twist motions and we maintain a general graph with priorities at each edge and provide an automated algorithm to compute the flexibility. We first describe the flexible data structure and then provide an automated algorithm to compute it for any given protein.

3.1 Hierarchical Domain Identification

Domains are considered to be rigid contiguous parts of a protein which show little movement in different conformations (obtained through molecular dynamics, normal mode analysis etc.). Given a threshold of rigidity, a protein decomposes in to different number of domains. Hence, we can automatically obtain, from dynamics or other simulations, a hierarchical domain decomposition for proteins. In particular, we keep a decomposition tree, with each discrete level representing the protein at a rigidity threshold. Since we provide a output ASCII file format, users can update any kind of flexibility at desired locations, with ranges. Below, we provide one method to automatically compute all the required quantities.

Connectors, Flexible Loops and Domain Model. At any given level, there are various domains which interact either through connected chain segments or large interfaces. In particular, we call these chain segments and areas as connectors. Domains and connectors form a complete description of the flexible protein at a given level. We also recognize that domains have flexible loops and chain ends on their surfaces. We identify

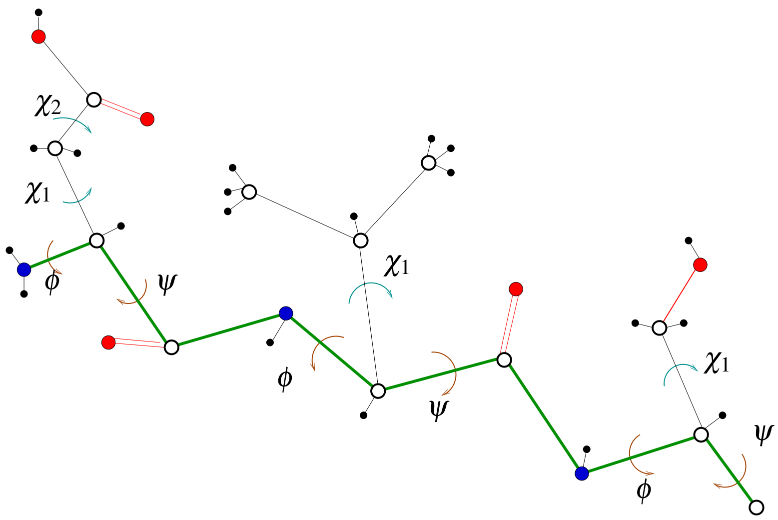


Figure 1: The first three residues of 1AY7.PDB: ASP, VAL, SER are shown schematically with the relevant backbone (ϕ, ψ) and side chain (χ_1, χ_2) torsion angles.

and mark these are flexible loops, apart from the rigid segments in a domain. To build a formal model, we define the following objects:

- **Segment.** A segment is a contiguous sequence of residues from the chain of a protein.
- **Flexible loops / Handles.** A segment on a rigid domain surface, not acting as a connector to another domain and also consists of large flexible side chains. It is useful to identify such flexible regions in a domain to provide a finer resolution flexibility model for the docking algorithm.
- **Domain.** A connected set of rigid segments and flexible loops. Using different rigidity thresholds, we obtain a hierarchy of sub-domains.
- **Connector / Linker.** Segment between two domains. We also consider large domain interfaces as connectors.
- **Flexor.** A set of connectors between a pair of domains, associated with certain flexibilities. The flexors are given priorities over all levels, to form a hierarchical description of protein flexibility. Unlike in the FlexTree, Flexors here need not form a cut.

3.1.1 Motions Allowed at Flexors

We provide shear, bending and twist motions at flexors. In [107] techniques to compute such motions for two domains linked by a single or double stranded linkers is provided.

Shear. This describes lateral movement along interfaces between domains. The magnitude of shear is limited by a maximum value chosen by the user and the length of the smallest connector between the two domains under consideration. In the absence of connectors, the line joining the centroids of the two domains is used to compute the normal of the shearing plane.

Bending. We apply the bending motion around three orthogonal axes when the domains are connected by at least one connector. The geometric center of the shortest connector between the two domains is chosen as the hinge point and the normal to the plain containing the geometric centers of the two domains and of the shortest connector is taken as the primary hinge axis. The secondary axis is orthogonal to the primary axis and also to the line connecting the two domain centers. The line through the domain centers is considered the third axis. First we compute the hinge-point and the primary axis. After applying the bending motion around this axis, we compute the secondary axis and recompute the hinge-point with respect to the new conformation, and apply motion around the new secondary axis. Then we compute the hinge-point again along with the third axis, and apply bending motion around this axis.

Twist. When a single physical connector exists between two domains, it is also given a twist motion by updating torsion angles along its backbone.

3.1.2 Normal Mode Analysis

Normal Mode Analysis for a given unbound structure of a protein is computed using Hinsen's DomainFinder program [67]. For a given deformation threshold and domain coarseness, a set of rigid domains with their similarity indices is computed. Their output defines the domains as a set of contiguous residues. These are collected as segments of the protein. Let $S_1,..,S_d$ be the set of segments in the d domains at a given level. By deleting these sets from the protein, we are left with segments which form either flexible loops, connectors between domains or ends of chains. We assume that a chain consists of atleast one domain. Virtual connectors are added to domains which share a common interface. If we are dealing with a large macromolecule, more than one level can be computed by varying the parameters to DomainFinder.

3.2 Flexible Docking Algorithm

The flexible docking algorithm consists of adaptively sampling conformation space. In the first step, the high priority flexors are used to compute a set of conformations, the size of which is given by the user. A low resolution representation of the proteins are used to compute docking at each of these conformations over all of orientation space (or limited by the active sites if known). Given a set of possible docking positions, the domain(s) in that regions is further subdivided and a new set of conformations are computed for docking. If the sub domains (whose union is not the parent domain due to the presence of flexible loops) are far away from the interaction region, then only conformation sampling of the flexible loops are considered. In the last step, we refit all the interface side chains using a greedy algorithm.

Multiresolution Sum-of-Gaussians Representation. The electron density of an atom at a point $\mathbf{x}$ is represented as a Gaussian function: $f(\mathbf{x}) = e^{\beta\left(\frac{|\mathbf{x}-\mathbf{c}|^2}{r^2}-1\right)}$ where $\mathbf{c}, r$ are the center and radius of the atom. and thus the electron density of a protein with M atoms at $\mathbf{x}$ is: $f(\mathbf{x}) = \sum_{i=0}^{M-1} e^{\beta\left(\frac{|\mathbf{x}-\mathbf{c}_i|^2}{r_i^2}-1\right)}$ where β is a parameter used to control the rate of decay of the Gaussian and known as the *blobbiness* of the Gaussian [12]. By clustering atoms and varying the rate of decay of clustered Gaussians, we can obtain a Sum-of-Gaussians representation for the protein at multiple resolution levels.

Soft Docking. We use our F^2dock algorithm described before in [11] for soft protein-protein docking. Given two conformations, the algorithm predicts a set of possible docking sites where the docking score is above a user defined threshold. In particular, given N scoring functions $f_{1,k}, f_{2,k}, k = 1..N$and a user defined score τ, we solve the equation:

$$\left\{ \;(\mathbf{t},\mathbf{r},s) : \left(s = Re\left(\sum_{k=1}^{N}\left(\int_{\mathbf{x}} f_{1,k,\beta}(\mathbf{x}) T_{\mathbf{t}}\left(\Delta_{\mathbf{r}}\left(f_{2,k,\beta}(\mathbf{x})\right)\right) d\mathbf{x}\right)\right)\right) \geq \tau, \forall(\mathbf{t},\mathbf{r})\; \right\}$$

In the above equation, $\mathbf{t}$ is the translational space we require to sample, and $\mathbf{r}$ is the rotational sampling space. The resolution of the maps is controlled by β. In particular, our soft docking algorithm can be restricted to the orientations we are interested in and the resolution of docking maps can be controlled.

3.2.1 Hierarchical Docking

Our flexible docking has three stages: Parallel docking of a global hierarchical conformational sampling, local flexible loop and large side chain sampling and interface fit using a greedy algorithm.

FCC Construction:

1. ***Input:*** For a given protein, Normal Mode Analysis is used to compute, for L levels, the domains D_i for each level.

   (a) $D_i = \bigcup S_{i,k}$, a set of k segments.

2. ***Compute flexible loops and Connectors:*** Follow each segment $s \in D_i$. If it terminates back in D_i without crossing into any other domain, or is the end of a chain, add it to D_i as a flexible segment f.

   (a) $D_i = D_i \bigcup F$, a set of flexible segments.

   (b) Any segment c that crosses from D_i to $D_j, i \neq j$ is added to a list of connectors: $C = C \bigcup c_{i,j}$.

3. ***Compute Labeled Flexors:*** For each pair of domains D_i, D_j, all c_{ij} are collected to form a Flexor between the domains. If $\{c_{i,j}\} = \Phi$, then the area of the interface is used to determine if we need a virtual flexor or not.

4. ***Compute Hierarchy:*** Steps **2**, **3** are repeated for all levels. Domains are broken up if necessary to maintain unique parent domain nodes.

5. ***Output:*** This labeled complex is printed out in a easy-to-use ASCII file. Users can intuitively add/delete new domains, connectors and flexors at will.

Adaptive Sampling of Conformations. The biased probability Monte Carlo sampling in [107] can be applied to our structure, but here we provide a random sampling followed by a steric collision test. There are two distinct types of flexors: Flexors which lead to a cut in the component and those which do not. For each flexor, we arbitrarily assign a left and right domain, and always update the right domain. For a flexor which defines a cut, all domains to the right are updated, while for the other case, only the right domain is updated. The connectors from the right to other domains are updated to maintain structural integrity of the protein. This reduces the range of motion at a flexor which is not a cut. Each flexor is given a score depending on the range of motions computed for its associated shear, primary/secondary bending and twist. The sampling is adaptively performed to reflect these scores.

Global Conformation Sampling and Low Resolution Search:

1. ***Input:*** The FCC of a ligand and a fixed number of global conformations N.

2. ***Allocate Conformations:*** Given the set of Flexors $\{f\}$ at the top level, a hierarchy of importance is built and the total number of conformations is divided among them.

   (a) Determine if each Flexor f is a cut or not. The domains, connectors at any level form a graph. Hence, each flexor, defined as the set of connectors between domains can possibly disconnect the graph.

   (b) If f is a cut, then let $\{d_l\}, \{d_r\}$ be the set of domains to the left and right. Let the sum of their weights be w_l, w_r. Then the score for f is $s_f = min(w_l, w_r)$.

   (c) If f is not a cut, let w_{dl}, w_{dr} be the weight of the left and right domain. Then the score for f is given as $s_f = min(w_{dl}, w_{dr})$.

   (d) N_f, the number of conformations allocated to f is N^{s_f/Σ_f}.

   (e) Each flexor is associated with at most 5 motions: shear, twist, and bending along 3 axes. Each motion is sampled using heuristics based on their computed range.

3. ***Compute Conformations:*** We recursively apply a new motion at each flexor.

   GetConformation (Flexor f_i, Molecule m)

   - Set $\hat{m} \leftarrow m$

   - For each motion t in f_i

   -     apply($f_i, t, \hat{m}$)

   -     If (i = Number of Flexors), Print($\hat{m}$)

   -     Else Call GetConformation($f_{i+1}, \hat{m}$)

   - End for

   Output: A set of at most N conformations of the ligand, with higher priority flexors given a higher resolution sampling.

4. ***Low Resolution Search:*** A 20 degree rotational sampling is used for soft docking. We use residue level parameters to represent the shape affinity functions.

   - Electron density can be represented as a sum of Gaussians, as described before. The FCC gives a clustering of atoms into residues. For this lower resolution search, we can either decrease the decay parameter, or use fewer Gaussians to represent a residue.
   - For every conformation $\hat{m}_i, i = 1..N$, call F^2Dock [11].

5. ***Output:*** Conformations and their orientations where the docking score exceeded a user defined threshold.

Finer Resolution Search. Given a set of promising orientations from the previous low resolution search, soft docking is again performed with high resolution affinity functions. For shape, we vary the rate of decay of Gaussians representing atoms to obtain a higher resolution map. We use a value of -2.3 to represent the atomic level resolution. For electrostatics, we use a charge from OPLS at each atom. Each conformation and orientation saved from the low resolution search is used as input. Hence, we adaptively sample orientation space and use a multiresolution representation of affinity functions. To further improve the docking score, we perform a refitting of side chains at potential interfaces.

Refitting Side Chains at Interfaces. We use the Dunbrack [41] backbone independent library to sample interface rotamers.

Given a certain conformation from soft docking, we would like to optimize the side chains in the interface to obtain a better fit of the proteins. Let there be N interface residues in the given conformation, and the residues be $R_i, i = 1..N$. Each residue is associated with a rotamer set $\{r^i\}$. The cardinality of this set depends on the type of the residue. Since we do not want to discard the current conformation for a given side chain, we include R_i into the set $\{r^i\}$. From the Dunbrack library, we also have $\{p^i\}$, probabilities for each rotamer for a given side chain. We set a probability for the current side chain as equal to the highest in its set of rotamers. Let the scoring function for the j^{th} rotamer for side chain i be $S(r^i_j)$. A solution is any set $\{r^i_j, i = 1..N\}$, and is the optimal solution when $\sum_i S(r^i_j)$ is maximized.

Intersection lists

Using the FCC of the second protein, residue intersections are computed for all surface residues. Surface residues are computed as those whose atoms (at least 1) are less than 4Å away from the other protein's surface. Given a residue, we traverse down the FCC hierarchy to adaptively compute all other intersecting residues. Let $\{I_i\}$ be the set of residues which intersect residue R_i and any of its rotamers. Assuming a maximum number of intersecting residues N_I, the cost of this algorithm is linear in the number of residues. Here, we are interested in computing intersection with neighbors, assuming that the current residue can be any of its rotamers.

Scoring

The addition of a residues rotamer into the current partially formed interface will influence both the current score of the docking and the scores of potential rotamers yet to be added. The score is currently computed as the sum of shape complementarity and electrostatics. First we compute the function ϕ_s, ϕ_e, density and electrostatics for the first protein. For any given atom in a rotamer under consideration, we calculate the approximate Lennard Jones score depending on its distance from the electron density, and the electrostatics score as a product of its negative charge and the field ϕ_e. A slightly higher rating is used for the electrostatics energy contribution. The addition of a new rotamer affect other rotamers yet to be added. The scores of those rotamers are updated using a simpler scheme based on steric overlaps with the newly added rotamer, to avoid costs of recomputing the fields.

Side chain repacking algorithm:

From the results of the previous docking steps, we are given a protein and ligand (in possibly new conformations) and a transformation between the proteins that yields a good docking score. Now we proceed to repack the side chains of the ligand to improve the fit. First, we compute the interface residues for the given transformation. This can be quickly done by a pre-computation of the signed distance function of the individual proteins. Next we compute all the rotamers of all the interface residues, by looking up appropriate entries in a Rotamer library. In the third preprocessing step, we compute all neighbors of a given residue. We can assume this to be a constant number. In the final step, we compute the current score for each residue and its set of rotamers. The score is currently a sum of both shape and electrostatic interactions.

Now we reinsert a new rotamer in place of each of the original interface residues. The choice is a greedy choice, with a backtracking option when the score is lower than a threshold. Hence, we first insert the rotamer with the highest docking score among all interface residues and all their rotamers. The potential costs of all neighboring residues and rotamers can now be updated (currently we use a simple steric test for performance reasons). In this greedy fashion, all interface residues are replaced by new rotamers. Since the current position of a residue may be its most stable state, we also include the current residue as one of its choice of rotamers. If the total score at any point is too unfavorable, we backtrack and use the probabilities in the Dunbrack rotamer library to pick a new choice.

Assuming that each residue has a fixed number of rotamers and a fixed number of neighbors, the cost of maintaining a dynamic list of scores for residues for N interface residues and their insertion into the ligand should be $O(N \log N)$, but our current implementation uses a simpler $O(N^2)$ update. The main steps in the algorithm is presented below.

1. ***Input:*** The FCCs of proteins A and B, with a list of transformations from soft docking, $\{X\}$, that lead to potential complexes.

2. ***Output:*** For each $X \in \{X\}$, multiple sets of new repacking of side chains at interface.

3. ***Preprocessing:***

   (a) Potential fields ϕ_s, ϕ_e.
   (b) Interface residues $\{R_i, i = 1..N\}$ of protein B.
   (c) Rotamer set $\{Rot_{i,j}\}, i = 1..N, j = 1..n(R_i)$ and probabilities $p_{i,j}$.
   (d) Neighboring residues $\{R_i^N\}$ for every interface residue R_i.

4. ***Fit interface for each*** X***:***

   (a) Transform second protein:
   $$Trans(B, X)$$
   (b) Compute relevant interface residues $R_i^X, i = 1..N_X$ from R_i.

(c) Compute initial scores for rotamers: $\{S(Rot^{X}_{i,j})\}$.

(d) ***Incremental greedy fit:*** Repeat till we get a desired number of (suboptimal) solutions within finite number of attempts.

i. Initialize $Sol = \Phi$.

ii. Choose next best fit: $R_b = argmax_{Rot_{i,j}}\{S(Rot_{i,j})\}$, $Sol = Sol \cup R_b$.

iii. Update scores:
$$S(R^{N}{}_{b}) -= StericScore(R_b, R^{N}_{b}).$$

iv. Discard $\{Rot_i\}$.

v. Feasibility test: If $S(Sol) < \tau$ discard Sol and restore StericScore of neighborhood residues R^N. Else, continue with step ii.

(e) ***Evaluate:*** For each solution, compute and print RMSD with true solution.

4 Results

In Sections 4.1 and 4.2 we provide preliminary analysis of the flexibility of molecules and perform conformational sampling for soft docking, and in Section 4.3 we present some additional results on ZDock benchmark suites [72, 112].

4.1 Flexibility Analysis of Adenylate Kinase

We obtained normal mode based domain decomposition using DomainFinder for adenylate kinase (chain A from 4AKE.pdb). Figure 2 shows the decomposition. The normal modes were chosen according to the suggestions provided in the DomainFinder documentation [66]. The decomposition we obtained using normal mode analysis, and the decomposition reported by Hayward [61] based on conformational change based analysis [62], identify more or less the same rigid regions as domains, and the domain sizes are also comparable. See Table 1 for more detailed comparison of the domains along with the DomainFinder parameters we used. In addition to the rigid segments of each domain identified by DomainFinder, we identified the handles and the linkers/flexors connecting the domains, and assigned appropriate motions (shear/hinge-bending/twist) to the flexors.

We briefly describe our flexibility analysis for adenylate kinase below, and in Figure 3 we show the motion graph. We identified three domains. The core domain (i.e., domain 1) contains 93 residues, 12 segments and 7 handles. The AMP-binding domain (i.e., domain 2) has 52 residues, 9 segments and 4 handles, while the ATP-lid domain (i.e., domain 3) has 36 residues, 4 segments and 3 handles. Domain 1 is connected to domains 2 and 3 with 3 and 2 linkers, respectively, and no linkers were identified between domains 2 and 3. The interface area between domains 1 and 2 is large (508 Å^2), and hence we assigned a shear motion to the flexor. Also since the flexor acts as a cut (i.e., removing it disconnects the two domains), it is given a hinge-bending motion. However, due to the very small interface (2 Å^2) between domains 1 and 3, the connecting flexor was assigned only bending motions.

In contrast, Hayward [61] also reports three domains for this enzyme containing 103 (core domain), 42 (AMP-binding domain) and 38 (ATP-lid domain) residues, and applies bending motions between the core and each of the remaining two domains.

In Figure 4 we show an example of the effectiveness of our motion assignment. We start with an open conformation of adenylate kinase (chain A from 4AKE.pdb) in (a), and apply a suitable compound bending motion (around three orthogonal axes) between the core (blue) and the ATP-lid (green) domains to obtain (b). The geometric center of the shortest linker between the two domains is chosen as the hinge point and the normal to the plain containing the geometric centers of the two domains and of the shortest linker is taken as the primary hinge axis. The secondary axis is orthogonal to the primary axis and also to the line connecting the two domain centers (i.e., the third axis). We applied 46° and -12° bending around the primary and the

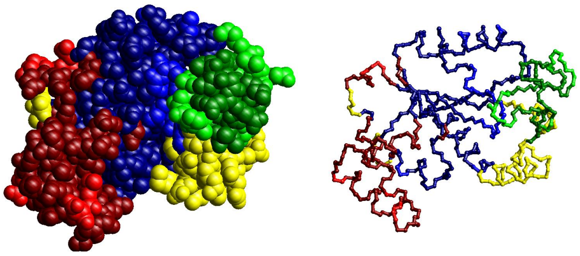


Figure 2: Domain decomposition using DomainFinder for adenylate kinase considered by Hayward [61] with its domains, rigid segments, flexible handles and linkers identified. Each domain is given a different color (domain 1: red, domain 2: green, domain 3: blue) with the rigid segments colored darker than the flexible handles. The linkers are colored yellow. We show both the atomic space-filling model, and the C_α-backbone structure. See Table 1 for more details.

Domain 1	Domain 2	Domain 3	Motions	Parameters (DomainFinder 1.1)
Residues (52): 26 - 27, 30 - 71, 75 - 80, 85, 99 Rigid Segments (9): 26 - 27, 30 - 43, 45 - 52, 55 - 61, 63 - 71, 75 - 77, 80, 85, 99 Flexible Handles (4): 44, 53 - 54, 62, 78 - 79	Residues (36): 120 - 155 Rigid Segments (4): 120, 122 - 126, 132 - 139, 144 - 155 Flexible Handles (3): 121, 127 - 131, 140 - 143	Residues (93): 1 - 25, 28 - 29, 72, 81 - 84, 86 - 97, 101 - 109, 175 - 214 Rigid Segments (12): 1 - 6, 16 - 25, 28 - 29, 72, 81 - 84, 86 - 94, 96 - 97, 101 - 109, 175 - 176, 178 - 202, 204 - 209, 212 - 214 Flexible Handles (7): 7 - 15, 95, 177, 203, 210 - 211	Domains 1 and 3 Interface Area: 508 Å^2 Shearing: Y Bending: Y Twisting: N Domains 2 and 3 Interface Area: 2 Å^2 Shearing: N Bending: Y Twisting: N	Normal Modes Calculated 642 Normal Modes Used 5 (modes 7 - 11) Deformation Threshold 1200 Domain Coarseness 14
Linkers (5) domains 1 and 3: 73 - 74, 98, 100 domains 2 and 3: 110 - 119, 156 - 174				
Residues (42) Reported by Hayward (using DynDom): 31 - 72	Residues (38) Reported by Hayward (using DynDom): 119 - 156	Residues (103) Reported by Hayward (using DynDom): 1 - 28, 80 - 112, 173 - 214		

Table 1: Domain decomposition (using DomainFinder 1.1) and flexibility analysis of the triple-domain enzyme adenylate kinase (chain A from 4AKE.pdb). We also compare our decomposition with the decomposition given by Hayward [61] using DynDom.

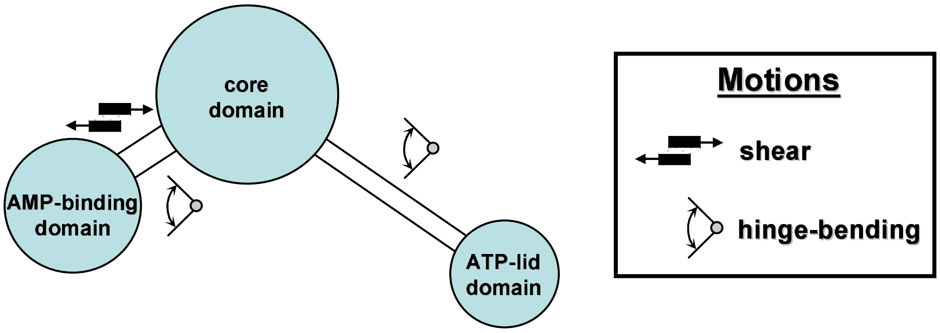


Figure 3: Motion graphs for adenylate kinase based on our flexibility analysis. The area of a circular domain is drawn proportional to its size (i.e., number of residues). If a flexor (i.e., set of linkers) exists between two domains, it is labeled with the motion(s) (shear and/or hinge-bending) assigned to it.

secondary axes, respectively, and -2° around the third to obtain (b). A closed conformation of adenylate kinase (chain B from 2ECK.pdb) is shown in (c). If the backbone atoms of the core domains of (a) and (c) are superimposed, the RMSD between the backbone atoms of the ATP-lid domains turns out to be 16.98 Å. But if we use (b) instead of (a) the RMSD reduces to 1.13 Å. Using HaywardŠs bending parameters [61, 168], on the other hand, we were able to reach a minimum RMSD of 1.17 Å at 53°.

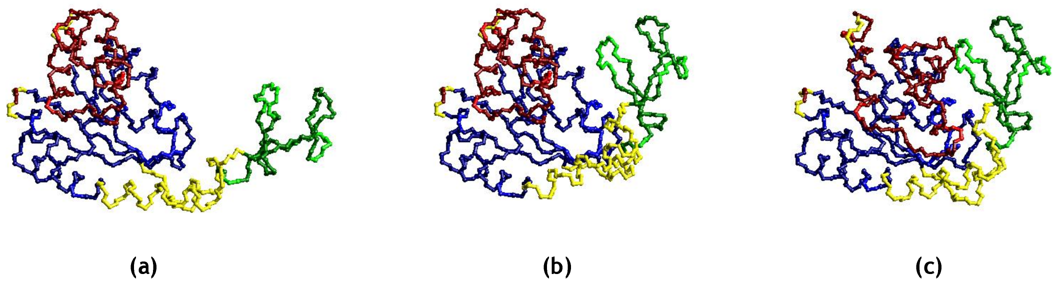


Figure 4: (a) An open conformation of adenylate kinase (chain A from 4AKE.pdb), (b) conformation we obtained from (a) after applying a suitable compound hinge-bending motion between the core (blue) and the ATP-lid (green) domains, and (c) a closed conformation (chain B from 2ECK.pdb). The AMP-binding domain is show in red, and the linkers in yellow. If the backbone atoms of the core domains of (a) and (c) are superimposed, the RMSD between the backbone atoms of the ATP-lid domains turns out to be 16.98 Å. But under the same superposition, the RMSD between the ATP-lid domains of (b) and (c) is only 1.13 Å.

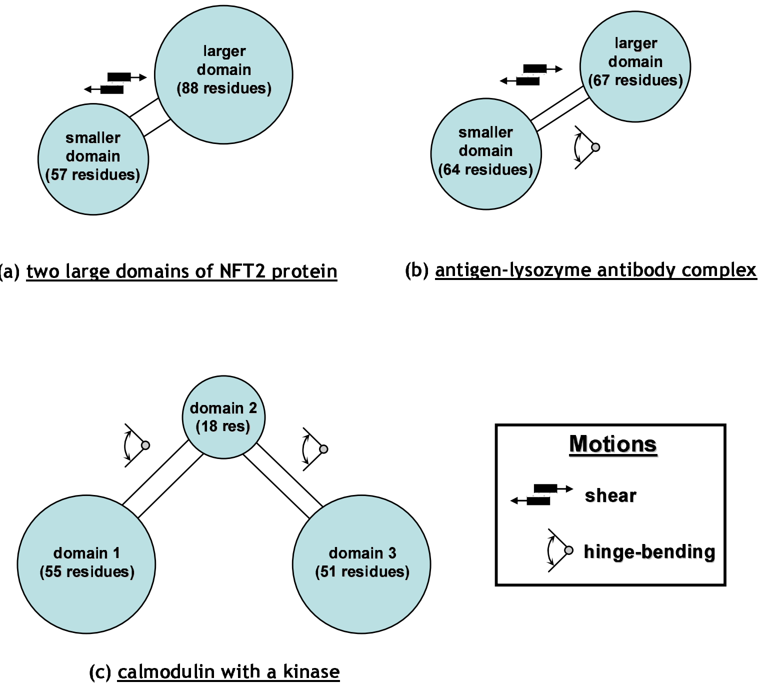


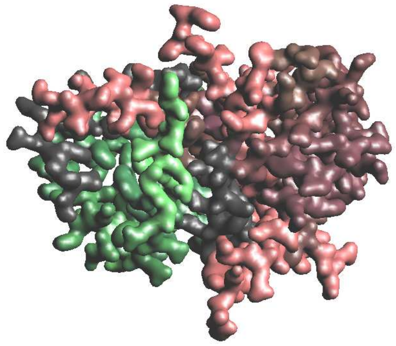
Figure 5: Motion graphs for (a) NFT2 protein, (b) antigen-lysozyme antibody complex, and (c) calmodulin with a kinase based on our flexibility analysis. The area of a circular domain is drawn proportional to its size (i.e., number of residues). If a flexor (i.e., set of linkers) exists between two domains, it is labeled with the motion(s) (shear and/or hinge-bending) assigned to it.

4.2 Flexibility Analysis and Conformational Sampling of Three Additional Complexes

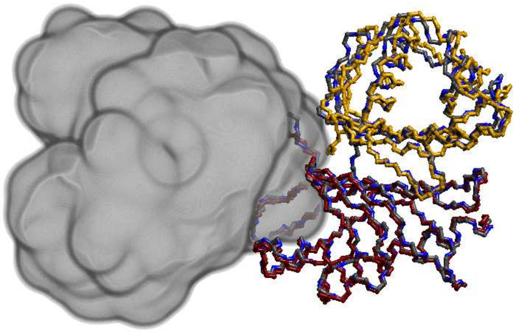
In this section we perform domain analysis using the algorithm in section §3.2.1 was performed on three additional complexes: 1A2K.pdb, 1VFB.pdb and calmodulin 2BBM.pdb. In Figures 6,7 and 8, we show the flexible regions and rigid domains identified using Normal Mode Analysis of DomainFinder and our clustering. We look at three primary motion types: shear, hinge and a combination of both. Shearing motion is shown at the large interface in 1A2K.pdb, the combination in 1VFB.pdb and a severe hinge motion in calmodulin 2BBM.pdb. In Figure 5 we show the corresponding motion graphs. Below we provide results from adaptive sampling to see how close we can get to a bound conformation from an unbound one using our model.

	RMSD to Bound Structure	
Complex	Rigid-body Docking	FCC Sampling
1A2K.pdb	1.453 Å	1.111 Å
1VFB.pdb	3.784 Å	0.586 Å
2BBM.pdb	14.429 Å	8.590 Å

Table 2: Improvements in RMSD between bound and unbound structures with and without FCC conformational sampling.



(a) Two large domains of NFT2 protein shown in green and red, consisting of 88, 57 residues each, 12,8 rigid segments and 10, 6 flexible loops respectively. The darker shades represent the rigid domains while the lighter shades the flexible loops.

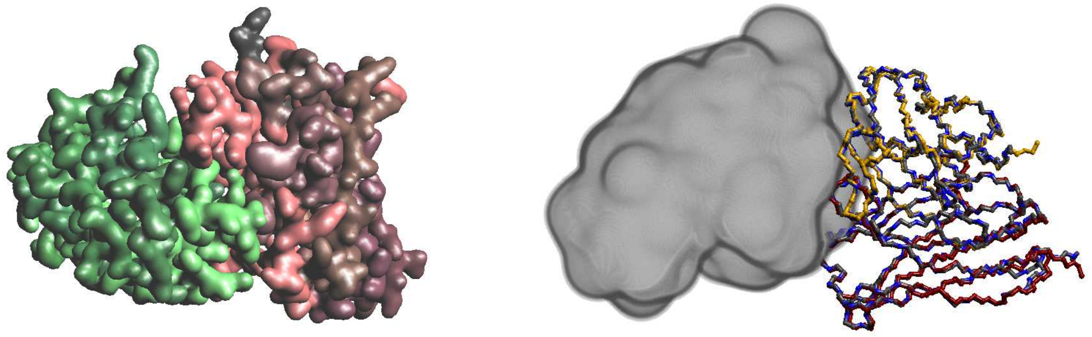


(b) The GDPRan protein is shown in gray transparency at 5 Å resolution while the bound NTF2 ligand's backbone is in blue. The two chains of the unbound structure are shown in gold and red. The original RMSD between these bound and unbound structures is 1.453Å.

Figure 6: GDPRan-NTF2 Complex

GDPRan-NTF2 Complex (1A2K.pdb). In the docking set given, only the C chain from the C,D and E chains of RAN is given. The A and B chains of the NTF2 protein are used as the flexible protein and domain analysis followed by adaptive conformation sampling is performed on it. From DomainFinder, we compute two domains with 88 and 57 residues. Using the above model, we build our FCC with the following information: The interface area at the flexor measures 436.612 Å^2. The flexor is a cut of our FCC. Due to the large interface area between the domains of the ligand at its only flexor, it is given a shear motion. Although we have a cut, the large interface limits the angle of search (This also prohibits any twisting motion in our model.). In figure 6(a), we show the NFT2 protein colored by the domains. In the right hand side, in figure 6(b), we overlap the bound and unbound proteins. To compute RMSD and to fit the proteins, we used backbone atoms from residues 4 to 126 from chain A and residues 4 to 124 from chain B. Using our FCC model and adaptive conformational search, we get unbound structures with RMSD ranging from 1.111 Å to 2.06 Å to the bound structure. The unbound crystal structure has a RMSD of 1.453 Å to the bound protein. Hence we can get conformations closer to the bound structure using our FCC sampling.

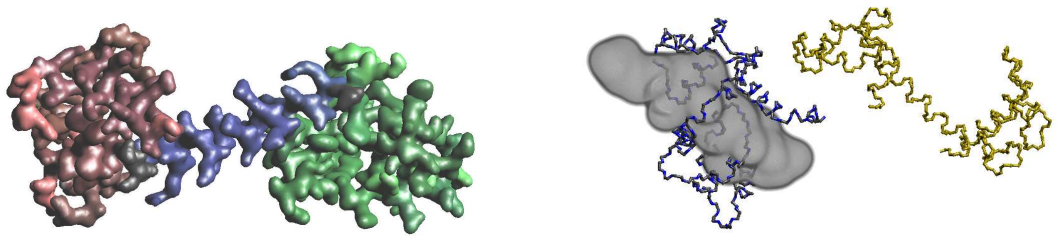
Immunoglobulin-Hen egg white lysozyme Complex (1VFB.pdb). The A B chains of the immunoglobulin is used as the flexible protein docking to the lysozyme. After domain analysis, we obtain two large domains at level 0. The flexor of this model is not associated with any shear as the interface area between the domains of the flexible immunoglobulin is only 101.433 Å^2. It is also a cut of the FCC and hence allows large bending motion at its hinge. Using this FCC model and rotations, we adaptively compute a set of conformations for the immunoglobulin. All backbone atoms of chain B and of residues from 1 to 107 of chain A were used in RMSD and fitting. We obtain RMSDs ranging from a high of 24.902 to a low of 0.586 Å. The original RMSD between the unbound and bound immunoglobulin is 3.784 Å. Hence we see a significantly closer conformation by our sampling.



(a) This consists of 2 distinct domains shown in green and red with 67, 64 residues, 12, 9 rigid segments and 11, 9 flexible loops.

(b) The colors represent the same structures as in figure 6. The original RMSD between the bound and unbound structures is 3.784 Å.

Figure 7: Antigen-lysozyme Antibody Complex



(a) Calmodulin consists of two domains (in red and green) linked by a third in blue through connectors in black. They have 55, 51, 18 residues and 4, 5, 0 flexible loops respectively.

(b) This molecule has been well studied due to its large conformation change on binding. Especially note the breakage of the central helix. We separate the bound and unbound structures to show the large conformational change.

Figure 8: Calmodulin with a kinase

Calmodulin bound to kinase (2BBM.pdb). We also decided to present calmodulin as it is known for its large conformational change. In this complex (2BBM.pdb), we use calmodulin and a target peptide given by a myosin light chain kinase. We let calmodulin be the flexible protein and to test the accuracy of Normal Mode Analysis based domain finding, we again use DomainFinder to compute domains. We obtain 3 domains for calmodulin. One of them contains the central helix. But from figure 8, we see that the central helix breaks into two during its conformational change for binding. Hence, we were unable to get a close RMSD to the bound protein from the unbound calmodulin (from 1CLL.pdb). The RMSD was computed using backbone atoms of residues 4 to 147. The RMSD of the unbound state was 14.429 Å!. The best RMSD found from rotating about the two flexors between domains 1,2 and domains 2,3 was just 8.590 Å. Hence we see that for large conformational changes, user input or analysis of more than one conformation is necessary.

4.3 Results on ZDock Benchmark

We ran F^3Dock on 3 of the 8 test cases that are tagged as the most difficult under ZDock Benchmark 2.0 [112]. The complexes we considered are: 1ATN, 1IBR and 2HMI. For each test case we took one of the two unbound proteins, and performed normal mode analysis using DomainFinder 1.1 in order to obtain a domain decomposition of the protein. See Table 3 for detailed information on the domains obtained and the parameters used. Using these domain definitions we obtained conformations of the corresponding unbound protein which are closer to the protein in the bound state. The best conformation obtained in each case under this metric is given in Table 4. These new conformations were then used for unbound-unbound docking with F^3Dock, and for each of the 3 test cases we were able to obtain docking results that are at least 0.8 Å closer to

Protein (From ZDock Benchmark 2.0)	Domains and Linkers		Parameters (DomainFinder1.1)
Actin [1IJJ.pdb (chain B) or 1ATN_r_u.pdb]	Domains (4):	Domain 1: 404 – 426, 478 – 543, 739 – 774; Domain 2: 436 – 464; Domain 3: 546 – 579, 671 – 736; Domain 4: 582 – 668	Normal Modes Calculated: 200 Normal Modes Used: 3 (modes 7 – 9) Deformation Threshold: 500 Domain Coarseness: 10
	Linkers (6):	Dom. 1 – Dom. 2: 427 – 435, 465 – 477; Dom. 1 – Dom. 3: 544 – 545, 737 – 738 Dom. 3 – Dom. 4: 580 – 581, 669 – 670	
Importin β [1F59.pdb (chain A) or 1IBR_l_u.pdb]	Domains (2):	Domain 1: 200 – 440; Domain 2: 2 – 194	Normal Modes Calculated: 210 Normal Modes Used: 3 (modes 7 – 9) Deformation Threshold: 200 Domain Coarseness: 45
	Linkers (1):	Dom. 1 – Dom. 2: 195 – 199	
FAB 28 [2HMI.pdb (chains C, D) or 2HMI_r_u.pdb]	Domains (3):	Domain 1: 429 – 552 (chain 1); Domain 2: 1 – 425 (chain 1) Domain 3: 1 – 427 (chain 2)	Normal Modes Calculated: 200 Normal Modes Used: 7 (modes 7 – 13) Deformation Threshold: 300 Domain Coarseness: 35
	Linkers (1):	Dom. 1 – Dom. 2: 426 – 428 (chain 1)	

Table 3: Domain decomposition (using DomainFinder 1.1) of three proteins from ZDock Benchmark 2.0. The domain definitions obtained from DomainFinder have been simplified (got rid of fragmentations) for convenience. The domain definition of Actin matches the one given in [81].

Protein 1					Docked Complex (computed by F3Dock)		
Reference (Bound) Conformation	Given Conformation		New Conformation (used by F3Dock)		Reference (Bound) Complex	Protein 2 (Unbound)	RMSD (Å) from Reference (Old RMSD)
	PDB (Unbound)	RMSD (Å) from Reference	Rotation Around the Shortest Linker between Domains 1 & 2	RMSD (Å) from Reference			
1ATN_r_b.pdb (1ATN: chain A)	1ATN_r_u.pdb (1IJJ: B)	2.70	Primary Axis: -5° Secondary Axis: -1° Third Axis: -30°	2.18	1ATN.pdb : A|D	1ATN_l_u.pdb (3DNI: _)	3.21 (4.01)
1IBR_l_b.pdb (1IBR: chain A)	1IBR_l_u.pdb (1F59: A)	2.94	Primary Axis: 0° Secondary Axis: 0° Third Axis: -20°	1.40	1IBR.pdb : A|B	1IBR_r_u.pdb (1QG4: A)	4.24 (5.88)
2HMI_r_b.pdb (2HMI: chains CD)	2HMI_r_u.pdb (2HMI: CD)	3.60	Primary Axis: -2° Secondary Axis: 12° Third Axis: 8°	2.94	2HMI.pdb : CD|AB	2HMI_l_u.pdb (1S6P: AB)	6.36 (7.23)

Table 4: Improved docking by F3Dock for three complexes from ZDock Benchmark 2.0. These are among the 8 complexes included in the benchmark that are categorized as difficult to dock. RMSD values are based on backbone atoms only.

the bound complex compared to rigid-body docking results obtained using original conformations. We used 6° of rotational sampling and 64^3 frquencies. Other details are given in Table 4.

In Figure 9 we show the docking results for 1IBR, which is a complex of Importin β and Ran GTPase. We performed flexibility analysis of both using DomainFinder [67, 68] and our flexibility model, and were able to decompose Importin β into two rigid domains (see Table 3 for details), while Ran GTPase turned out to be mostly rigid. In Figure $9(a)$ we show the conformation of Importin β in this complex, while Figure $9(b)$ shows the complex itself. As the figures show the two domains of Importin β form a grip-like conformation that holds Ran GTPase tightly between them. Figure $9(b)$ shows another conformation of Importin β which appears as chain A of 1F59.pdb, and Figure $9(d)$ shows the best rigid-body docking we were able to obtain between this conformation of Importin β and Ran GTPase (chain A of 1QG4.pdb). The RMSD distance (based on backbone atoms) between the two conformations of Importin β is 2.94 Å, while the two corresponding complexes (in Figures $9(b)$ and $9(d)$) are 5.88 Å apart. As Figure $9(d)$ shows the space inside the grip formed by the two domains of Importin β is too small for Ran GTPase to fit. In Figure $9(e)$ we show another conformation of Importin β that was obtained by applying a -20° bending motion around the third axis defined for the connector between its two domains (see Section 3.1.1 for the definitions of our axes of bending motion). In this new conformation the space inside the grip is larger than that in Figure $9(c)$, and it is 1.4 Å away from the reference conformation in Figure $9(a)$. Now if we dock Ran GTPase with this new

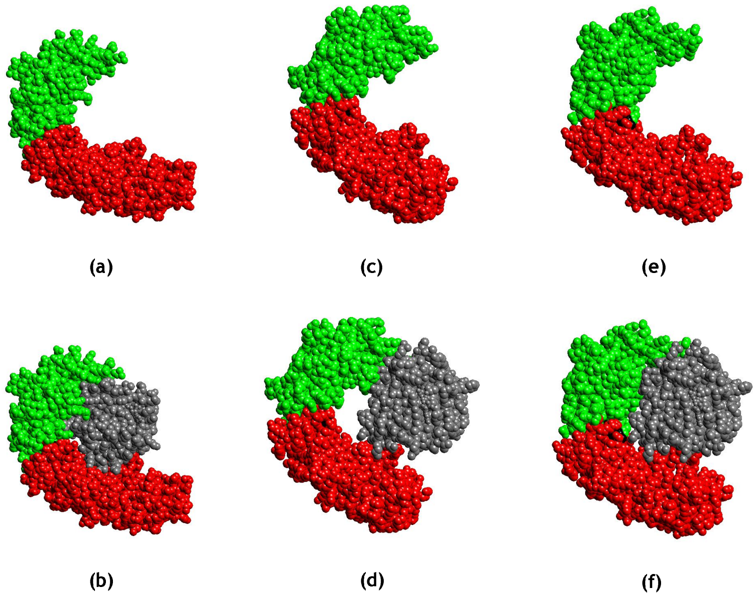


Figure 9: Flexible docking of Importin β (i.e., 1IBR_l_u.pdb or chain A of 1F59.pdb) and Ran GTPase (i.e., 1IBR_r_u.pdb or chain A of 1QG4.pdb) from ZDock Benchmark 2.0: **(a)** Conformation of protein 1 (i.e., 1IBR_l_u.pdb) in bound state (given as 1IBRł_b.pdb or 1IBR.pdb: chain B). **(b)** Bound complex of 1IBR_l_b.pdb (1IBR.pdb: chain B) and 1IBR_r_b.pdb (1IBR.pdb: chain A). **(c)** Domain decomposition of unbound protein 1 (i.e., 1IBR_l_u.pdb, or 1F59.pdb: chain A). Domains 1 and 2 are colored in red and green, respectively, and the linkers are colored in black. **(d)** Rigid-body docking of unbound proteins 1IBR_l_u.pdb (i.e., 1F59.pdb: chain A) and 1IBR_r_u.pdb (i.e., 1QG4.pdb: chain _). Protein 2 (i.e., 1IBR_r_u.pdb) is colored grey. **(e)** A new conformation generated from the given unbound protein 1IBR_l_u.pdb (i.e., 1F59.pdb: chain A). See Section 4.3 and Table 3 for details. **(f)** Flexible docking of the new conformation of 1IBR_l_u.pdb from (e) and given unbound protein 1IBR_r_u.pdb. See Table 4 for details.

conformation the docked complex, shown in Figure 9(f), is 4.24 Å away from the bound complex in Figure 9(b), which is an improvement of 1.64 Å from the rigid-body docking result shown in Figure 9(d).

5 Conclusions

Our algorithms are based on representing affinity functions in a multi-resolution radial basis function format. The smoothed particle representation, together with non-equispaced Fast Fourier transforms allows us to design and analyze our algorithm without the use of a sampling grid. The soft protein-protein docking algorithm is built upon accurate construction of molecular surfaces and properties. Its efficiency and multiresolution nature is utilized to sample conformational space and allow flexible docking. Flexibility models of proteins were created using Normal Mode Analysis. This was used to compute a diverse set of conformations which can be used in docking and effectively sampling orientation space. A simple greedy hueristic based algorithm for finer refitting of side chains at interfaces has also been presented. All of these interactions can be better studied by visual inspection of surfaces, functional properties and interfaces.

Acknowledgment

This work was supported in part by NSF grants IIS-0325550, CNS-0540033 and grants from the NIH 0P20 RR020647, R01 GM074258, R01-GM073087,R01-EB004873. We are grateful to Dr. Art Olson and Dr. Michel Sanner of TSRI, for many helpful discussions related to this project. Our in-house molecular modeling

and visualization software tool, called TexMol, was used as the visualization front end for docking. The TexMol program is in the public domain and can be freely downloaded from our center's software website (`http://www.ices.utexas.edu/CVC/software/`). Our rigid body docking algorithm F^2Dock [11] has been implemented as a web-based service (`http://cvcweb.ices.utexas.edu/cvc/f2dock/`) which is undergoing extensive tests at present.

References

[1] ABAGYAN, R., TOTROV, M., AND KUZNETSOV, D. ICM - new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry 15*, 5 (1994), 488–506.

[2] AKUTSU, T. NP-hardness results for protein side-chain packing. In *Genome Informatics* (1997), S. Miyano and T. Takagi, Eds., vol. 8, pp. 180–186.

[3] ALEXANDROV, N., AND SHINDYALOV, I. PDP: Protein Domain Parser. *Bioinformatics 19*, 3 (2003), 429–30.

[4] ALEXANDROV, V., LEHNERT, U., ECHOLS, N., MILBURN, D., ENGELMAN, D., AND GERSTEIN, M. Normal modes for predicting protein motions: A comprehensive database assessment and associated web tool. *Protein Sci 14*, 3 (2005), 633–643.

[5] ALTHAUS, E., KOHLBACHER, O., LENHOF, H., AND MÜLLER, P. A branch and cut algorithm for the optimal solution of the side-chain placement problem. Technical Report MPI-I-2000-1-001, Max-Planck-Institute für Informatik, 2000.

[6] ALTHAUS, E., KOHLBACHER, O., LENHOF, H.-P., AND MÜLLER, P. A combinatorial approach to protein docking with flexible side chains. *Journal of Computational Biology 9*, 4 (August 2002), 597–612.

[7] AMADEI, A., LINSSEN, A. B., AND BERENDSEN, H. J. Essential dynamics of proteins. *Proteins 17*, 4 (1993), 412–425.

[8] APOSTOLAKIS, J., PLÜCKTHUN, A., AND CAFLISCH, A. Docking small ligands in flexible binding sites. *Journal of Computational Chemistry 19*, 1 (1998), 21–37.

[9] BAHADUR, K. C. D., TOMITA, E., SUZUKI, J., AND AKUTSU, T. Protein side-chain packing problem: a maximum edge-weight clique algorithmic approach. *Journal of Bioinformatics and Computational Biology 3* (2005), 103–126.

[10] BAJAJ, C., BERNARDINI, F., AND SUGIHARA, K. A geometric approach to molecular docking and similarity. Technical Report CSD-TR-94-017, Computer Sciences, Purdue University, March 1994.

[11] BAJAJ, C., AND SIDDAVANAHALLI, V. F2Dock: A Fast and Fourier Based Error-Bounded Approach to Protein-Protein Docking. CS Technical report TR-06-57, The University of Texas at Austin, Austin, TX, USA 78712., November 2006.

[12] BAJAJ, C., AND SIDDAVANAHALLI, V. Fast error-bounded surfaces and derivatives computation for volumetric particle data. ICES Report 06-03, The University of Texas at Austin, January 2006.

[13] BAKER, T. S., OLSON, N. H., AND FULLER, S. D. Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiology and Molecular Biology Reviews 63*, 4 (1999), 862–922.

[14] BARRETT, C. P., HALL, B. A., AND NOBLE, M. E. M. Dynamite: a simple way to gain insight into protein motions. *Acta Crystallographica Section D 60*, 12 Part 1 (2004), 2280–2287.

[15] BARRETT, C. P., AND NOBLE, M. E. M. Dynamite extended: two new services to simplify protein dynamic analysis. *Bioinformatics 21*, 14 (2005), 3174–3175.

[16] BASTARD, K., PRÉVOST, C., AND ZACHARIAS, M. Accounting for loop flexibility during protein-protein docking. *Proteins 62*, 4 (2006), 956–969.

[17] BASTARD, K., THUREAU, A., LAVERY, R., AND PRÉVOST, C. Docking macromolecules with flexible segments. *Journal of Computational Chemistry 24*, 15 (2003), 1910–1920.

[18] BELNAP, D. M., KUMAR, A., FOLK, J. T., SMITH, T. J., AND BAKER, T. S. Low-resolution density maps from atomic models: How stepping "back" can be a step "forward". *J Struct Biol. 125*, 2-3 (1999), 166–175.

[19] BEN-ZEEV, E., KOWALSMAN, N., BEN-SHIMON, A., SEGAL, D., ATAROT, T., NOIVIRT, O., SHAY, T., AND EISENSTEIN, M. Docking to single-domain and multiple-domain proteins: Old and new challenges. *Proteins: Structure, Function, and Bioinformatics 60*, 2 (2005), 195–201.

[20] BONVIN, A. M. J. J. Flexible protein-protein docking. *Current Opinion in Structural Biology 16*, 2 (2006), 194–200.

[21] BOUTONNET, N., ROOMAN, M., AND WODAK, S. Automatic analysis of protein conformational changes of by multiple linkage clustering. *Journal of Molecular Biology 253*, 4 (1995), 633–647.

[22] BOWER, M. J., COHEN, F. E., AND DUNBRACK, R. L. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *Journal of Molecular Biology 267*, 5 (1997), 1268–1282.

[23] BROOKS, B., AND KARPLUS, M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc Natl Acad Sci USA 82*, 15 (1985), 4995–9.

[24] CAFLISCH, A., FISCHER, S., AND KARPLUS, M. Docking by monte carlo minimization with a solvation correction: Application to an fkbp-substrate complex. *Journal of Computational Chemistry 18*, 6 (1997), 723–743.

[25] CAMACHO, C. J. Modeling side-chains using molecular dynamics improve recognition of binding region in CAPRI targets. *Proteins: Structure, Function, and Bioinformatics 60*, 2 (2005), 245–251.

[26] CAMACHO, C. J., AND GATCHELL, D. W. Successful discrimination of protein interactions. *Proteins: Structure, Function, and Genetics 52*, 1 (2003), 92–97.

[27] CAMACHO, C. J., AND VAJDA, S. Protein docking along smooth association pathways. *Proceedings of the National Academy of Sciences 98*, 19 (2001), 10636–10641.

[28] CANUTESCU, A. A., SHELENKOV, A. A., AND DUNBRACK, R. L. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci 12*, 9 (2003), 2001–2014.

[29] CHAZELLE, B., KINGSFORD, C., AND SINGH, M. A semidefinite programming approach to side chain positioning with new rounding strategies. *Informs Journal on Computing 16*, 4 (2004), 380–392.

[30] CHEN, R., AND WENG, Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins: Structure, Function, and Genetics 47*, 3 (2002), 281–294.

[31] CHENG, Y., ZHANG, Y., AND MCCAMMON, J. A. How does activation loop phosphorylation modulate catalytic activity in the cAMP-dependent protein kinase: A theoretical study. *Protein Sci 15*, 4 (2006), 672–683.

[32] COLLINS, J. R., BURT, S. K., AND ERICKSON, J. W. Flap opening in hiv-1 protease simulated by 'activated' molecular dynamics. *Nature Structural Biology. 2*, 4 (April 1995), 334–338.

[33] DESJARLAIS, R. L., SHERIDAN, R. P., DIXON, J. S., KUNTZ, I. D., AND VENKATARAGHAVAN, R. Docking flexible ligands to macromolecular receptors by molecular shape. *Journal of Medicinal Chemistry 29*, 11 (November 1986), 2149–2153.

[34] DESMET, J., MAEYER, M. D., HAZES, B., AND LASTERS, I. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature 356*, 6369 (1992), 539–542.

[35] DESMET, J., MAEYER, M. D., AND LASTERS, I. Theoretical and algorithmical optimization of the dead-end elimination theorem. In *Pac. Symp. Biocomput.* (1997), pp. 122–133.

[36] DIJK, A. D. J. V., VRIES, S. J. D., DOMINGUEZ, C., CHEN, H., ZHOU, H. X., AND BONVIN, A. M. J. J. Data-driven docking: HADDOCK's adventures in CAPRI. *Proteins: Structure, Function, and Bioinformatics 60*, 2 (2005), 232–238.

[37] DOMINGUEZ, C., BOELENS, R., AND BONVIN, A. M. J. J. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc. 125*, 7 (2003), 1731–1737.

[38] DOMINGUEZ, C., BONVIN, A. M. J. J., WINKLER, G. S., VAN SCHAIK, F. M. A., TIMMERS, H. T. M., AND BOELENS, R. Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis, and docking approaches. *Structure 12*, 4 (2004), 633–644.

[39] DUNBRACK, R. L. Comparative modeling of casp3 targets using psi-blast and scwrl. *Protein: Structure, Function, and Genetics 3* (1999), 81–87.

[40] DUNBRACK, R. L. Rotamer libraries in the 21st century. *Current Opinion in Structural Biology 12*, 4 (2002), 431–440.

[41] DUNBRACK, R. L., AND COHEN, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci 6*, 8 (1997), 1661–1681.

[42] DUNBRACK, R. L., AND KARPLUS, M. Backbone-dependent rotamer library for proteins application to side-chain prediction. *Journal of Molecular Biology 230*, 2 (1993), 543–574.

[43] DUNBRACK, R. L., AND KARPLUS, M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat Struct Mol Biol 1*, 5 (1994), 334–340.

[44] EHRLICH, L. P., NILGES, M., AND WADE, R. C. The impact of protein flexibility on protein-protein docking. *Proteins: Structure, Function, and Bioinformatics 58*, 1 (2005), 126–133.

[45] EISENSTEIN, M., AND KATCHALSKI-KATZIR, E. On proteins, grids, correlations, and docking. *Comptes Rendus Biologies 327*, 5 (2004), 409–420.

[46] ELCOCK, A. H., SEPT, D., AND MCCAMMON, J. A. Computer simulation of protein-protein interactions. *J. Phys. Chem. B 105*, 8 (2001), 1504–1518.

[47] ERIKSSON, O., ZHOU, Y., AND ELOFSSON, A. Side chain-positioning as an integer programming problem. In *Algorithms in Bioinformatics*. 2001, pp. 128–141.

[48] FERNANDEZ-RECIO, J., TOTROV, M., AND ABAGYAN, R. Soft protein-protein docking in internal coordinates. *Protein Sci 11*, 2 (2002), 280–291.

[49] FERNÁNDEZ-RECIO, J., TOTROV, M., AND ABAGYAN, R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins: Structure, Function, and Genetics 52*, 1 (2003), 113–117.

[50] FITZJOHN, P. W., AND BATES, P. A. Guided docking: First step to locate potential binding sites. *Proteins: Structure, Function, and Genetics 52*, 1 (2003), 28–32.

[51] FLORY, P. J. Statistical thermodynamics of random networks. *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences 351*, 1666 (1976), 351–380.

[52] GABB, H. A., JACKSON, R. M., AND STERNBERG, M. J. E. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology 272*, 1 (1997), 106–120.

[53] GEHLHAAR, D. K., VERKHIVKER, G., FOGEL, P. A. R. D. B., FOGEL, L. J., AND FREER, S. T. Docking conformationally flexible small molecules into a protein binding site through simulated evolution. In *Evolutionary Programming IV: The Proc. of Fourth Annual Conference on Evolutionary Programming* (MIT Press, Cambridge, MA, 1995), J. McDonnell, R. Reynolds, and D. Fogel, Eds., pp. 615–627.

[54] GOLDSTEIN, R. F. Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys. J. 66*, 5 (1994), 1335–1340.

[55] GOODSELL, D. S., AND OLSON, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins:Structure, Function and Genetics 8*, 3 (1990), 195–202.

[56] GORDON, D. B., AND MAYO, S. L. Branch-and-terminate: a combinatorial optimization algorithm for protein design. *Structure 7*, 9 (1999), 1089–1098.

[57] GRAY, J. J. High-resolution protein-protein docking. *Current Opinion in Structural Biology 16*, 2 (2006), 183–193.

[58] GRAY, J. J., MOUGHON, S., WANG, C., SCHUELER-FURMAN, O., KUHLMAN, B., ROHL, C. A., AND BAKER, D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of Molecular Biology 331*, 1 (2003), 281–299.

[59] GRÜNBERG, R., LECKNER, J., AND NILGES, M. Complementarity of structure ensembles in protein-protein binding. *Structure 12*, 12 (2004), 2125–2136.

[60] HALPERIN, I., MA, B., WOLFSON, H., AND NUSSINOV, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Structure, Function, and Genetics 47*, 4 (2002), 409–443.

[61] HAYWARD, S. Identification of specific interactions that drive ligand-induced closure in five enzymes with classic domain movements. *Journal of Molecular Biology 339* (2004), 1001–1021.

[62] HAYWARD, S., AND BERENDSEN, H. J. C. Systematic analysis of domain motions in proteins from conformational change; new results on citrate synthase and T4 lysozyme. *Proteins, Structure, Function and Genetics 30*, 2 (February 1998), 144–154.

[63] HAYWARD, S., KITAO, A., AND BERENDSEN, H. J. C. Model-free methods of analyzing domain motions in proteins from simulation: A comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins: Structure, Function, and Genetics 27*, 3 (1997), 425–437.

[64] HAYWARD, S., AND LEE, R. A. Improvements in the analysis of domain motions in proteins from conformational change: Dyndom version 1.50. *J Mol Graph Model 21*, 3 (2002), 181–3.

[65] HEIFETZ, A., AND EISENSTEIN, M. Effect of local shape modifications of molecular surfaces on rigid-body protein-protein docking. *Protein Eng. 16*, 3 (2003), 179–185.

[66] HINSEN, K. Domainfinder user's manual. url: http://dirac.cnrs-orleans.fr/DomainFinder/documentation.html.

[67] HINSEN, K. Analysis of domain motions by approximate normal mode calculations. *Proteins: Structure, Function, and Genetics 33* (1998), 417–429.

[68] HINSEN, K. The molecular modeling toolkit: A new approach to molecular simulations. *Journal of Computational Chemistry 21*, 2 (2000), 79–85.

[69] HINSEN, K., REUTER, N., NAVAZA, J., STOKES, D. L., AND LACAPERE, J.-J. Normal mode-based fitting of atomic structure into electron density maps: Application to sarcoplasmic reticulum ca-atpase. *Biophys. J. 88*, 2 (2005), 818–827.

[70] HOLM, L., AND SANDER, C. Database algorithm for generating protein backbone and side-chain co-ordinates from a c[alpha] trace : Application to model building and detection of co-ordinate errors. *Journal of Molecular Biology 218*, 1 (1991), 183–194.

[71] HOLM, L., AND SANDER, C. Parser for protein folding units. *Proteins 19*, 3 (July 1994), 256–268.

[72] HWANG, H., PIERCE, B., MINTSERIS, J., JANIN, J., AND WENG, Z. Protein-protein docking benchmark version 3.0 (to appear). *Proteins: Structure, Function, and Bioinformatics* (2008).

[73] JACKSON, R. M., GABB, H. A., AND STERNBERG, M. J. E. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *Journal of Molecular Biology 276*, 1 (1998), 265–285.

[74] JACOBS, D. J., RADER, A. J., KUHN, L. A., AND THORPE, M. F. Protein flexibility predictions using graph theory. *Proteins: Structure, Function, and Genetics 44*, 2 (2001), 150–165.

[75] JANIN, J., HENRICK, K., MOULT, J., EYCK, L. T., STERNBERG, M. J. E., VAJDA, S., VAKSER, I., AND WODAK, S. J. CAPRI: A critical assessment of predicted interactions. *Proteins: Structure, Function, and Genetics 52*, 1 (2003), 2–9.

[76] JIANG, L., KUHLMAN, B., KORTEMME, T., AND BAKER, D. A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. *Proteins: Structure, Function, and Bioinformatics 58*, 4 (2005), 893–904.

[77] JONES, G., WILLETT, P., AND GLEN, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology 245*, 1 (January 1995), 43–53.

[78] JONES, G., WILLETT, P., GLEN, R. C., LEACH, A. R., AND TAYLOR, R. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology 267*, 3 (April 1997), 727–748.

[79] JONES, S., AND THORNTON, J. M. Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences 93*, 1 (1996), 13–20.

[80] JUDSON, R. S., JAEGER, E. P., AND TREASURYWALA, A. M. A genetic algorithm based method for docking flexible molecules. *Journal of Molecular Structure: THEOCHEM 308* (May 1994), 191–206.

[81] KABSCH, W., MANNHERZ, H. G., SUCK, D., PAI, E. F., AND HOLMES, K. C. Atomic structure of the actin: DNase I complex. *Nature 347* (1990), 37–44.

[82] KALE, L., SKEEL, R., BHANDARKAR, M., BRUNNER, R., GURSOY, A., KRAWETZ, N., PHILLIPS, J., SHINOZAKI, A., VARADARAJAN, K., AND SCHULTEN, K. Namd2: greater scalability for parallel molecular dynamics. *J. Comput. Phys. 151*, 1 (1999), 283–312.

[83] KELLER, D. A., SHIBATA, M., MARCUS, E., ORNSTEIN, R. L., AND REIN, R. Finding the global minimum: a fuzzy end elimination implementation. *Protein Eng. 8*, 9 (1995), 893–904.

[84] KESKIN, O., JERNIGAN, R. L., AND BAHAR, I. Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophysical Journal 78*, 4 (April 2000), 2093–2106.

[85] KINGSFORD, C. L., CHAZELLE, B., AND SINGH, M. Solving and analyzing side-chain positioning problems using linear and integer programming. *Bioinformatics 21*, 7 (2005), 1028–1039.

[86] KROEMER, R. T., VULPETTI, A., MCDONALD, J. J., ROHRER, D. G., TROSSET, J.-Y., GIORDANETTO, F., COTESTA, S., MCMARTIN, C., KIHLÉN, M., AND STOUTEN, P. F. W. Assessment of docking poses: interactions-based accuracy classification (ibac) versus crystal structure deviations. *Journal of Chemical Information and Computer Sciences 44* (2004), 871–888.

[87] KUNTZ, I. D., BLANEY, J. M., OATLEY, S. J., LANGRIDGE, R., AND FERRIN, T. E. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology 161*, 2 (October 1982), 269–288.

[88] LASTERS, I., AND DESMET, J. The fuzzy-end elimination theorem: correctly implementing the side chain placement algorithm based on the dead-end elimination theorem. *Protein Eng. 6*, 7 (1993), 717–722.

[89] LASTERS, I., MAEYER, M. D., AND DESMET, J. Enhanced dead-end elimination in the search for the global minimum energy conformation of a collection of protein side chains. *Protein Eng. 8*, 8 (1995), 815–822.

[90] LAUGHTON, C. A. Prediction of protein side-chain conformations from local three-dimensional homology relationships. *Journal of Molecular Biology 235*, 3 (1994), 1088–1097.

[91] LEACH, A. R. Ligand docking to proteins with discrete side-chain flexibility. *Journal of Molecular Biology 235*, 1 (January 1994), 345–356.

[92] LEACH, A. R., AND KUNTZ, I. D. Conformational analysis of flexible ligands in macromolecular receptor sites. *Journal of Computational Chemistry 13*, 6 (1992), 730–748.

[93] LEACH, A. R., AND LEMON, A. P. Exploring the conformational space of protein side chains using dead-end elimination and the a* algorithm. *Proteins: Structure, Function, and Genetics 33*, 2 (1998), 227–239.

[94] LEE, C., AND SUBBIAH, S. Prediction of protein side-chain conformation by packing optimization. *Journal of Molecular Biology 217*, 2 (1991), 373–388.

[95] LEE, J., SCHERAGA, H. A., AND RACKOVSKY, S. New optimization method for conformational energy calculations on polypeptides: Conformational space annealing. *Journal of Computational Chemistry 18*, 9 (1997), 1222–1232.

[96] LEE, K., CZAPLEWSKI, C., KIM, S.-Y., AND LEE, J. An efficient molecular docking using conformational space annealing. *Journal of Computational Chemistry 26*, 1 (2005), 78–87.

[97] LEE, K., SIM, J., AND LEE, J. Study of protein-protein interaction using conformational space annealing. *Proteins: Structure, Function, and Bioinformatics 60*, 2 (2005), 257–262.

[98] LEECH, J., PRINS, J. F., AND HERMANS, J. Smd: Visual steering of molecular dynamics for protein design. *IEEE Computational Science and Engineering 3*, 4 (Winter 1996), 38–45.

[99] LEI, M., ZAVODSZKY, M. I., KUHN, L. A., AND THORPE, M. F. Sampling protein conformations and pathways. *Journal of Computational Chemistry 25*, 9 (2004), 1133–1148.

[100] LI, L., CHEN, R., AND WENG, Z. RDOCK: Refinement of rigid-body protein docking predictions. *Proteins: Structure, Function, and Genetics 53*, 3 (2003), 693–707.

[101] LI, X., KESKIN, O., MA, B., NUSSINOV, R., AND LIANG, J. Protein-protein interactions: Hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: Implications for docking. *Journal of Molecular Biology 344*, 3 (2004), 781–795.

[102] LOOGER, L. L., AND HELLINGA, H. W. Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *Journal of Molecular Biology 307*, 1 (2001), 429–445.

[103] LOVELL, S. C., WORD, J. M., RICHARDSON, J. S., AND RICHARDSON, D. C. The penultimate rotamer library. *Proteins: Structure, Function, and Genetics 40*, 3 (2000), 389–408.

[104] MA, B., ELKAYAM, T., WOLFSON, H., AND NUSSINOV, R. Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proceedings of the National Academy of Sciences 100*, 10 (2003), 5772–5777.

[105] MA, J., AND KARPLUS, M. The allosteric mechanism of the chaperonin groel: A dynamic analysis. *Proceedings of the National Academy of Sciences USA. 95*, 15 (July 1998), 8502–8507.

[106] MAEYER, M. D., DESMET, J., AND LASTERS, I. All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of sidechains by dead-end elimination. *Folding and Design 2*, 1 (1997), 53–66.

[107] MAIOROV, V. N., AND ABAGYAN, R. A. A new method for modeling large-scale rearrangements of protein domains. *Proteins 27* (1997), 410–424.

[108] MANDELL, J. G., ROBERTS, V. A., PIQUE, M. E., KOTLOVYI, V., MITCHELL, J. C., NELSON, E., TSIGELNY, I., AND TEN EYCK, L. F. Protein docking using continuum electrostatics and geometric fit. *Protein Eng. 14*, 2 (2001), 105–113.

[109] MAY, A., AND ZACHARIAS, M. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics 1754*, 1-2 (2005), 225–231.

[110] MCGREGOR, M. J., ISLAM, S. A., AND STERNBERG, M. J. E. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *Journal of Molecular Biology 198*, 2 (1987), 295–310.

[111] MCMARTIN, C., AND BOHACEK, R. S. QXP: powerful, rapid computer algorithms for structure-based drug design. *Journal of Computer-Aided Molecular Design 11* (1997), 333–344.

[112] MINTSERIS, J., WIEHE, K., PIERCE, B., ANDERSON, R., CHEN, R., JANIN, J., AND WENG, Z. Protein-protein docking benchmark 2.0: An update. *Proteins: Structure, Function, and Bioinformatics 60*, 2 (2005), 214–216.

[113] MIZUTANI, M. Y., TOMIOKA, N., AND ITAI, A. Rational automatic search method for stable docking models of protein and ligand. *Journal of Molecular Biology 243*, 2 (October 1994), 310–326.

[114] MÉNDEZ, R., LEPLAE, R., LENSINK, M. F., AND WODAK, S. J. Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins: Structure, Function, and Bioinformatics 60*, 2 (2005), 150–169.

[115] MÉNDEZ, R., LEPLAE, R., MARIA, L. D., AND WODAK, S. J. Assessment of blind predictions of protein-protein interactions: Current status of docking methods. *Proteins: Structure, Function, and Genetics 52*, 1 (2003), 51–67.

[116] MORRIS, G. M., GOODSELL, D. S., HALLIDAY, R. S., HUEY, R., HART, W. E., BELEW, R. K., AND OLSON, A. J. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry 19*, 14 (1998), 1639–1662.

[117] MUSTARD, D., AND RITCHIE, D. W. Docking essential dynamics eigenstructures. *Proteins: Structure, Function, and Bioinformatics 60*, 2 (2005), 269–274.

[118] NAKAGAWA, A., MIYAZAKI, N., TAKA, J., NAITOW, H., OGAWA, A., FUJIMOTO, Z., MIZUNO, H., HIGASHI, T., WATANABE, Y., OMURA, T., CHENG, R., AND TSUKIHARA, T. The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins. *Structure 11* (2003), 1227–1238.

[119] NOLA, A. D., ROCCATANO, D., AND BERENDSEN, H. J. C. Molecular dynamics simulation of the docking of substrates to proteins. *Proteins 19*, 3 (July 1994), 174–182.

[120] OSHIRO, C. M., KUNTZ, I. D., AND DIXON, J. S. Flexible ligand docking using a genetic algorithm. *Journal of Computer-aided Molecular Design 9*, 2 (April 1995), 113–130.

[121] PALMA, P. N., KRIPPAHL, L., WAMPLER, J. E., AND MOURA, J. J. G. BiGGER: A new (soft) docking algorithm for predicting protein interactions. *Proteins: Structure, Function, and Genetics 39*, 4 (2000), 372–384.

[122] PIERCE, N. A., SPRIET, J. A., DESMET, J., AND MAYO, S. L. Conformational splitting: A more powerful criterion for dead-end elimination. *Journal of Computational Chemistry 21*, 11 (2000), 999–1009.

[123] PIERCE, N. A., AND WINFREE, E. Protein design is NP-hard. *Protein Eng. 15*, 10 (2002), 779–782.

[124] PONDER, J. W., AND RICHARDS, F. M. Tertiary templates for proteins : Use of packing criteria in the enumeration of allowed sequences for different structural classes. *Journal of Molecular Biology 193*, 4 (1987), 775–791.

[125] RAJAMANI, D., THIEL, S., VAJDA, S., AND CAMACHO, C. J. Anchor residues in protein-protein interactions. *Proceedings of the National Academy of Sciences 101*, 31 (2004), 11287–11292.

[126] RAREY, M., KRAMER, B., AND LENGAUER, T. J. Multiple automatic base selection: Protein-ligand docking based on incremental construction without manual intervention. *Journal of Computer-Aided Molecular Design 11* (1997), 369–384.

[127] RAREY, M., KRAMER, B., LENGAUER, T. J., AND KLEBE, G. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology 261*, 3 (August 1996), 470–489.

[128] ROHL, C. A., STRAUSS, C. E. M., CHIVIAN, D., AND BAKER, D. Modeling structurally variable regions in homologous proteins with rosetta. *Proteins: Structure, Function, and Bioinformatics 55*, 3 (2004), 656–677.

[129] ROSSMANN, M. G. Fitting atomic models into electron-microscopy maps. *Acta Crystallographica D56* (2000), 1341–1349.

[130] RUSSELL, R. B., ALBER, F., ALOY, P., DAVIS, F. P., KORKIN, D., PICHAUD, M., TOPF, M., AND SALI, A. A structural perspective on protein-protein interactions. *Current Opinion in Structural Biology 14*, 3 (2004), 313–324.

[131] SAIBIL, H. R. Conformational changes studied by cryo-electron microscopy. *Nat Struct Mol Biol 7*, 9 (2000), 711–714. 1072-8368 10.1038/78923 10.1038/78923.

[132] SALI, A., AND BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology 234*, 3 (1993), 779–815.

[133] SANDAK, B., NUSSINOV, R., AND WOLFSON, H. J. An automated computer vision and robotics-based technique for 3-d flexible biomolecular docking and matching. *Computer Applications in the Biosciences : CABIOS 11*, 1 (February 1995), 87–99.

[134] SCHNEIDMAN-DUHOVNY, D., INBAR, Y., NUSSINOV, R., AND WOLFSON, H. J. Geometry-based flexible and symmetric protein docking. *Proteins: Structure, Function, and Bioinformatics 60*, 2 (2005), 224–231.

[135] SCHNEIDMAN-DUHOVNY, D., NUSSINOV, R., AND WOLFSON, H. J. Predicting molecular interactions in silico: Ii. protein-protein and protein- drug docking. *Current Medicinal Chemistry 11* (2004), 91–107.

[136] SIDDIQUI, A. S., AND BARTON, G. J. Continuous and discontinuous domains: an algorithm for the automatic generation of reliable protein domain definitions. *Protein Science 4*, 5 (May 1995), 872–884.

[137] SKEEL, R. D., TEZCAN, I., AND HARDY, D. J. Multiple grid methods for classical molecular dynamics. *Journal of Computatioanl Chemistry 23*, 6 (2002), 673–684.

[138] SMITH, G. R., FITZJOHN, P. W., PAGE, C. S., AND BATES, P. A. Incorporation of flexibility into rigid-body docking: Applications in rounds 3-5 of CAPRI. *Proteins: Structure, Function, and Bioinformatics 60*, 2 (2005), 263–268.

[139] SMITH, G. R., AND STERNBERG, M. J. E. Prediction of protein-protein interactions by docking methods. *Current Opinion in Structural Biology 12*, 1 (2002), 28–35.

[140] SMITH, G. R., STERNBERG, M. J. E., AND BATES, P. A. The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking. *Journal of Molecular Biology 347*, 5 (2005), 1077–1101.

[141] STERNBERG, M. J. E., GABB, H. A., AND JACKSON, R. M. Predictive docking of protein–protein and protein–dna complexes. *Current Opinion in Structural Biology 8*, 2 (1998), 250–256.

[142] STONE, J. E., GULLINGSRUD, J., AND SCHULTEN, K. A system for interactive molecular dynamics simulation. In *Proceedings of the 2001 symposium on Interactive 3D graphics* (2001), ACM Press, pp. 191–194.

[143] SUHRE, K., AND SANEJOUAND, Y.-H. Elnemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucl. Acids Res. 32*, suppl 2 (2004), W610–614.

[144] TAI, K. Conformational sampling for the impatient. *Biophysical Chemistry 107*, 3 (2004), 213–220.

[145] TAMA, F., GADEA, F. X., MARQUES, O., AND SANEJOUAND, Y. H. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins 41*, 1 (October 2000), 1–7.

[146] TAMA, F., MIYASHITA, O., AND BROOKS, C. L. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *Journal of Molecular Biology 337*, 4 (2004), 985–999.

[147] TEODORO, M., GEORGE N. PHILLIPS, J., AND KAVRAKI., L. E. Molecular docking: A problem with thousands of degrees of freedom. *IEEE International Conference on Robotics and Automation, Seoul, Korea* (2001).

[148] TEODORO, M. L., PHILLIPS, JR., G. N., AND KAVRAKI, L. E. A dimensionality reduction approach to modeling protein flexibility. In *Proceedings of the sixth annual international conference on Computational biology* (2002), ACM Press, pp. 299–308.

[149] TIRION, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical Review Letters 77*, 9 (1996), 1905.

[150] TOTROV, M., AND ABAGYAN, R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 åaccuracy. *Nat Struct Mol Biol 1*, 4 (1994), 259–263.

[151] TUFFERY, P., ETCHEBEST, C., HAZOUT, S., AND LAVERY, R. A new approach to the rapid determination of protein side chain conformations. *J. Biomol.Struct. Dyn. 8* (1991), 1267–1289.

[152] VAKSER, I. A. Protein docking for low-resolution structures. *Protein Eng. 8*, 4 (1995), 371–378.

[153] VOIGT, C. A., GORDON, D. B., AND MAYO, S. L. Trading accuracy for speed: a quantitative comparison of search algorithms in protein sequence design. *Journal of Molecular Biology 299*, 3 (2000), 789–803.

[154] VOLKMANN, N., AND HANEIN, D. Quantitative fitting of atomic models into observed densities derived by electron microscopy. *Journal of Structural Biology 125* (1999), 176–184.

[155] WANG, C., SCHUELER-FURMAN, O., AND BAKER, D. Improved side-chain modeling for protein-protein docking. *Protein Sci 14*, 5 (2005), 1328–1339.

[156] WANG, J., KOLLMAN, P. A., AND KUNTZ, I. D. Flexible ligand docking: A multistep strategy approach. *Proteins: Structure, Function, and Genetics 36*, 1 (October 1999), 1–19.

[157] WELCH, W., RUPPERT, J., AND JAIN, A. N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chemistry & Biology 3*, 6 (June 1996), 449–462.

[158] WOLFSON, H. J., SHATSKY, M., SCHNEIDMAN-DUHOVNY, D., DROR, O., SHULMAN-PELEG, A., MA, B., AND NUSSINOV, R. From structure to function: Methods and applications. *Current Protein and Peptide Science 6* (2005), 171–183.

[159] WRIGGERS, W., AND SCHULTEN, K. Protein domain movements: Detection of rigid domains and visualization of effective rotations in comparisons of atomic coordinates. *Proteins: Structure, Function, and Genetics 29* (1997), 1–14.

[160] XIANG, Z., AND HONIG, B. Extending the accuracy limits of prediction for side-chain conformations. *Journal of Molecular Biology 311*, 2 (2001), 421–430.

[161] XU, J. Rapid protein side-chain packing via tree decomposition. In *Research in Computational Molecular Biology*. 2005, pp. 423–439.

[162] XU, J., AND BERGER, B. Fast and accurate algorithms for protein side-chain packing. *Journal of the ACM 53*, 4 (2006), 533–557.

[163] Xu, Y., Xu, D., and Gabow, H. N. Protein domain decomposition using a graph-theoretic approach. *Bioinformatics 16*, 12 (2000), 1091–1104.

[164] Zacharias, M. Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci 12*, 6 (2003), 1271–1282.

[165] Zacharias, M. Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: Binding of FK506 to FKBP. *Proteins: Structure, Function, and Bioinformatics 54*, 4 (2004), 759–767.

[166] Zavodszky, M. I., Lei, M., Thorpe, M. F., Day, A. R., and Kuhn, L. A. Modeling correlated main-chain motions in proteins for flexible molecular recognition. *Proteins: Structure, Function, and Bioinformatics 57*, 2 (2004), 243–261.

[167] Zehfus, M. H., and Rose, G. D. Compact units in proteins. *Biochemistry 25*, 19 (September 1986), 5759–5765.

[168] Zhao, Y., Stoffler, D., and Sanner, M. Hierarchical and multi-resolution representation of protein flexibility. *Bioinformatics 22*, 22 (2006), 2768–2774.

[169] Zheng, W., and Brooks, B. R. Normal-modes-based prediction of protein conformational changes guided by distance constraints. *Biophys. J. 88*, 5 (2005), 3109–3117.

[170] Zhou, Z. H., Baker, M. L., Jiang, W., Dougherty, M., Jakana, J., Dong, G., Lu, G., and Chiu, W. Electron cryomicroscopy and bioinformatics suggest protein fold models for rice dwarf virus. *Nature Structural Biology 8*, 10 (2001), 868–873.