F^{2}Dock: Fast Fourier Protein-Protein Docking

Chandrajit Bajaj, *Member, IEEE,* Rezaul Chowdhury, *Student Member, IEEE,*
Vinay Siddavanahalli, *Student Member, IEEE*

Abstract—The functions of proteins is often realized through their mutual interactions. Determining a relative transformation for a pair of proteins and their conformations which form a stable complex, reproducible in nature, is known as docking. It is an important step in drug design, structure determination and understanding function and structure relationships. In this paper we extend our non-uniform fast Fourier transform docking algorithm to include an adaptive search phase (both translational and rotational) and thereby speed up its execution. We have also implemented a multithreaded version of the adaptive docking algorithm for even faster execution on multicore machines. We call this protein-protein docking code F^{2}Dock (F^2 = Fast Fourier) . We have calibrated F^{2}Dock based on an extensive experimental study on a list of benchmark complexes and conclude that F^{2}Dock works very well in practice. Though all docking results reported in this paper use shape complementarity and Coulombic potential based scores only, F^{2}Dock is structured to incorporate Lennard-Jones potential and re-ranking docking solutions based on desolvation energy .

Index Terms—Computational Structural Biology, Protein-Protein Interactions, Fast Fourier Methods, Algorithms, Docking, Redocking

1 INTRODUCTION

PROTEINS are stable, folded chains of amino acid polymers, and together with lipids (fats and oils), carbohydrates (e.g., sugars) and nucleic acids (DNA and RNA) form the structural and functional building blocks in our cells. Functions of these building blocks, and particularly those of proteins are expressed through their mutual structural interactions. For example, inhibitors bind to enzymes to limit their rate of reaction. Another example is the attachment of immunoglobins to antigens like viruses, in order to signal that these antigens are foreign objects in our cells. Hence the study of protein-protein interactions plays an important role in understanding the processes of life [1]. In particular, as the two preceding examples suggest, protein-protein interaction is at the core of structure-based drug design. Though advancements in X-ray crystallography and other imaging techniques have lead to the extraction of near atomic resolution information for numerous individual proteins, the creation, crystallization and imaging of macromolecular complexes, as extensively required for drug design, still remains a difficult task. Flexibility of proteins makes the search for the required conformation through experimentation even more difficult. Hence, the need for fast and robust computational approaches to predicting the structures of protein-protein interactions is growing[2]. An important step towards understanding protein-protein interactions is *protein-protein docking* which can be defined as computationally finding the best relative transformation and conformation of two proteins that results in a stable complex, reproducible in nature (if one exists). If only large, fairly inflexible proteins are involved, *rigid protein-protein docking* can be performed as an initial step. Rigid docking based on structure alone has shown to be adequate for a range of proteins[3].

There are two main aspects of a docking algorithm:

(1) scoring or measuring the quality of any given docked complex, and
(2) searching for the highest scoring or a pool of high quality docking conformations

Shape complementarity along the docked interface is seen to one of the primary measure of docking quality. Other factors which contribute to the formation of stable complexes include electrostatics, hydrophobicity, hydrogen bonds, solvation energy etc. [2], [4]. These, together with shape complementarity are known as *affinity functions*. The docking problem can be viewed as the search for stable minimum energy complexes. The energy function has several major terms.

(i) The *Lennard-Jones* 12-6 dispersion-repulsion potential is given by $\sum_{i,j}\left(\frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^{6}}\right)$, where r_{ij} is the distance between two given atoms, and a_{ij} and b_{ij} are constants based on atom types.
(ii) The *electrostatic potential* is given by $\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}}$, where q_i and q_j are Coulombic charges, and $\varepsilon(r_{ij})$ is a distance dependant dielectric constant. Electrostatics plays a role in long range interaction due to partially charged protein and solvent atoms.
(iii) *Desolvation energy* is defined as the change in energy due to the displacement of solvent molecules from the interface. The desolvation free energy for moving an atom of charge q and radius r from a region of dielectric ε_1 to a region of dielectric ε_2, is given by $\frac{q^2}{r}(\frac{1}{\varepsilon_1} - \frac{1}{\varepsilon_2})$. The total desolvation energy is the sum of desolvation energies of individual atoms involved.
(iv) Docking energy computations also involve change in energy due to hydrophobicity, hydrogen bond formation and conformational changes. Given the affinity functions,

• *C. Bajaj and R. Chowdhury are with the Computational Visualization Center, Department of Computer Sciences and The Institute of Computational Engineering and Sciences, The University of Texas at Austin, 1 University Station C0500, Austin, Texas 78712, USA.*
E-mails: {bajaj,shaikat}@cs.utexas.edu
• *V. Siddavanahalli is with Google Inc., Mountain View, California.*

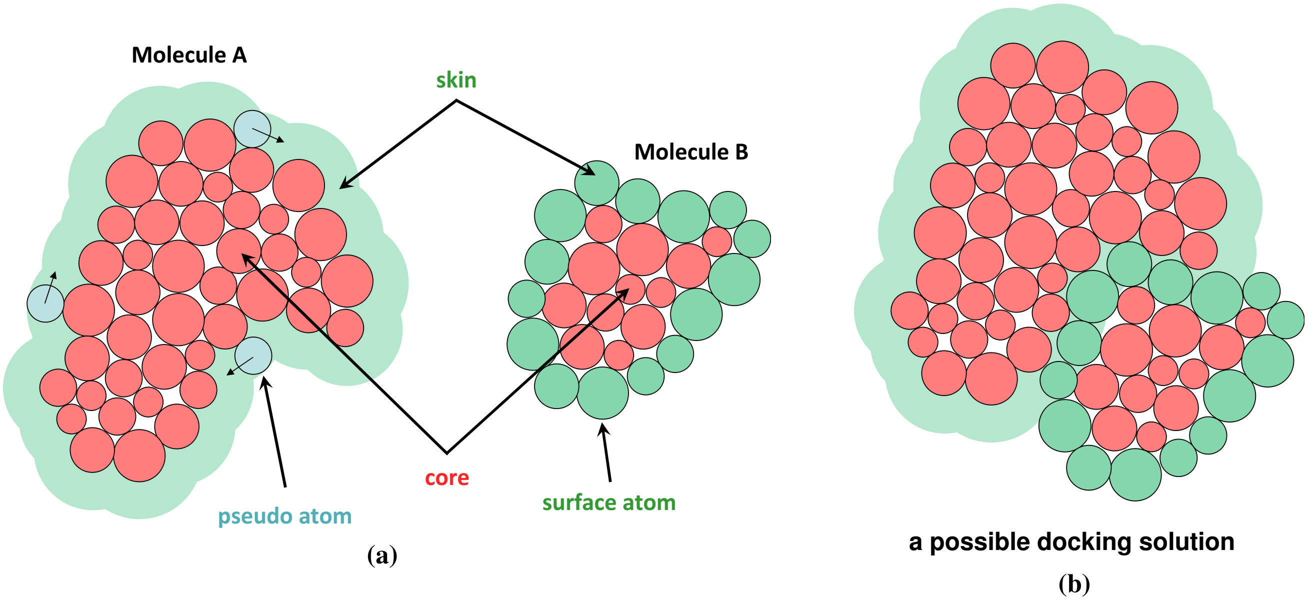


Fig. 1. (a) Skin and Core regions for complementary space docking. Atoms are drawn as solid circles. The skins regions are colored green while the core regions are red. The skin volume of molecule A is obtained by rolling a solvent ball over its surface. (b) A possible docking of the molecules show a large overlap between the grown layer of molecule A and the surface atoms of molecule B.

and a scoring method, a search is performed over all of transformation and conformation spaces to find where the two given proteins fit best.

Shape based complementarity, coupled with electrostatic compatibility is typically used as an initial step to obtain possible docking sites. These sites are further ranked using other energy terms. The few remaining potential docking sites are then tested using energy minimization routines.

In [5] we described a Non-equispaced Fast Fourier (NFFT) based algorithm for efficiently performing the initial docking search (based on shape and electrostatics complementarity). We presented a sum of Gaussians based model for proteins, and described a new specification of the rigid protein-protein docking problem. Given two proteins A and B with M_A and M_B atoms, respectively, our algorithm spends $O(max(M_A, M_B) + n^3 \log n + \rho n^3)$ time to find the top ρ peaks in the docking profile, and n is a parameter chosen to satisfy a user required accuracy in the docking profile. We showed that for a summation of Gaussians model for the molecule where atoms are represented as Gaussian kernels, n^3 varies as $O(max(M_A, M_B))$. Compared to traditional grid based Fourier docking algorithms, the algorithm was shown to have lower computational complexity and memory requirement.

In this paper we extend our non-uniform fast Fourier transform(NFFT) based docking algorithm to include an adaptive search phase (both translational and rotational) and thus speed up its execution. We have also implemented a multithreaded version of the adaptive docking algorithm for even faster execution on multicore machines. We call this protein-protein docking code F^{2}Dock (F^2 = Fast Fourier) . We have calibrated F^{2}Dock based on an extensive experimental study on a list of benchmark complexes and conclude that F^{2}Dock works very well in practice. Though all docking results reported in this paper use shape complementarity and Coulombic potential based scores only, F^{2}Dock is structured to incorporate Lennard-Jones potential and re-ranking docking solutions based on desolvation energy . In our consider three scenarios of pairwise rigid protein-protein docking. The first is known as redocking, where a given complex of two proteins, are first separated, randomly rotated and translated, and then redocked. In this case the top docking solutions are compared with the original complex, and the RMSD (root mean square deviation) error measure computed. The second scenario is known as bound-unbound docking, where one of the two proteins is in the same conformation as in a complex, while the conformation of the second protein is independent and unknown from the one in the complex. Again the RMSD of the solution dockings are computed with respect to the original complex. The third and final docking scenario is the unbound-unbound case, where both proteins are in unknown conformations with respect to those in the complex. All three docking scenarios have the same computational complexity.

The rest of the paper is organized as follows. In Section 2 we include a review of prior work on rigid protein-protein docking. In Section 3 we describe our new algorithm with adaptive translational and rotational search. We include our experimental results with F^{2}Dock on ZDock Benchmark Suite 2.0 [6] in Section 4. Finally, in Section 5 we include some concluding remarks and plans for future research.

2 Related Work

There have been a wide range of work on both flexible and rigid-body docking. In this Section we discuss some relevant prior work on rigid-body docking. Please see the technical report on our flexible docking algorithm F^{3}Dock [7] for a review of known techniques for docking flexible molecules.

Graph theory based docking methods [8], [9], [10] reduce the shape complementarity based molecular fitting problems into combinatorial search that have well developed algorithms. However, some good potential matches may be ignored during search due to the use of pruning for reducing the cost of

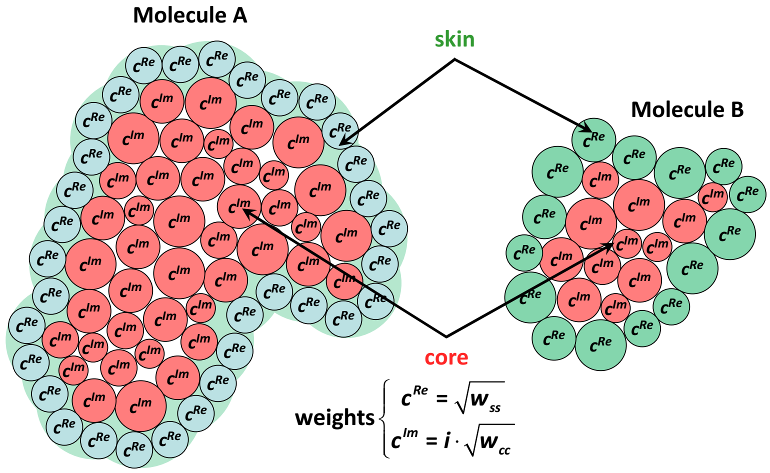


Fig. 2. For shape-complementarity scoring skin atoms are assigned a weight of $c^{Re} = \sqrt{w_{ss}}$, and core atoms are assigned weight $c^{Im} = i \cdot \sqrt{w_{cc}}$, where w_{ss} is the reward factor for skin-skin overlaps, and w_{cc} is the penalty factor for core-core overlaps.

combinatorial search. Geometry-based docking methods use a first level assumption that molecules will 'dock' if the receptor and the ligand exhibit very high shape (surface and volume) complementarity. Point-wise spherical approximations, surface normals, etc. have also been considered in characterizing shape complementarity. In [11], [12] spheres are used to represent grooves in one protein and the density of the other. It was later used in a geometric hashing scheme [13], [14], [15], [16], [17], [18] where a search strategy based on matching pairs of consistent spheres, one from each protein was used, instead of a full combinatorial search. In [19] the combinatorial search was reduced to a clique finding problem by considering pairwise distances among atoms. A knob and hole detection and matching algorithm was used in [20], [21] where an optimization is performed using a *grid-based double skin layer approach* in 2D. We shall further discuss this double skin layer approach later as we use a variation of it in our algorithm. A full 6D grid based search was used in [22] which also provides a method to uniformly sample 3D rotational space. Using geometric features such as pockets, holes, and surface normals, these methods attempt to constrain the search areas to relatively small portions of the receptorŠs surface. Geometric signatures/feature points were also used in earlier geometry-based docking methods [13], [23]. However, geometric signature based approaches often have difficulties in dealing with molecular surfaces without notable features such as flat regions. These methods are also quite sensitive to small geometric feature changes, and a large amount of hashing of storage space is needed for complicated ligand/receptor geometries. Some relatively recent surface and 3-D shape matching methods could be customized to improve the efficiency of geometric surface-surface docking. For example, including molecular properties into the scoring function would necessarily move the geometry matching problem to higher than three dimensions. Belongie et al. [24] calculate shape matches by using shape contexts to describe the relation of the shape to a certain point on the shape. Since corresponding points on two similar shapes will have similar shape contexts, the matching problem is reduced to an optimal point pair assignment problem between two shapes. This technique has reduced sensitivity to small variations in the two shapes.

Using some representation of molecular surface boundary (skin), and a correlation/scoring function based on cumulative overlap of characteristic (electron density) functions of molecular shape, rigid docking can be performed by conducting a combinatorial search in a six dimensional parameter space of all possible translations and orientations of a rigid protein relative to another rigid protein. In [25] coarse grids and rotational angles are used to reduce the combinatorics of the search. The combinatorics of possible relative conformations can be reduced by using a priori knowledge of suitable binding site locations on the proteins [3]. Fast Fourier Transforms can be used to speed up the cumulative scoring function computations [25], [3], [26]. The grid based double skin layer approach became the base of many variations and software, e.g., DOT [27], ZDOCK [28], [29], [30] and RDOCK [31]. Hydrogen bonds were used in [32] to reduce the rotational sampling space and improve the scoring function. Spherical harmonics based approached were studied in [33], [34], [26], [35], [36], [37], [38]. We have compared our algorithm to previous grid based Fourier transform and Spherical harmonics approaches in [5].

There have also been other approaches including building webs over the surfaces and matching them using least squares fit [39], a slice based matching scheme [40], mapping surfaces to 2D matrices and detection of matching sub matrices [41] and fixing anchors and searching over other degrees of freedom (TreeDock [42]). A simulated annealing method, by choosing angles in discrete 45 degree steps and translations of 2Å is used in [43] to perform a random walk and dock proteins. In [44], a coarse approximation of the protein is obtained by approximating each residue by a single spheres, and furthermore the 6D docking search space is parameterized by 5 rotations and 1 translation. The 5D rotational space is further sampled using simulated annealing techniques.

3 Algorithm Details

Consider two proteins A and B, with M_A and M_B atoms respectively. We represent the molecules using Gaussian kernels,

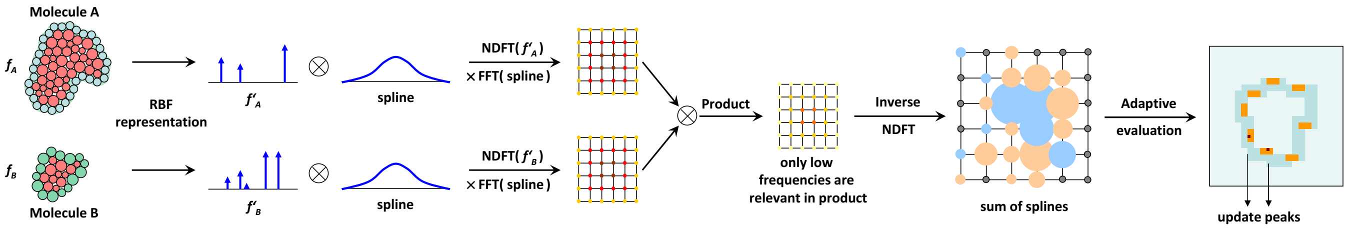


Fig. 3. Overview of the translational search phase of the F^{2}Dock algorithm. Here f_A and f_B are affinity functions of molecule A and B, respectively. We assume that a given rotation has already been applied on molecule B.

construct double skin layers used for complementary space docking and derive a new model for docking.

3.1 Affinity Functions

The affinity functions are modeled as Radial Basis Functions (RBFs) to facilitate using Fourier transforms to efficiently solve the docking problem.

We use the sum of Gaussian's representation to model our proteins. An atom centered at $\mathbf{x}_c$, with a van der Waal's radius of r, is modeled as an isotropic Gaussian kernel: $g(\mathbf{x}-\mathbf{x}_c) = e^{-\beta\left(\frac{(\mathbf{x}-\mathbf{x}_c)^2}{r^2}-1\right)}$. The decay rate of the kernel is controlled by the blobbiness parameter β. A value of 2.3 is used in the literature [45] to approximate the solvent excluded surface at an isovalue of 1. By lowering this parameter, we can model molecules at lower resolutions [46].

3.1.1 Shape Complementarity

For shape based docking we maximize the overlap of the surface of protein B with the complementary space of A. The *double skin layer* approach is used here. It was introduced in [21] for 2D, [22] for 3D, sped up using Fast Fourier Transforms in [47], and extended to complex space in [29]. We define two *skin regions*:

1. The complementary region of A, defined by a *grown skin region*, by introducing a 1-layer of pseudo-atoms on the surface of A. Typically each pseudo-atoms has the same radius which is chosen to make its size comparable to that of a solvent molecule.
2. The *surface skin* of B, which is the density function of the set of surface atoms of B.

The atoms of A and the inner atoms of B form *core regions*. These regions are shown in Figure 1. We use an adaptive grid based algorithm to construct these regions [5].

To maximize skin overlaps and to minimize overlaps of the cores, we assign positive imaginary weights to the core atoms and positive real weights to the skin atoms/pseudo-atoms (see Figure 2). An integral of the superposition of the molecules has two real contributions: the core overlaps contribute negatively and the skin overlaps contribute positively. The magnitude of the imaginary part of the integral due to skin-core clashes (caused by psuedo-atom vs atom overlaps) are also non-desirable and assigned a 'smaller' negative weight in the accumulated score.

The weighted sum of Gaussians function definition of a molecule $P \in \{A, B\}$ with M_P atoms be expressed as follows:

$$\begin{aligned} f_P^{SC}(\mathbf{x}) &= \sum_{k\in skin(P)} c^{Re} g_k(\mathbf{x}-\mathbf{x}_k) + \sum_{k\in core(P)} c^{Im} g_k(\mathbf{x}-\mathbf{x}_k) \\ &= \sum_{k=1}^{M_P} c_k g_k(\mathbf{x}-\mathbf{x}_k), \end{aligned}$$

where, g is the Gaussian function located at each atom (or pseudo atom) and (SC) stands for shape complementarity. The weights $\{c_k \in \{c^{Im}, c^{Re}\}, k = 1, \ldots, M_P\}$ are either positive imaginary or positive real. See also [30] for an extension of shape complementarity to *pairwise shape complementarity*.

3.1.2 Electrostatics Interactions

Similar to the procedure used for shape complementarity, Gabb et. al. [3] have shown how to introduce the electrostatics term. The first protein's electric potential is computed and matched against the charges in the other. This can also be sped up using a Fourier based algorithm. Charge assignments are made using PDB2PQR [48]). We define two new affinity functions f_A^E and f_B^E for molecule A and B, respectively.

$$f_A^E(\mathbf{x}) = \sum_{k=1}^{M_A} q_k \frac{1}{E(\mathbf{x}-\mathbf{x}_k)(\mathbf{x}-\mathbf{x}_k)}$$

$$\text{and} \quad f_B^E(\mathbf{x}) = \sum_{k=1}^{M_B} q_k \delta(\mathbf{x}-\mathbf{x}_k),$$

where, q_k is the Coulombic charge on atom k, $\delta(\mathbf{x})$ is the Kronecker delta function with value 1 at $||\mathbf{x}|| = 0$, and 0 everywhere else, and $E(\mathbf{x})$ is the distance dependent dielectric constant [3] as given below.

$$E(\mathbf{x}) = \begin{cases} 4 & \text{if } ||\mathbf{x}|| \leq 6\text{Å}, \\ 80 & \text{if } ||\mathbf{x}|| > 8\text{Å}, \\ 38 \cdot ||\mathbf{x}|| - 224 & \text{otherwise.} \end{cases}$$

3.2 Rigid Docking Model Specification

Let T and Δ denote the translational and the rotational operators, respectively. If the user considers a potential docking site as one where the overlap potential (plus electrostatics potential if electrostatics interactions are used) is over a threshold τ, then the rigid protein-protein docking solution, using our affinity functions definition, is expressed as the set of triplets:

$$\left\{ (\mathbf{t},\mathbf{r},s) : \left(\begin{array}{l} s = Re\left(F_{A,B}^{SC}(\mathbf{t},\mathbf{r}) - w_E \cdot F_{A,B}^{E}(\mathbf{t},\mathbf{r})\right) \\ \quad - \frac{w_{SC}}{\sqrt{w_{SS}\cdot w_{CC}}} \cdot Im\left(F_{A,B}^{SC}(\mathbf{t},\mathbf{r})\right) \end{array} \right) \geq \tau \right\} \quad (1)$$

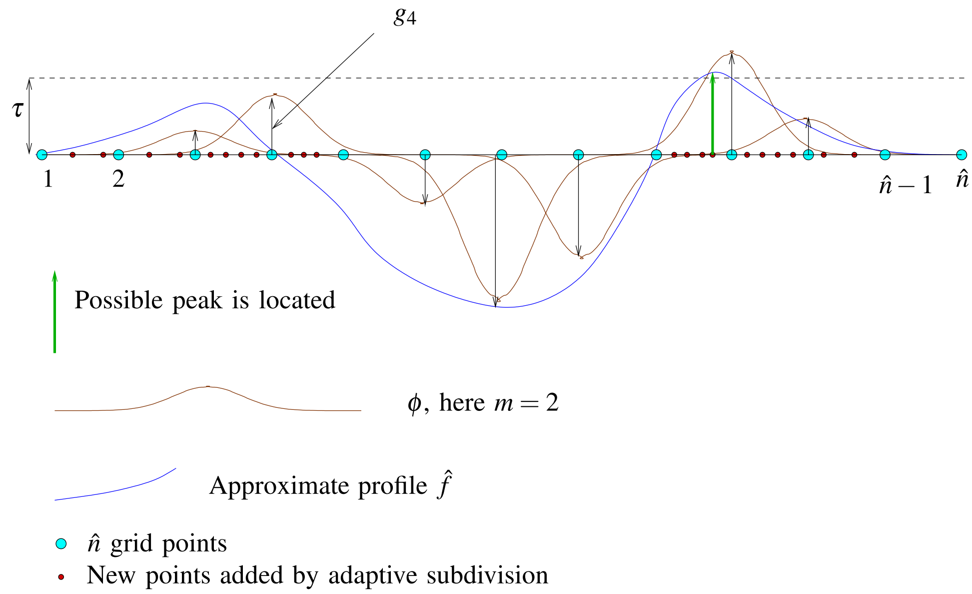


Fig. 4. The docking peak search can be represented as finding the peak positions and values in a grid of overlapping splines.

where,

$$F^{SC}_{A,B}(\mathbf{t},\mathbf{r}) = \int_{\mathbf{x}} f^{SC}_A(\mathbf{x})\, T_{\mathbf{t}}\left(\Delta_{\mathbf{r}}\left(f^{SC}_B(\mathbf{x})\right)\right) d\mathbf{x},$$

$$F^{E}_{A,B}(\mathbf{t},\mathbf{r}) = \int_{\mathbf{x}} f^{E}_A(\mathbf{x})\, T_{\mathbf{t}}\left(\Delta_{\mathbf{r}}\left(f^{E}_B(\mathbf{x})\right)\right) d\mathbf{x},$$

w_{ss} = reward for (unit) skin-skin overlap,
w_{cc} = penalty for (unit) core-core overlap,
w_{sc} = penalty for (unit) skin-core overlap, and
w_E = reward for (unit) charge-complementarity.

This model assumes that each skin atom is assigned a positive real weight of $c^{Re} = \sqrt{w_{ss}}$, and each core atom is assigned a positive imaginary weight of $c^{Im} = \sqrt{w_{cc}}$ (see Figure 2).

3.3 Search

We solve Equation 1 using Fourier series expansions. Shape complementarity scores and electrostatics scores are computed separately, and then combined. For simplicity of exposition, we describe below our search algorithm for the following simpler case where both w_{sc} and w_E are set to 0. Generalization to Equation 1 is straight-forward.

$$\left\{(\mathbf{t},\mathbf{r},s) : \left(s = Re\left(F^{SC}_{A,B}(\mathbf{t},\mathbf{r})\right)\right) \geq \tau\right\} \quad (2)$$

We express the integral as a sum of compactly supported radial basis functions and provide an adaptive algorithm to search for regions where the scoring function exceeds the threshold provided by the user.

3.3.1 Fourier Series Expansions

Any periodic integrable function can be expanded as a Fourier series. For example, a periodic function in $[-1/2, 1/2]$ can be expressed as: $q(x) = \sum_{j=-\infty}^{\infty} \omega_j e^{2\pi i j x}$, where the coefficients $\omega_j = \int_{-1/2}^{1/2} q(x) e^{-2\pi i j x} dx$. Let I_n denote a 3D grid of integer indices: $\{k : [-n/2..n/2)^3, k \in \mathscr{Z}^3\}$. Let us expand the kernel function in its Fourier series form: $g(\mathbf{x}-\mathbf{x}_k) = \sum_{\boldsymbol{\omega}\in I_\infty} G_{\boldsymbol{\omega}} e^{2\pi i(\mathbf{x}-\mathbf{x}_k).\boldsymbol{\omega}}$. Hence, the affinity function $f^{SC}_P(\mathbf{x}) = \sum_{k=1}^{M_P} c_k g(\mathbf{x}-\mathbf{x}_k)$ can be expressed as $f^{SC}_P(\mathbf{x}) = \sum_{k=1}^{M_P} c_k (\sum_{\boldsymbol{\omega}\in I_\infty} G_{\boldsymbol{\omega}} e^{2\pi i(\mathbf{x}-\mathbf{x}_k).\boldsymbol{\omega}})$. Rearranging terms, we obtain: $f^{SC}_P(\mathbf{x}) = \sum_{\boldsymbol{\omega}\in I_\infty} G_{\boldsymbol{\omega}} e^{2\pi i\mathbf{x}.\boldsymbol{\omega}} \sum_{k=1}^{M_P} c_k e^{-2\pi i \mathbf{x}_k.\boldsymbol{\omega}}$. Let us denote the second terms by $C_{\boldsymbol{\omega}}$. Hence, $f^{SC}_P(\mathbf{x}) = \sum_{\boldsymbol{\omega}\in \mathbf{I}_\infty} \mathbf{G}_{\boldsymbol{\omega}} \mathbf{C}_{\boldsymbol{\omega}} \mathbf{e}^{\mathbf{2\pi i x.}\boldsymbol{\omega}}$. Similarly: $f^{SC}_P(\mathbf{x}-\mathbf{y}) = \sum_{\boldsymbol{\omega}\in \mathbf{I}_\infty} \mathbf{G}_{\boldsymbol{\omega}} \mathbf{C}_{\boldsymbol{\omega}} \mathbf{e}^{\mathbf{2\pi i(x-y).}\boldsymbol{\omega}}$.

Expanding f^{SC}_A and f^{SC}_B using the above series, for a given rotation $\mathbf{r}$, with the molecules scaled to lie in $\pi^3 = (-0.5..0.5]^3$ for simpler mathematical notation, the scoring integral in Equation 2 reduces to

$$\forall\, \mathbf{x} : \int_{\mathbf{y}\in\pi^3} f^{SC}_A(\mathbf{y})(\Delta_{\mathbf{r}}(f^{SC}_B))(\mathbf{x}-\mathbf{y}) d\mathbf{y}$$
$$= \int_{\mathbf{y}\in\pi^3} \sum_{\boldsymbol{\omega}_A\in I_\infty} G_{\boldsymbol{\omega}_A} C_{\boldsymbol{\omega}_A} e^{2\pi i\mathbf{y}.\boldsymbol{\omega}_A} \sum_{\boldsymbol{\omega}_B\in I_\infty} G_{\boldsymbol{\omega}_B} C'_{\boldsymbol{\omega}_B} e^{2\pi i(\mathbf{x}-\mathbf{y}).\boldsymbol{\omega}_B} d\mathbf{y}$$

Since $\int_{-1/2}^{1/2} e^{2\pi i y(a-b)} = 1$ if $a = b$ and 0 otherwise, the integral reduces to $\sum_{\boldsymbol{\omega}\in I_\infty} G^2_{\boldsymbol{\omega}} C_{\boldsymbol{\omega}} C'_{\boldsymbol{\omega}} e^{2\pi i\mathbf{x}.\boldsymbol{\omega}}$.

3.3.2 Approximations

We make three approximations in computing the above coefficients. Since the truncated Gaussian is a decaying kernel, we choose to compute only the first $(-n/2..n/2]^3$ Fourier coefficients. The parameter n is chosen to satisfy a user required accuracy in the docking profile. If we include electrostatics, the decay should be even slower, and hence, the same bounds derived for shape complementarity should be sufficient. The current analysis, though, is based on shape complementarity. The Fourier coefficients of the atoms centers, $C_{\boldsymbol{\omega}}, C'_{\boldsymbol{\omega}}$ are approximated as $\hat{C}_{\boldsymbol{\omega}}, \hat{C}'_{\boldsymbol{\omega}}$, computed using a Nonequispaced

Algorithm 1 Inverse adaptive peak search

Inputs :
  -$\hat{n}^3$: number of frequencies
  -h: accuracy of peak position
  -ϕ: Compactly supported smooth decaying function at each $k \in I_{\hat{n}}$
  -τ: threshold for docking score
  -$\{(val, pos)\}$: Current output peak regions and scores
Preprocessing: [Interval set: $I = intervals(k)$]
while $I \neq \emptyset$ **do**
  $interval \leftarrow I.next(\ )$
  if interval.isLowRes( ) **then**
    $t \leftarrow 0,\ \ \{\phi\} \leftarrow interval.overlapping\phi(\ )$
    for $\phi \in \{\phi\}$ **do**
      if $\phi > 0$ **then**
        if $interval.isOutside(\phi)$ **then**
          $t \leftarrow t + \phi(interval.fIdx(\phi.center))$
        else
          $t \leftarrow t + \phi_{max}$
        end if
      else
        $t \leftarrow t - \phi(interval.fIdx(\phi.center))$
      end if
    end for
    if (t $> \tau$) **then**
      $I \leftarrow I \cup interval.subIntervals(\ )$ [midpoint subdivision based on h]
    end if
  else
    $update(\{(val, pos)\}, interval)$
  end if
end while
Output: $[\{(val, pos)\}]$

Fast Fourier Transform (NFFT) algorithm given in [49] (Very briefly, the NFFT algorithm computes an approximation to Fourier coefficients when input data is not uniformly sampled). The truncated Gaussian is a tensor product kernel. The Fourier coefficients of the truncated Gaussians are now approximated as the tensor product $\hat{G}_{\boldsymbol{\omega}}$. Hence, we approximate the scoring integral as $\sum_{\boldsymbol{\omega} \in I_n} \hat{G}^2_{\boldsymbol{\omega}} \hat{C}_{\boldsymbol{\omega}} \hat{C}'_{\boldsymbol{\omega}} e^{2\pi i \mathbf{x}.\boldsymbol{\omega}} = \sum_{\boldsymbol{\omega} \in I_n} \hat{F}_{\boldsymbol{\omega}} e^{2\pi i \mathbf{x}.\boldsymbol{\omega}}$.

3.3.3 Inverse Peak Search

Given the function $\hat{f}(\mathbf{x}) = \sum_{\boldsymbol{\omega} \in I_n} \hat{F}_{\boldsymbol{\omega}} e^{2\pi i \mathbf{x}.\boldsymbol{\omega}}$, we are required to compute $\{(\mathbf{x}, s) : s = Re(\hat{f}(\mathbf{x})) \geq \tau\}$. A 3D IFFT (Inverse nonequispaced fast Fourier transform) of $\hat{F}_{\boldsymbol{\omega}}$ yields the docking profile $\hat{f}(\mathbf{x})$ at a uniform sampling. If we have prior knowledge on the smoothness of the profile, we can zero pad $\hat{F}_{\boldsymbol{\omega}}$ (if necessary) and obtain the profile at a sufficient sampling. This would generally lead to higher computational and memory requirements. Instead, we perform an adaptive computation of $\hat{F}_{\boldsymbol{\omega}}$, progressively zooming in on regions where the threshold τ is satisfied. Using the NFFT algorithm in [49], we make the following approximation: $\hat{f}(\mathbf{x}) \approx \hat{g}(\mathbf{x}) = \sum_{\mathbf{k} \in I_{\hat{n},m}(\boldsymbol{\omega}_\mathbf{j})} g_k \phi(\boldsymbol{\omega}_\mathbf{j} - \mathbf{k}/\hat{n})$, $(\mathbf{j} \in I_n,\ \hat{n} = \alpha n,\ \alpha \approx 2, I_{\hat{n},m}(\boldsymbol{\omega}_\mathbf{j}) = \{\mathbf{l} \in I_{\hat{n}} : \hat{n}\boldsymbol{\omega}_\mathbf{j} - m \leq \mathbf{l} \leq \hat{n}\boldsymbol{\omega}_\mathbf{j} + m\})$. This is schematically represented in 1D in Figure 4. Obtaining regions which are above a certain threshold is now reduced to finding roots of the polynomial $Re(\hat{g}(\mathbf{x})) = \tau$ If we use a cubic Bspline function for ϕ with a support width of 5, it requires the root of a 7x7x7 system of degree 5 equations. We instead adaptively compute regions which satisfy our docking threshold using an adaptive search algorithm. We initially start with the $\hat{n}^3$ grid of ϕ as a set of intervals. We determine using a simple procedure if any interval can potentially contain a value greater than the docking threshold and, if so, subdivide and recursively search the sub intervals. Consider any interval I. There are multiple ϕ functions whose summation determine the function in I. If we change these ϕ, such that positive ones centered outside I come closer by one interval width, negative ones shift away from I by one interval width and positive ones centered inside I are given its maximum value, the sum of the new function (called ψ) at the interval endpoints defines an upper bound for the original function ϕ and $\hat{g}(\mathbf{x})$ inside I. This upper bound function yields an approximate profile to our score $\hat{f}(\mathbf{x})$ and provides us with a test function for determining where to further subdivide and refine an interval as we locate the positive peaks of the scoring function.

The docking score profile is usually large in a thin closed region (as skin-skin overlaps occur in a relatively small subset of 3D space) with zeros on the outside and large negatives on the inside. Hence, in the very first step of the algorithm, a large number of regions are removed from further consideration. We are able to reduce the full 3D inverse FFT of $\hat{F}_{\boldsymbol{\omega}}$ which yields the docking profile $\hat{f}(\mathbf{x})$ in the first step of our adaptive search into an inverse FFT of size $\hat{n}^3$. This is an efficient way of speeding up the overall inverse peak search algorithm 1. We provide an analysis in 1D, which can be easily extended to 3D. Consider an interval $[i, i+1]$, with B-spline functions ϕ_k, where $i - m \leq k \leq i + 1 + m$, capturing both positive and negative peaks of $\hat{F}_{\boldsymbol{\omega}}$. Let the extent of the ϕ_k be m on each side of k. We construct a new upper bound function ψ_k (to construct an approximate scoring profile , by raising the value of ϕ_k to $max(\phi_k, \phi_{k+1}, \phi_{k-1})$ on the $\hat{n}^3$ grid. This gives us the following simple observation:

Lemma 3.1. *The summation of ψ values at a point k in the low resolution grid of the Gaussian centers is always greater than the summation of ϕ values at any point in any interval which includes k.*

The approximate docking profile, $\hat{f}(\mathbf{x}) \approx \hat{g}(\mathbf{x}) = \sum_{\mathbf{k} \in I_{\hat{n},m}(\boldsymbol{\omega}_\mathbf{j})} g_k \psi(\boldsymbol{\omega}_\mathbf{j} - \mathbf{k}/\hat{n})$ is a summation of smooth functions, and is now computed over a uniform interval of $\hat{n}^3$ points. This summation of smooth functions is equivalent to a convolution of a discretely sampled kernel function ψ with discrete values of g, namely g_k. The convolution of ψ and g is, as is well known, equivalent to the inverse Fourier transform, of the product of the Fourier transforms of ψ and g respectively and hence computable using 3D FFT in $O(n^3 \log n)$ as the first step of our algorithm. This initial uniform coarse approximation of the docking profile eliminates most regions outside the overlap of skin and core clashes. Hence, our adaptive search is then limited to a narrower region where the skin-skin overlaps occur, which yield the maximum positive values to the docking profile.

Figure 3 gives an overview of the adaptive translation search phase of F^2Dock.

3.3.4 Rotational Sampling

For the orientational degrees of freedom we use the optimized and uniform sampling described in [27]. The sampling is based on Euler angles, and the rotations are applied on molecule B. Each rotational step is followed by a 3D translational search as described in preceding sections. For 20° of mean rotational spacing the number of samples obtained is 1,800, while for 6° there are 54,000 sample rotations. Rotational search can also be made adaptive as follows. We first perform a low resolution rotational search, say, of mean rotational spacing of R_1, and retain only those rotations for which translational search yield solutions above a user-specified threshold. Then for each of these retained coarse rotations we perform a finer rotational search, say, of mean rotational spacing of $R_2 < R_1/4$, within a cone of angular radius $R_1/2$ around the coarse rotational sample under consideration. As before we retain only rotations that produce solutions above the given threshold during translational search. Such adaptive refinement steps can be repeated with finer and finer rotational samplings until some given level of accuracy is reached.

4 Experimental Results

We have computed docking predictions for a set of 84 complexes obtained from the ZDock Benchmark Suite 2.0 [6]. For soft docking we first use shape complementarity (i.e. van der Waal's interactions) as the affinity function in scoring. Then we investigate the effects of introducing electrostatics interactions.

We performed three types of docking experiments:

Bound-bound (Redocking). Both molecules A and B are taken from the bound complex involving A and B, and they are then computationally redocked.

Bound-unbound. One molecule, say A, is taken from the bound complex involving A and B, and the other one, i.e., B, is taken from another known independent structure of B.

Unbound-unbound. Neither A nor B is taken from the bound complex involving A and B, that is, each of them comes from an independent structure that does not include the other molecule.

In all experiments, we measured the quality of our docking solution based on its RMSD distance from the known bound structure of the two molecules involved. RMSD was calculated using the C_α atoms within 5Å of the interface of the bound structure. We used Kabsch's optimal vector alignment algorithm [50], [51] for aligning the two sets of interface atoms during RMSD computation. We had F^2Dock output the top 50,000 solutions ranked based on the score it assigns to each solution. We claimed a 'hit' if there was a solution with RMSD less than 5 Å among the top 2,000 solutions returned by F^2Dock. A rotational sampling of 6 degrees was used, and unless specified otherwise, the number of frequencies extracted by FFT is 32^3. Adaptive search was not used for obtaining the results reported in this section.

4.1 Unbound-unbound Docking

Tables 1 and 2 shows the results of running F^2Dock on the 84 complexes of ZDock Benchmark Suite 2.0 [6] for unbound-unbound docking using shape complementarity only. We used four different sets of weight values given to the skin-skin (w_{ss}), core-core (w_{cc}) and skin-core (w_{sc}) overlap costs. In the tables 'Rank' is the best rank among all predicted positions whose RMSD from the known bound structure was less than 5Å. 'Good Peaks' is the number of peaks in the predicted set which were less than 5Å RMSD from the known position. In the 'RMSD' column in the tables we report the lowest RMSD among all peaks that were retained. We also list the ZDock results in the last column. ZDock used 6° rotational sampling like F^2Dock, but retained 54,000 peaks. The RMSD computation procedure is also based on C_α atoms within 5Å of the interface.

We observe from Tables 1 and 2 that the number of hits slightly increased as w_{cc} is increased from 5 to 10 (with w_{ss} and w_{sc} held constant at 1.0 and 0.5, respectively), and increased even further if w_{sc} is increased from 0.5 to 1.0. However, increasing w_{cc} further to 20 did not seem to increase the number of hits anymore. Moreover, increasing w_{cc} from 5 to 10 generally improved the lowest RMSD value of the predictions, but increasing w_{cc} even further or increasing w_{sc} from 0.5 to 1.0 generally worsened the lowest RMSD. We also observe that ZDock performed better than F^2Dock in most cases under these parameter settings.

In Figure 5 we show the best docking positions we obtained during unbound-unbound docking of the following four complexes: (a) Ribonuclease A complexed with Rnase inhibitor, (b) Epstein-Barr virus receptor CR2 complexed with Complement C3, (c) Cyt C peroxidase complexed with Cytochrome C, and (d) Colicin E7 nuclease complexed with Im7 immunity protein.

In Table 3 we report the results of incorporating the approximate electrostatics interactions score computed by our method into the docking score. We used 1.0, 10.0 and 1.0 as skin-skin (w_{ss}), core-core (w_{cc}) and skin-core (w_{sc}) weights, respectively. Electrostatics based affinity function is defined using a model by Gabb [3]. The dielectric value is set to 4 for distances less than 6 Å from the center of atoms, 80 for greater than 8 Å and a linear interpolation in between. The electrostatics weight (w_E) was set to an empirically determined value of 350 which seems to improve the 'Rank' for the largest number of complexes when w_{ss}, w_{cc} and w_{sc} are set to 1.0, 10.0 and 1.0, respectively. We observe that adding the electrostatics score improved the 'Rank' of 45 out of 84 complexes ($\approx 53\%$), while for 24 complexes ($\approx 29\%$) solutions actually degraded. Among the complexes with improved 'Rank' values, 42 had their 'Rank' improved by at least 10, 30 by at least 100, and 15 by at least 1,000. There are 2 complexes ((1) 1K5D: Ran GTPase complexed with Ran GAP, and (2) 1ML0: Viral

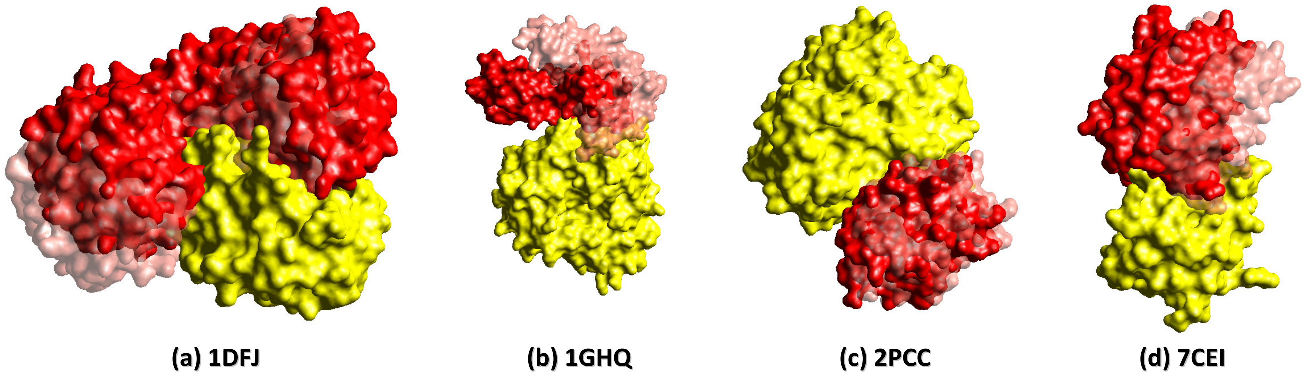


Fig. 5. Unbound-unbound docking: **(a)** **(1DFJ: Ribonuclease A complexed with Rnase inhibitor)** Docking the unmarked chain of 2BNH.pdb (Rnase inhibitor) on chain B (Ribonuclease A) of 9RSA.pdb, **(b)** **(1GHQ: Epstein-Barr virus receptor CR2 complexed with Complement C3)** Docking chain A (Complement C3) of 1LY2.pdb on the unmarked chain (Epstein-Barr virus receptor CR2) of 1C3D.pdb, **(c)** **(2PCC: Cyt C peroxidase complexed with Cytochrome C)** Docking the unmarked chain (Cytochrome C) of 1YCC.pdb on the unmarked chain (Cyt C peroxidase) of 1CCP.pdb, and **(d)** **(7CEI: Colicin E7 nuclease complexed with Im7 immunity protein)** Docking chain B (Im7 immunity protein) of 1M08.pdb on chain D (Colicin E7 nuclease) of 1UNK.pdb. In all cases the first chain is static (colored yellow), and the other chain is moved around for docking. The position of the moving molecule shown in pink corresponds to the true solution (obtained by the best superimposition of each molecule on the corresponding molecule in the bound structure) while red is our final docked position.

			F^{2}Dock Results ($w_{ss}=1.0$, frequencies = 32^3)												ZDock
Data			$w_{cc}=5.0$ $w_{sc}=0.5$			$w_{cc}=10.0$ $w_{sc}=0.5$			$w_{cc}=10.0$ $w_{sc}=1.0$			$w_{cc}=20.0$ $w_{sc}=1.0$			Results
Bound Complex	Unbound Mol 1	Unbound Mol 2	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)	RMSD (Å)
1A2K_C:AB	1QG4_A	1OUN_AB	2	15,258	4.37	29	19,083	3.02	36	8,100	3.02	29	5,565	3.19	1.61
1ACB_E:I	2CGA_B	1EGL_	1,913	361	2.55	1,117	480	2.89	569	803	3.08	328	1,282	3.08	2.54
1AHW_AB:C	1FGN_LH	1TFH_A	1	46,475	4.77	23	13,916	1.65	36	6,516	1.65	44	3,844	1.65	0.89
1AK4_A:D	2CPL_	1E6J_P	604	84	3.43	248	91	3.49	110	160	3.49	95	207	3.49	2.01
1AKJ_AB:DE	2CLR_DE	1CD8_AB	1,412	16	1.54	961	165	1.45	679	102	1.45	381	79	1.45	1.24
1ATN_A:D	1IJJ_B	3DNI_	8	8,017	4.68	8	3,889	4.68	4	19,423	4.72	1	32,962	4.72	3.87
1AVX_A:B	1QQU_A	1BA7_B	725	408	1.58	470	723	1.58	339	1,769	1.75	198	870	1.88	0.76
1AY7_A:B	1RGH_B	1A19_B	491	156	0.80	420	100	0.69	303	94	0.87	237	360	1.04	1.08
1B6C_A:B	1D6O_A	1IAS_A	166	3,278	1.70	157	1,844	1.70	127	1,862	1.96	77	1431	2.18	2.05
1BGX_HL:T	1AY1_HL	1CMW_A	3	21,434	4.54	-	-	6.03	-	-	6.54	-	-	6.57	5.69
1BJ1_HL:VW	1BJ1_HL	2VPF_GH	-	-	7.31	-	-	7.31	-	-	6.81	1	49,034	4.45	0.87
1BUH_A:B	1HCL_	1DKS_A	6,060	154	1.04	5,244	107	0.97	4,505	65	0.75	3,825	20	0.87	1.00
1BVK_DE:F	1BVL_BA	3LZT_	9	18,274	3.97	61	3,692	2.88	139	801	2.21	173	234	2.21	1.49
1BVN_P:T	1PIG_	1HOE_	1,566	1	1.58	1,087	9	1.58	685	72	1.58	442	117	1.62	1.00
1CGI_E:I	2CGA_B	1HPT_	3,533	29	2.53	2,736	14	2.53	1,859	39	2.55	1,167	4	2.57	2.08
1D6R_A:I	2TGT_	1K9B_A	3,923	48	1.45	2,858	477	1.43	2,419	177	1.45	2,252	164	1.49	2.61
1DE4_AB:CF	1A6Z_AB	1CX8_AB	131	4,182	2.98	40	34,372	2.81	110	607	2.81	81	1,059	2.81	2.65
1DFJ_E:I	9RSA_B	2BNH_	1,198	154	1.07	640	75	1.07	318	243	1.15	112	1,093	1.15	1.35
1DQJ_AB:C	1DQQ_CD	3LZT_	-	-	8.78	-	-	6.67	-	-	5.80	50	17,605	2.83	1.63
1E6E_A:B	1E1N_A	1CJE_D	136	9,817	2.15	141	5,428	2.26	47	12,176	3.38	61	4,953	3.84	1.18
1E6J_HL:P	1E6O_HL	1A43_	-	-	9.85	-	-	8.31	-	-	7.03	36	32,782	3.05	1.28
1E96_A:B	1MH1_	1HH8_A	104	768	2.08	196	725	1.79	175	300	1.79	195	684	1.50	1.68
1EAW_A:B	1EAX_A	9PTI_	1,088	35	1.22	1,146	478	1.22	913	517	1.70	636	760	2.40	0.66
1EER_A:BC	1BUY_A	1ERN_AB	512	20	2.47	250	7	2.47	112	4	2.80	33	2	3.11	3.24
1EWY_A:C	1GJR_A	1CZP_A	3,055	172	1.08	2,608	30	1.08	1,567	4	1.21	791	2	1.27	1.49
1EZU_C:AB	1TRM_A	1ECZ_AB	266	630	2.48	86	412	2.94	42	826	3.40	21	2,762	3.81	1.35
1F34_A:B	4PEP_	1F32_A	972	484	1.23	783	156	1.23	570	98	1.34	396	35	1.90	1.23
1F51_AB:E	1IXM_AB	1SRR_C	-	-	-	-	-	-	-	-	-	-	-	-	0.83
1FAK_HL:T	1QFK_HL	1TFH_B	-	-	8.30	-	-	8.26	-	-	8.43	-	-	8.67	6.85
1FC2_C:D	1BDD_	1FC1_AB	-	-	5.95	-	-	5.86	1	45,800	4.98	20	13,678	4.16	2.23
1FQ1_A:B	1FPZ_F	1B39_A	62	652	4.01	53	706	3.89	42	970	4.01	20	2,950	4.03	3.52
1FQJ_A:B	1TND_C	1FQI_A	558	79	1.90	345	20	1.90	288	27	2.12	162	179	2.14	2.75
1FSK_BC:A	1FSK_BC	1BV1_	-	-	8.58	8	38,144	2.88	39	14,829	2.19	58	5,874	2.19	0.66
1GCQ_B:C	1GRI_B	1GCP_B	-	-	14.19	-	-	14.19	-	-	14.19	-	-	14.19	1.17
1GHQ_A:B	1C3D_	1LY2_A	159	1,253	2.75	211	181	3.05	245	101	2.85	226	58	2.85	3.60
1GP2_A:BG	1GIA_	1TBG_DH	-	-	7.05	-	-	7.05	-	-	7.05	-	-	7.38	2.02
1GRN_A:B	1A4R_A	1RGP_	486	1,600	2.26	357	1,418	2.26	349	1,264	2.23	297	1,605	2.23	1.62
1H1V_A:G	1IJJ_B	1D0N_B	-	-	13.45	-	-	13.46	-	-	13.47	-	-	13.48	9.58
1HE1_C:A	1MH1_	1HE9_A	3,492	25	1.12	1,866	3	1.12	1,116	1	1.12	592	5	1.12	1.16
1HE8_B:A	821P_	1E8Z_A	64	11,791	2.98	4	41,665	4.60	-	-	5.14	-	-	5.40	3.24
1HIA_AB:I	2PKA_XY	1BXB_	749	88	3.09	590	103	3.09	488	453	3.10	284	570	3.35	2.60
1I2M_A:B	1QG4_A	1A12_A	210	574	2.74	181	1,133	2.86	137	1,352	3.06	70	1,411	3.51	2.31

TABLE 1

Unbound-unbound docking results using shape complementarity only, where we use four different sets of skin-skin (w_{ss}), core-core (w_{cc}) and skin-core (w_{sc}) weight values for F^{2}Dock. 'Rank' is the best rank among all predicted positions whose RMSD was less than 5Å. 'Good Peaks' is the number of peaks in the predicted set which were less than 5Å RMSD from the known position. 'RMSD' is the lowest RMSD among all peaks that were retained. Both F^{2}Dock and ZDock use 6° rotational sampling. F^{2}Dock and ZDock retained 50,000 and 54,000 peaks, respectively. RMSD was calculated using the C_α atoms near the interface of the known bound conformation (within 5Å of the interface for F^{2}Dock).

chemokine binding p.M3 complexed with Chemokine Mcp1) for which we did not have a single solution with RMSD less than 5 Å in the top 50,000 without electrostatics, but with w_E set to 350 we had several such solutions for each. For one of the complexes (2PCC: Cyt C peroxidase complexed with Cytochrome C) while we did not have a hit (i.e., at least one solution with RMSD less than 5 Å in the top 2,000) when electrostatics was not used, it was a hit when w_E was set to 350. On the other hand, for 1FC2 (i.e., Staphylococcus protein A complexed with Human Fc fragment) we had a solution

			F^{2}Dock Results (w_{ss} = 1.0, frequencies = 32^3)													
Data			$w_{cc} = 5.0$, $w_{sc} = 0.5$			$w_{cc} = 10.0$, $w_{sc} = 0.5$			$w_{cc} = 10.0$, $w_{sc} = 1.0$			$w_{cc} = 20.0$, $w_{sc} = 1.0$			ZDock Results
Bound Complex	Unbound Mol 1	Unbound Mol 2	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)	RMSD (Å)
1I4D_D:AB	1MH1_	1I49_AB	42	6,391	3.58	-	-	-	96	6,940	3.41	-	-	-	1.74
1I9R_HL:ABC	1I9R_HL	1ALY_ABC	13	13,814	2.31	109	4043	1.60	129	2,739	1.51	149	842	1.51	1.49
1IB1_AB:E	1QJB_AB	1KUY_A	66	18,213	3.66	54	13,593	3.66	18	20,918	3.66	-	-	5.19	3.97
1IBR_A:B	1QG4_A	1F59_A	6	13,885	4.41	-	-	7.38	-	-	6.89	-	-	6.78	4.71
1IJK_BC:A	1FVU_AB	1AUQ_	289	3,414	2.54	228	3,514	2.54	197	2,221	2.54	113	3,036	2.55	1.11
1IQD_AB:C	1IQD_AB	1D7P_M	-	-	8.65	9	33,186	1.34	31	8,909	1.34	53	3,551	1.34	0.75
1JPS_HL:T	1JPT_HL	1TFH_B	71	5,846	3.25	174	1,733	1.29	265	484	1.24	322	799	1.21	0.86
1K4C_AB:C	1K4C_AB	1JVM_ABCD	167	74	3.02	147	13	3.02	115	64	3.02	55	1,569	3.02	0.64
1K5D_AB:C	1RRP_AB	1YRG_B	13	1,203	4.52	6	18,833	4.34	-	-	5.06	3	27,117	4.49	1.81
1KAC_A:B	1NOB_F	1F5W_B	301	2,005	1.42	375	941	1.42	380	747	1.67	341	431	1.67	1.34
1KKL_ABC:H	1JB1_ABC	2HPR_	-	-	5.75	-	-	5.62	-	-	6.07	-	-	5.02	2.35
1KLU_AB:D	1H15_AB	1STE_	47	2,582	4.09	19	3,276	4.31	8	20,914	4.36	22	6,464	3.45	0.87
1KTZ_A:B	1TGK_	1M9Z_A	-	-	5.03	2	33,047	4.89	3	26,751	4.89	14	14,660	4.78	0.76
1KXP_A:D	1IJJ_B	1KW2_B	223	418	1.59	178	226	2.01	138	306	2.01	82	70	2.01	1.58
1KXQ_H:A	1KXQ_H	1PPI_	160	1,502	1.36	279	2,270	1.36	303	646	1.36	263	302	1.36	0.85
1M10_A:B	1AUQ_	1MOZ_B	146	3,412	2.99	90	3,593	2.99	42	7,365	3.36	37	6,232	3.67	4.29
1MAH_A:F	1J06_B	1FSC_	-	-	5.50	7	30,532	2.16	39	6,598	2.07	77	2,628	2.07	0.86
1ML0_AB:D	1MKF_AB	1DOL_	186	4,634	2.62	40	9,643	3.57	-	-	5.22	1	48,211	3.38	1.25
1MLC_AB:E	1MLB_AB	3LZT_	-	-	9.96	-	-	5.48	-	-	5.12	-	-	5.12	0.83
1N2C_ABCD:EF	3MIN_ABCD	2NIP_AB	9	11,739	3.70	-	-	-	2	16,076	4.82	-	-	-	3.03
1NCA_HL:N	1NCA_HL	7NN9_	2	46,528	4.50	32	7,060	1.50	37	7,406	1.50	51	3,765	0.86	0.60
1NSN_HL:S	1NSN_HL	1KDC_	29	29,539	2.31	90	9,501	2.13	69	7,846	2.09	31	4,773	2.09	0.94
1PPE_E:I	1BTP_	1LU0_A	3,425	118	1.12	2,574	210	1.12	1,634	355	1.12	1,007	165	1.12	0.58
1QA9_A:B	1HNF_	1CCZ_A	4	35,505	4.45	11	12,385	3.37	23	9,957	3.37	49	6,689	2.03	1.38
1QFW_IM:AB	1QFW_IM	1HRP_AB	12	34,831	2.43	27	5,651	1.34	35	1,372	1.34	46	391	1.34	1.13
1RLB_ABCD:E	2PAB_ABCD	1HBP_	25	7,151	3.53	35	19,653	4.29	26	6,480	3.82	33	3,088	2.85	1.11
1SBB_A:B	1BEC_	1SE4_	-	-	5.43	4	25,893	4.80	19	6,270	4.06	8	3,717	4.34	1.36
1TMQ_A:B	1JAE_	1B1U_A	564	9	1.63	379	18	1.63	233	247	1.63	175	1,652	1.97	1.43
1UDI_E:I	1UDH_	2UGI_B	352	5,597	1.46	236	3,693	1.60	113	5,438	1.98	121	1,817	1.99	1.24
1VFB_AB:C	1VFA_AB	8LYZ_	50	4,533	3.26	135	863	0.75	243	310	0.75	259	96	0.75	1.42
1WEJ_HL:F	1QBL_HK	1HRC_	-	-	6.91	-	-	7.03	-	-	6.44	4	44,648	3.24	0.51
1WQ1_R:G	6Q21_D	1WER_	1,039	327	1.58	809	132	1.95	503	96	1.95	392	52	2.01	1.55
2BTF_A:P	1IJJ_B	1PNE_	1	41,750	2.96	13	13,803	2.31	7	17,075	2.31	8	5,799	2.96	0.88
2HMI_CD:AB	2HMI_CD	1S6P_AB	7	18,636	3.73	13	4,480	3.73	10	884	4.15	10	303	4.15	2.58
2JEL_HL:P	2JEL_HL	1POH_	-	-	10.62	-	-	-	-	-	-	-	-	-	0.72
2MTA_HL:A	2BBK_JM	2RAC_A	358	882	2.35	434	1,489	2.25	384	1,378	1.58	619	304	1.58	0.74
2PCC_A:B	1CCP_	1YCC_	245	5,259	1.55	88	8,369	1.64	73	19,509	1.10	79	8,413	1.60	1.46
2QFW_HL:AB	1QFW_HL	1HRP_AB	113	6,453	1.75	193	1,308	1.18	239	525	1.18	223	595	1.18	1.48
2SIC_E:I	1SUP_	3SSI_	352	1,978	2.35	293	936	1.79	226	1,072	1.79	213	773	1.79	0.43
2SNI_E:I	1UBN_A	2CI2_I	827	291	1.63	421	359	1.63	257	362	1.92	168	1,739	2.28	1.05
2VIS_AB:C	1GIG_LH	2VIU_ACE	-	-	8.07	-	-	-	-	-	7.74	-	-	-	1.24
7CEI_A:B	1UNK_D	1M08_B	279	1,182	1.22	262	845	0.95	318	1,188	1.04	378	516	1.04	0.80

TABLE 2

Unbound-unbound docking results using shape complementarity only (continued), where we use four different sets of skin-skin (w_{ss}), core-core (w_{cc}) and skin-core (w_{sc}) weight values for F^{2}Dock. 'Rank' is the best rank among all predicted positions whose RMSD was less than 5Å. 'Good Peaks' is the number of peaks in the predicted set which were less than 5Å RMSD from the known position. 'RMSD' is the lowest RMSD among all peaks that were retained. Both F^{2}Dock and ZDock use 6° rotational sampling. F^{2}Dock and ZDock retained 50,000 and 54,000 peaks, respectively. RMSD was calculated using the C_α atoms near the interface of the known bound conformation (within 5Å of the interface for F^{2}Dock).

with RMSD less than 5 Å in the top 50,000 when w_E was set to 0, but lost it when w_E was set to 350. Electrostatics scores did not seem to have as much impact on the minimum RMSD value as they had on 'Rank'. For only 16 complexes the minimum RMSD improved by at least 0.05 Å, while for 9 it degraded by at least 0.05 Å. For 52 complexes the minimum RMSD did not change. Overall, electrostatics was most effective on inhibitors or enzyme-substrate and antigen-bound antibody complexes (improving results in more than 60% of the 35 cases), and least effective on antibody-antigens (marginally improving results for only 3 out of 10 complexes). For the remaining 39 complexes, however, electrostatics was effective in more than 70% of the cases.

4.2 Bound-unbound Docking

Table 4 shows the results of increasing the number of frquencies extracted by FFT from 32^3 to 64^3 when performing bound-unbound docking on the complexes of the ZDock benchmark suite. The weight values are the same as in Table 3, and electrostatics interactions were not considered. We observe that increasing the number of frequencies generally improved the lowest RMSD considerably. For 45 complexes the lowest RMSD improved by at least 0.05 Å.

In Figure 6(b) we show our docking of chains A & B (nuclear transport factor 2) obtained from 1OUN.pdb on chain C (Ran GTPase) of 1A2K.pdb (i.e., docking the unbound nuclear transport factor 2 from 1OUN.pdb instead of the same protein already docked on Ran GTPase of 1A2K.pdb). In Figure 6(d) we show the docking of PSTI obtained from 1HPT.pdb on chain E (Bovine chymotrypsinogen) of 1CGI.pdb replacing the PSTI (chain I) already docked there.

4.3 Bound-bound Docking or Redocking

In Table 5 we report our bound-bound docking results on ZDock benchmark 2.0 [6]. We use the same weight values as in Table 4, and show results both with and without electrostatics. We did not move molecule B (the moving molecule) to a random location at the beginning of the experiment since F^{2}Dock initially centers both molecules at the origin anyway. We also did not rotate molecule B by a random amount initially since we are using rotations sampled uniformly at random and the identity matrix (i.e., 0° rotation) was not included as a rotation matrix separately. For 27 complexes the lowest RMSD was less than 1 Å, and for 47 it was less than 1.5 Å. The impact of including electrostatics was almost similar to the unbound-unbound case. For example, electrostatics improved the 'Rank' value for around 54% of the complexes, while for around 34% of the complexes 'Rank' degraded.

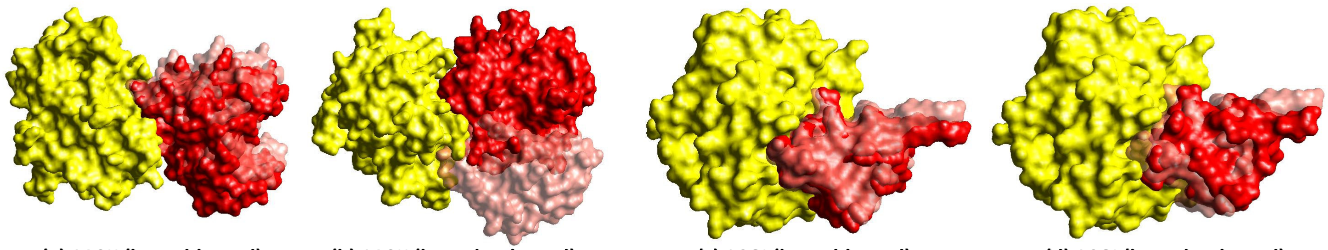
Figure 6(a) shows our redocking of chains A & B (nuclear transport factor 2) of 1A2K.pdb on its chain C (Ran GTPase),

Data			F^{2}Dock Results, Weights: $w_{ss} = 1.0$, $w_{cc} = 10.0$, $w_{sc} = 1.0$, Frequencies = 32^3 — Without Electrostatics $w_E = 0$			With Electrostatics $w_E = 350$		
Bound Complex	Unbound Mol 1	Unbound Mol 2	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)
1A2K_C:AB	1QG4_A	1OUN_AB	36	8,100	3.02	75	4,374	3.02
1ACB_E:I	2CGA_B	1EGL_	569	803	3.08	501	849	3.20
1AHW_AB:C	1FGN_LH	1TFH_A	36	6,516	1.65	36	5,396	1.65
1AK4_A:D	2CPL_	1E6J_P	110	160	3.49	139	128	3.48
1AKJ_AB:DE	2CLR_DE	1CD8_AB	679	102	1.45	907	46	1.45
1ATN_A:D	1IJJ_B	3DNI_	4	19,423	4.72	4	14,779	4.72
1AVX_A:B	1QQU_A	1BA7_B	339	1,769	1.75	326	1,909	1.75
1AY7_A:B	1RGH_B	1A19_B	303	94	0.87	474	32	0.98
1B6C_A:B	1D6O_A	1IAS_A	127	1,862	1.96	144	1,687	1.96
1BGX_HL:T	1AY1_HL	1CMW_A	-	-	6.54	-	-	6.54
1BJ1_HL:VW	1BJ1_HL	2VPF_GH	-	-	6.81	-	-	7.19
1BUH_A:B	1HCL_	1DKS_A	4,505	65	0.75	4,569	64	0.75
1BVK_DE:F	1BVL_BA	3LZT_	139	801	2.21	177	560	2.21
1BVN_P:T	1PIG_	1HOE_	685	72	1.58	608	54	1.58
1CGI_E:I	2CGA_B	1HPT_	1,859	39	2.55	1,762	45	2.55
1D6R_A:I	2TGT_	1K9B_A	2,419	177	1.45	2,480	170	1.45
1DE4_AB:CF	1A6Z_AB	1CX8_AB	110	607	2.81	131	589	2.81
1DFJ_E:I	9RSA_B	2BNH_	318	243	1.15	881	22	1.14
1DQJ_AB:C	1DQQ_CD	3LZT_	-	-	5.80	-	-	5.80
1E6E_A:B	1E1N_A	1CJE_D	47	12,176	3.38	210	3,526	2.41
1E6J_HL:P	1E6O_HL	1A43_	-	-	7.03	-	-	7.00
1E96_A:B	1MH1_	1HH8_A	175	300	1.79	218	193	1.79
1EAW_A:B	1EAX_A	9PTI_	913	517	1.70	1,265	454	1.52
1EER_A:BC	1BUY_A	1ERN_AB	112	4	2.80	142	1	2.84
1EWY_A:C	1GJR_A	1CZP_A	1,567	4	1.21	2,308	4	1.17
1EZU_C:AB	1TRM_A	1ECZ_AB	42	826	3.40	42	763	3.40
1F34_A:B	4PEP_	1F32_A	570	98	1.34	625	60	1.34
1F51_AB:E	1IXM_AB	1SRR_C	-	-	-	-	-	-
1FAK_HL:T	1QFK_HL	1TFH_B	-	-	8.43	-	-	8.43
1FC2_C:D	1BDD_	1FC1_AB	1	45,800	4.98	-	-	5.12
1FQ1_A:B	1FPZ_F	1B39_A	42	970	4.01	-	-	-
1FQJ_A:B	1TND_C	1FQI_A	288	27	2.12	326	30	2.10
1FSK_BC:A	1FSK_BC	1BV1_	39	14,829	2.19	37	14,873	2.19
1GCQ_B:C	1GRI_B	1GCP_B	-	-	14.19	-	-	14.19
1GHQ_A:B	1C3D_	1LY2_A	245	101	2.85	190	431	2.85
1GP2_A:BG	1GIA_	1TBG_DH	-	-	7.05	-	-	6.97
1GRN_A:B	1A4R_A	1RGP_	349	1,264	2.23	504	674	2.23
1H1V_A:G	1IJJ_B	1D0N_B	-	-	13.47	-	-	13.47
1HE1_C:A	1MH1_	1HE9_A	1,116	1	1.12	1,253	1	1.12
1HE8_B:A	821P_	1E8Z_A	-	-	5.14	-	-	5.14
1HIA_AB:I	2PKA_XY	1BXB_	488	453	3.10	718	220	2.98
1I2M_A:B	1QG4_A	1A12_A	137	1,352	3.06	349	381	2.86

Data			F^{2}Dock Results, Weights: $w_{ss} = 1.0$, $w_{cc} = 10.0$, $w_{sc} = 1.0$, Frequencies = 32^3 — Without Electrostatics $w_E = 0$			With Electrostatics $w_E = 350$		
Bound Complex	Unbound Mol 1	Unbound Mol 2	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)
1I4D_D:AB	1MH1_	1I49_AB	96	6,940	3.41	94	7,033	3.41
1I9R_HL:ABC	1I9R_HL	1ALY_ABC	129	2,739	1.51	185	2,090	1.51
1IB1_AB:E	1QJB_AB	1KUY_A	18	20,918	3.66	13	22,719	3.73
1IBR_A:B	1QG4_A	1F59_A	-	-	6.89	-	-	6.26
1IJK_BC:A	1FVU_AB	1AUQ_	197	2,221	2.54	299	1,426	2.43
1IQD_AB:C	1IQD_AB	1D7P_M	31	8,909	1.34	50	6,412	1.34
1JPS_HL:T	1JPT_HL	1TFH_B	265	484	1.24	265	702	1.17
1K4C_AB:C	1K4C_AB	1JVM_ABCD	115	64	3.02	114	87	3.02
1K5D_AB:C	1RRP_AB	1YRG_B	-	-	5.06	64	8,013	2.79
1KAC_A:B	1NOB_F	1F5W_B	380	747	1.67	377	672	1.67
1KKL_ABC:H	1JB1_ABC	2HPR_	-	-	6.07	-	-	6.07
1KLU_AB:D	1H15_AB	1STE_	8	20,914	4.36	6	33,414	4.36
1KTZ_A:B	1TGK_	1M9Z_A	3	26,751	4.89	4	20,866	4.89
1KXP_A:D	1IJJ_B	1KW2_B	138	306	2.01	168	157	2.01
1KXQ_H:A	1KXQ_H	1PPI_	303	646	1.36	353	528	1.39
1M10_A:B	1AUQ_	1MOZ_B	42	7,365	3.36	115	3,138	2.99
1MAH_A:F	1J06_B	1FSC_	39	6,598	2.07	89	3,327	2.07
1ML0_AB:D	1MKF_AB	1DOL_	-	-	5.22	3	33,027	4.50
1MLC_AB:E	1MLB_AB	3LZT_	-	-	5.12	-	-	5.33
1N2C_ABCD:EF	3MIN_ABCD	2NIP_AB	2	16,076	4.82	2	8,637	4.82
1NCA_HL:N	1NCA_HL	7NN9_	37	7,406	1.50	29	8,944	1.65
1NSN_HL:S	1NSN_HL	1KDC_	69	7,846	2.09	68	8,340	2.09
1PPE_E:I	1BTP_	1LU0_A	1,634	355	1.12	1,450	392	1.12
1QA9_A:B	1HNF_	1CCZ_A	23	9,957	3.37	24	9,730	3.37
1QFW_IM:AB	1QFW_IM	1HRP_AB	35	1,372	1.34	45	1,212	1.34
1RLB_ABCD:E	2PAB_ABCD	1HBP_	26	6,480	3.82	28	4,843	3.77
1SBB_A:B	1BEC_	1SE4_	19	6,270	4.06	19	6,146	4.06
1TMQ_A:B	1JAE_	1B1U_A	233	247	1.63	238	241	1.63
1UDI_E:I	1UDH_	2UGI_B	113	5,438	1.98	217	3,043	1.74
1VFB_AB:C	1VFA_AB	8LYZ_	243	310	0.75	269	213	0.75
1WEJ_HL:F	1QBL_HK	1HRC_	-	-	6.44	-	-	6.44
1WQ1_R:G	6Q21_D	1WER_	503	96	1.95	608	62	1.95
2BTF_A:P	1IJJ_B	1PNE_	7	17,075	2.31	8	13,957	2.31
2HMI_CD:AB	2HMI_CD	1S6P_AB	10	884	4.15	10	836	4.15
2JEL_HL:P	2JEL_HL	1POH_	-	-	-	57	11,932	2.58
2MTA_HL:A	2BBK_JM	2RAC_A	384	1,378	1.58	811	1,124	1.58
2PCC_A:B	1CCP_	1YCC_	73	19,509	1.10	1,574	843	0.66
2QFW_HL:AB	1QFW_HL	1HRP_AB	239	525	1.18	307	427	1.18
2SIC_E:I	1SUP_	3SSI_	226	1,072	1.79	180	1,429	2.35
2SNI_E:I	1UBN_A	2CI2_I	257	362	1.92	246	377	1.92
2VIS_AB:C	1GIG_LH	2VIU_ACE	-	-	7.74	-	-	7.74
7CEI_A:B	1UNK_D	1M08_B	318	1,188	1.04	958	598	0.85

TABLE 3

Effect of using electrostatics on shape-complementarity-based unbound-unbound docking with F^{2}Dock. 'Rank' is the best rank among all predicted positions whose RMSD was less than 5Å. 'Good Peaks' is the number of peaks in the predicted set which were less than 5Å RMSD from the known position. 'RMSD' is the lowest RMSD among all peaks that were retained. In both cases we used 6° rotational sampling, and retained 50,000. RMSD was calculated using the C_α atoms near the interface of the known bound conformation (within 5Å of the interface).



(a) 1A2K (bound-bound) **(b) 1A2K (bound-unbound)** **(c) 1CGI (bound-bound)** **(d) 1CGI (bound-unbound)**

Fig. 6. **(a & b) Docking 1A2K (Ran GTPase complexed with nuclear transport factor 2):** (a) (Bound-Bound) Redocking chains A & B (nuclear transport factor 2) of 1A2K.pdb on it's chain C (Ran GTPase), (b) (Bound-Unbound) Docking chains A & B (nuclear transport factor 2) of 1OUN.pdb on chain C of 1A2K.pdb. **(c & d) Docking 1CGI (Bovine chymotrypsinogen complxed with PSTI):**: (c) (Bound-Bound) Redocking chain I (PSTI) of 1CGI.pdb on it's chain E (Bovine chymotrypsinogen), (d) (Bound-Unbound) Docking the unmarked chain (PSTI) of 1HPT.pdb on chain E of 1CGI.pdb. In (a) & (b) chain C is static (colored yellow), and in (c) & (d) chain E is static, and in all cases the other chain(s) is (are) moved around for docking (the true position in the bound complex is pink, and our final docked position is red).

while Figure 6(c) shows our redocking of chain I (PSTI) of 1CGI.pdb on its chain E (Bovine chymotrypsinogen).

Figure 7 shows the distribution of electrostatics potential on the molecular surfaces of Ran GTPase and Ran GAP, and also how the distribution changes when they form a complex (1K5D.pdb). In Figure 8 we show the electrostatics complementarity at the interface when Ran GTPase and Ran GAP dock at three different locations and orientations. The electrostatics potential for all of these examples, were computed using our CVC in-house software called PBEM3D (Molecular Poisson Boltzmann Boundary Element Electrostatics Potential calculation in 3D [52]). Figures (visualization) were created using CVC software TexMol.

5 Conclusion

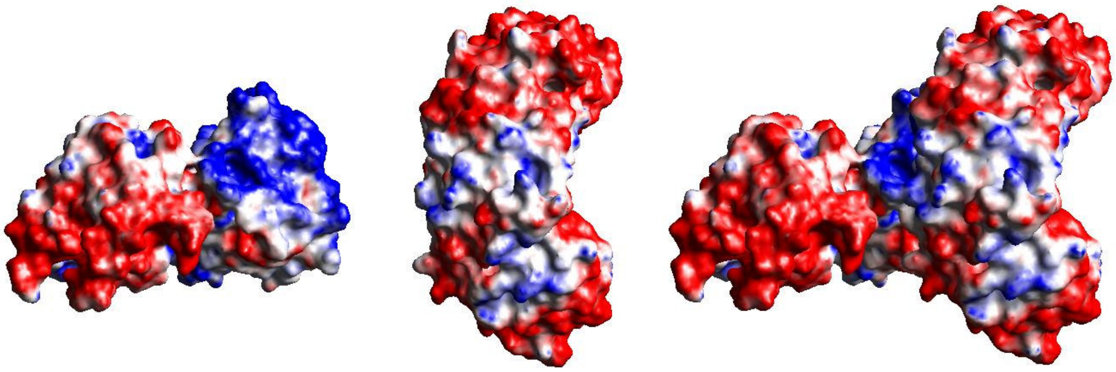
We have presented a fast, and practical adaptive algorithm for rigid protein-protein docking. Our algorithm is based on representing affinity functions in a multi-resolution radial basis function format. The smoothed particle protein representation, together with nonequispaced Fast Fourier transforms allows us several advantages of efficiency and accuracy tradeoffs visavis traditional FFT based docking approaches. Our contributions

Data		F^{2}Dock Results, weights $w_{ss}=1.0$, $w_{cc}=1.0$, $w_{sc}=1.0$: Frequencies = 32^3			Frequencies = 64^3		
Bound Complex	Unbound Mol 2	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)
1A2K_C:AB	1OUN_AB	40	5,240	3.01	26	2,329	3.17
1ACB_E:I	1EGL_	581	130	1.90	594	50	1.93
1AHW_AB:C	1TFH_A	42	5,742	1.24	94	1,001	1.27
1AK4_A:D	1E6J_P	58	785	4.09	82	3,480	3.97
1AKJ_AB:DE	1CD8_AB	427	320	1.26	532	286	1.26
1ATN_A:D	3DNI_	3	17,662	4.61	1	25,273	1.57
1AVX_A:B	1BA7_B	588	262	1.70	781	176	1.40
1AY7_A:B	1A19_B	121	2,607	1.48	109	45	1.41
1B6C_A:B	1IAS_A	92	2,059	2.08	66	7,647	1.56
1BGX_HL:T	1CMW_A	-	-	5.21	12	2,049	3.51
1BJ1_HL:VW	2VPF_GH	2	43,036	4.69	-	-	6.02
1BUH_A:B	1DKS_A	6,041	8	0.46	5,723	9	0.22
1BVK_DE:F	3LZT_	97	3,687	1.58	61	842	1.72
1BVN_P:T	1HOE_	719	36	1.27	1,255	14	1.03
1CGI_E:I	1HPT_	3,289	5	0.75	4,752	14	1.20
1D6R_A:I	1K9B_A	2,508	170	1.11	2,469	200	1.10
1DE4_AB:CF	1CX8_AB	206	1,296	1.61	113	878	2.09
1DFJ_E:I	2BNH_	512	65	0.86	637	732	0.64
1DQJ_AB:C	3LZT_	8	3,5060	3.15	16	18,100	2.24
1E6E_A:B	1CJE_D	212	4,586	2.27	319	175	1.29
1E6J_HL:P	1A43_	-	-	6.99	23	23,314	1.93
1E96_A:B	1HH8_A	252	514	1.62	150	2,084	1.74
1EAW_A:B	9PTI_	837	203	2.21	1,460	149	1.54
1EER_A:BC	1ERN_AB	112	29	2.86	534	47	1.79
1EWY_A:C	1CZP_A	2,253	129	1.14	2,160	1	1.04
1EZU_C:AB	1ECZ_AB	61	24	3.23	113	51	3.36
1F34_A:B	1F32_A	528	65	1.28	875	15	1.13
1F51_AB:E	1SRR_C	168	2,553	3.05	351	499	1.63
1FAK_HL:T	1TFH_B	39	1,391	2.41	58	2,184	2.72
1FC2_C:D	1FC1_AB	-	-	5.61	-	-	6.04
1FQ1_A:B	1B39_A	15	4,591	4.23	1	28,985	4.87
1FQJ_A:B	1FQI_A	325	21	1.75	277	124	1.99
1FSK_BC:A	1BV1_	39	14,829	2.19	27	8,442	1.75
1GCQ_B:C	1GCP_B	1,280	20	1.18	1,263	2	1.30
1GHQ_A:B	1LY2_A	239	11	2.90	368	190	2.77
1GP2_A:BG	1TBG_DH	42	1,990	1.35	14	10,191	1.61
1GRN_A:B	1RGP_	171	3,286	1.59	239	708	1.23
1H1V_A:G	1D0N_B	-	-	13.33	-	-	13.49
1HE1_C:A	1HE9_A	1,134	27	0.88	1,400	40	0.91
1HE8_B:A	1E8Z_A	9	28,558	3.50	62	4,239	2.14
1HIA_AB:I	1BXB_	454	90	2.61	641	1	2.20
1I2M_A:B	1A12_A	532	48	0.84	576	27	0.87

Data		F^{2}Dock Results, weights $w_{ss}=1.0$, $w_{cc}=1.0$, $w_{sc}=1.0$: Frequencies = 32^3			Frequencies = 64^3		
Bound Complex	Unbound Mol 2	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)
1I4D_D:AB	1I49_AB	35	4,657	4.08	227	353	2.68
1I9R_HL:ABC	1ALY_ABC	108	3,983	0.85	123	1,782	0.84
1IB1_AB:E	1KUY_A	75	589	1.79	107	3,166	1.35
1IBR_A:B	1F59_A	1	49,336	4.98	3	31,965	3.43
1IJK_BC:A	1AUQ_	56	2,647	1.72	18	7,958	1.77
1IQD_AB:C	1D7P_M	31	8,909	1.34	9	25,042	1.74
1JPS_HL:T	1TFH_B	178	1,689	0.93	142	1,195	0.75
1K4C_AB:C	1JVM_ABCD	115	64	3.02	357	31	2.84
1K5D_AB:C	1YRG_B	7	34,601	1.80	3	7,478	4.73
1KAC_A:B	1F5W_B	465	340	1.53	319	804	1.73
1KKL_ABC:H	2HPR_	24	30,156	2.09	94	7,376	2.27
1KLU_AB:D	1STE_	31	7,312	4.04	9	11,638	4.30
1KTZ_A:B	1M9Z_A	-	-	5.15	-	-	5.05
1KXP_A:D	1KW2_B	221	102	1.35	345	126	1.16
1KXQ_H:A	1PPI_	249	1,020	1.69	295	1,758	0.65
1M10_A:B	1MOZ_B	91	5,622	3.09	26	5,628	3.65
1MAH_A:F	1FSC_	25	16,095	3.39	73	3,508	1.58
1ML0_AB:D	1DOL_	-	-	5.34	34	621	1.86
1MLC_AB:E	3LZT_	-	-	5.43	-	-	5.11
1N2C_ABCD:EF	2NIP_AB	13	797	4.44	10	2,936	4.41
1NCA_HL:N	7NN9_	37	7,406	1.50	67	3,133	0.91
1NSN_HL:S	1KDC_	69	7,846	2.09	106	1,996	2.09
1PPE_E:I	1LU0_A	2,994	205	1.68	3,171	18	1.27
1QA9_A:B	1CCZ_A	26	15,078	2.59	40	4,334	1.57
1QFW_IM:AB	1HRP_AB	35	1,371	1.34	11	4,852	1.57
1RLB_ABCD:E	1HBP_	30	10,452	2.20	10	16,389	2.16
1SBB_A:B	1SE4_	9	30,808	4.24	4	18,560	4.07
1TMQ_A:B	1B1U_A	309	9	1.60	504	12	1.33
1UDI_E:I	2UGI_B	398	1,071	1.51	509	192	1.06
1VFB_AB:C	8LYZ_	129	8,387	2.53	96	2,511	1.84
1WEJ_HL:F	1HRC_	-	-	6.57	4	27,001	3.62
1WQ1_R:G	1WER_	868	379	1.40	1,080	93	1.44
2BTF_A:P	1PNE_	126	7,748	1.57	89	3,769	0.87
2HMI_CD:AB	1S6P_AB	-	-	5.73	-	-	5.97
2JEL_HL:P	1POH_	46	14,110	2.76	6	25,303	3.29
2MTA_HL:A	2RAC_A	171	6,357	3.36	333	1,273	1.09
2PCC_A:B	1YCC_	200	9,587	0.62	85	5,616	1.56
2QFW_HL:AB	1HRP_AB	239	525	1.18	209	3,715	1.06
2SIC_E:I	3SSI_	328	550	1.59	207	838	2.39
2SNI_E:I	2CI2_I	234	855	2.53	262	2,688	1.87
2VIS_AB:C	2VIU_ACE	-	-	7.02	-	-	7.01
7CEI_A:B	1M08_B	582	67	1.25	725	19	1.56

TABLE 4

Effect of changing the number of frequencies extracted by FFT during Bound-unbound docking with F^{2}Dock. 'Rank' is the best rank among all predicted positions whose RMSD was less than 5Å. 'Good Peaks' is the number of peaks in the predicted set which were less than 5Å RMSD from the known position. 'RMSD' is the lowest RMSD among all peaks that were retained. F^{2}Dock used 6° rotational sampling, and retained 50,000 peaks. RMSD was computed using the C_α atoms near the interface of the known bound conformation (within 5Å of the interface).



(a) Ran GTPase **(b) Ran GAP** **(c) Ran GTPase complexed with Ran GAP**

Fig. 7. Poisson-Boltzmann electrostatics potential on the surface of **(a)** Ran GTPase, **(b)** Ran GAP, and **(c)** complex of Ran GTPase and Ran GAP (1K5D.pdb). The potential ranges from $-3.8\ k_bT/e_c$ (red) to $+3.8\ k_bT/e_c$ (blue).

are also in scoring of docked conformations as a convolution of complex affinity functions, and providing approximation algorithms to detect peaks in the docking scoring profiles. Both shape complementarity and electrostatics are used for scoring and to obtain the top docking conformations. Our implementation of F^{2}Dock speeds up computation even further by executing multiple concurrent threads on multicore machines. The rotation matrices are evenly distributed among the threads. When electrostatics is not used we use on the average, around 15 mins for computing docking positions (with 6° rotational sampling and 32^3 frequencies) per typical protein complex on a quad-core linux desktop (3.0GHz) with 4GB RAM. The running time approximately doubles when electrostatics is used. We used the FFTW package [53] for computing FFT and the inverse FFT. We are also working on an MPI [54] based distributed implementation of F^{2}Dock capable of running on Linux clusters. This implementation will be available as a web-based docking server. Jobs can

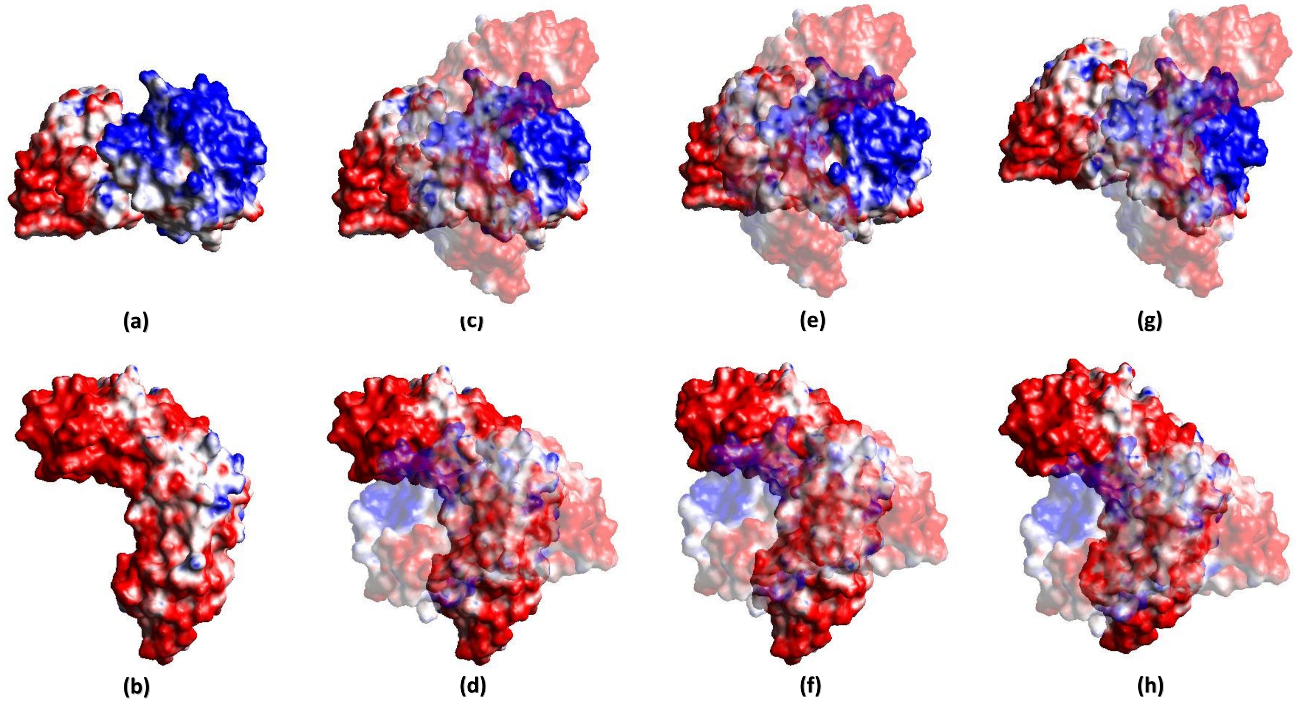


Fig. 8. Figures (a) and (b) show Poisson-Boltzmann electrostatics potential on the surface of Ran GTPase and Ran GAP, respectively. The potential ranges from $-3.8\ k_bT/e_c$ (red) to $+3.8\ k_bT/e_c$ (blue). Figures (c) and (d) show the bound complex of Ran GTPase and Ran GAP (1K5D.pdb). In (c) Ran GAP is drawn semi-transparent while in (d) Ran GTPase is drawn semi-transparent in order to show the electrostatics complementarity at the interface. Figures (e) and (f) show the solution with the lowest RMSD (1.66 Å) from the bound complex among the top 2,000 solutions returned by F^{2}Dock when electrostatics weight was set to 350. Figures (g) and (h) show the solution with the lowest RMSD (2.90 Å) from the bound complex among the top 2,000 solutions returned by F^{2}Dock when electrostatics weight was set to 0.

also be launched on the server from our in-house molecular modeling and visualization client software tool, called TexMol [55]. The TexMol client tool is in the public domain and can be freely downloaded from our center's software website (http://www.ices.utexas.edu/CVC/software/).

We are also in the process of extending F^{2}Dock to F^{3}Dock which is capable of handling flexible molecules. Some preliminary results on F^{3}Dock are available as a technical report [7].

Acknowledgments

This work was supported in part by NSF grants IIS-0325550, CNS-0540033 and grants from the NIH R01 GM074258, R01-GM073087, R01-EB004873. Particular thanks to Albert Chen, Maysam Moussalem, Wenqi Zhao for all their immense help in the development, and use of our CVC software, namely, the fast linear Poisson Boltzmann solver (PBEM3D) the molecular modeling and visualization tool (TexMol), and of course F^{2}Dock, used in producing the figures and the ZDock benchmark comparisons captured in the various tables. We are also grateful to Dr. Art Olson and Dr. Michel Sanner of The Scripps Research Institute (TSRI), for many helpful suggestions related to protein-protein docking.

References

[1] J. S. Fruton, *Proteins, Enzymes, Genes: The Interplay of Chemistry and Biology*. Yale University Press, 1999.

[2] A. R. Leach, *Molecular Modelling: Principles and Applications, second edition*. Pearson Education EMA, 2001.

[3] H. A. Gabb, R. M. Jackson, and M. J. E. Sternberg, "Modelling protein docking using shape complementarity, electrostatics and biochemical information," *Journal of Molecular Biology*, vol. 272, no. 1, pp. 106–120, September 1997.

[4] I. M. Klotz, *Ligand-receptor energetics: A guide for the perplexed*. John Wiley and Sons, Inc, 1997.

[5] J. Castrillon-Candas, V. Siddavanahalli, and C. Bajaj, "Nonequispaced fourier transforms for protein-protein docking," The University of Texas at Austin, Austin TX USA, ICES Report 05-44, October 2005.

[6] J. Mintseris, K. Wiehe, B. Pierce, R. Anderson, R. Chen, J. Janin, and Z. Weng, "Protein-protein docking benchmark 2.0: An update," *Proteins: Structure, Function, and Bioinformatics*, vol. 60, no. 2, pp. 214–216, 2005. [Online]. Available: http://zlab.bu.edu/zdock/benchmark.shtml

[7] C. Bajaj, R. A. Chowdhury, and V. Siddavanahalli, "F3Dock: A fast, flexible and Fourier based approach to protein-protein docking," The University of Texas at Austin, ICES Report 08-01, January 2008.

[8] T. Ewing, "Automated molecular docking: Development and evaluation of new search methods," PhD Thesis, University of California, 1997.

[9] M. D. Miller, S. K. Kearsley, D. J. Underwood, and R. P. Sheridan, "FLOG: A system to select Šquasi-flexibleŠ ligands complementary to a receptor of known three-dimensional structure," *Journal of Computer-Aided Molecular Design*, vol. 8, no. 2, pp. 153–174, 1994.

[10] B. K. Shoichet, D. L. Bodian, and I. D. Kuntz, "Molecular docking using shape descriptors," *Journal of Computational Chemistry*, vol. 13, no. 3, pp. 380–397, 1992.

[11] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin, "A geometric approach to macromolecule-ligand interactions," *Journal of Molecular Biology*, vol. 161, no. 2, pp. 269–288, October 1982.

[12] B. K. Shoichet and I. D. Kuntz, "Protein docking and complementarity," *Journal of Molecular Biology*, vol. 221, no. 1, pp. 327–346, September 1991.

[13] D. Fischer, R. Norel, R. Nussinov, and H. J. Wolfson, "3-d docking of protein molecules," in *CPM '93: Proceedings of the 4th Annual Symposium on Combinatorial Pattern Matching*. London, UK: Springer-Verlag, 1993, pp. 20–34.

[14] R. Norel, D. Fischer, H. J. Wolfson, and R. Nussinov, "Molecular surface recognition by a computer vision-based technique," *Protein engineering*, vol. 7, no. 1, pp. 39–46, January 1994.

[15] R. Norel, S. L. Lin, H. J. Wolfson, and R. Nussinov, "Molecular surface

	F^2Dock Results Weights: $w_{ss} = 1.0$, $w_{cc} = 10.0$, $w_{sc} = 1.0$ Frequencies = 32^3							F^2Dock Results Weights: $w_{ss} = 1.0$, $w_{cc} = 10.0$, $w_{sc} = 1.0$ Frequencies = 32^3					
Data	Without Electrostatics $w_E = 0$			With Electrostatics $w_E = 350$			Data	Without Electrostatics $w_E = 0$			With Electrostatics $w_E = 350$		
Bound Complex	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)	Bound Complex	Good Peaks	Rank	RMSD (Å)	Good Peaks	Rank	RMSD (Å)
1A2K_C:AB	240	232	0.60	440	50	0.60	1I4D_D:AB	12	25,200	1.75	8	26,792	2.16
1ACB_E:I	2,005	1	0.45	2,731	1	0.45	1I9R_HL:ABC	37	2,794	1.69	79	1,189	1.69
1AHW_AB:C	29	5,807	0.79	46	5,542	0.79	1IB1_AB:E	141	181	0.91	190	56	0.91
1AK4_A:D	1,417	13	0.34	2,665	5	0.34	1IBR_A:B	120	398	1.87	289	166	1.74
1AKJ_AB:DE	286	32	0.93	607	12	0.93	1IJK_BC:A	194	277	1.00	38	8,490	3.09
1ATN_A:D	10	11,589	3.81	16	12,168	3.81	1IQD_AB:C	85	772	0.99	315	81	0.99
1AVX_A:B	729	46	0.64	1,114	10	0.64	1JPS_HL:T	346	1,414	1.51	458	666	0.85
1AY7_A:B	111	1,867	0.55	145	941	0.55	1K4C_AB:C	53	4,338	1.31	49	5,984	1.31
1B6C_A:B	108	911	0.94	86	1,588	0.94	1K5D_AB:C	79	1,370	0.83	324	42	0.69
1BGX_HL:T	33	35	1.40	29	44	1.40	1KAC_A:B	187	1,018	0.55	311	341	0.55
1BJ1_HL:VW	-	-	7.39	-	-	7.47	1KKL_ABC:H	322	1,097	1.38	437	297	1.38
1BUH_A:B	3,367	8	0.33	3,106	2	0.26	1KLU_AB:D	43	424	1.13	41	1,558	1.13
1BVK_DE:F	72	1,831	0.66	279	310	0.41	1KTZ_A:B	64	2,965	0.80	1,323	190	0.61
1BVN_P:T	552	3	0.98	154	44	0.98	1KXP_A:D	70	203	0.98	84	54	0.98
1CGI_E:I	1,622	1	0.40	2,132	1	0.40	1KXQ_H:A	104	1,511	1.70	238	563	1.69
1D6R_A:I	2,086	40	0.35	1,947	41	0.35	1M10_A:B	81	197	0.93	726	11	0.84
1DE4_AB:CF	282	51	1.36	299	38	1.36	1MAH_A:F	58	6,719	3.48	634	768	2.74
1DFJ_E:I	248	1	0.61	3,156	1	0.61	1ML0_AB:D	26	17,851	3.56	180	4,134	2.67
1DQJ_AB:C	112	3,336	2.23	31	10,128	3.16	1MLC_AB:E	12	27,310	1.04	5	31,822	3.31
1E6E_A:B	251	34	1.18	873	3	1.02	1N2C_ABCD:EF	-	-	6.71	-	-	6.71
1E6J_HL:P	9	6,805	4.35	18	4,873	4.15	1NCA_HL:N	40	6,351	1.57	25	8,636	1.57
1E96_A:B	139	946	1.26	174	1,053	1.26	1NSN_HL:S	42	5,504	2.85	19	8,735	3.15
1EAW_A:B	451	59	1.14	1,851	10	1.14	1PPE_E:I	1,767	1	0.77	630	1	0.77
1EER_A:BC	29	5,727	1.56	159	531	1.55	1QA9_A:B	701	77	1.25	1,471	22	0.84
1EWY_A:C	657	779	0.73	1,285	447	0.62	1QFW_IM:AB	226	433	0.89	332	147	0.89
1EZU_C:AB	148	24	1.09	145	9	1.09	1RLB_ABCD:E	24	5,651	1.74	10	7,951	1.74
1F34_A:B	577	1	1.35	297	1	1.35	1SBB_A:B	64	9,509	1.42	103	9,156	1.42
1F51_AB:E	264	642	2.21	112	782	2.51	1TMQ_A:B	55	302	1.06	59	254	1.08
1FAK_HL:T	29	974	1.89	28	818	1.89	1UDI_E:I	135	324	1.15	977	18	0.94
1FC2_C:D	307	2,530	0.49	130	3,749	1.18	1VFB_AB:C	156	349	0.59	271	159	0.59
1FQ1_A:B	143	187	0.73	-	-	-	1WEJ_HL:F	484	2,266	1.36	389	2,778	1.36
1FQJ_A:B	71	2,220	3.22	220	1,376	2.76	1WQ1_R:G	447	10	0.49	1,127	2	0.49
1FSK_BC:A	206	1,030	1.89	233	994	1.89	2BTF_A:P	24	18,464	1.47	86	9,529	1.31
1GCQ_B:C	1,149	11	0.40	311	328	0.43	2HMI_CD:AB	-	-	5.91	-	-	5.34
1GHQ_A:B	171	16	2.84	33	2,742	3.83	2JEL_HL:P	44	3,029	1.05	89	3,124	0.86
1GP2_A:BG	6	2,224	1.85	12	1,277	1.42	2MTA_HL:A	330	269	1.58	834	305	1.41
1GRN_A:B	147	329	1.21	377	39	1.20	2PCC_A:B	216	503	1.36	4,634	16	0.60
1H1V_A:G	23	6,904	1.38	11	16,219	1.38	2QFW_HL:AB	170	1,106	0.91	243	364	0.91
1HE1_C:A	1,098	3	0.59	1,438	1	0.59	2SIC_E:I	570	1	0.64	173	7	0.64
1HE8_B:A	-	-	5.17	-	-	5.17	2SNI_E:I	889	1	0.81	809	1	0.81
1HIA_AB:I	1,853	1	0.52	3,731	1	0.52	2VIS_AB:C	8	12,239	2.17	8	12,678	2.17
1I2M_A:B	129	433	0.99	1,633	2	0.98	7CEI_A:B	518	162	0.34	2,468	58	0.34

TABLE 5

Shape-complementarity-based bound-bound docking results with and without electrostatics using F^2Dock. 'Rank' is the best rank among all predicted positions whose RMSD was less than 5Å. 'Good Peaks' is the number of peaks in the predicted set which were less than 5Å RMSD from the known position. 'RMSD' is the lowest RMSD among all peaks that were shortlisted. F^2Dock used use 6° rotational sampling, and retained 50,000 peaks. RMSD was calculated using the C_α atoms near the interface of the known bound conformation (within 5Å of the interface).

complementarity at protein-protein interfaces: The critical role played by surface normals at well placed, sparse, points in docking," *Journal of Molecular Biology*, vol. 252, no. 2, pp. 263–273, September 1995.

[16] D. Fischer, S. L. Lin, H. L. Wolfson, and R. Nussinov, "A geometry-based suite of moleculardocking processes," *Journal of Molecular Biology*, vol. 248, no. 2, pp. 459–477, 1995.

[17] D. Duhovny, R. Nussinov, and H. J. Wolfson, "Efficient unbound docking of rigid molecules," in *Proceedings of the Fourth International Workshop on Algorithms in Bioinformatics*, R. Guigo and D. Gusfield, Eds., Springer-Verlag GmbH Rome, Italy, September 2002, pp. 185–200.

[18] H.-P. Lenhof, "New contact measures for the protein docking problem," in *RECOMB '97: Proceedings of the first annual international conference on Computational molecular biology*. New York, NY, USA: ACM Press, 1997, pp. 182–191.

[19] F. S. Kuhl, G. M. Crippen, and D. K. Friesen, "A combinatorial algorithm for calculating ligand binding," *Journal of Computational Chemistry*, vol. 5, no. 1, pp. 24–34, 1984.

[20] M. L. Connolly, "Shape complementarity at the hemoglobin $\alpha_1\beta_1$ subunit interface," *Biopolymers*, vol. 25, no. 7, pp. 1229–1247, February 1986.

[21] H. Wang, "Grid-search molecular accessible surface algorithm for solving the protein docking problem," *Journal of Computational Chemistry*, vol. 12, no. 6, pp. 746–750, 1991.

[22] F. Jianga and S.-H. Kim, ""soft docking": Matching of molecular surface cubes," *Journal of Molecular Biology*, vol. 219, no. 1, pp. 79–102, May 1991.

[23] N. C. J. Strynadka, M. Eisenstein, E. Katchalski-Katzir, B. K. Shoichet, I. D. Kuntz, R. Abagyan, M. Totrov, J. Janin, J. Cherfils, F. Zimmerman, A. Olson, B. Duncan, M. M. Rao, R. Jackson, M. Sternberg, and M. N. G. James, "Molecular docking programs successfully predict the binding of a β-lactamase inhibitory protein to tem-1 β-lactamase," *Nature Struct. Biol.*, vol. 3, pp. 233–239, 1996.

[24] S. Belongie, J. Malik, and J. Puzicha, "Matching shapes," *Computer Vision, IEEE International Conference on*, vol. 1, p. 454, 2001.

[25] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo, and I. A. Vakser, "Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 6, pp. 2195–2199, 1992.

[26] D. W. Ritchie and G. J. Kemp, "Protein docking using spherical polar fourier correlations," *Proteins: Structure, Function, and Genetics*, vol. 39, no. 2, pp. 178–194, March 2000.

[27] J. G. Mandell1, V. A. Roberts, M. E. Pique, V. Kotlovyi, J. C. Mitchell, E. Nelson, I. Tsigelny, and L. F. T. Eyck, "Protein docking using continuum electrostatics and geometric fit," *Protein Engineering*, vol. 14, no. 2, pp. 105–113, February 2001.

[28] R. Chen and Z. Weng, "Docking unbound proteins using shape complementarity, desolvation, and electrostatics," *Proteins: Structure, Function, and Genetics*, vol. 47, no. 3, pp. 281–294, March 2002.

[29] R. Chen, L. Li, and Z. Weng, "Zdock: An initial-stage protein-docking algorithm," *Proteins: Structure, Function, and Genetics, Special Issue: CAPRI - Critical Assessment of PRedicted Interactions . Issue Edited by Joël Janin*, vol. 52, no. 1, pp. 80–87, May 2003.

[30] R. Chen and Z. Weng, "A novel shape complementarity scoring function for protein-protein docking," *Proteins: Structure, Function, and Genetics*, vol. 51, no. 3, pp. 397–408, March 2003.

[31] L. Li, R. Chen, and Z. Weng, "Rdock: Refinement of rigid-body protein docking predictions," *Proteins: Structure, Function, and Genetics*, vol. 53, no. 3, pp. 693–707, September 2003.

[32] M. Meyer, P. Wilson, and D. Schomburg, "Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking," *Journal of Molecular Biology*, vol. 264, no. 1, pp. 199–210, November 1996.

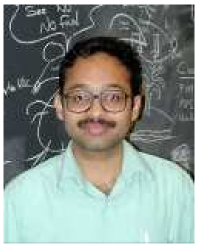
[33] D. W. Ritchie, "Parametric protein shape recognition," PhD Thesis, Departments of Computer Science & Molecular and Cell Biology, University of Aberdeen, King's College, Aberdeen, UK, September 1998.

[34] D. W. Ritchie and G. J. L. Kemp, "Fast computation, rotation, and

comparison of low resolution spherical harmonic molecular surfaces," *Journal of Computational Chemistry*, vol. 20, no. 4, pp. 383–395, February 1999.
[35] B. S. Duncan and A. J. Olson, "Approximation and characterization of molecular surfaces," *Biopolymers*, vol. 33, pp. 219–229, 1993.
[36] N. L. Max and E. D. Getzoff, "Spherical harmonic molecular surfaces," *IEEE Computer Graphics & Applications*, vol. 8, pp. 42–50, 1988.
[37] J. A. Kovacs and W. Wriggers, "Fast rotational matching," *Acta Crystallographica, Biological Crystallography*, vol. D58, no. 8, pp. 1282–1286, August 2002.
[38] J. A. Kovacs, P. Chacón, Y. Cong, E. Metwally, and W. Wriggers, "Fast rotational matching of rigid bodies by fast fourier transform acceleration of five degrees of freedom," *Acta Crystallographica, Biological Crystallography*, vol. D59, no. 8, pp. 1371–1376, August 2003.
[39] D. J. Bacon and J. Moult, "Docking by least-squares fitting of molecular surface patterns," *Journal of Molecular Biology*, vol. 225, no. 3, pp. 849–858, June 1992.
[40] P. H. Walls and M. J. E. Sternberg, "New algorithm to model protein-protein recognition based on surface complementarity. applications to antibody-antigen docking," *Journal of Molecular Biology*, vol. 228, no. 1, pp. 277–297, November 1992.
[41] M. Helmer-Citterich and A. Tramontano, "Puzzle: A new method for automated protein docking based on surface shape complementarity," *Journal of Molecular Biology*, vol. 235, no. 3, pp. 1021–1031, January 1994.
[42] A. Fahmy and G. Wagner, "Treedock: A tool for protein docking based on minimizing van der waals energies," *Journal of the American Chemical Society*, vol. 124, no. 7, pp. 1241–1250, February 2002.
[43] S.-Y. Yue, "Distance-constrained molecular docking by simulated annealing," *Protein engineering.*, vol. 4, no. 2, pp. 177–184, December 1990.
[44] J. Cherfils, S. Duquerroy, and J. Janin, "Protein-protein recognition analyzed by docking simulation," *Proteins: Structure, Function, and Genetics*, vol. 11, no. 4, pp. 271–280, 1991.
[45] R. Gabdoulline and R. Wade, "Analytically defined surfaces to analyze molecular interaction properties," *J. of Molecular Graphics*, vol. 14, no. 6, pp. 341–353, December 1996.
[46] C. Bajaj, G. Xu, and Q. Zhang, "A fast variational method for the construction of resolution adaptive c^2 smooth molecular surfaces," *Computer Methods in Applied Mechanics and Engineering*, vol. 198, pp. 1684–1690, 2009.
[47] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo, and I. A. Vakser, "Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 6, pp. 2195–2199, March 1992.
[48] "PDB2PQR: An automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations," http://pdb2pqr.sourceforge.net/.
[49] D. Potts, G. Steidl, and M. Tasche, *Fast fourier transform for nonequispaced data: A tutorial, in Modern Sampling Theory: Mathematics and Applications*, 1998, ch. 12, pp. 249–274.
[50] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A*, vol. 32, pp. 922–923, 1976.
[51] ——, "A discussion of the solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A*, vol. 34, no. 5, pp. 827–828, September 1978.
[52] C. Bajaj and S.-C. Chen, "Efficient and accurate higher-order fast multipole boundary element method for poisson boltzmann electrostatics," The University of Texas at Austin, ICES Report 09-xx, April 2009.
[53] M. Frigo and S. G. Johnson, "The design and implementation of FFTW3," *Proceedings of the IEEE. Invited paper, Special Issue on Program Generation, Optimization, and Platform Adaptation*, vol. 93, no. 2, pp. 216–231, february 2005.
[54] "The Message Passing Interface (MPI) Standard," http://www-unix.mcs.anl.gov/mpi/.
[55] C. Bajaj, P. Djeu, V. Siddavanahalli, and A. Thane, "TexMol: Interactive visual exploration of large flexible multi-component molecular complexes," in *Proc. of the Annual IEEE Visualization Conference*, Austin, Texas, 2004, pp. 243–250.

Chandrajit Bajaj graduated from the Indian Institute of Technology, Delhi with a Bachelor's Degree in Electrical Engineering, in 1980 and received his M.S. and Ph.D. degrees in Computer Science from Cornell University, in 1983, and 1984 respectively. Bajaj is currently the Computational Applied Mathematics Chair in Visualization Professor of Computer Sciences at the University of Texas at Austin, as well as the director of the Center for Computational Visualization, in the Institute for Computational and Engineering Sciences (ICES). His research areas of interest include Computational Biology and Biophysics, Computational Geometry, Computer Graphics, Geometric Modeling, Image Processing, and Visualization. He has over 200 publications, has written one book and edited three other books in his areas of expertise. He is on the editorial boards for the ACM Transactions on Graphics, ACM Computing Surveys, the International Journal of Computational Geometry & Applications, and SIAM Journal of Imaging Sciences. He is on numerous national and international committees, and has served as a scientific consultant to national labs and industry. He is also a fellow of The American Association for the Advancement of Science.

Rezaul Chowdhury received his BS degree in Computer Science & Engineering from Bangladesh University of Engineering & Technology, and PhD in Computer Science from the University of Texas at Austin in 2007. Currently he is a postdoctoral fellow at the Institute for Computational Engineering & Sciences, UT Austin. His research interests include design and analysis of combinatorial algorithms & data structures, bioinformatics & computational biology, algorithms for massive datasets, cache-efficient algorithms, and parallel & multicore computing.



Vinay Siddavanahalli obtained a Bachelor of Engineering degree in Computer Science from Sri Jayachamarajendra College of Engineering in Mysore, and also worked for a year on "MPEG text detection and recognition", as part of his thesis in The Indian Institute of Science, Bangalore, India. He received his MS and PhD degree in Computer Science from the University of Texas at Austin in 2006. Currently he is working as a software engineer at Google. His research interests include Computer Graphics, Computational Geometry and Computational Biology.