Protein-Protein Docking with F²Dock 2.0 and GB-rerank

Supplement
Rezaul Chowdhury¹, Muhibur Rasheed¹, Donald Keidel², Maysam Moussalem¹, Arthur Olson², Michel Sanner², Chandrajit Bajaj²,*

¹ Department of Computer Science, University of Texas at Austin, 1 University Station C0500, Austin, Texas 78712, USA
² The SCRIPPS Research Institute, Torrey Pines Rd N, La Jolla, CA 92037, USA
* E-mail: bajaj@cs.utexas.edu

Contents

1 Details on Affinity Function Computations 2
  1.1 Shape Complementarity .............................................. 2
    1.1.1 Skin-core Definition and Weighting .............................. 2
    1.1.2 FFT based formulation ......................................... 4
  1.2 Electrostatics (E) .................................................. 5
    1.2.1 FFT based formulation ......................................... 5
  1.3 Interface Propensity (IP) and Hydrophobicity (HP) ............ 6
    1.3.1 FFT based formulation ......................................... 6

2 Details on Filters 8
  2.1 Lennard-Jones Filter ............................................... 8
  2.2 Interface Area Filter ............................................... 9
  2.3 Interface Propensity Filter ...................................... 9
  2.4 Clash Filter ........................................................ 10
  2.5 Residue-Residue Contact Filter .................................. 10
  2.6 Glycine Filter ..................................................... 11
  2.7 Antibody-Antigen Contact Filter ................................ 11
  2.8 Overall Cost of Filtering ......................................... 11

3 Solvation Energy Based Reranking 12
  3.1 Approximating $E_{pol}$ (Polarization Energy) .................... 12
  3.2 Approximating $E_{vdw(s-s)}$ (Dispersion Energy) ................ 13
  3.3 Approximating $E_{cav}$ (Cavity Forming Energy) ............... 13
  3.4 Overall Cost of Reranking ......................................... 14

4 Additional Results on the Effects of Various Affinity Functions and Filters 15

5 Additional Results Comparing ZDock 3.0.2 and F²Dock 2.0 25
1 Details on Affinity Function Computations

This section adds detail on the FFT based implementation and parameter selection of three affinity functions: shape complementarity, electrostatics and interface propensity (or hydrophobicity).

Figure S1: Traditional vs improved double-skin definitions. (Left) Traditional double skin-layer approach for shape complementarity, (Right) Curvature-based weighting of skin atoms and depth dependent weighting of core atoms of molecule A, and depth dependent weighting of the atoms of B

1.1 Shape Complementarity

First we present the new skin-core definitions in detail.

1.1.1 Skin-core Definition and Weighting

- Floating Receptor (i.e., Stationary Molecule A) Skin: Unlike the traditional approach [1] the receptor skin layer does not touch the receptor van der Waals surface (i.e., there is a gap between the skin and the core regions), and also the radius of those skin atoms differ from that in the traditional approach. The plots in Figure S2 justify this skin-core gap which, for each of 60 rigid-body test cases chosen from Zlab Benchmark Suite 2.0 [2], plots average distance and standard deviation of ligand atom centers from the receptor van der Waals surface when the ligand is bound to the receptor. We have considered only those ligand atoms that lie within 3Å from the receptor surface in bound condition. For the bound-bound case (when both molecules are crystallized together in the complex form) the average distance lies between 1.6Å and 2.0Å, while for unbound-unbound case (when the two molecules are crystallized separately) it lies between 1.4Å and 2.0Å most of the time. The standard deviation lies between 0.55Å and 0.75Å for the bound-bound case, and between 0.6Å and 1Å for the unbound-unbound case. In our experiments pseudo skin atoms of radius 1.1Å with their centers placed at a distance of 1.7Å from the receptor vdW surface (so there is a gap of 0.6Å) performed much better than the traditional approach where pseudo atoms of radius 1.4Å are placed touching the receptor vdW surface.
Figure S2: Interface gap for bound and unbound complexes in benchmark 4.0. Plots show average distance (in red) and standard deviation of ligand atom centers from the receptor van der Waals surface when the ligand is bound to the receptor. Only the ligand atoms that lie within 3Å from the receptor surface in bound condition are considered. The plots are for the rigid-body test cases from Zlab Benchmark Suite 2.0 [2]. Top: Bound-bound. Bottom: Unbound-unbound

- CURVATURE DEPENDENT WEIGHTING OF RECEPTOR SKIN: The weights assigned to receptor skin atoms are computed based on the curvature of the skin around that atom (as opposed to
the constant weights assigned to skin atoms in the traditional approach [1]). Such weighting encourages convex-concave and concave-convex interfaces as opposed to large flat interfaces.

We know that the centers of molecule A’s skin atoms are placed at a constant distance \(d\) from the receptor vdW surface \(S\). Let us enlarge \(S\) outward by distance \(d\) so that it passes through the skin atom centers, and let \(V\) be the volume enclosed by the enlarged surface. Let \(B_k\) be a ball of radius \(\delta\) at the center of skin atom \(k\), and \(B_{k}^{in} = B \cap V\) and \(B_{k}^{out} = V \setminus B_{k}^{in}\). We define:

\[
\rho(k) = \max\left\{\frac{|B_{k}^{in}|}{|B_{k}^{out}|}, \frac{|B_{k}^{out}|}{|B_{k}^{in}|}\right\}.
\]

Then the weight assigned to skin atom \(k\) of molecule A is \(c_{SC,A,k} = \rho(k)\).

We approximate \(\rho(k)\) from the 3D grid embedding of molecule A along with its skin region, and \(\delta = 3.6\) produced good results in our experiments.

- **Depth dependent Weighting of Receptor Core:** The core atoms of molecule A are assigned weights using an increasing function of depth (distance of the atom center from the surface of A). Such weightings discourage deeper core-core overlaps more compared to shallower ones by making it difficult to compensate for deep core-core overlaps with wide skin-skin overlaps.

The core atom weights are computed based on *atomic layers*. Any atom which contributes to the solvent excluded surface (SES) is considered an exposed atom. The exposed atoms of a molecule lie in layer 1. When atoms in layer 1 are removed, the new exposed atoms make layer 2. In general, when atoms in layers 1 to \(l \geq 1\) are removed, the exposed atoms among the remaining ones make layer \(l + 1\). The atomic layer of core atom \(k\) of A is denoted by \(L_{A}(k)\), and the weight assigned to that atom is \(c_{SC,A,k} = i \cdot \alpha^{L_{A}(k)}\), where \(\alpha \geq 1\) is a constant. A value of \(\alpha = 1.1\) worked quite well in our experiments.

Hence overall, the weight assigned to atom/pseudo-atom \(k\) of molecule A:

\[
c_{SC,A,k} = \begin{cases} 
\rho(k) & \text{if atom } k \in \text{skin}(A), \\
 i \cdot \alpha^{L_{A}(k)} & \text{otherwise}. 
\end{cases}
\]

- **Thicker Skin of the Ligand (i.e., Moving Molecule B) with Skin-Core Overlap:** Since in the traditional approach [1] the ligand skin is defined using its surface atoms, the skin thickness varies and can be too thin in some areas. It turns out that a double layer ligand skin with skin-core overlap performs better than a single skin layer disjoint from the core. The weights assigned to the atoms of B are as follows.

\[
c_{SC,B,k} = \begin{cases} 
1 & \text{if } L_{B}(k) = 1, \\
1 + i & \text{if } L_{B}(k) = 2, \\
i & \text{otherwise}. 
\end{cases}
\]

### 1.1.2 FFT based formulation

Now suppose the user-specified relative weights for various types of overlaps are as follows:

\(w_{ss} = \text{reward for (unit) skin-skin overlap,}\)
Let $A'$ denote molecule $A$ with its grown skin layer. If the weight assigned to an atom/pseudo-atom $k$ (of molecule $P \in \{A', B\}$) is $c^SC_k = c^{SC}_{k,Re} + i \cdot c^{SC}_{k,Im}$, then the affinity function for shape complementarity is:

$$f^{SC}_{P}(x) = \sum_{k \in P} \left( \sqrt{w_{ss}} \cdot c^{SC}_{k,Re} + i \cdot \sqrt{w_{cc}} \cdot c^{SC}_{k,Im} \right) \cdot g^{SC}_{P,k}(x),$$

where, $g^{SC}_{P,k}(x) = e^{-\beta \left( \frac{\|x-c_k\|^2}{r^2_k} - 1 \right)}$ with blobbiness $\beta = 2.3$.

The overall shape complementarity score for translation $t$ and rotation $r$ is:

$$H^{SC}_{A,B}(t,r) = \text{Re} \left( f^{SC}_{A,B}(t,r) \right) + \frac{w_{sc}}{\sqrt{w_{ss} \cdot w_{cc}}} \cdot \text{Im} \left( f^{SC}_{A,B}(t,r) \right),$$

where, $f^{SC}_{A,B}(t,r) = \int x f^{SC}_{A}(x) T_t(\Delta_r(f^{SC}_{B}(x)))$ with $T$ and $\Delta$ being the translation and the rotation operator, respectively.

### 1.2 Electrostatics (E).

In [3] Gabb et. al. described a simplified model for electrostatics which allows efficient FFT-based computation of the term during docking search. We used this simple electrostatics model in the original version of F$^2$Dock 2.0 [1]. The first protein’s electric potential is computed and matched against the charges in the other. Charge assignments are made using PDB2PQR [4]).

Two affinity functions $f^{E}_{A}$ and $f^{E}_{B}$ are defined for molecule $A$ and $B$, respectively.

$$f^{E}_{A}(x) = \sum_{k \in A} q_k E(x - c_k)(x - c_k) \cdot g^{E}_{A,k}(x)$$

and

$$f^{E}_{B}(x) = \sum_{k \in B} q_k \delta(x - c_k) \cdot g^{E}_{B,k}(x),$$

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

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

#### 1.2.1 FFT based formulation

In the current version of F$^2$Dock 2.0, we define for $P \in A, B$,

$$g^{E}_{P,k}(x) = e^{-\beta \left( \frac{\|x-c_k\|^2}{r^2_k} - 1 \right)},$$
where, $\gamma > 0$ is a constant, and blobbiness $\beta = 2.3$. In our experiments the new definition of $g^E_{P,k}(x)$ with $\gamma = 3.4$ performed much better than the original Gabb et al. formulation. Smoothing with the Gaussian function as above has the effect of reduced discretization error on the grid.

The overall electrostatics score for translation $t$ and rotation $r$ is:

$$H^E_{A,B}(t, r) = w_E \cdot \text{Re} \left( F^E_{A,B}(t, r) \right),$$

where, $F^E_{A,B}(t, r) = \int_x f^E_A(x) T_t \left( \Delta_r \left( f^E_B(x) \right) \right)$, and $w_E$ is the user-specified weight given to electrostatics interaction.

### 1.3 Interface Propensity (IP) and Hydrophobicity (HP).

F^2Dock 2.0 scores the interfaces between molecules $A$ and $B$ using the per-residue interface propensity values computed in [5] based on the relative frequencies of different residues in the interfaces of a set of 63 protein-protein complexes from [6]. Let $IP(R)$ denote the natural logarithm of the interface propensity value of a residue $R$. Alternatively, the user can choose to use per-residue Hydrophobicity values from [7]. Note that, Hydrophobic residues are more likely to be found on the interface, and hence Hydrophobicity and Interface propensity model the same chemical property and hence are used as substitutes in F^2Dock 2.0, as opposed to using both at the same time.

The IP values for the 20 amino acid residues lie between -0.38 (for ASP) and 0.83 (for TRP). A residue with a higher IP value is likely to occur more frequently in a protein-protein interface than one with a lower IP value. Hydrophobic residues Phe, Met, Ile, Leu and Val, polar aromatic residues Trp, Tyr and His, and the charged residue Arg favor interface locations, and they have positive IP values. Additionally, Cys and Asn also have positive IP. The IP value of an atom is set to the IP value of the residue it belongs to.

Suppose $A + B_{t,r}$ is a docking pose obtained by rotating molecule $B$ by $r$ and translating by $t$. Let $iAtom^+_t(P)$ and $iAtom^-_t(P)$ denote the set of atoms in the interface of $P \in \{A, B\}$ in this docking pose that have positive and negative IP values, respectively. Also let $iAtom^+_{t,r}(A, B) = iAtom^+_t(A) + iAtom^-_t(B)$, for $s \in \{+, -\}$. Then we assign the following interface propensity score to the docking pose:

$$IP-score_{t,r}(A, B) = \frac{-\sum_{a \in iAtom^+_t(A, B)} IP(a)}{\min\left(\sum_{a \in iAtom^-_t(A, B)} IP(a)\right)},$$

where $IP_t = \max_{IP(R) > 0} IP(R)$.

#### 1.3.1 FFT based formulation

We show below how we approximate $IP-score_{t,r}(A, B)$. For this approximation, all atoms of $A$ and $B$ are treated as solid spheres.

For $Q \in \{A, B\}$, we define two affinity functions $f^{IP}_A$ and $f^{IP}_B$ for $A$ and $B$, respectively, and evaluate $F^{IP,Q}_{A,B}(t, r) = \int_x f^{IP,Q}_A(x) T_t \left( \Delta_r \left( f^{IP,Q}_B(x) \right) \right)$. Let $Q' \in \{A, B\} \setminus \{Q\}$.

We set the radius of each atom of $Q$ to its van der Waals radius, while the van der Waals radius of each atom of $Q'$ is extended by a non-negative constant $r_{ext}^1$.

We assign weight $c^{IP,Q}_{P,k}$ to each atom $k \in P = \{Q, Q'\} = \{A, B\}$ as follows:

$\footnote{r_{ext} = 1.5\AA$ produced good results in our experiments.} $
\[ c_{Q,k}^{IP} = \begin{cases} IP(k) & \text{if } k \text{ surface atom, } IP(k) > 0, \\ -i \cdot IP(k) & \text{if } k \text{ surface atom, } IP(k) < 0, \\ 0 & \text{otherwise.} \end{cases} \]

\[ c_{Q',k}^{IP} = \begin{cases} 1 & \text{if } k \text{ is a surface atom,} \\ 0 & \text{otherwise.} \end{cases} \]

Then we define:

\[ f_{Q}^{IP,Q}(x) = \sum_{k \in Q} c_{Q,k}^{IP} \cdot g_{Q,k}^{IP,Q}(x) \]

where \( g_{Q,k}^{IP,Q}(x) = 1 \) if \(|x - c_k| \leq r_k\) and 0 otherwise; and

\[ f_{Q'}^{IP,Q}(x) = \bigvee_{k \in Q'} c_{Q',k}^{IP} \cdot g_{Q',k}^{IP,Q}(x) \]

where \( g_{Q',k}^{IP,Q}(x) = 1 \) if \(|x - c_k| \leq r_k + r_{ext}\) and 0 otherwise.

The overall interface propensity score for translation \( t \) and rotation \( r \) (i.e., approximated \( IP\)-score \( t, r(A, B) \)) is:

\[ H_{A,B}^{IP}(t, r) = w_{IP} \cdot \frac{Re \left( F_{A,B}^{IP,A}(t, r) + F_{A,B}^{IP,B}(t, r) \right)}{IP + Im \left( F_{A,B}^{IP,A}(t, r) + F_{A,B}^{IP,B}(t, r) \right)}, \]

where \( w_{IP} \) is the user-specified weight given to interface propensity.
2 Details on Filters

In this section we describe the implementations of various filters available in F^2Dock 2.0. Each filter has a preprocessing step (e.g., construction of octrees and initialization of other relevant data structures) which is executed only once before F^2Dock 2.0 enters the rotations loop in phase I, and so not included in the running times of the filters which are executed in each iteration of the loop. Also octrees constructed for the initial configuration of molecule B can be easily transformed on-the-fly so that it correctly matches any rotated configuration of B by applying the corresponding rotation matrix on the node coordinates (e.g., node center) and atoms/points stored in them.

2.1 Lennard-Jones Filter.

We approximate the Lennard-Jones (LJ) potential between molecules A and B\textsubscript{i,r} given by the following expression.

\[ LJ(A, B_{i,r}) = \sum_{i \in A, j \in B_{i,r}} \left( \frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^{6}} \right), \]

where \( r_{ij} \) is the distance between atoms \( i \in A \) and \( j \in B_{i,r} \), constants \( a_{ij} \) and \( b_{ij} \) depend on the type (e.g., C, H, O, etc.) of the two atoms involved. For any fixed pair of atom types \( a_{ij} \) and \( b_{ij} \) are fixed, and are calculated from the Amber force field using well depths \( \mu \) and equivalence contact distances of homogeneous pairs \( r_{eqm} \) as follows (assuming \( X = \text{atomType}(i \in A) \) and \( Y = \text{atomType}(j \in B_{i,r}) \)) [8, 9].

\[ a_{ij} = \mu_{XY} r_{eqm,XY}^{12}, \quad b_{ij} = 2\mu_{XY} r_{eqm,XY}^{6}, \]

\[ \mu_{XY} = \sqrt{\mu_{XX} + \mu_{YY}}, \quad r_{eqm,XY} = \frac{1}{2}(r_{eqm,XX} + r_{eqm,YY}) \]

We assume that \( X, Y \in \{C, H, N, O, P, S\} \).

Our approach is to identify potential docking solutions based on a cutoff potential. Assuming that the docking solution can not have steric clashes at the interface, we can filter (or penalize) all poses with positive LJ potential. However, though this assumption is not unreasonable for bound-bound docking, in the case of bound-unbound and unbound-unbound docking we need to allow some steric clashes. Hence we soften the potential by reducing the \( r_{eqm,XX} \) values by a constant factor \( \delta_{LJ} \) which has the effect of reducing the inter-atomic clash distances\(^2\) by the same factor. As a result docking poses with negative LJ values can now have some clashes, and the maximum number of allowed clashes will depend on the value of \( \delta_{LJ} \). In our experiments \( \delta_{LJ} = 0.3 \) produced good results.

Observe that direct computation of \( LJ(A, B_{i,r}) \) requires \( O(M_A M_B) \) time, where \( M_A \) (resp. \( M_B \)) is the number of atoms in molecule A (resp. B). Instead we use our octree-based fast multipole-type algorithm described in [10] to obtain a \( 1 + \epsilon \) approximation of the potential in \( O \left( \frac{1}{\epsilon^2}(M_A + M_B) \log (M_A + M_B) \right) \) time and \( O(M_A + M_B) \) space, where \( \epsilon \in (0, 1] \) is a user-defined approximation parameter.

\(^2\)i.e., \( r_{ij} \) values at which two atoms are considered to be in a clash
2.2 Interface Area Filter.

We sample weighted quadrature/integration points from the surface of each molecule as described in [11, 10]. Sum of the weights of all such integration points is equal to the molecular surface area up to quadrature approximation error, and each integration point can be treated as the center of a small surface patch with its weight being the patch area. Each such patch \( \pi \) with center \( c_\pi \) and area \( a_\pi \) is represented by the tuple \( \langle c_\pi, a_\pi \rangle \).

We then evaluate the following expression as an approximation of the interface area between molecules \( A \) and \( B_{t,r} \).

\[
\text{InterfaceArea}(A, B_{t,r}) = \sum_{\pi \in IP_{A}(B_{t,r}) \cup IP_{B_{t,r}}(A)} a_\pi,
\]

where,

\[
IP_{P_1}(P_2) = \left\{ \pi_1 | (\pi_1 \in S_{P_1}) \land \exists \pi_2 \in S_{P_2} (\text{dist}(c_{\pi_1}, c_{\pi_2}) < \mu_{IA}) \right\}
\]

and \( \mu_{IA} \) is a user-defined constant.

Using our algorithm described in [10] based on octrees [12] and our Dynamic Packing Grid (DPG) data structure [13], the expression above can be evaluated in \( O((M_A + M_B) \log (M_A + M_B)) \) (w.h.p.)\(^3\) time and \( O(M_A + M_B) \) space. However, this algorithm runs much faster in practice, i.e., in \( O(N_{int}) \) time (w.h.p.), where \( N_{int} \) is the total number of patches in the interface.

2.3 Interface Propensity Filter.

As in Section 2.2 we assume that we have a patch decomposition of the surface of each molecule \( P \in \{ A, B \} \). However, each patch \( \pi \) is now represented as a triple \( \langle c_\pi, a_\pi, h_\pi \rangle \) where the additional value \( h_\pi \) is a measure of interface propensity of \( \pi \) in the following sense. We assume that each atom \( k \in P \) is assigned an interface propensity value \( h_k \) as in Section 1.3, and \( h_\pi = \frac{1}{|S_{P,\pi}|} \sum_{k \in S_{P,\pi}} h_k \), where \( S_{P,\pi} = \{ k | (k \in P) \land (\text{dist}(c_k, c_\pi) < \nu r_k) \} \), \( c_k \) is the center of atom \( k \), \( r_k \) is its van der Waals radius, and \( \nu \) is a user-defined constant.

This filter computes \( IPVec(A, B_{t,r}) \) which is defined as the following vector:

\[
\begin{bmatrix}
\sum_{\pi \in IP_{A}(B_{t,r}) \land h_\pi < 0} h_\pi a_\pi, & \sum_{\pi \in IP_{A}(B_{t,r}) \land h_\pi > 0} h_\pi a_\pi, \\
\sum_{\pi \in IP_{B_{t,r}}(A) \land h_\pi < 0} h_\pi a_\pi, & \sum_{\pi \in IP_{B_{t,r}}(A) \land h_\pi > 0} h_\pi a_\pi
\end{bmatrix}
\]

where \( IP_{P_1}(P_2) \) is as defined in Section 2.2.

\(^3\)For an input of size \( n \), an event \( E \) occurs w.h.p. (with high probability) if, for any \( \alpha \geq 1 \) and \( c \) independent of \( n \), \( Pr(E) \leq 1 - \frac{c}{n^\alpha} \).
This vector can be computed simultaneously with the approximate interface area in Section 2.2 incurring negligible overhead. The algorithm is given in [10].

Once we have $IPVec(A, B_{t,r}) = (v_1, v_2, v_3, v_4)$, we approximate $IP-score_{t,r}(A, B)$ defined in Section 1.3 as follows:

$$IP-score_{t,r}(A, B) \approx -\frac{v_2 + v_4}{\min(IP'_t, v_1 + v_3)},$$

where, $IP'_t = (\max_{h_k < 0} h_k) (\min_{\pi \in A \cup B} a_\pi)$.

We also observed that rewarding docking poses with larger positive interface propensity weighted area $v_2 + v_4$ generally improves the ranks of near native poses. However, rewarding based on the product of $IP-score_{t,r}(A, B)$ and $v_2 + v_4$ seems to work even better in practice. This composite term which we define as $IP-product_{t,r}(A, B)$ below, rewards a docking pose with higher $v_2 + v_4$ value more than one with lower such value when both have the same $IP-score$.

$$IP-product_{t,r}(A, B) = IP-score_{t,r}(A, B) \times (v_2 + v_4) \approx -\frac{(v_2 + v_4)^2}{\min(IP'_t, v_1 + v_3)}$$

The interface propensity filter in F$^2$Dock 2.0 penalizes a docking pose provided its $IP-score$ is below a user-defined threshold $\mu_{IP}$. The $IP-product$ value, after weighting it by a user-specified weight $w'_{IP}$, is added to the overall score of the pose.

2.4 Clash Filter.

Two atoms $a \in A$ and $b \in B$ with van der Waals radii $r_a$ and $r_b$, respectively, are said to be in a clash provided the distance between their centers is smaller than $\alpha(r_a + r_b)$, where $\alpha$ is a user-defined positive constant. F$^2$Dock 2.0 counts the total number of atomic clashes between molecules $A$ and $B_{t,r}$ as follows.

$$Clash_{t,r}(A, B) = |\{b(b \in B_{t,r}) \land \exists a \in A(r_{a,b} < \alpha(r_a + r_b))\}|$$

Direct computation of this number requires $O(M_AM_B)$ time. So we use an octree-based spatial subdivision approach that is similar to the interface area approximation algorithm used in Section 2.2 and has similar running time.

F$^2$Dock 2.0 penalizes a docking pose $A + B_{t,r}$ provided $Clash_{t,r}(A, B) > \mu_C$, where $\mu_C$ is a user-defined constant. A value of around 10 for $\mu_C$ seems to work well in practice.

2.5 Residue-Residue Contact Filter.

F$^2$Dock 2.0 uses the pairwise contact preferences listed in either Table III (without volume normalization) or Table IV (normalized w.r.t. residue volumes) of [14] based on user preference with Table III being the default. The contact preferences were derived as follows. Two residues are considered to be in contact if the distance between their $C_\beta$ atoms ($C_\alpha$ for Gly) is less than 6 Å. The normalized number of contacts between residue types $i$ and $j$ is defined as $Q_{ij} = C_{ij}/\sum_{k<l} C_{kl}$, where $C_{ij}$ is the actual number of contacts observed between residue types $i$ and $j$. The expected number of contacts $W_{ij}$ is obtained by assuming that there are no contact preferences between residues of
of all positive and negative two molecules using a fast algorithm similar to the one used in Section 2.3, and compute the sum of all positive and negative \( G_{ij} \) values (or \( G_{ij}(v) \) values if chosen by the user) denoted by \( G^+ \) and \( G^- \), respectively. Then we compute the following ratio:

\[
RC\text{-score}_{t,r}(A,B) = \frac{G^+}{\min(G^-, RC_t)}.
\]

where \( RC_t = \max_{G_{ij} < 0} G_{ij} \) (or \( RC_t = \max_{G_{ij}(v) < 0} G_{ij}(v) \) if the user chooses volume-normalized contact preferences).

\( F^2 \)Dock 2.0 penalizes a docking pose if its \( RC\text{-score} \) is below a user-specified constant \( \mu_R C \). Without volume normalization \( \mu_R C = 3.0 \) seems to work well in practice. Docking poses can also be penalized based on user-specified lower and upper bounds for \( G^+ \) and \( G^- \), respectively.

### 2.6 Glycine Filter.

We mark the oligopeptides with the properties described above on the enzyme surface, and for any given docking pose count the number of these motifs occurring at the interface using a fast algorithm similar to the one used for counting atomic clashes in Section 2.4.

### 2.7 Antibody-Antigen Contact Filter.

We say that an atom \( a \) is in close neighborhood of another atom \( b \) provided the distance between the centers of \( a \) and \( b \) is at most \( 2(r_a + r_b) \). Given a potential antibody-antigen docking pose, \( F^2 \)Dock 2.0 computes three quantities: \( N_{L1∪H1} \), \( N_L3 \) and \( N_{H3} \), denoting the number of antigen atoms that are in the close neighborhood of any atom in the antibody regions CDR-L1/CDR-H1, CDR-L3 and CDR-H3, respectively. These counts are obtained using a fast algorithm similar to the one used in Section 2.4. \( F^2 \)Dock 2.0 penalizes a docking pose provided \( N_{L1∪H1} < \mu_{L1∪H1} \) or \( N_L3 < \mu_{L3} \) or \( N_{H3} < \mu_{H3} \), otherwise it adds \( w_{AC} (N_{L1∪H1} + N_L3 + N_{H3}) \) to the total score, where \( \mu_{L1∪H1} \), \( \mu_{L3} \), \( \mu_{H3} \) and \( w_{AC} \) are user-specified quantities. \( F^2 \)Dock 2.0 uses \( \mu_{L1∪H1} = 100 \), \( \mu_{L3} = \mu_{H3} = 150 \) and \( w_{AC} = 10 \) by default. The CDR (Complementarity Determining Region) loops are identified using the method described in [15].

### 2.8 Overall Cost of Filtering.

Summing up the cost of all filters described above for a single run, we obtain

\[
\mathcal{O} \left( \frac{1}{2} (M_A + M_B) \log (M_A + M_B) \right) \text{ (w.h.p.)}
\]

Assuming that each filter is applied on at most \( N_F \) configurations, the total time taken by all filters is

\[
\mathcal{O} \left( \frac{1}{2} N_F (M_A + M_B) \log (M_A + M_B) \right) \text{ (w.h.p.)}
\]

Assuming \( N_F = \mathcal{O} (N_R) \), where \( N_R \) is the number of samples in the rotations space, the running time reduces to

\[
\mathcal{O} \left( \frac{1}{2} N_R (M_A + M_B) \log (M_A + M_B) \right) \text{ (w.h.p.)}
\]
3 Solvation Energy Based Reranking

The solvation energy $E_{\text{sol}}$ consists of the energy to form cavity in the solvent ($E_{\text{cav}}$), the solute-solvent van der Waals interaction energy ($E_{\text{vdw(s-s)}}$), and the electrostatic potential energy change due to the solvation (also known as the polarization energy, $E_{\text{pol}}$) [16, 17, 18, 19].

$$E_{\text{sol}} = E_{\text{cav}} + E_{\text{vdw(s-s)}} + E_{\text{pol}}$$

The first two terms are often modeled as [16, 20]

$$E_{\text{cav}} = pV + \sum_i \gamma_i A_i$$

and

$$E_{\text{vdw(s-s)}} = \rho_0 \sum_i \int_{\text{ex}} u_i^{(\text{att})}(x_i, r) d^3r$$

where $p$ is the solvent pressure, $V$ is the molecular volume, $A_i$ is the solvent accessible surface area of atom $i$ and $\gamma_i$ is its solvation parameter, $\rho_0$ is the bulk density, and $u_i^{(\text{att})}$ is the van der Waals dispersive component of the interaction between atom $i$ and the solvent.

The last term, $E_{\text{pol}}$, can be approximated using Generalized Born (GB) theory [21].

$$E_{\text{pol}} = -\tau^2 \sum_{i,j} \frac{q_i q_j}{\sqrt{r_{i,j}^2 + R_i R_j e^{-r_{i,j}^2/\epsilon}}}$$

where $\tau = 1 - \frac{1}{\epsilon}$, and $R_i$ is the effective Born radius of atom $i$.

GB-rerank approximates each of these terms as described in the following sections, and reranks the list of top docking poses produced by F$^2$Dock 2.0 based on the resulting $\Delta E_{\text{sol}}$ values. In order to approximate $\Delta E_{\text{sol}}$, GB-rerank precomputes the $E_{\text{sol}}$ values for molecules $A$ and $B$, and then computes $E_{\text{sol}}$ for each docking pose.

3.1 Approximating $E_{\text{pol}}$ (Polarization Energy).

We first estimate all Born radii of the molecule/complex using our fast approximation scheme described in [10] followed by another of our algorithms from [10] for estimating $E_{\text{pol}}$ from the approximated Born radii. Given a molecule/complex consisting of $M$ atoms, both algorithms run in $O\left(\frac{\epsilon}{\epsilon^3} M \log M\right)$ time using $O(M)$ space, where $\epsilon > 0$ is an approximation parameter that provides a speed-accuracy tradeoff. The smaller the value of $\epsilon$ the more accurate the algorithms are, and the larger the $\epsilon$ value the faster they run.

GB-rerank uses either the $r^4$- or the $r^6$-approximation of $R_i$ (see [10]) based on user preference. By default, it uses the $r^6$-approximation [22] which shows better accuracy for spherical solutes as well as for proteins [23].

$$\frac{1}{R_i^3} = \frac{3}{4\pi} \int_{\text{ex}} \frac{1}{|r - x_i|^6} d^3r, \quad i \in [1, M].$$

where $\Gamma$ is the molecular surface, $n(r)$ is the unit outward normal on the molecular surface at $r$, and $x_i$ is the center of atom $i$. 

12
We obtain the following discrete surface formulation of Equation (2) by applying the divergence theorem and Gaussian quadrature.

\[
\frac{1}{R_i^3} \approx \frac{1}{4\pi} \sum_{k=1}^{N} w_k \frac{(r_k - x_i) \cdot \hat{n}_k}{|r_k - x_i|^6}.
\]  

(3)

where, the \( r_k \)'s are \( N = O(M) \) Gauss quadrature/integration points [11] on the molecular surface, and \( w_k \) is a weight assigned to \( r_k \) in order to achieve higher order of accuracy for small \( N \).

In order to avoid the slowdown due to the repeated computation of quadrature points for each docking pose \( A + B_{t,r} \). \( \text{F}^2\text{Dock 2.0} \) precomputes the set of quadrature points \( Q_A \) and \( Q_B \) for \( A \) and \( B \), respectively, and chooses all quadrature points from \( Q_A \cup Q_{B_{t,r}} \) that are not on the interface of \( A \) and \( B_{t,r} \) for use during the Born radii estimation of \( A + B_{t,r} \).

3.2 Approximating \( E_{\text{vdw}(s-s)} \) (Dispersion Energy).

The solute-solvent van der Waals interaction energy (also known as dispersion energy) is modeled as [16, 20]:

\[
E_{\text{vdw}(s-s)} = \rho_0 \sum_{i=1}^{M} \int_{\text{ex}} u_i^{(\text{att})}(x_i, r) d^3r.
\]

where \( \rho_0 \) is the bulk density, and \( u_i^{(\text{att})} \) is the van der Waals dispersive component of the interaction between atom \( i \in [1, M] \) and the solvent which is given as follows.

\[
u_i^{(\text{att})}(x_i, r) = \frac{1}{|r - x_i|^6}
\]

Thus

\[
E_{\text{vdw}(s-s)} = \rho_0 \sum_{i=1}^{M} \int_{\text{ex}} \frac{1}{|r - x_i|^6} d^3r
\]

(4)

If \( R_i \) is the Born radius of atom \( i \) calculated using the \( r^6 \)-approximation (i.e., Equation 2/3), then Equation 4 can be rewritten as:

\[
E_{\text{vdw}(s-s)} \approx \rho_0 \frac{4\pi}{3} \sum_{i=1}^{M} \frac{1}{R_i^3}
\]

(5)

Therefore, \( E_{\text{vdw}(s-s)} \) can be approximated in \( O(M) \) time once the Born radii of all atoms are available.

3.3 Approximating \( E_{\text{cav}} \) (Cavity Forming Energy).

Instead of computing \( E_{\text{cav}} \) for \( A, B \) and each \( A + B_{t,r} \) separately, we approximate \( \Delta E_{\text{cav}} \) with the buried surface area of \( A + B_{t,r} \) which is approximated using the same algorithm used for interface area filter in Section 2.2. The algorithm runs in \( O((M_A + M_B) \log (M_A + M_B)) \) (w.h.p.) time (much faster in practice; see Section 2.2) using \( O(M_A + M_B) \) space.
3.4 Overall Cost of Reranking.

Assuming that GB-rerank is applied on $N_G$ docking poses, its total running time is

$$O\left(\frac{1}{\epsilon^3} N_G (M_A + M_B) \log (M_A + M_B) \right) \text{(w.h.p.)}.$$ 

Typically $N_G = O(N_R)$, where $N_R$ is the number of samples in the rotations space, and so the running time reduces to

$$O\left(\frac{1}{\epsilon^3} N_R (M_A + M_B) \log (M_A + M_B) \right) \text{(w.h.p.)}.$$
4 Additional Results on the Effects of Various Affinity Functions and Filters

Figures S3 to S10 track the changes in the rank of the top hit, RMSD of the top hit, number of hits in top 1000, and the minimum RMSD in the top 1000 predictions, respectively, as the various F²Dock 2.0 options are activated one after another on the rigid-body test cases from Zlab benchmark 2.0 [2].

When we activate Lennard-Jones filter, clash filter and proximity clustering after shape complementarity we get hits for 4 new test cases, and the rank of the top hit improves for 15 more (see Figure S3(top)). However, we also lose hits in top 1000 for 3 test cases, and the rank of the top hit degrades for one test case. Overall, the application of these filters and clustering seem largely beneficial. The best results are obtained for enzyme-inhibitor/enzyme-substrate complexes, as for more than 50% of these complexes rank of the top hit improve.

The impact on the RMSD of the top hit is not clear though it has improved for 5 test cases in the enzyme-inhibitor/enzyme-substrate group (see Figure S5(top)). Similar is the impact on the minimum RMSD in the top 1000 predictions (see Figure S9(top)).

The number of hits in the top 1000 predictions is reduced in majority of the cases (see Figure S7(top)) as the two filters penalize predictions that are close to the target complex but have steric clashes. The proximity clustering step also penalizes poses that are very similar to poses with higher scores.

When electrostatics is turned on we get hits in top 1000 for 9 test cases for which we did not have a single hit before, and for 14 other cases rank of the top hit improve (see Figure S3(middle)). However, we lose all hits for 1 test case, and for 4 others rank of the top hit degrades.

Though there does not seem to be any particular trend in the change of the RMSD of the top hit except for the 9 cases with new first hits as mentioned above (see Figure S5(middle)), for antibody-antigen and antigen-bound antibody complexes the minimum RMSD in the top 1000 seems to have generally improved (see Figure S9(middle)). For 22 test cases the number of hits in the top 1000 predictions has improved while for 5 cases this number has dropped (see Figure S7(middle)).

The FFT-based interface propensity scoring is activated next which improves the rank of the top hit for 30 test cases (i.e., for around 50% of all cases) among which 7 cases did not have a single hit before (see Figure S3(bottom)). Among these 7 cases with new first hits 5 are antibody-antigen or antigen-bound antibody complexes, and none are enzyme-inhibitor or enzyme-substrate. Ranks have degraded for 5 test cases. Minimum RMSD in top 1000 has improved for 20 test cases degrading for only 1 (see Figure S9(bottom)), and RMSD of top hit has also degraded for only 1 test case (see Figure S5(bottom)). The number of hits has increased for 28 test cases and decreased for 7 (see Figure S7(bottom)).

The interface propensity filter is turned on next. It improves the rank of the top hit for 25 complexes, and degrades for 5 (see Figure S4(top)). For 3 test cases we did not have a single hit in top 1000 before among which 2 are antibody-antigens. The number of hits has increased for 16 cases, and decreased for 5 (see Figure S8(top)). The impacts on the RMSD of the top hit and the minimum RMSD in top 1000 are not significant in general (see Figures S6(top) and S10(top)).

The residue-residue contact filter which is activated next improves the rank of the top hit for 27 test cases, and degrades for none (see Figure S4(middle)). The enzyme-inhibitor and enzyme-substrate complexes seem to have benefited the least from this filter. The impact on the number of hits in top 1000 is not significant and is mixed (see Figure S8(middle)). The RMSD of the top hit

15
and the minimum RMSD in top 1000 do not change except for the 3 cases for which we get new first hits (see Figures S6(middle) and S10(middle)).

Next we apply the antibody contact filter and the Glycine filter. The antibody contact filter improves the rank of the top hit for 9 antibody-antigen and antigen-bound antibody test cases, and degrades for 3, while the Glycine filter slightly improves the same for 4 enzyme-inhibitor/enzyme-substrate complexes (see Figures S4(bottom)). The antibody contact filter improves the number of hits in the top 1000 predictions for 5 complexes (see Figure S8(bottom)).

All other changes are insignificant.
Figure S3: Changes in the rank of top hit as various options in F²Dock 2.0 are activated one after another (on the rigid-body test cases from Zlab benchmark 2.0 [2]). 

**Top:** Lennard-Jones Filter (LJ), clash filter (CL) and proximity clustering (PC) are activated after shape complementarity (SC), 

**Middle:** electrostatics & charge complementarity (EL) after SC+LJ+CL+PC, 

**Bottom:** interface propensity (IP) after SC+LJ+CL+PC+EL.
Figure S4: Changes in the rank of top hit as various options in F²Dock 2.0 are activated one after another (on the rigid-body test cases from Zlab benchmark 2.0 [2]). Continued from S3. **Top:** interface propensity filter (PF) after SC+LJ+CL+PC+EL+IP, **Middle:** residue-residue contact filter (RC) after SC+LJ+CL+PC+EL+IP+PF, and **Bottom:** antibody contact filter (AF) or glycine filter (GF) after SC+LJ+CL+PC+EL+IP+PF+RC.
Figure S5: Changes in the RMSD of the top hit as various options in F^2Dock 2.0 are activated one after another (on the rigid-body test cases from Zlab benchmark 2.0 [2]). **Top:** Lennard-Jones Filter (LJ), clash filter (CL) and proximity clustering (PC) are activated after shape complementarity (SC), **Middle:** electrostatics & charge complementarity (EL) after SC+LJ+CL+PC, **Bottom:** interface propensity (IP) after SC+LJ+CL+PC+EL.
Figure S6: Changes in the RMSD of the top hit as various options in F$^2$Dock 2.0 are activated one after another (on the rigid-body test cases from Zlab benchmark 2.0 [2]). Continued from S5. **Top:** interface propensity filter (PF) after SC+LJ+CL+PC+EL+IP, **Middle:** residue-residue contact filter (RC) after SC+LJ+CL+PC+EL+IP+PF, and **Bottom:** antibody contact filter (AF) or glycine filter (GF) after SC+LJ+CL+PC+EL+IP+PF+RC.
Figure S7: Changes in the number of hits in the top 1000 docking poses generated by F^2Dock 2.0 as various options in F^2Dock 2.0 are activated one after another (on the rigid-body test cases from Zlab benchmark 2.0 [2]). **Top:** Lennard-Jones Filter (LJ), clash filter (CL) and proximity clustering (PC) are activated after shape complementarity (SC). **Middle:** electrostatics & charge complementarity (EL) after SC+LJ+CL+PC. **Bottom:** interface propensity (IP) after SC+LJ+CL+PC+EL.
Figure S8: Changes in the number of hits in the top 1000 docking poses generated by F$^2$Dock 2.0 as various options in F$^2$Dock 2.0 are activated one after another (on the rigid-body test cases from Zlab benchmark 2.0 [2]). Continued from S7. **Top:** interface propensity filter (PF) after SC+LJ+CL+PC+EL+IP. **Middle:** residue-residue contact filter (RC) after SC+LJ+CL+PC+EL+IP+PF. **Bottom:** antibody contact filter (AF) or glycine filter (GF) after SC+LJ+CL+PC+EL+IP+PF+RC.
Figure S9: Changes in the minimum RMSD in the top 1000 docking poses generated by F^2Dock 2.0 as various options in F^2Dock 2.0 are activated one after another (on the rigid-body test cases from Zlab benchmark 2.0 [2]). **Top:** Lennard-Jones Filter (LJ), clash filter (CL) and proximity clustering (PC) are activated after shape complementarity (SC), **Middle:** electrostatics & charge complementarity (EL) after SC+LJ+CL+PC, **Bottom:** interface propensity (IP) after SC+LJ+CL+PC+EL.
Figure S10: Changes in the minimum RMSD in the top 1000 docking poses generated by F^2Dock 2.0 as various options in F^2Dock 2.0 are activated one after another (on the rigid-body test cases from Zlab benchmark 2.0 [2]). Continued from S9. **Top:** interface propensity filter (PF) after SC+LJ+CL+PC+EL+IP, **Middle:** residue-residue contact filter (RC) after SC+LJ+CL+PC+EL+IP+PF, and **Bottom:** antibody contact filter (AF) or glycine filter (GF) after SC+LJ+CL+PC+EL+IP+PF+RC.
5 Additional Results Comparing ZDock 3.0.2 and F²Dock 2.0

| Antibody | Class | Difficulty | Complex | ZDock (Top) | ZDock (Best) | F²Dock (Top) | F²Dock (Best) |
|----------|-------|------------|---------|-------------|-------------|--------------|--------------|
|          |       |            |         | Rank | RMSD | E_nat | VarOut | Rank | RMSD | E_nat | VarOut | Rank | RMSD | E_nat | VarOut |
|          |       |            |         | 1AHW | 354 | 4.5 | 0.30 | Acc | 8 | 4.4 | 0.47 | Acc | 1242 | 0.9 | 0.77 | High |
|          |       |            |         | 1BJ1 | 1 | 1.9 | 0.69 | Med | 63 | 2.3 | 0.53 | Acc | 1 | 1.8 | 0.69 | Med |
|          |       |            |         | 1BVK | 184 | 3.6 | 0.54 | Med | 205 | 4.9 | 0.23 | Acc | 358 | 1.9 | 0.67 | Med |
|          |       |            |         | 1DQJ | 274 | 4.0 | 0.56 | Acc | 74 | 4.9 | 0.21 | Acc | 1787 | 3.3 | 0.44 | Med |
|          |       |            |         | 1E6J | 3 | 4.1 | 0.20 | Acc | 126 | 5 | 0.19 | Acc | 181 | 2.7 | 0.58 | Med |
|          |       |            |         | 1FSK | 1 | 2.9 | 0.52 | Med | 1 | 3.2 | 0.46 | Acc | 2 | 1.8 | 0.58 | Med |
|          |       |            |         | 19R  | 9 | 3.9 | 0.42 | Acc | 9 | 3.9 | 0.42 | Acc | 9 | 3.9 | 0.42 | Acc |
|          |       |            |         | 1QD  | 18 | 4.3 | 0.48 | Acc | 4 | 3 | 0.38 | Med | 68 | 1.7 | 0.68 | Med |
|          |       |            |         | 1JPS | 1266 | 2.1 | 0.65 | Med | 186 | 2.7 | 0.61 | Acc | 1266 | 2.1 | 0.65 | Med |
|          |       |            |         | 1KC  | 583 | 2.9 | 0.56 | Med | 105 | 4.4 | 0.47 | Acc | 583 | 2.9 | 0.56 | Med |
|          |       |            |         | 1KQ  | 2 | 1.2 | 0.78 | Med | 1 | 1.6 | 0.75 | Med | 2 | 1.2 | 0.78 | Med |
|          |       |            |         | 1LC  | 57 | 2.0 | 0.41 | Acc | 11 | 3.8 | 0.32 | Acc | 57 | 2.0 | 0.41 | Acc |
|          |       |            |         | 1NCA | 11 | 1.7 | 0.58 | Med | 168 | 3.7 | 0.55 | Acc | 11 | 1.7 | 0.56 | Med |
|          |       |            |         | 1SN  | 1267 | 1.6 | 0.67 | Med | 1267 | 1.6 | 0.67 | Med | 1267 | 1.6 | 0.67 | Med |
|          |       |            |         | 1QFW | 80 | 1.9 | 0.57 | Med | 80 | 1.9 | 0.57 | Med | 80 | 1.9 | 0.57 | Med |
|          |       |            |         | 1VFB | 250 | 3.1 | 0.43 | Acc | 191 | 4.8 | 0.34 | Acc | 560 | 2.9 | 0.46 | Med |
|          |       |            |         | 1WEJ | 9 | 1.5 | 0.75 | Med | 5 | 3.2 | 0.65 | Med | 9 | 1.5 | 0.75 | Med |
|          |       |            |         | 2FD6 | 3 | 5.0 | 0.18 | Inc | 62 | 4.4 | 0.26 | Acc | 282 | 3.3 | 0.25 | Acc |
|          |       |            |         | 2G5  | 2 | 3.0 | 0.55 | Med | 122 | 3.9 | 0.61 | Med | 40 | 1.7 | 0.54 | Med |
|          |       |            |         | 2JEL | 4 | 3.5 | 0.50 | Med | 1 | 3.3 | 0.45 | Med | 753 | 2.6 | 0.56 | Med |
|          |       |            |         | 2V1S | 9QFW | 2 | 4.0 | 0.67 | Inc | 1 | 3.9 | 0.31 | Acc | 48 | 1.9 | 0.71 | Med |

Figure S11: Comparison of the performance of F²Dock 2.0 and ZDock 3.0.2 for each of the 25 antibody-antigen and antigen-bound antibody complexes from ZLab’s benchmark 4.0 in terms of the rank, RMSD, E_nat and CAPRI classification of the top hit and the best hit.
Figure S12: Comparison of the performance of F$^2$Dock 2.0 and ZDock 3.0.2 for each of the 52 enzyme-inhibitor and enzyme-substrate complexes from ZLab’s benchmark 4.0 in terms of the rank, RMSD, $F_{nat}$ and CAPRI classification of the top hit and the best hit.

| Enzyme | Easy | Top ranked near-native solution | Best ranked near-native solution |
|--------|------|---------------------------------|---------------------------------|
|        |      | Rank | RMSD | AzP | Acc | Rank | RMSD | AzP | Acc |
| 1AVX   | 1    | 25   | 3.5  | 0.60 | Med | 1    | 4.5  | 0.40 | Acc |
| 1AY7   | 2    | 577  | 2.5  | 0.67 | Med | 2    | 4.0  | 0.36 | Acc |
| 1BVM   | 3    | 1    | 1.2  | 0.78 | Med | 3    | 1    | 1.0  | 0.55 | Med |
| 1CGI   | 4    | 10   | 4.0  | 0.54 | Acc | 4    | 7.6  | 3.4  | 0.53 | Med |
| 1CIV   | 5    | 2    | 3.2  | 0.49 | Med | 5    | 2    | 3.2  | 0.49 | Med |
| 1CIV   | 6    | 68   | 1.7  | 0.84 | Med | 6    | 1.7  | 0.84 | Med |
| 1CIV   | 7    | 68   | 1.7  | 0.84 | Med | 7    | 1.7  | 0.84 | Med |
| 1CIV   | 8    | 53   | 1.2  | 0.48 | Med | 8    | 1.2  | 0.48 | Med |
| 1CIV   | 9    | 841  | 4.9  | 0.32 | Med | 9    | 841  | 4.9  | 0.32 | Med |
| 1F34   | 10   | 62   | 3.4  | 0.60 | Med | 10   | 3.4  | 0.60 | Med |
| 1F34   | 11   | 31   | 1.0  | 0.26 | Acc | 11   | 1.0  | 0.26 | Acc |
| 1F4L   | 12   | 45   | 1.1  | 1.1  | Med | 12   | 1.1  | 1.1  | Med |
| 1AHA   | 13   | 26   | 2.6  | 0.74 | Med | 13   | 2.6  | 0.74 | Med |
| 1JTG   | 14   | 1    | 2.6  | 0.60 | Med | 14   | 2.6  | 0.60 | Med |
| 1JTG   | 15   | 1    | 3.1  | 0.57 | Med | 15   | 3.1  | 0.57 | Med |
| 1NRO   | 16   | 7    | 3.4  | 0.48 | Acc | 16   | 3.4  | 0.48 | Acc |
| 1OC0D  | 17   | 1950 | 48.0 | 0.29 | Acc | 17   | 48.0 | 0.29 | Acc |
| 1OHH   | 18   | 1694 | 4.9  | 0.48 | Acc | 18   | 4.9  | 0.48 | Acc |
| 1UOV   | 19   | 15   | 4.9  | 0.39 | Acc | 19   | 4.9  | 0.39 | Acc |
| 1UPE   | 20   | 1    | 2.9  | 0.45 | Acc | 20   | 2.9  | 0.45 | Acc |
| 1URX   | 21   | 138  | 2.2  | 0.67 | Acc | 21   | 2.2  | 0.67 | Acc |
| 1UQX   | 22   | 18   | 3.6  | 0.52 | Med | 22   | 3.6  | 0.52 | Med |
| 1UQX   | 23   | 25   | 3.5  | 0.45 | Med | 23   | 3.5  | 0.45 | Med |
| 1VVB   | 24   | 1    | 2.4  | 0.63 | Med | 24   | 2.4  | 0.63 | Med |
| 1VVB   | 25   | 1    | 4.2  | 0.51 | Acc | 25   | 4.2  | 0.51 | Acc |
| 1VVB   | 26   | 1    | 4.2  | 0.51 | Acc | 26   | 4.2  | 0.51 | Acc |
| 2BMZ   | 27   | 76   | 4.4  | 0.30 | Acc | 27   | 4.4  | 0.30 | Acc |
| 2BMZ   | 28   | 29   | 5.0  | 0.27 | Acc | 28   | 5.0  | 0.27 | Acc |
| 2BMZ   | 29   | 76   | 4.4  | 0.30 | Acc | 29   | 4.4  | 0.30 | Acc |
| 2BMZ   | 30   | 18   | 2.9  | 0.47 | Med | 30   | 2.9  | 0.47 | Med |
| 2BMZ   | 31   | 76   | 4.4  | 0.30 | Acc | 31   | 4.4  | 0.30 | Acc |
| 2BMZ   | 32   | 20   | 2.6  | 0.54 | Med | 32   | 2.6  | 0.54 | Med |
| 2BMZ   | 33   | 20   | 2.6  | 0.54 | Med | 33   | 2.6  | 0.54 | Med |
| 2BMZ   | 34   | 5   | 1.1  | 0.74 | Med | 34   | 1.1  | 0.74 | Med |
| 2BMZ   | 35   | 177  | 3.8  | 0.46 | Med | 35   | 3.8  | 0.46 | Med |
| 2BMZ   | 36   | 793  | 2.4  | 0.39 | Med | 36   | 2.4  | 0.39 | Med |
| 2BMZ   | 37   | 428  | 4.0  | 0.37 | Med | 37   | 4.0  | 0.37 | Med |
| 2BMZ   | 38   | 1    | 4.4  | 0.32 | Acc | 38   | 4.4  | 0.32 | Acc |
| 2BMZ   | 39   | 1    | 4.4  | 0.32 | Acc | 39   | 4.4  | 0.32 | Acc |
| 2BMZ   | 40   | 1    | 4.4  | 0.32 | Acc | 40   | 4.4  | 0.32 | Acc |
| 2BMZ   | 41   | 1    | 4.4  | 0.32 | Acc | 41   | 4.4  | 0.32 | Acc |
| 2BMZ   | 42   | 1    | 4.4  | 0.32 | Acc | 42   | 4.4  | 0.32 | Acc |
| 2BMZ   | 43   | 1    | 4.4  | 0.32 | Acc | 43   | 4.4  | 0.32 | Acc |
| 2BMZ   | 44   | 1    | 4.4  | 0.32 | Acc | 44   | 4.4  | 0.32 | Acc |
| 2BMZ   | 45   | 220  | 3.0  | 0.51 | Med | 45   | 3.0  | 0.51 | Med |
| 2BMZ   | 46   | 126  | 4.4  | 0.23 | Acc | 46   | 4.4  | 0.23 | Acc |
| 2BMZ   | 47   | 81   | 3.0  | 0.61 | Med | 47   | 3.0  | 0.61 | Med |
| 2BMZ   | 48   | 321  | 4.9  | 0.42 | Med | 48   | 4.9  | 0.42 | Med |

References

[1] Bajaj C, Chowdhury RA, Siddavanahalli V (2011) F2Dock: Fast Fourier Protein-Protein Docking. IEEE/ACM Transactions on Computational Biology and Bioinformatics 8: 45–58.

[2] Mintseris J, Wiehe K, Pierce B, Anderson R, Chen R, et al. (2003) Protein-protein docking
Figure S13: Comparison of the performance of F$^2$Dock 2.0 and ZDock 3.0.2 for each of the 99 other type of complexes from ZLab’s benchmark 4.0 in terms of the rank, RMSD, $F_{\text{nat}}$ and CAPRI classification of the top hit and the best hit. Continued as Figure S14.

**benchmark 2.0: an update.** Proteins 60: 214–6.

[3] Gabb HA, Jackson RM, Sternberg MJE (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. Journal of Molecular Biology 272: 106–120.

---

### Table: Comparison of F$^2$Dock 2.0 and ZDock 3.0.2 performance

| Complex | F$^2$Dock | ZDock | Others |
|---------|-----------|-------|--------|
| | $F_{\text{nat}}$ | $C_{\text{nat}}$ | $C_{\text{nat}}$ |
| | | | |
| 1A29 | 1348 | 4.3 | 0.54 | 44 | 2.4 | 0.48 | Med |
| 1AK4 | 1090 | 3.5 | 0.52 | 964 | 4.3 | 0.43 | Inc |
| 1AKJ | 548 | 2.9 | 0.48 | 39 | 3.4 | 0.43 | Med |
| 1AZS | 42 | 2.9 | 0.59 | Acc |
| 1BEC | 1 | 2.9 | 0.51 | Med |
| 1BHU | 30 | 3.6 | 0.41 | Med |
| 1EF6 | 1171 | 3.8 | 0.57 | Acc |
| 1EFN | 589 | 4.6 | 0.39 | Med |
| 1F1C | 1190 | 5 | 0.31 | Inc |
| 1FC | 73 | 4.5 | 0.29 | Med |
| 1FCQ | 1105 | 4.1 | 0.65 | Med |
| 1GCO | 1708 | 3.9 | 0.58 | Med |
| 1GQA | 3 | 3.6 | 0.56 | Acc |
| 1GQ | 1006 | 4.5 | 0.37 | Med |
| 1HF | 175 | 4.0 | 0.48 | Acc |
| 1HFL | 1141 | 4.7 | 0.53 | Med |
| 1HOF | 571 | 4.2 | 0.43 | Acc |
| 1L2 | 182 | 4.6 | 0.45 | Med |
| 1L7 | 7 | 3.6 | 0.51 | Acc |
| 1L7A | 3 | 2.7 | 0.79 | Med |
| 1KAC | 592 | 4.5 | 0.39 | Acc |
| 1KLU | 1957 | 3.4 | 0.51 | Acc |
| 1KZT | 535 | 2.8 | 0.73 | Med |
| 1KP | 1 | 1.6 | 0.71 | Med |
| 1MLE | 4 | 3.1 | 0.58 | Med |
| 1OFU | 84 | 4.5 | 0.30 | Inc |
| 1PH | 748 | 4.5 | 0.59 | Acc |
| 1QA9 | 3 | 4.6 | 0.55 | Acc |
| 1Q1 | 2 | 1.3 | 0.78 | Med |
| 1QIQ | 756 | 1.9 | 0.58 | Med |
| 1SS | 58 | 3.6 | 0.52 | Acc |
| 1ST7 | 74 | 1.1 | 0.76 | Med |
| 1TDD | 74 | 1.1 | 0.76 | Med |
| 1UXI | 912 | 5.0 | 0.33 | Acc |
| 1ZOK | 8 | 3.3 | 0.49 | Med |
| 1ZSY | 20 | 3.4 | 0.45 | Acc |
| 1Z4H | 65 | 4.4 | 0.40 | Med |
| 2KST | 268 | 3.6 | 0.50 | Med |
| 2K6 | 558 | 3.4 | 0.46 | Acc |
| 2AJP | 475 | 3.6 | 0.47 | Med |
| 2AYO | 37 | 3.3 | 0.47 | Med |
| 2BAA | 53 | 4.7 | 0.42 | Med |
| 2FJU | 261 | 3.2 | 0.68 | Acc |

---
Figure S14: Comparison of the performance of F$^2$Dock 2.0 and ZDock 3.0.2 for each of the 99 other type of complexes complexes from ZLab’s benchmark 4.0 in terms of the rank, RMSD, $F_{nat}$ and CAPRI classification of the top hit and the best hit.

[4] Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: An automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Research 32: w665–w667.

[5] Jones S, Thornton JM (1997) Analysis of protein-protein interaction sites using surface patches. Journal of Molecular Biology 272: 121–132.
[6] Jones S, Thornton JM (1996) Principles of protein-protein interactions. Proceedings of the National Academy of Sciences of the United States of America 93: 13–20.

[7] Black S, Mould D (1991) Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. Analytical Biochemistry 193: 72–82.

[8] Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, et al. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. Journal of the American Chemical Society 106: 765–784.

[9] Morris GM, Goodsell DS, Huey R, Olson AJ (1996) Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. Journal of Computer-Aided Molecular Design 10: 293–304.

[10] Chowdhury RA, Bajaj C (2010) Algorithms for faster molecular energetics, forces and interfaces. ICES report 10–32, Institute for Computational Engineering & Science, The University of Texas at Austin.

[11] Bajaj C, Zhao W (2010) Fast molecular solvation energetics and forces computation. SIAM Journal on Scientific Computing 31: 4524–4552.

[12] Jackins CL, Tanimoto SL (1980) Oct-trees and their use in representing three-dimensional objects. Computer Graphics and Image Processing 14: 249–270.

[13] Bajaj C, Chowdhury RA, Rasheed M (2010) A dynamic data structure for flexible molecular maintenance and informatics. Bioinformatics 27: 55–62.

[14] Glaser F, Steinberg DM, Vakser IA, Ben-Tal N (2001) Residue frequencies and pairing preferences at protein-protein interfaces. Proteins: Structure, Function, and Genetics 43: 89–102.

[15] Martin A (2010) Protein sequence and structure analysis of antibody variable domains. Antibody engineering : 33–51.

[16] Eisenberg D, McLachlan A (1986) Solvation energy in protein folding and binding. Nature 319: 199–203.

[17] Gilson M, Davis M, Luty B, McCammon J (1993) Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation. Journal of Physical Chemistry 97: 3591–3600.

[18] Hermann RB (1972) Theory of hydrophobic bonding. ii. correlation of hydrocarbon solubility in water with solvent cavity surface area. Journal of Physical Chemistry 76: 2754–2759.

[19] Simonson T, Bruenger AT (1994) Solvation free energies estimated from macroscopic continuum theory: An accuracy assessment. Journal of Physical Chemistry 98: 4683–4694.

[20] Wagener J, Baker N (2006) Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms. Proceedings of the National Academy of Sciences of the USA 103: 8331–8336.
[21] Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. Journal of the American Chemical Society 112: 6127–6129.

[22] Grycuk T (2003) Deficiency of the coulomb-field approximation in the generalized born model: An improved formula for born radii evaluation. Journal of Chemical Physics 119: 4817–4826.

[23] Tjong H, Zhou HX (2007) Gbr6: A parameterization-free, accurate, analytical generalized born method. Journal of Physical Chemistry B 111: 3055–3061.