Electrostatics in protein–protein docking

ALEXANDER HEIFETZ,[1] EPHRAIM KATCHALSKI-KATZIR,[1]
AND MIRIAM EISENSTEIN[2]

[1]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel
[2]Department of Chemical Services, The Weizmann Institute of Science, Rehovot 76100, Israel

(RECEIVED June 27, 2001; FINAL REVISION November 9, 2001; ACCEPTED November 21, 2001)

Abstract

A novel geometric-electrostatic docking algorithm is presented, which tests and quantifies the electrostatic complementarity of the molecular surfaces together with the shape complementarity. We represent each molecule to be docked as a grid of complex numbers, storing information regarding the shape of the molecule in the real part and information regarding the electrostatic character of the molecule in the imaginary part. The electrostatic descriptors are derived from the electrostatic potential of the molecule. Thus, the electrostatic character of the molecule is represented as patches of positive, neutral, or negative values. The potential for each molecule is calculated only once and stored as potential spheres adequate for exhaustive rotation/translation scans. The geometric-electrostatic docking algorithm is applied to 17 systems, starting form the structures of the unbound molecules. The results—in terms of the complementarity scores of the nearly correct solutions, their ranking in the lists of sorted solutions, and their statistical uniqueness—are compared with those of geometric docking, showing that the inclusion of electrostatic complementarity in docking is very important, in particular in docking of unbound structures. Based on our results, we formulate several "good electrostatic docking rules": The geometric-electrostatic docking procedure is more successful than geometric docking when the potential patches are large and when the potential extends away from the molecular surface and protrudes into the solvent. In contrast, geometric docking is recommended when the electrostatic potential around the molecules to be docked appears homogenous, that is, with a similar sign all around the molecule.

Keywords: Molecular docking; molecular recognition; electrostatic complementarity; surface matching; electrostatic patches; grid representation by complex numbers

Living organisms rely on the specific recognition of pairs of molecules in practically every biological process. Hence, the importance of understanding molecular recognition and determining the structures of molecular complexes cannot be overestimated. The immense amount of sequence and structure data owing to the genome and structural genome projects can be exploited to predict the structures of many new proteins and to investigate their relations with other molecules. Thus, reliable theoretical tools for predicting the structures of molecular complexes (docking procedures) are needed. Such tools must be able to deal with molecules for which the activity is not fully understood and with modeled structures, for which the accuracy may be limited. Therefore, incorporation of all the available knowledge regarding intermolecular interfaces is important. An adequate representation and quantification of this information must be formulated to reduce the sensitivity of the prediction procedure to structural errors.

Analyses of experimentally determined structures indicate that intermolecular recognition is facilitated by a myriad of weak, noncovalent interactions that together promote specificity at different levels. The prediction of the structures of complexes is therefore a difficult, multidimensional problem that attempts to solve simultaneously the relative position of the molecules in the complex and their conformation. The docking problem can, however, be

Reprint requests to: Miriam Eisenstein, Weizmann Institute Of Science, Chemical Services Unit, Rehovot 76100, Israel; e-mail: miriam.eisenstein@weizmann.ac.il; fax: 972-8-9344136.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.26002.

solved in a series of steps: the first of which is the determination of the relative position of the molecules in the complex. This is a six-dimensional problem in which the docked molecules are treated as rigid bodies. Comparisons of the structures of bound and unbound molecules indicate that in many cases, these structures are very similar (Conte et al. 1999), supporting the rigid body approximation. Once the relative positions of the molecules in a complex are determined, attempts can be made to modify and refine their structures in a manner that improves the interaction (Robert and Janin 1998; Oliva and Moult 1999).

Interfaces are characterized by complementarity of the shape and the chemical character of the interacting surfaces (Jones and Thornton 1996; Tsai et al. 1996; Conte et al. 1999). Recent analyses of molecular interfaces indicate that in nonpermanent complexes composed of molecules that can exist individually in solution, electrostatic contacts are abundant (Xu et al. 1997; Conte et al. 1999). In permanent complexes, that is, oligomers, hydrophobic contacts are dominant (Jones and Thornton 1996; Tsai et al. 1996). Nevertheless, in most oligomers the hydrophobic patches at the interface are mixed with hydrophilic ones (Larsen et al. 1998). It appears that despite the small and often unfavorable contribution of electrostatics to the stabilization of molecular complexes (Sheinerman et al. 2000), there is electrostatic complementarity at the interface. Evidently, electrostatic contacts should be included in docking algorithms, in particular when applied to nonpermanent complexes. Yet, attempts in this direction, in which the electrostatic energy is calculated, show only marginal improvement in the docking results (Gabb et al. 1997; M. Eisenstein, unpubl.). This is most likely caused by the acute sensitivity of the electrostatic interaction energy to the details of the structure of the complex. In docking, this sensitivity is problematic in two ways: First, the conformation of the unbound protein differs from that of the same protein in the complex. Second, in many docking algorithms, the relative positions of the two molecules are determined only approximately, for example, by a stepwise sampling of the translation/rotation space, in which the exact relative position is usually not included. The structural errors caused by the conformation change and the mispositioning reduce the usefulness of electrostatic energy computations in docking (Robert and Janin 1998). Mandell et al. (2001) use a continuum model to calculate the potential around one of the molecules and place charges on the other one. They report improved ranking of the nearly correct solution in geometric-electrostatic docking for two of the three unbound systems that they tested.

Shape complementarity is the core of most docking algorithms, and some of them rely only on geometric recognition (for example, Jiang and Kim 1991; Katchalski-Katzir et al. 1992; Walls and Sternberg 1992; Norel et al. 1994). Other algorithms also include electrostatic interactions (Ausiello et al. 1997; Gabb et al. 1997; Mandell et al. 2001), hydrogen bonding (Meyer et al. 1996; Ausiello et al. 1997), and hydrophobic contacts (Vakser and Aflalo 1994; Ackermann et al. 1998). Previously, our group has presented a docking algorithm based only on shape complementarity, which successfully reassembled binary complexes (Katchalski-Katzir et al. 1992) and helical aggregates (Eisenstein et al. 1997) and predicted the structures of new complexes (Strynadka et al. 1996a; Dixon 1997; Eisenstein and Katchalski-Katzir 1998).

Here we present a geometric-electrostatic docking algorithm in which the effects of electrostatics are combined with geometric surface complementarity. Instead of calculating electrostatic interaction energies for the different putative complexes formed by a given pair of molecules, we correlated their tendencies to form good electrostatic contacts. Recently, it has been shown that the electrostatic potentials at the interfaces of interacting molecules are anticorrelated (McCoy et al. 1997). This means that at the interface, there is a good chance to find a patch of positive electrostatic potential on the surface of one molecule positioned next to a negative patch on the surface of the adjacent molecule and vise versa. We represent the electrostatic potential of each molecule as positive, neutral, or negative patches. These patches are combined with the geometric representation of the molecule in a three-dimensional (3D) matrix of complex numbers. The correlation of such matrices provides a measure of the geometric and electrostatic complementarity of the molecular surfaces. We also present an algorithm for rotating of the electrostatic potential by translating it into potential spheres. These spheres are treated as atoms; they are rotated to new orientations and then translated back into potential patches.

The new method is applied to a selection of known binary complexes, starting from the structures of the unbound molecules. The inclusion of electrostatic complementarity significantly improves the docking results for most systems by ranking the nearly correct solutions as highly probable. This is important because once a nearly correct solution is ranked as highly probable, additional screening methods and refinement algorithms, which can only be applied to a limited set of solutions, can be used to indicate the correct structure (Robert and Janin 1998). Our statistical analyses of the docking results for each system and for all the selection of systems further emphasize the contribution of electrostatic complementarity to the successful prediction of the structures of complexes. Finally, we sum up our results and analyses in several "good electrostatic docking rules," which depict systems in which geometric-electrostatic docking is likely to be more successful than geometric docking and vise versa.

Algorithms

In designing the combined geometric-electrostatic algorithm, we had several requirements: (1) The first is a full

rotation/translation scan, which does not assume prior knowledge of the interaction site. Such knowledge, if available, can be introduced at a later stage. (2) For every point in the rotation/translation space, a combined geometric plus electrostatic score is determined. Thus, electrostatics is an integral part of the scan and not a pre- or postscan filter. (3) Another requirement is reduced sensitivity to conformation changes and to errors caused by the stepwise sampling of the rotation/translation space. This requirement ensures that the algorithm is applicable to 'real' situations in which the structures of unbound or modeled molecules are docked. (4) The algorithm should include only a few adjustable parameters, which are independent of the system under investigation.

The electrostatic potential

The electrostatic potential around each of the docked molecules is calculated by solving the linearized Poisson-Boltzman equation, using the finite-differences method as implemented in the program Delphi (Klapper et al. 1986; Honig and Nicholls 1995). The calculations are performed on a fine grid, 0.5 Å, producing accurate estimates of the potential. For each system, we ascertain that the Delphi grid extent is large enough to encompass all the potential points, with absolute values exceeding a given minimum, $P_{\min}$ (see below).

The calculation of the potential is separate from our docking procedure (implemented in a computer program named MolFit), which reads the potential files for the two molecules, together with the necessary data regarding the grid interval and the origin of the potential grid. Thus, in principle, potentials calculated with other programs or with different forcefield parameters can be read in and used for describing the electrostatic character of the docked molecules.

Calculation of the geometric-electrostatic correlation function

In MolFit, the common 3D atomic representation of the molecules to be docked is replaced by a 3D grid representation. Each molecule is projected onto a 3D grid, such that grid points outside the molecule are given the value 0, points on the surface of the molecule are given the value 1, and those in the interior of the molecule are given either the negative value ρ or the positive value δ for molecules **a** and **b**, respectively. The grids are then correlated using discrete Fourier transformations (Katchalski-Katzir et al. 1992; Eisenstein et al. 1997). In these transformations, complex numbers are involved, which can be exploited for describing the electrostatic character of each molecule as follows:

Each grid point in the representation of molecule **a** is given the complex value

$$A_{l,m,n} = G^a_{l,m,n} + i\sqrt{w}\, E^a_{l,m,n}. \quad (1)$$

$G^a_{l,m,n}$ is the geometric descriptor of molecule **a**, and its value is either 0, 1, or ρ; $E^a_{l,m,n}$ is the electrostatic descriptor of molecule **a** (its values are discussed below), i is the square root of −1 and, w is an adjustable scale factor. Similarly, each grid point in the representation of molecule **b** is given by the complex value

$$B_{l,m,n} = G^b_{l,m,n} + i\sqrt{w}\, E^b_{l,m,n}. \quad (2)$$

In equation 2, $G^b_{l,m,n}$ is the geometric descriptor of molecule **b**, and its value is 0, 1, or δ, and $E^b_{l,m,n}$ is its electrostatic descriptor (see below). The correlation function $C_{\alpha,\beta,\gamma}$ can be calculated by a triple summation as follows:

$$\begin{aligned} C_{\alpha,\beta,\gamma} &= \sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} A_{l,m,n} \cdot B_{l+\alpha,m+\beta,n+\gamma} \\ &= \sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} [G^a_{l,m,n} \cdot G^b_{l+\alpha,m+\beta,n+\gamma} \\ &\quad - wE^a_{l,m,n} \cdot E^b_{l+\alpha,m+\beta,n+\gamma}] \\ &\quad + i\sqrt{w}\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} [G^a_{l,m,n} \cdot E^b_{l+\alpha,m+\beta,n+\gamma} \\ &\quad + E^a_{l,m,n} \cdot G^b_{l+\alpha,m+\beta,n+\gamma}]. \end{aligned} \quad (3)$$

The indices α,β,γ in equation (3) are the translations of molecule **b** with respect to molecule **a**, along three perpendicular axes.

The real part of $C_{\alpha,\beta,\gamma}$ is the complementarity score for the given relative orientation and translation of the two molecules. It consists of two terms. The first term,

$$\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} G^a_{l,m,n} \cdot G^b_{l+\alpha,m+\beta,n+\gamma}, \quad (4)$$

is the geometric correlation term. It is equivalent to the geometric score in our original algorithm and it reflects the extent of surface complementarity offset by the amount of inter-penetration for the translation vector α,β,γ. The values of ρ and δ are −15 and 1, respectively (Katchalski-Katzir et al., 1992). The second term,

$$\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} - wE^a_{l,m,n} \cdot E^b_{l+\alpha,m+\beta,n+\gamma}, \quad (5)$$

is the electrostatic correlation term. The negative sign reflects the preference of positive electrostatic patches to face

negative patches in the other molecule and vise versa. The values of E are discussed in the next section.

Instead of the lengthy summation in equation 3, the correlation function is calculated via fast Fourier transformations (Bringham 1988), and then the real part of the correlation matrix is extracted. Our procedure requires a single $N \times N \times N$ matrix to describe both the geometric and the electrostatic character of each molecule. Only one series of forward Fourier transformation, multiplication, and inverse Fourier transformation is performed for each orientation, as in the geometric docking (Katchalski-Katzir et al. 1992). A geometric-electrostatic rotation/translation scan, which uses a grid of $128 \times 128 \times 128$ points and a rotation interval of 12°, requires approximately 9 h on a SGI Octane with a single R10000 processor.

Grid representation of the potential via potential spheres

The values of $E^a_{l,m,n}$ and $E^b_{l,m,n}$ are derived from the potentials $P^a_{i,j,k}$ and $P^b_{i,j,k}$ for molecules **a** and **b**, respectively. The program Delphi uses a grid to calculate the potential, and the indices i, j, and k identify the potential grid points. Notably, these do not necessarily correspond to the indices l, m, and n, which identify the MolFit grid points. This is a result of the different requirements of Delphi and MolFit regarding the grid interval. As mentioned above, potentials are calculated using a fine grid (usually 0.5 Å) to attain better accuracy. Such a grid is, however, too fine for the geometric representation of the molecules in MolFit, in which an interval of 1.0 to 1.2 Å was found adequate (Katchalski-Katzir et al. 1992). The different grid intervals make it impossible to map the potential grid directly onto the MolFit grid. Therefore, we translate the potential grid into potential spheres with a radius

$$r_{pot} = \frac{\sqrt{3}}{2} h \cdot f. \quad (6)$$

In equation 6, h is the potential grid interval, and f is an adjustable parameter. When $f = 1.0$, the volume of the potential sphere equals the volume of a sphere that circumscribes a potential grid cube. The center of the potential sphere is at the position of the grid point from which it is derived, and it is assigned the potential value of that grid point (see Fig. 1). To reduce the number of potential spheres, we omit points with small absolute values of the potential. The minimum absolute value $P_{\min}$ is another adjustable parameter.

The potential spheres are projected onto the MolFit grid. Thus, every potential sphere contributes to the MolFit grid points within its volume. In cases in which several potential spheres contribute to a given grid point, the contributions

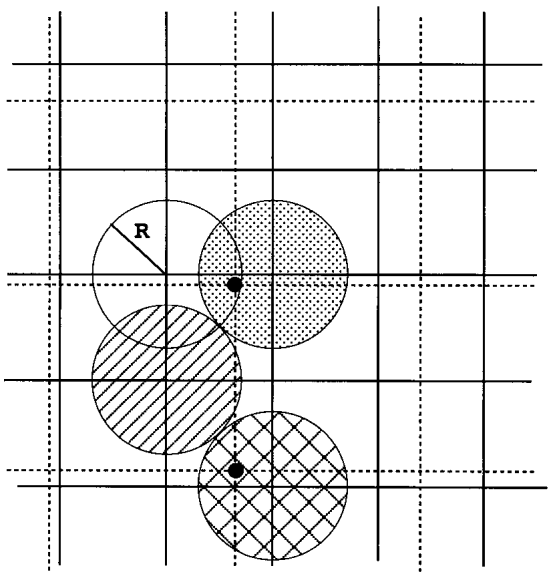

Fig. 1. A two-dimensional illustration defining the potential spheres and their projection on the MolFit grid. The potential grid is shown in solid lines; the MolFit grid, in dashed lines. Each potential sphere is centered on a potential grid point, and its radius R is chosen such that it circumscribes a grid cube. Some MolFit grid points (small solid circles) are either inside one potential sphere (hashed circle) and assigned the potential value of that sphere or are inside several potential spheres (dotted circle and hollow circle) and assigned the average of the potential values of these spheres. Some potential spheres do not contribute to any MolFit grid point (hatched circle).

are averaged. We tested two approaches for assigning values to $E^a_{l,m,n}$. In the first approach, the values of $E^a_{l,m,n}$ are equal to the average potentials provided, $|E_{l,m,n}| \geq P_{min}$. This is the continuous representation of the potential. In the second approach, only the sign of the average potential is kept for grid points with $|E_{l,m,n}| \geq P_{min}$. Thus, grid points are assigned potential values 1, −1, or 0. This is the single-value step representation of the potential. In both approaches, the potential values of grid points in the interior of molecule **a**, where $G^a_{l,m,n}$ is negative, are 0. Notably, the potential is not limited to the geometric surface of the molecule, and nonzero potential values may occur in grid points in which $G_{l,m,n}$ is 0.

All the potential values are scaled by the adjustable factor $\sqrt{w}$, and the combined score is a linear combination of the geometric score and the electrostatic score. The value of w is determined empirically by optimizing the combined geometric-electrostatic score for several known complexes (see below).

Rotation of molecule **b** *and its potential*

The computation of the correlation function must be repeated for many relative orientations of the two molecules. When molecule **b** is rotated with respect to molecule **a**, both

its geometric and electrostatic representations must be recalculated. However, the calculation of the electrostatic potential is time consuming, and in principle, there is no need to recalculate the potential because it is invariant to rotations. The sphere representation of the potential is very helpful here. Thus, we apply the rotation matrix to the potential spheres in the same manner as to the atomic coordinates. The rotated atomic coordinates and potential spheres are then projected onto the MolFit grid, producing the new (rotated) geometric and electrostatic representation of molecule **b**.

Rotation/translation scans and determination of the rank of nearly correct solutions

Geometric and geometric-electrostatic rotation/translation scans are performed using a grid interval between 1.0 and 1.2 Å and a rotation interval of 12°. Only one solution is saved for each orientation, resulting in 8760 putative binary complexes sorted by their complementarity scores (Eisenstein et al. 1997). All the MolFit solutions are compared with the experimental structure of the complex by calculating the root mean square differences (RMSDs) between the positions of the common C_α atoms. The rank of the nearly correct solution is its position in the sorted list of solutions, and it is determined by searching for the highest scoring solution with a RMSD <3 Å. The only exception to this rule is the system 1bth, in which the RMSD value between the contact residues of the enzyme in the bound and unbound structure is exceedingly large (see Table 1). In many cases, the score of the nearly correct solution is identical to that of other solutions. In such cases, the rank is given as a range of numbers representing the ranks of all these solutions.

Statistical analysis

The 8760 solutions produced by each scan are statistically analyzed to obtain estimates of the mean score for the given scan and the standard deviation. Thus, the number of solutions with a given score range is plotted as a function of the score. Then an extreme-value distribution function (Levitt and Gerstein 1998) is fitted to the distribution of scores, providing estimates for the mean score, μ, and the standard deviation, σ. These values are used to calculate a uniqueness value (Z_i) for each docking solution, i, the score of which is S_i, as follows:

$$Z_i = (S_i - \mu)/\sigma \quad (7)$$

Prediction of the ranks of the nearly correct solutions for different values of w (virtual scans)

The parameter w (equations 1, 2), which determines the relative contributions of the geometric and electrostatic terms to the combined complementarity scores, was optimized for several systems as described in Materials and Methods. To further test the adequacy of w for all the systems and in particular to examine the sensitivity of the geometric-electrostatic docking results to the value of w, we designed the following analysis: The combined geometric-electrostatic score for each solution, j, is a linear combination of the geometric and the electrostatic contributions, in which $a_j \times w$ is the slope. The value of a_j differs for each solution because it depends on the character of the interface in solution j. The a_j can be estimated from the results of at least two real scans with different w values and used to predict the score of each solution for another value of w. The prediction is valid only when the rotations and translations for the given solution in the two real scans are very close. This requirement is satisfied when the two w values are close to one another, and therefore, an additional rotation/translation scan is needed. The results of the geometric scan ($w = 0.$) and the two geometric-electrostatic scans are used to estimate the slopes, a_j, for all j solutions. Next, the a_j values are used to predict the geometric-electrostatic complementarity scores for other w values (virtual scans). The solutions in each virtual scan are sorted according to the predicted scores, and finally the predicted rank of the nearly correct solution is determined.

Results

Throughout this study, we used coordinates from the Protein Data Bank (PDB; Berman et al. 2000). Notably, all the water molecules in the experimental structures in both the bound and the unbound systems were omitted because we could not assume that the arrangement of water molecules around an unbound molecule resembles their arrangement at the interface of a complex.

The structures and their PDB codes are listed in Table 1. The table also lists the MolFit translation grid intervals used in the rotation/translation scans and the RMSDs between bound and unbound structures, calculated for all the common C_α atoms and for the residues at the interface. Four of the unbound structures are incomplete, and the coordinates of several exposed side-chains are not listed in the PDB file. This reflects high thermal motion or disorder and is characteristic of long and exposed side-chains. However, such side-chains are often charged side-chains (arginine, lysine, glutamate), and we suspect that their omission (by using the PDB coordinates as they are) may affect the docking results, particularly the geometric-electrostatic docking. Therefore, for four structures (1bni, 1avu, 4htc, and 1ace), we modeled the missing side-chains. This was performed automatically, using the MSI molecular graphics package (MSI Inc.). The added side-chains usually have an extended conformation (as long as it does not clash with the rest of the molecule),

Table 1. *List of the systems used to develop and verify the geometric-electrostatic docking algorithm*

System	PDB codes (complex; unbound)	RMSD (Å)[1] All Cα atoms/Interface residues		MolFit grid interval (Å)	References
		Mol. A	Mol. B		
β-trypsin/BPT1	2ptc;			1.176	Marquart et al. 1983
	2ptn/4pti	0.63/0.51	1.70/1.76	1.109	Walter et al. 1982; Marquart et al. 1983
barnase/barstar	1brs;			1.139	Buckle et al. 1994
	1bni[2]/1bta	0.77/1.81	1.44/1.03	1.197	Buckle et al. 1993; Lubienski et al. 1994
subtilisin/eglin-C	2sec;			1.146	McPhalen and James 1988
	1scd/1tec[3]	0.74/0.56	1.02/0.61	1.079	Fitzpatrick et al. 1994; Gros et al. 1994
α-chymotrypsin/	1cho;			1.124	Fujinaga et al. 1987
HPTI	5cha/1ovo	0.98/1.21	1.53/1.04	1.136	Papamokos et al. 1982; Blevins and Tulinsky 1985
Trypsin/	1avw;			1.160	Song and Suh 1998
soy-bean inhibitor	1ept/1avu[2]	0.81/0.91	1.79/1.41	1.115	Huang et al. 1994; Song and Suh 1998
thermitase/eglin-C	1tec;			1.110	Gros et al. 1994
	1thm/2sec[3]	0.80/0.70	1.02/0.73	1.068	McPhalen and James 1988; Teplyakov et al. 1990
trypsin/Bowman-	1smf;			1.115	Li et al. 1994
Birk inhibitor	2ptn/1pi2	0.75/0.68	1.14/1.30	1.125	Walter et al. 1982; Chen et al. 1992
thrombin/BPTI	1bth;			1.142	van de Locht et al. 1997
	4htc[3]/4pti	2.52/4.26	2.32/1.74	1.119	Marquart et al. 1983; Rydel et al. 1991
trypsin/leech derived	1ldt;			1.090	Stubbs et al. 1997
trypsin inhibitor	1ept/1ldt[4]	0.94/1.17	—	1.178	Huang et al. 1994; Stubbs et al. 1997
acetylcholinesterase/	1fss;			1.158	Harel et al. 1995
fasciculin-II	2ace/1fsc	0.91/1.56	1.35/1.19	1.182	Raves et al. 1997; LeDu et al. 1992
β-lactamase TEM1/				1.121	Strynadka et al. 1996b
BLIP		1.20/1.36	1.39/1.74	1.117	Strynadka et al. 1992; Strynadka et al. 1994
Ab Hyhel-10 (Fv)/	3hfm;			1.116	Padlan et al. 1989
lysozyme	—	—	—	—	
Ab Hyhel-10 (Fab)/	3hfm;			1.176	Padlan et al. 1989
lysozyme	3hfm[4]/1lza	—	1.24/1.52	1.144	Padlan et al. 1989; Maenaka et al. 1995
Ab Hyhel-10 (Fab)/	3hfm;			1.176	Padlan et al. 1989
lysozyme	3hfm[4]/1hel	—	1.14/1.32	1.152	Padlan et al. 1989; Wilson et al. 1992
Ab Hyhel-5 (Fab)/	3hfl;			1.157	Cohen et al. 1996
lysozyme	3hf1[4]/1lza	—	1.21/1.29	1.209	Maenaka et al. 1995; Cohen et al. 1996
Ab Hyhel-5 (Fab)/	3hf1;			1.157	Cohen et al. 1996
lysozyme	3hfl[4]/1hel	—	1.08/1.23	1.217	Wilson et al. 1992; Cohen et al. 1996
Ab D1.3 (Fv)/	1vfb;			1.135	Bhat et al. 1994
lysozyme	1vfa/1lza	0.95/0.93	1.52/1.75	1.160	Bhat et al. 1994; Maenaka et al. 1995
Ab D1.3 (Fv)/	1vfb;			1.135	Bhat et al. 1994
lysozyme	1vfa/1hel	0.95/0.82	1.42/1.43	1.169	Wilson et al. 1992; Bhat et al. 1994
Jel42 (Fab)/HPR	2jel;			1.099	Prasad et al. 1998
	2jel[4]/1poh	—	1.64/1.20	1.093	Jia et al. 1993; Prasad et al. 1998
Igg2A (Fab)/	1cft;				Keitel et al. 1997
Peptide	1cfq/1cft[4]	0.83/1.12	—		Keitel et al. 1997
Ab D1.3 (Fv)/	1dvf;			1.092	Braden et al. 1996
Ab E5.2 (Fv)	1vfa/1dvf[4]	0.98/0.99	—	1.150	Bhat et al. 1994; Braden et al. 1996
α-hemoglobin/	2hhb	—	—	1.126	Fermi et al. 1984
β-hemoglobin					

[1] This column lists the root mean square differences (RMSDs) between disassembled and unbound structures for molecules **a** and **b**.
[2] Several side-chains on the surface of the molecule are not complete. These side-chains were completed automatically (see text).
[3] The structure of the unbound molecule is not known. We chose a structure from another complex to represent a situation in which the conformation of the docked proteins is not that of the bound molecules.
[4] The structure of the unbound molecule is not known; therefore, the structure of the bound molecule was used.

which is not related to the conformation of this side-chain in the bound system.

Eleven disassembled enzyme/inhibitor systems were tested. For eight of these systems, we also docked the unbound molecules; in two cases, we docked the unbound enzyme to the matching inhibitor from a different complex, imitating an unbound situation; and in only one case, we docked the bound structure of the inhibitor (leech-derived tryptase inhibitor) to the unbound enzyme. We also tested six disassembled antibody/antigen systems. The unbound

structures of four of the antibodies were not known: Hyhel-10 (3hfm), Hyhel-5 (3hfl), Jel42 (2jel), and antibody E5.2 (1dvf). We therefore used the bound structure when necessary. There were two entries for the unbound structure of lysozyme in the PDB: 1lza and 1hel. Both were docked to the appropriate antibodies in the unbound docking tests.

Optimization of the adjustable parameters

Implementation of the geometric-electrostatic docking algorithm described above requires the optimization of several adjustable parameters: P_{min}, which is the minimum absolute value of the potential considered by the program (in kT/e); *f*, which determines the radius of the potential spheres in equation 6; and *w*, which determines the relative contributions of the geometric and electrostatic terms in equation 3. In addition, we tested the suitability of potentials calculated with either the Formal or the PARSE set of electrostatic charges, to our docking procedure. All the parameters were optimized for the two representations of the potential, the step representation and the continuous representation.

The details of the optimization procedures are described in Materials and Methods. We find that electrostatic potentials calculated with the PARSE set of charges are more adequate for our electrostatic representations than those obtained using Formal charges. The optimal value of the parameter *f* is 1.0 for both representations of the electrostatic potential. In contrast, the optimal values of P_{min} and *w* depend on the electrostatic representation. Thus, P_{min} is 3.0 kT/e for the step representation and 2.0 for the continuous representation, and *w* is 0.25 or 0.35 for the step representation and 0.0015 for the continuous representation.

Geometric-electrostatic docking with the step representation of the potential

In this series of computations positive, neutral and negative single-value patches represent the electrostatic potential. The results of geometric and geometric-electrostatic scans for disassembled systems are summarized in Table 2. The geometric-electrostatic scores for the nearly correct solutions are usually higher than the geometric scores, and the ranking of these solutions is also higher. The general increase in the complementarity scores observed on the introduction of electrostatic complementarity is in agreement with the results of McCoy et al. (1997), who showed that there is significant anticorrelation between the electrostatic potentials at the interface of interacting molecules.

We tested two approaches in the docking of lysozyme to

Table 2. *Comparison of the geometric and geometric-electrostatic docking results obtained for disassembled structures with the step representation of the electrostatic potential*

System	Buried surface area[1] (Å^2)	Geometric[3]			Geometric-electrostatic[3]		
		Rank	Score	Z, Z_1	Rank	Score	Z, Z_1
2ptc	1358	1	523	7.9, 7.9	1	600	9.4, 9.4
1brs	1533	1	626	11.5, 11.5	1	874	7.1, 7.1
2sec	1406	5	536	8.3, 9.9	5	546	7.0, 9.2
1cho	1362	3	522	6.6, 8.5	2	545	7.1, 8.2
1avw	1661	1	589	7.7, 7.7	1	589	8.0, 8.0
1tec	1434	9	522	6.7, 8.8	8	523	6.7, 8.5
1smf	931	1	387	8.3, 8.3	1	385	8.2, 8.2
1bth	2195	1	839	17.7, 17.7	1	955	16.8, 16.8
1ldt	1255	7	518	6.9, 8.9	2	608	7.5, 8.6
TEM1/BLIP	1912	1	795	10.8, 10.8	1	726	8.9, 8.9
1fss	1861	2	665	8.5, 8.7	1	1193	5.4, 5.4
3hfm (Fv)	1567	5	590	7.2, 7.6	4	620	5.6, 7.8
3hfm (Fab)[2]	1567	1	607	8.4, 8.4	3	642	5.3, 5.5
3hf1 (Fab)[2]	1563	1	642	8.8, 8.8	9	648	6.0, 8.1
1dvf	1448	23	605	5.4, 8.7	9–10	633	5.9, 8.4
1vfb	1321	34–35	519	4.4, 8.3	19	543	5.0, 7.7
2jel[2]	1453	2	503	8.0, 8.7	1	509	8.2, 8.2
1cft[2]	731	86–92	277	3.7, 6.1	32–33	292	4.7, 6.6
2hhb	1637	1	627	9.6, 9.6	1	631	9.4, 9.4

[1] The buried surface area was calculated by subtracting the accessible surface area of the complex from the sum of accessible surface areas for the individual molecules.
[2] The whole Fab fragment of the antibody was used in the docking scan; however, the surface of the constant domain was modified as described in the text.
[3] Z and Z_1 are the statistical uniqueness values (see Algorithms) for the nearly correct and the top-ranking solutions. Z values are measured in σ units.

Hyhel-10: Either the ligand was docked to the Fv fragment of the antibody or it was docked to the Fab fragment in which the part of the surface belonging to the Fc fragment was made negative (−15), preventing favorable contacts with the ligand. The results of the second docking approach are somewhat better than those obtained when the Fv fragment of the antibody was used. Hence, the latter practice was adopted when possible.

Table 3 presents the docking results for unbound structures. Geometric docking of unbound structures ranks a nearly correct solution among the top 10 in four systems out of the 17 that we tested. The introduction of electrostatic complementarity significantly improves the ranking of a nearly correct solution for 10 out of the 17 systems, and in some cases, the improvement is dramatic. For example, the rank of a nearly correct solution for the acetylcholinesterase/fasciculin-II system is elevated from 689–705 to 3, for the trypsin/BPTI system, it is elevated from 271–278 to 63–64, and in the 1bth system, the rank of the nearly correct solution is elevated from 508–531 to 99–102 on introduction of electrostatic complementarity. The geometric-electrostatic results are worse than the geometric results for the TEM1/BLIP system and for some of the antibody/antigen systems. These results are discussed below.

The step representation versus the continuous representation of the potential

Five disassembled (2ptc, 1brs, 2sec, 1cho, 2hhb) and two unbound systems (5cha/1ovo and 1bni/1bta) are tested with both the continuous and the step representations of the potential. The simplified step representation of the potential does not reduce the ability of the algorithm to identify the correct solution. In fact, the results obtained with the step representation are superior to those obtained with the continuous representation. Thus, the ranks of the nearly correct solutions for 5cha/1ovo and 1bni/1bta are 66–68 and 31–32, respectively, worse than the ranks obtained with the step representation: 27–29 and 1, respectively. The geometric-electrostatic complementarity scores for these systems, calculated with either the continuous or the step representation, are higher than the geometric scores. Hence, the lower ranking of the nearly correct solutions obtained with the continuous representation of the potential is caused by the appearance of additional false-positive solutions. Inspection of the values of the potential on the surfaces of the molecules involved shows that the range of the absolute values of the potential is very large and spans two orders of magnitude. The products of the potentials, which are summed up to

Table 3. *Comparison of the geometric and geometric-electrostatic docking results obtained for unbound structures with the step representation of the electrostatic potential*

	Geometric[4]				Geometric-electrostatic[4]			
System	Rank	Score	Z, Z_1	RMSD	Rank	Score	Z, Z_1	RMSD
2ptc:2ptn/4pti	271–278	423	3.2, 7.9	2.55	63–64	516	5.0, 8.3	2.55
1brs:1bni[1]/1bta	10–11	462	6.3, 8.3	1.82	1	659	5.2, 5.2	1.82
2sec:1scd/1tec	918–949	423	2.1, 9.1	1.02	961–992	427	2.0, 7.6	1.02
1cho:5cha/1ovo	131–136	423	3.8, 9.1	0.93	27–29	483	5.4, 8.3	0.93
1avw:1ept/1avu[1]	2288–2348	410	1.2, 7.7	2.12	1688–1729	425	1.5, 8.1	2.12
1tec:1thm/2sec	180–186	480	3.5, 7.1	2.60	170–176	482	3.5, 8.0	2.60
1smf:2ptn/1pi2	8	520	6.1, 7.3	1.29	6	532	5.8, 7.5	1.29
1bth:4htc[1]/4pti	508–531	434	2.6, 7.7	4.20	99–102	559	4.5, 8.2	3.87
1ldt:1ept/1ldt	27	416	5.4, 8.9	1.00	7–8	502	6.5, 9.4	1.00
1fss:2ace[1]/1fsc	689–705	435	2.2, 7.9	0.70	3	978	3.2, 3.3	0.88
TEM1/BLIP[2]	1	688	8.2, 8.2	1.77	98–101	610	4.0, 5.7	1.77
3hfm:3hfm[3]/1lza	102–105	481	3.9, 6.9	2.80	2330–2351	579	1.1, 4.1	2.80
3hfm:3hfm[3]/1hel	119–128	488	4.0, 6.7	2.70	1096–1103	636	1.8, 4.3	2.70
3hfl:3hfl[3]/1lza	199–207	440	3.5, 6.1	1.25	3514–3552	470	0.6, 5.1	1.25
3hfl:3hfl[3]/1hel	1	580	7.9, 7.9	0.24	521–525	594	2.7, 5.6	0.24
1vfb:1vfa/1lza	No solution		—, 12.0	—	4673–4738	351	0.5, 11.2	1.91
1vfb:1vfa/1hel	932–961	397	2.0, 9.8	1.71	894–920	400	2.1, 9.2	1.71
2jel:2jel[3]/1poh	286–292	425	3.0, 6.6	2.42	381–389	422	2.8, 6.4	2.42
1dvf:1vfa/1dvf	2725–2793	417	0.9, 6.6	0.37	2599–2665	433	1.0, 8.8	0.37
1cft:1cfq/1cft[3]	329–354	249	2.8, 5.8	0.21	168–188	261	3.4, 6.4	0.21

The RMSD values listed denote the difference between the predicted structure of the complex and the experimental structure calculated for the common Cα atoms in the complex.
[1] Several missing side-chains in this structure were modeled (see text).
[2] The results presented here slightly differ from the results in Strynadka et al. (1996a) and Eisenstein and Katzir (1998) because different grid intervals were used.
[3] See footnote 2 in Table 2.
[4] See footnote 3 in Table 2.

provide the electrostatic contribution to the complementarity score (equation 5), span up to four orders of magnitude, and in some cases, a single contribution may dominate the electrostatic complementarity score. This high sensitivity to details is undesirable, as discussed in the Introduction, and we therefore prefer the step representation of the potential.

Statistical significance

Tables 2 and 3 also list Z values for the nearly correct solution and for the top-ranking solution (Z_1) in each geometric and geometric-electrostatic scan. The Z_1 values for the geometric scans of disassembled systems range from 6.1 to 17.7. The corresponding range for unbound systems is 5.8 to 9.8. The distributions of scores in the geometric-electrostatic scans are consistently wider than those in the corresponding geometric scans (larger σ values; data not shown). Hence, the ranges of Z_1 values in such scans are shifted toward smaller values: 3.3 to 11.2 for unbound systems versus 5.4 to 16.8 for disassembled systems.

Discussion

In this study, we present a new algorithm for geometric-electrostatic docking of proteins, which is an extension of our geometric docking algorithm (Katchalski-Katzir et al. 1992; Eisenstein et al. 1997). The molecules to be docked are represented by 3D grids of complex numbers. We exploit the vector nature of complex numbers and store information regarding the shape of the molecule in the real part and its electrostatic character in the imaginary part. The representation of the electrostatic character is approximate and derived from the electrostatic potential. It allows correlation of the gross electrostatic features, which are not sensitive to small conformation changes and small mispositioning of the molecules in the complex. Indeed, a comparison of the electrostatic patches in bound and unbound structures supports this notion, showing almost no difference in the shape and size of these patches (Fig. 2).

The use of the imaginary part of the complex numbers to store information is a novel feature in protein-protein docking. In this study, we combine geometric and electrostatic complementarity by storing information regarding the electrostatic character of the molecule in the imaginary part of the grid. However, other parameters, for example, the degree of hydrophobicity or aromaticity of the exposed side-chains, can be stored there. Alternatively, biological, biochemical, and evolutionary information, which reflects the tendency of given amino acids to be involved in interactions with another molecule, can be stored in the imaginary part

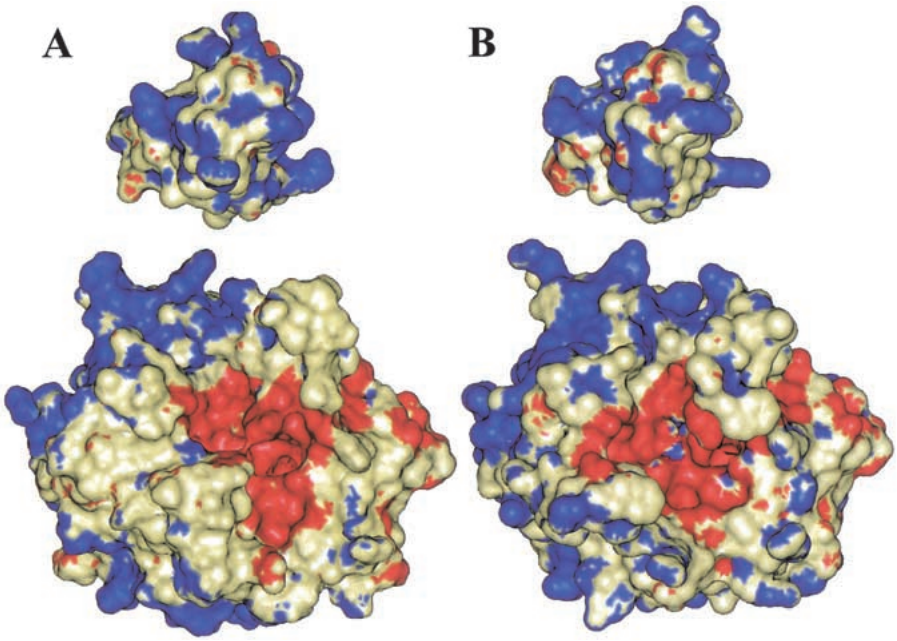

Fig. 2. Comparison of the electrostatic representation for the unbound structures of thrombin and BPTI (*A*) and for the disassembled structures (*B*). The inhibitor in the disassembled structure was rotated by 180° on the horizontal axis and translated to expose the binding sites of the two molecules. The orientation of the unbound molecules is the same as that of the disassembled molecules. We choose the pair thrombin/BPTI, which undergoes a relatively large conformation change on complex formation, to illustrate the stability of our electrostatic representation. The solvent-accessible surface of the molecules was calculated using the MSI package. The red patches on the surface represent areas with a negative potential <−3.0 kT/e, and the blue patches represent areas with a positive potential >3.0 kT/e.

of our complex-numbers representation. We are currently exploring all these directions.

Another novel feature of our docking algorithm is the translation of the electrostatic-potential to potential spheres, which are rotated together with the atoms in the molecule and projected onto the MolFit grid. Hence, only a single computation of the electrostatic potential is necessary for each molecule involved.

The geometric-electrostatic algorithm was applied to 17 disassembled systems and to the corresponding unbound systems, providing very significant improvement in the ranking of the nearly correct solution in a considerable number of cases. We compare the results to the results of geometric docking and analyze them in terms of the rank of the nearly correct solution (the position of this solution in the sorted list of solutions produced by MolFit). The ranks reflect the success of docking tests more effectively than the complementarity scores, which cannot be compared between different systems.

Comparison of the geometric and geometric-electrostatic docking results

Electrostatics consistently improves the docking results for disassembled systems (see Table 2): The complementarity scores increase (except for the TEM1/BLIP system, which is discussed below), and the ranks of the nearly correct solutions are higher, when possible. More importantly, electrostatics elevates the scores of the nearly correct solutions for 18 of the 20 unbound systems and improves their ranking for 13 of these systems (see Table 3). At first glance, the enzyme/inhibitor and antibody/antigen groups of systems present a distinctly different behavior with respect to the electrostatic contribution to the complementarity score: In the enzyme/inhibitor group, the scores increase and improve the ranking of nearly correct solutions, whereas in the antibody/antigen group, the increase in score is accompanied by an increase in the number of false-positive solutions, resulting in inferior ranking of the nearly correct solutions. The different behavior of enzyme/inhibitor and antibody/antigen systems has been noticed before (Sternberg et al. 1998). It is discussed in more detail and challenged below.

The statistical analysis of the distribution of scores (Table 4) provides an additional view of the results. For each scan, we compare the Z value of the nearly correct solution to Z_1, which is the Z value of the top-ranking solution. All the 17 disassembled systems have a nearly correct solution within the top 4σ range in the geometric scans (ranks 92 and less) and the top 3σ range in the geometric-electrostatic scans (ranks 33 and less). One expects that the conformation differences between disassembled and unbound structures will obscure the uniqueness of the correct structure. Indeed, the Z values for the nearly correct solutions obtained in docking of unbound structures are generally lower than those obtained for disassembled systems. We find a nearly correct solution in the top 5σ range for 12 of the 20 systems in geometric docking and for 15 systems in the geometric-electrostatic scans. Thus, our geometric-electrostatic docking appears to be more successful than the geometric docking in most cases.

Parameters affecting the geometric docking results

It is important to pinpoint the factors that determine both qualitatively and quantitatively the geometric and electro-

Table 4. *Statistics of the nearly correct solutions in different series of scans*

σ range	Z_1–1	Z_1–2	Z_1–2.5	Z_1–3	Z_1–3.5	Z_1–4	Z_1–5
Geometric disassembled (11, 6)	10 (7, 3)	13 (10, 3)	15 (11, 4)	15 (11, 4)	16 (11, 5)	17 (11, 6)	
Geometric- electrostatic disassembled (11, 6)	10 (7, 3)	13 (10, 3)	16 (11, 5)	17 (11, 6)			
Geometric unbound (11, 9)	2 (1, 1)	4 (3, 1)	4 (3, 1)	7 (3, 4)	8 (4, 4)	11 (5, 6)	12 (6, 6)
Geometric- electrostatic unbound (11, 9)	2 (2, 0)	4 (4, 0)	5 (4, 1)	10 (6, 4)	11 (7, 4)	13 (8, 5)	15 (9, 6)

In each cell, we list the numbers of systems for which a nearly correct solution is found in the given σ range. Z_1 is the statistical uniqueness value for the solution with the highest score in the given scan. Z values are measured in σ units; hence, the Z_1–1 range is the top 1σ range. The values in parenthesis are for the enzyme/inhibitor and antibody/antigen groups. These groups contain 11 and 6 disassembled systems (excluding 3hfm (Fv)/lysozyme), respectively, and 11 and 9 unbound systems, respectively.

static contributions to the score. One may expect that the geometric part of the complementarity score be affected by the degree of structural change, which occurs on complex formation. However, the correlation is weak at most. For example, the geometric docking results for the five cases in which an unbound structure is docked to a disassembled structure are not better than those for cases of unbound/unbound docking. Another demonstration of this lack of correlation is a comparison of the case of thrombin/BPTI, in which exceptionally large structural changes are observed for the contact residues (4.26 and 1.74 Å, respectively), and the case of trypsin/soy bean inhibitor, in which the structural changes are moderate (0.91 and 1.41 Å, respectively). The rank of the nearly correct solution in the first case is significantly higher than in the second case.

The geometric complementarity scores are roughly related to the size of the interface, and therefore, the size and possibly the shape of the interface may contribute to the success of the prediction by geometric docking. However, the buried surface areas, listed in Table 2, appear to be unrelated to the success of the geometric docking, in terms of the ranking of the nearly correct solutions, of either disassembled or unbound structures (see Tables 2, 3).

It appears that the results of the geometric docking of unbound structures do not correlate with obvious features such as the amount of structural change on complex formation and the area of the interface. Possibly the success or failure of the geometric docking is related to the omission of water molecules in the docking procedure. In most systems, there are several buried water molecules, which fill gaps between the interacting molecular surfaces. Our algorithm ignores these water molecules, and therefore, in some cases conformation changes are accommodated more easily than in other cases. The complex between barnase and barstar provides a good example for a beneficial effect because of the omission of water. The shape complementarity in this system is not perfect because several water molecules cluster on one side of the interface (Buckle et al. 1994). Our analyzes of the top-ranking solutions in the geometric and geometric-electrostatic scans for disassembled structures and for unbound structures identify clusters of nearly correct solutions. Such clustering is most likely caused by the omission of water molecules, which allows limited rotational freedom at the interface.

Parameters affecting the geometric-electrostatic docking results

In most cases, the geometric-electrostatic docking is superior to geometric docking, as described in the results section and summarized in Table 4. Our 'imperfect success' raises two related questions: What are the factors that determine the success or failure of the geometric-electrostatic docking compared with the geometric docking, and can we predict which algorithm should be applied to a given system? In attempt to answer these questions, we inspected the potential patches on the surfaces of the molecules in all the systems under consideration and compared their size and distribution at the interface to other portions of the surface (the noninteracting surface). We also calculated the molecular dipole moments and the angles between these vectors for each pair of interacting molecules.

The 10 systems for which the introduction of electrostatic complementarity significantly improves the ranking of the nearly correct solution (2ptc, 1brs, 1cho, 1avw, 1smf, 1bth, 1ldt, 1fss, 1vfb, and 2jel) are all characterized by one or two very large potential patches on the interacting surface of the enzyme or antibody and a very large or several large potential patches of the opposite sign on the interacting surface of the ligand. Moreover, the potentials at the interface often protrude into the solvent (Fig. 3). This increases the number of MolFit grid points with an imaginary part that is nonzero (the real part may be zero) and adds to the electrostatic complementarity score. In contrast to the above, the potentials on the noninteracting portions of the surfaces are characterized, in most cases, by a mixture of positive and negative potential patches of different sizes.

Another factor that probably contributes to the success of the geometric-electrostatic docking is a dipolar character of the molecule. The most striking example is the acetylcholinesterase/fasciculin-II system in which the contribution of electrostatic complementarity is exceedingly large (Fig. 3). In addition to the occurrence of large potential patches with opposite signs on the interacting surfaces, which extend away from the molecular surface, both molecules are dipolar in nature, with large dipole moments that are parallel in the complex. The dipolar disposition of potential patches is likely to reduce the number of false-positive docking solutions.

Three systems—1tec, 2sec, and 2jel—are only slightly affected by the introduction of electrostatic complementarity. They are characterized by relatively small potential patches on the interacting surfaces of the enzyme or antibody and the ligand. The noninteracting surfaces of these molecules display a mixture of positive and negative patches.

The TEM1/BLIP system, in which the electrostatic contribution to the complementarity score is negative, stands out in the group of 11 enzyme/inhibitor systems that we tested. Inspection of the structures and the potentials around TEM1 and BLIP reveals that there is only limited electrostatic complementarity at the interface (see Fig. 4). Thus, a large negative electrostatic patch is found on the binding surface of TEM1 next to an assembly of small positive patches. The inhibitor has a small positive patch at the interface surrounded by negative patches, which in the complex are in contact with the negative patch on the surface of

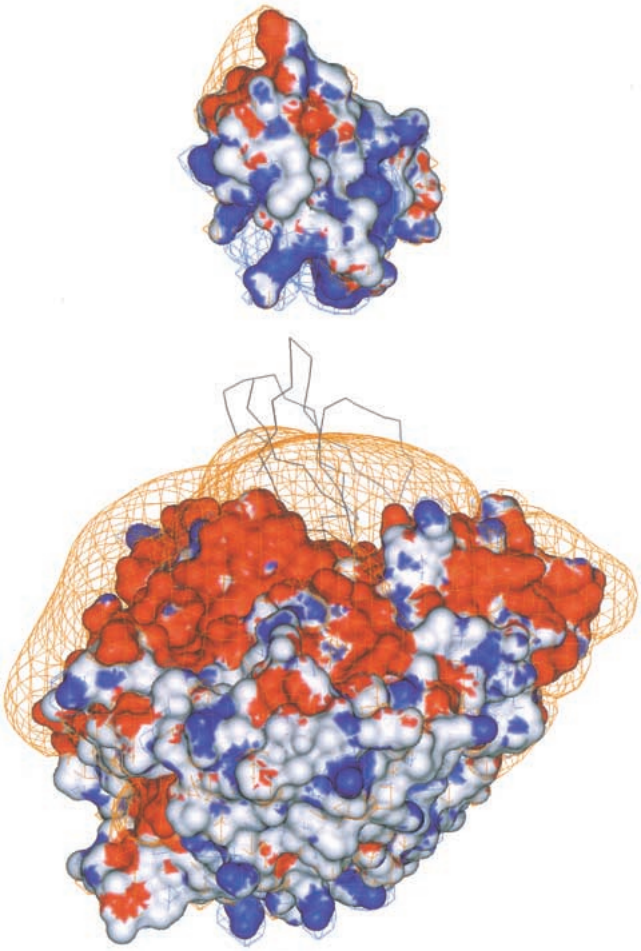

Fig. 3. The electrostatic patches on the solvent accessible surfaces of acetylcholinesterase and fasciculine-II (1fss) and the electrostatic potential contours at 3 kT/e (blue) and −3 kT/e (orange), illustrating the large potential patch on the interacting surface of the enzyme, the protrusion of the potential into the solvent, and the dipolar disposition of the electrostatic patches on the surfaces of both molecules. The inhibitor was shifted away from the binding site for clarity. Its position in the complex is labeled by the black trace. The color code of the surface is as in Fig. 2. The molecules in this complex comply with our good electrostatic docking rules.

the enzyme, leading to a negative contribution to the electrostatic complementarity score (see Tables 2, 3).

A recent study by Selzer et al. (2000) shows that mutation of aspartate 163 in BLIP to a neutral or a positively charged residue significantly increases the rate of TEM1-BLIP association. The effect is larger for the D163K mutant than for the D163A mutant. We replaced aspartate 163 in BLIP by alanine or lysine (the side-chain of the latter was given an extended conformation), recalculated the electrostatic potential around the molecule, and performed new geometric and geometric-electrostatic rotation/translation scans. The results are shown in Figure 4. The replacement by alanine does not affect significantly the geometric docking results. In contrast, the geometric-electrostatic results for the D163A mutant are significantly improved. The mutation to alanine causes shrinkage of the negative electrostatic patch on the binding surface of BLIP (see Fig. 4), and therefore, the geometric-electrostatic complementarity score is considerably higher than the corresponding value for wild-type BLIP (717 versus 610 score units). The rank of the nearly correct solution improves from 98–101 to 9. Geometric docking of the mutant D163K to TEM1 gives slightly worse results than the docking of wild-type BLIP (score, 612; rank, 11) because the lysine side-chain clashes with the enzyme. Nevertheless, the geometric-electrostatic docking results for this mutant are much better than those for wild-type BLIP because of the high electrostatic complementarity score (95), ranking the nearly correct solution 17 instead of 98–101. Our docking results agree with the experimental data (Selzer et al. 2000), despite the approximate evaluation of the electrostatic complementarity score and the approximate nature of rigid body docking. The electrostatic repulsion that we observe for TEM1/BLIP is a genuine effect particular to this system.

Introduction of electrostatic complementarity has a nega tive effect on the docking results for several antibody/antigen systems. The inferior ranking of the nearly correct solutions in the geometric-electrostatic docking for these systems cannot be related to a lack of electrostatic anticorrelation at the interaction sites because in most cases there is an increase in the complementarity score when either disassembled or unbound structures are docked. We note, however, that three of the six antibody/antigen complexes involve lysozyme (3hfm, 3hfl, and 1vfb). Two of the three anti-lysozyme antibodies (Hyhel-10 in 3hfm and Hyhel-5 in 3hfl) are characterized by negative potential patches at the interface and on the noninteracting surface. The antigen, lysozyme, is characterized by positive potential patches all around, which extend away from the molecular surface in several places. It is most likely that the electrostatic homogeneity of the molecules that form complexes 3hfm and 3hfl allows formation of many false-positive docking solutions and reduces the ability of the geometric-electrostatic docking algorithm to identify the correct solution. In contrast, inclusion of electrostatic complementarity improves the docking results for the third antibody/lysozyme system, 1dvf. The antilysozyme antibody in this complex, D1.3, displays a large negative potential patch at the interface, whereas on its noninteracting surface positive and negative potential patches are mixed.

Conclusions: The good electrostatic docking rules

In view of the analyses above, one may conclude that geometric-electrostatic docking should be preferred over geometric docking when the molecules involved display large positive or negative potential patches on a part of the surface, and the potential next to some of these patches pro-

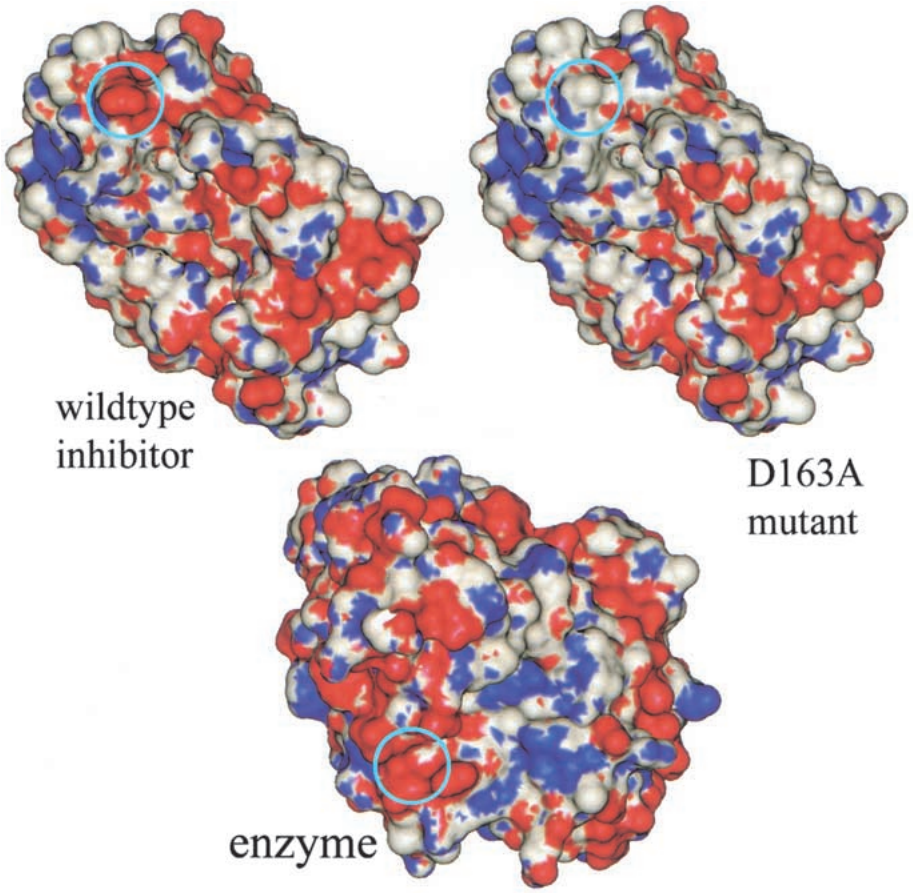

Fig. 4. The electrostatic patches on the solvent-accessible surface of TEM1, wild-type BLIP, and the mutated D163A BLIP, showing the change in the electrostatic potential in the mutant, which significantly affects the geometric-electrostatic docking of these molecules. The color code is as in Fig. 2. The cyan circles highlight the position of the D163A mutation in the inhibitor and its binding site in the enzyme.

trudes into the solvent. On the other hand, if the electrostatic potential on the surface of one or both molecules is homogenous, geometric docking is likely to produce better results. These rules originate from the nature of our docking algorithm and of essentially all other protein-protein docking algorithms. Large potential patches are less sensitive to structural differences between bound and unbound structures and to mispositioning because of the stepwise sampling of the rotation/translation space. Therefore, when the electrostatic potentials are characterized by large features geometric-electrostatic docking is successful. In contrast, when the electrostatic potential appears homogenous, that is, with the same sign all around the molecule, the nearly correct solution is less distinguishable, many false-positive solutions are produced, and the rank deteriorates.

To test these rules, we searched for an antibody/antigen system with an adequate potential. In the Igg2a/peptide system (1cft), the Fv fragment of the antibody displays a mixture of large positive and negative patches, and the potential next to one of the negative patches extends away from the molecular surface. The surface of the antigen in this system is characterized by two large positive and two large negative potential patches. Geometric and geometric-electrostatic docking results for 1cft are shown in Tables 2 and 3. The geometric-electrostatic docking results are considerably better than the geometric docking results, significantly improving the ranking of the nearly correct solution. This is observed in the docking of the disassembled structures as well as for the unbound structures and supports our good electrostatic docking rules. Moreover, it appears that the same rules apply to antibody/antigen systems and enzyme/inhibitor systems, and the different behavior with regard to electrostatic complementarity noted before is a result of the limited choice of antibody/protein systems in the PDB, many of which involve lysozyme as a ligand.

We can further examine our rules against the results of Mandell et al. (2001). They studied three unbound systems and noticed a significant electrostatic contribution in two of them: mouse acetylcholinesterase/fasciculin 2 and cytochrome c peroxidase/cytochrome c. The electrostatic char-

acter of the first system is discussed above. Cytochrome c peroxidase displays a large negative patch at the interface, whereas on the noninteracting surface, positive and negative patches are mixed. Cytochrome c has a cluster of positive patches at the interface and mixed positive and negative patches on the noninteracting surface. Hence, this system too complies with our good electrostatic docking rules. In contrast, the third unbound system tested by Mandell et al., uracil-DNA glycosylase (UDG) and UDG inhibitor, is less appropriate for electrostatic docking. Thus, for UDG we note a large positive patch at the interface and a large negative patch on the noninteracting surface. However, the inhibitor is mostly neutral, that is, with a weak homogeneous potential. Indeed, the rank of the nearly correct solution, obtained by Mandell et al. for this system, deteriorates when electrostatics is included.

Clustering of nearly correct solutions

Clustering of similar solutions is mentioned above for the barnase/barstar system. In this system, we analyze the 10 top-ranking solutions obtained in the docking of either disassemble or unbound structures. In the first case, three solutions in the geometric scan and five in the geometric-electrostatic scan are variations of the nearly correct solution. Similarly, in the docking results for unbound structures, there are two nearly correct solutions among the top 10 in the geometric scan and four in the geometric-electrostatic scan. Interestingly, the clustering of nearly correct solutions is more effective in the geometric-electrostatic scans than in geometric docking.

We performed a limited cluster analysis for all the other systems, comparing only the translations of molecule **b** with respect to molecule **a**, in the 300 top-ranking docking solutions for unbound structures. Such an analysis indicates positions on the surface of molecule **a**, which are preferred by molecule **b** (at any orientation). We found that for 15 of the 20 systems studied here, one to three distinct clusters were formed. In 14 of the cases, the nearly correct solution was included in one of the clusters. Similarly, the geometric-electrostatic docking solutions form clusters in 17 systems out of the 20, and often the number of clusters is smaller than in the corresponding geometric scans. Again, in 16 cases the nearly correct solution is included in one of the clusters. It appears that clustering analysis can help in identifying the correct docking solution. A more detailed clustering analysis is underway. It is noteworthy that similar results, that is, more effective clustering of solutions in geometric-electrostatic docking, were previously reported by Mandell et al. (2001).

Docking of "residents" and "strangers"

In view of the constantly increasing number of new sequences and structures, one may wish to answer the following question: Do molecules **a** and **b** form a complex? The underlying assumption of the statistical analysis described above is that the interacting surfaces of molecules posses some unique structural and electrostatic features, which are rare among all other possible contacts. Therefore, we want to compare the distributions of scores obtained in the docking of unbound structures ("residents") with the distributions obtained in docking of unrelated system ("strangers"), expecting that the latter are wider. Hence, their Z_1 values will be lower, and there will be no clustering of solutions. To this end, we performed geometric and geometric-electrostatic rotation/translation scans for four pairs of unrelated molecules: trypsin/barstar, acetylcholinesterase/eglin-C, antibody D1.3(Fv)/ovomucoid inhibitor, and subtilisin/β-lactamase. The Z_1 values for the geometric scans for these systems are 7.0, 6.8, 5.9, and 8.6, respectively (7.1 on the average), and for the geometric-electrostatic scans, they are 5.0, 7.3, 6.1, and 8.3, respectively (6.7 on the average). As expected, the average Z_1 values obtained for the docking of strangers are lower than the corresponding values obtained for docking of disassembled and unbound structures: 9.3 and 7.8, respectively, for geometric docking and 8.4 and 7.1, respectively, for geometric-electrostatic docking. However, the ranges of the Z_1 values obtained for the docking of strangers overlap the corresponding ranges for the docking of unbound structures. Interestingly, the docking results for strangers form clusters in only one case, the trypsin/barstar case, and the number of clusters increases when electrostatic complementarity is included. This is unlike the results obtained for the docking of unbound structures. Hence, the combination of statistical analysis and cluster analysis of the docking results may help to distinguish between biologically related and biologically unrelated molecules.

Summary

The results presented in this article indicate that the inclusion of electrostatic complementarity is important in docking, in particular when applied to unbound structures. Interestingly, the geometric-electrostatic docking procedure is more successful than geometric docking when the potential patches are large and the potential protrudes into the solvent. When the electrostatic potential around the molecules to be docked is homogenous, geometric docking is recommended. The exhaustive rotation/translation search used by our algorithm often produces clusters of similar solutions. Clustering may shorten the list of solutions and possibly improve the rank of the nearly correct solution. Moreover, the statistical analysis of the docking solutions together with cluster analysis can help to answer the question "Do molecules **a** and **b** form a complex?" Finally, additional terms will be considered in future docking studies: the importance of hydrophobic complementarity, in particular in permanent complexes (oligomers), and of charge-aromatic interactions,

as well as the inclusion of data from biological, biochemical, and bioinformatics studies.

Materials and methods

Optimization of the value of P_{min}

The fine grid used in the Delphi computations (range of 121^3 to 149^3 grid points) results in a large number of potential spheres, and the rotation and projection of all these spheres is time consuming. Most of the spheres, however, represent very small absolute values of the potential that can be omitted. We chose three systems with distinctly different shapes of the interface: trypsin/BPTI (deep cleft), barnase/barstar (wide concave site), and α/β-hemoglobin (large and flat). In each system, we placed opposite charges on atoms at the interface, calculated the electrostatic potentials emanating from these charges, and then computed the combined geometric-electrostatic correlation scores at the experimental orientation for different values of $P_{\min}$. Notably, in the α/β-hemoglobin system, there are no ion pairs at the interface. Therefore, only for these computations, we placed a negative charge on glutamine 127 in the β-chain.

The electrostatic complementarity score is lower when $P_{\min}$ increases, that is, when more potential spheres are omitted (data not shown). However, the dependence appears to be linear, indicating that docking calculations can be performed with a $P_{\min}$ value of ~3.0 kT/e (for the step representation of the potential), and application of an appropriate scale factor will restore the electrostatic score. Moreover, the slopes of the least-squares lines for four charge pairs from the three systems mentioned above are similar, indicating that this scale factor is independent of the absolute values of the charges, the distance between them and the shape of the interacting surfaces. Similar analyses using the continuous representation of the potential produce $P_{\min} = 2.0$.

Choosing the set of electrostatic charges and atomic radii in the Delphi computation of the electrostatic potential

We compared two sets of charges: the Formal set (charges placed on the formally charged side-chains lysine, arginine, histidine, glutamate, and aspartate and on the carboxy and amino termini) and the PARSE set of charges (Sitkoff 1994), in which partial charges are assigned to all the atoms. Starting from the coordinates of the disassembled trypsin/BPTI complex, a geometric rotation/translation scan was performed. Two nearly correct solutions were obtained, ranking 1 and 4 (12° deviation). Then the electrostatic potentials for each of the eight top-ranking solutions were calculated with either the Formal or the PARSE set of charges, and the combined geometric-electrostatic scores were calculated for several values of *w*.

It appears that when the potentials are calculated with the Formal charges, the electrostatic score is small and positive for the nearly correct solutions and negative for the false-positive solutions (data not shown). When PARSE charges are used, the electrostatic score is large and positive for the nearly correct solutions and small positive or negative for the false-positive solutions. Both sets of charges point out the nearly correct solutions. However, electrostatic potentials calculated with the PARSE charges allow more freedom in the choice of the value of *w* than potentials calculated with Formal charges. In addition, the number of potential spheres is considerably smaller when the PARSE potentials are used: 263,125 and 138,425 for trypsin and BPTI, respectively, compared with 419,561 and 204,967 for the Formal charges potentials. These results indicate that for our docking procedure the PARSE charges produce more adequate potentials than the Formal charges.

Optimization of f, the radius of the potential spheres

The series of computations described above was also used to verify our potential rotation procedure and to determine the value of *f* in equation 6. As mentioned above, the electrostatic potential for molecule **b** was calculated for each of the eight top-ranking solutions. It was translated into potential spheres and projected onto the MolFit grid using three values of *f*: 0.9, 1.0, and 1.1. Alternatively, the electrostatic potential was calculated only once and translated into potential spheres, which were rotated as necessary and projected with the three different values of *f*. The linear correlation coefficients between the scores obtained with the exact and the rotated potentials are close to 1.0 (0.99, 0.99, and 0.97 for $f = 0.9$, 1.0, and 1.1, respectively), and the RMSD values are small (8.5, 7.0, and 11.6 score units, respectively), indicating that the rotation procedure is adequate. In addition, both measures indicate that the most appropriate value for *f* is 1.0. The same optimal value was obtained when the potentials were calculated with either the step representation or the continuous representation of the potential.

Optimization and verification of w

To obtain an estimate of *w*, which weighs the geometric and electrostatic contributions, we compared the top-ranking solutions from the full rotation/translation geometric scan and two geometric-electrostatic scans with different values of *w*. The scores of the nearly correct solutions increase as *w* increases. At the same time, the scores of several top-ranking false-positive solutions from the geometric scan decrease, or increase less steeply. However, often the scores of several low-ranking solutions from the geometric scan increase very steeply when electrostatics are included. These are new false-positive solutions that can be eliminated by limiting the value of *w*. Tests for the trypsin/BPTI, αβ-hemoglobin, and barnase/barstar indicate that when the continuous representation of the potential is used, $w = 0.0015$ is adequate for all three systems.

Similar analyses for the step representation of the electrostatic potential indicate an optimal value of 0.25 for *w*. This value is used in the rotation/translation scans for disassembled structures. However, analyses of the complementarity scores versus *w* for all the disassembled systems indicated that a larger value of *w* (0.35) might be more adequate. The larger value is used in all the rotation/translation scans for unbound systems.

The value of *w* for the step representation of the potential was reexamined for five unbound systems (1thm+2sec, 5cha+1ovo, 1bni+1bta, 2ptn+4pti, and 2ace+1fsc). For each system, we predicted the results of rotation/translation scans with different values of *w*, as described in the Algorithms section. The dependence of the predicted rank of the nearly correct solution on *w* for the five systems is shown in Figure 5. The value $w = 0.35$ appears to be adequate. Moreover, the graphs have very wide minima, indicating that the geometric-electrostatic docking results are stable, and variations in *w* will not significantly improve the ranking of the nearly correct solutions. To ascertain the prediction procedure, we compare the predicted scores and ranks to corresponding values from real scans. For two systems, 1thm+2sec and 5cha+1ovo, we performed additional scans with *w* values predicted to be best for these systems (0.25 and 0.60, respectively, according to Fig. 5).

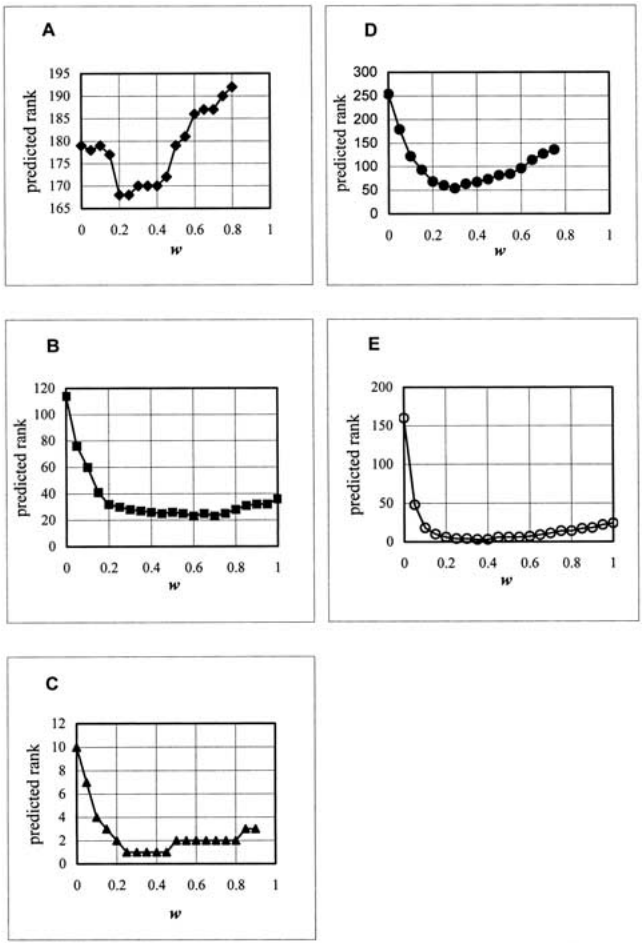

Fig. 5. The predicted ranks of the nearly correct solution in several virtual scans with different values of *w* (see Materials and Methods). The systems are 1thm/2sec (*A*), 5cha/1ovo (*B*), 1bni/1bta (*C*), 2ptn/4pti (*D*), and 2ace/1fsc (*E*).

For the other three systems, 1bni+1bta, 2ptn+4pti, and 2ace+1fsc, in which $w = 0.35$ is optimal, additional scans were performed with nonoptimal *w* values (0.1, 0.45, and 0.50, respectively). The comparison in Table 5 shows that both the scores and the ranks of the nearly correct solutions were predicted correctly, supporting our verification procedure and the choice of the value of *w*.

Table 5. *Real and predicted geometric-electrostatic scores and ranks for different values of the weight parameter* ***w*** *(see Materials and Methods)*

		Real scan		Predicted scan	
System	*w*	Rank	Score	Rank	Score
1thm + 2sec	0.25	168–176	482	167–171	481
5cha + 1ovo	0.60	23–24	526	23–24	526
1bni + 1bta	0.10	4	512	4	512
2ptn + 4pti	0.45	73–74	542	73–74	542
2ace + 1fsc	0.50	3	1210	6	1210

Acknowledgments

This study was supported in part by the Da'at Consortium for Developing Generic Technologies for Design and Development of Drugs and Diagnostics. We thank Dr. I. Shariv and Dr. S. Pietrokovski for helpful discussion. The Program MolFit is available through personal contact with M. Eisenstein.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Ackermann, F., Herrmann, G., Posch, S., and Sagerer, G. 1998. Estimation and filtering of potential protein-protein docking positions. *Bioinformatics* **14:** 196–205.

Ausiello, G., Cesareni, G., and Helmer-Citterich, M. 1997. ESCHER: A new docking procedure applied to the reconstruction of protein tertiary structure. *Proteins* **28:** 556–567.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28:** 235–242.

Bhat, T.N., Bentley, GA, Boulot, G., Greene, M.I., Tello, D., Dall'Acqua, W., Souchon, H., Schwarz, F.P., Mariuzza, R.A., and Poljak, R.J. 1994. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci.* **91:** 1089–1993.

Blevins, R.A. and Tulinsky A. 1985. The refinement and the structure of the dimer of α-chymotrypsin at 1.67-Å resolution. *J. Biol. Chem.* **260:** 4264–4275.

Braden, B.C., Fields, B.A., Ysern, X., Dall'Acqua, W., Goldbaum, F.A., Poljak, R.J., and Mariuzza, R.A. 1996. Crystal structure of an Fv-Fv idiotope–anti-idiotope complex at 1.9 Å resolution. *J. Mol. Biol.* **264:** 137–151.

Bringham, E.O. 1988. *The fast Fourier transform and its applications.* Prentice-Hall, Englewood Cliffs, NJ.

Buckle, A.M., Henrick, K., and Fersht, A.R. 1993. Crystal structural analysis of mutations in the hydrophobic cores of barnase. *J. Mol. Biol.* **234:** 847–860.

Buckle, A.M., Schreiber, G., and Fersht, A.R. 1994. Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution. *Biochemistry* **33:** 8878–8889.

Chen, P., Rose, J., Love, R., Wei, C.H., and Wang, B.C. 1992. Reactive sites of an anticarcinogenic Bowman-Birk proteinase inhibitor are similar to other trypsin inhibitors. *J. Biol. Chem.* **267:** 1990–1994.

Cohen, G.H., Sheriff, S., and Davies, D.R. 1996. Refined structure of the monoclonal antibody HyHEL-5 with its antigen hen egg-white lysozyme. *Acta Crystallogr. D* **52:** 315–326.

Conte, L.L., Chothia, C., and Janin, J. 1999. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285:** 2177–2198.

Dixon, J.S. 1997. Evaluation of the CASP2 docking section. *Proteins* **Suppl 1:** 198–204.

Eisenstein, M. and Katchalski-Katzir, E. 1998. Geometric recognition as a tool for predicting structures of molecular complexes. *Lett. Pept. Sci.* **5:** 365–369.

Eisenstein, M., Shariv, I., Koren, G., Friesem, A.A., and Katchalski-Katzir, E. 1997. Modeling supra-molecular helices: Extension of the molecular surface recognition algorithm and application to the protein coat of the tobacco mosaic virus. *J. Mol. Biol.* **266:** 135–143.

Fermi, G., Perutz, M.F., Shaanan, B., and Fourme, R. 1984. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.* **175:** 159–174.

Fitzpatrick, P.A., Ringe, D., and Klibanov, A.M. 1994. X-ray crystal structure of cross-linked subtilisin Carlsberg in water vs. acetonitrile. *Biochem. Biophys. Res. Commun.* **198:** 675–681.

Fujinaga, M., Sielecki, A.R., Read, R.J., Ardelt, W., Laskowski, Jr., M., and James, M.N. 1987. Crystal and molecular structures of the complex of α-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. *J. Mol. Biol.* **195:** 397–418.

Gabb, H.A., Jackson, R.M., and Sternberg, M.J. 1997. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **272:** 106–120.

Gros, P., Fujinaga, M., Dijkstra, B.W., Kalk, K.H., and Hol, W.G.J. 1994. Crystallographic refinement by incorporation of molecular dynamics: The

thermostable serine protease thermitase complexed with eglin-c. *Acta Crystallogr. B* **45:** 488–499.
Harel, M., Kleywegt, G.J., Ravelli, R.B., Silman, I., and Sussman, J.L. 1995. Crystal structure of an acetylcholinesterase-fasciculin complex: Interaction of a three-fingered toxin from snake venom with its target. *Structure* **3:** 1355–1366.
Honig, B. and Nicholls, A., 1995. Classical electrostatics in biology and chemistry. *Science* **268:** 1144–1149.
Huang, Q., Wang, Z., Li, Y., Liu, S., and Tang, Y. 1994. Refined 1.8 Å resolution crystal structure of the porcine ε-trypsin. *Biochim. Biophys. Acta* **1209:** 77–82.
Jia, Z., Quail, J.W., Waygood, E.B., and Delbaere, L.T. 1993. The 2.0-Å resolution structure of *Escherichia coli* histidine-containing phosphocarrier protein HPr: A redetermination. *J. Biol. Chem.* **268:** 22490–22501.
Jiang, F. and Kim, S.H. 1991. "Soft docking": matching of molecular surface cubes. *J. Mol. Biol.* **219:** 79–102.
Jones, S. and Thornton, J.M. 1996. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci.* **93:** 13–20.
Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C., and Vakser, I.A. 1992. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci.* **89:** 2195–2199.
Keitel, T., Kramer, A., Wessner, H., Scholz, C., Schneider-Mergener, J., and Hohne, W. 1997. Crystallographic analysis of anti-p24 (HIV-1) monoclonal antibody cross-reactivity and polyspecificity. *Cell* **91:** 811–820.
Klapper, I., Hagstrom, R., Fine, R., Sharp, K., and Honig, B. 1986. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins* **1:** 47–59.
Larsen, T.A., Olson, A.J., and Goodsell, D.S. 1998. Morphology of protein-protein interfaces. *Structure* **6:** 421–427.
LeDu, M.H., Housset, D., Marchot, P., Bougis, P.E., Navaza, J., and FontecillaCamps, C. 1992. Structure of fasciculin 2 from green mamba snake venom: Evidence for unusual loop flexibility. *Acta Crystallogr. D* **52:** 87–92.
Levitt, M. and Gerstein, M. 1998. A unified statistical framework for sequence comparison and structure comparison. *Proc. Natl. Acad. Sci.* **95:** 5913–5920.
Li, Y., Huang, Q., Zhang, S., Liu, S., Chi, C., and Tang, Y. 1994. Studies on an artificial trypsin inhibitor peptide derived from the mung bean trypsin inhibitor: Chemical synthesis, refolding, and crystallographic analysis of its complex with trypsin. *J. Biochem. (Tokyo)* **116:** 18–25.
Lubienski, M.J., Bycroft, M., Freund, S.M., and Fersht, A.R. 1994. Three-dimensional solution structure and 13C assignments of barstar using nuclear magnetic resonance spectroscopy. *Biochemistry* **33:** 8866–8877.
Maenaka, K., Matsushima, M., Song, H., Sunada, F., Watanabe, K., and Kumagai, I. 1995. Dissection of protein-carbohydrate interactions in mutant hen egg-white lysozyme complexes and their hydrolytic activity. *J. Mol. Biol.* **247:** 281–293.
Mandell, J.G., Roberts, V.A., Pique, M.E., Kotlovyi, V., Mitchell, J.C., Nelson, E., Tsigelny, I., and Ten Eyck, L.F. 2001. Protein docking using continuum electrostatics and geometric fit. *Protein Eng.* **14:** 105–113.
Marquart, M., Walter, J., Deisenhofer, J., Bode, W., and Huber, R. 1983. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr. B* **39:** 480–490.
McCoy, A.J., Chandana Epa, V., and Colman, P.M. 1997. Electrostatic complementarity at protein/protein interfaces. *J. Mol. Biol.* **268:** 570–584.
McPhalen, C.A. and James, M.N. 1988. Structural comparison of two serine proteinase-protein inhibitor complexes: Eglin-c-subtilisin Carlsberg and CI-2-subtilisin Novo. *Biochemistry* **27:** 6582–6598.
Meyer, M., Wilson, P., and Schomburg, D. 1996. Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking. *J. Mol. Biol.* **264:** 199–210.
Norel, R., Lin, S.L., Wolfson, H.J., and Nussinov, R. 1994. Shape complementarity at protein-protein interfaces. *Biopolymers* **34:** 933–940.
Oliva, M.T. and Moult, J. 1999. Local electrostatic optimization in proteins. *Protein Eng.* **12:** 727–735.
Padlan, E.A., Silverton, E.W., Sheriff, S., Cohen, G.H., Smith-Gill, S.J., and Davies, D.R. 1989. Structure of an antibody-antigen complex: Crystal structure of the HyHEL-10 Fab-lysozyme complex. *Proc. Natl. Acad. Sci.* **86:** 5938–5942.
Papamokos, E., Weber, E., Bode, W., Huber, R., Empie, M.W., Kato, I., and Laskowski, Jr., M. 1982. Crystallographic refinement of Japanese quail ovomucoid, a Kazal-type inhibitor, and model building studies of complexes with serine proteases. *J. Mol. Biol.* **158:** 515–537.
Prasad, L., Waygood, E.B., Lee, J.S., and Delbaere, L.T. 1998. The 2.5 Å resolution structure of the jel42 Fab fragment/HPr complex. *J. Mol. Biol.* **280:** 829–845.
Raves, M.L., Harel, M., Pang, Y.P., Silman, I., Kozikowski, A.P., and Sussman, J.L. 1997. Structure of acetylcholinesterase complexed with the nootropic alkaloid, (—)-huperzine A. *Nat. Struct. Biol.* **4:** 57–63.
Robert, C.H. and Janin, J. 1998. A soft, mean-field potential derived from crystal contacts for predicting protein-protein interactions. *J. Mol. Biol.* **283:** 1037–1047.
Rydel, T.J., Tulinsky, A., Bode, W., and Huber, R. 1991. Refined structure of the hirudin-thrombin complex. *J. Mol. Biol.* **221:** 583–601.
Selzer, T., Albeck, S., and Schreiber, G. 2000. Rational design of faster associating and tighter binding protein complexes. *Nat. Struct. Biol.* **7:** 537–541.
Sheinerman, F.B., Norel, R., and Honig, B. 2000. Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* **10:** 153–159.
Sitkoff, D. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98:** 1978–1988.
Song, H.K. and Suh, S.W. 1998. Kunitz-type soybean trypsin inhibitor revisited: Refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J. Mol. Biol.* **275:** 347–363.
Sternberg, M.J., Gabb, H.A., and Jackson, R.M. 1998. Predictive docking of protein-protein and protein-DNA complexes. *Curr. Opin. Struct. Biol.* **8:** 250–256.
Strynadka, N.C., Adachi, H., Jensen, S.E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K., and James, M.N. 1992. Molecular structure of the acyl-enzyme intermediate in β-lactam hydrolysis at 1.7 Å resolution. *Nature* **359:** 700–705.
Strynadka, N.C., Jensen, S.E., Johns, K., Blanchard, H., Page, M., Matagne, A., Frere, J.M., and James, M.N. 1994. Structural and kinetic characterization of a β-lactamase-inhibitor protein. *Nature* **368:** 657–660.
Strynadka, N.C., Eisenstein, M., Katchalski-Katzir, E., Shoichet, B.K., Kuntz, I.D., Abagyan, R., Totrov, M., Janin, J., Cherfils, J., Zimmerman, F., Olson, A., Duncan, B., Rao, M., Jackson, R., Sternberg, M., and James, M.N. 1996a. Molecular docking programs successfully predict the binding of a β-lactamase inhibitory protein to TEM-1 β-lactamase. *Nat. Struct. Biol.* **3:** 233–239.
Strynadka, N.C., Jensen, S.E., Alzari, P.M., and James, M.N. 1996b. A potent new mode of β-lactamase inhibition revealed by the 1.7 Å X- ray crystallographic structure of the TEM-1-BLIP complex. *Nat. Struct. Biol.* **3:** 290–297.
Stubbs, M.T., Morenweiser, R., Sturzebecher, J., Bauer, M., Bode, W., Huber, R., Piechottka, G.P., Matschiner, G., Sommerhoff, C.P., Fritz, H., and Auerswald, E.A. 1997. The three-dimensional structure of recombinant leech-derived tryptase inhibitor in complex with trypsin: Implications for the structure of human mast cell tryptase and its inhibition. *J. Biol. Chem.* **272:** 19931–19937.
Teplyakov, A.V., Kuranova, I.P., Harutyunyan, E.H., Vainshtein, B.K., Frommel, C., Hohne, W.E., and Wilson, K.S. 1990. Crystal structure of thermitase at 1.4 Å resolution. *J. Mol. Biol.* **214:** 261–279.
Tsai, C.J., Lin, S.L., Wolfson, H.J., and Nussinov, R. 1996. Protein-protein interfaces: Architectures and interactions in protein- protein interfaces and in protein cores: Their similarities and differences. *Crit. Rev. Biochem. Mol. Biol.* **31:** 127–152.
Vakser, I.A. and Aflalo, C. 1994. Hydrophobic docking: A proposed enhancement to molecular recognition techniques. *Proteins* **20:** 320–329.
van de Locht, A., Bode, W., Huber, R., Le Bonniec, B.F., Stone, S.R., Esmon, C.T., and Stubbs, M.T. 1997. The thrombin E192Q-BPTI complex reveals gross structural rearrangements: Implications for the interaction with antithrombin and thrombomodulin. *EMBO J.* **16:** 2977–2984.
Walls, P.H. and Sternberg, M.J. 1992. New algorithm to model protein-protein recognition based on surface complementarity: Applications to antibody-antigen docking. *J. Mol. Biol.* **228:** 277–297.
Walter, J., Steigemann, W., Singh, T.P., Bartunik, H., Bode, W., and Huber, R. 1982. On the disordered activation domain intrypsinogen: Chemical labelling and low-temperature crystallography. *Acta Crystallogr. B* **38:** 1462–1472.
Wilson, K.P., Malcolm, B.A., and Matthews, B.W. 1992. Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme. *J. Biol. Chem.* **267:** 10842–10849.
Xu, D., Tsai, C.J., and Nussinov, R. 1997. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng.* **10:** 999–1012.