Protein Engineering vol.14 no.2 pp.105–113, 2001

Protein docking using continuum electrostatics and geometric fit

Jeffrey G.Mandell[1,2], Victoria A.Roberts[3], Michael E.Pique[3], Vladimir Kotlovyi[2], Julie C.Mitchell[2], Erik Nelson[2], Igor Tsigelny[1] and Lynn F.Ten Eyck[1,2,4]

[1]Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0654, [2]San Diego Supercomputer Center, 9500 Gilman Drive, La Jolla, CA 92093-0505 and [3]Department of Molecular Biology, MB4, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037-1000, USA

[4]To whom correspondence should be addressed

The computer program DOT quickly finds low-energy docked structures for two proteins by performing a systematic search over six degrees of freedom. A novel feature of DOT is its energy function, which is the sum of both a Poisson–Boltzmann electrostatic energy and a van der Waals energy, each represented as a grid-based correlation function. DOT evaluates the energy of interaction for many orientations of the moving molecule and maintains separate lists scored by either the electrostatic energy, the van der Waals energy or the composite sum of both. The free energy is obtained by summing the Boltzmann factor over all rotations at each grid point. Three important findings are presented. First, for a wide variety of protein–protein interactions, the composite-energy function is shown to produce larger clusters of correct answers than found by scoring with either van der Waals energy (geometric fit) or electrostatic energy alone. Second, free-energy clusters are demonstrated to be indicators of binding sites. Third, the contributions of electrostatic and attractive van der Waals energies to the total energy term appropriately reflect the nature of the various types of protein–protein interactions studied.

Keywords: convolution/partition function/Poisson–Boltzmann/protein–protein interactions/structure prediction

Introduction

Predicting protein–protein interactions has long been a goal of computational chemistry. Reliable predictive docking algorithms will provide researchers with a substantial head start in their efforts to study novel protein complexes. Such predictions have been challenging owing to the difficulty in modeling the many forces contributing to protein–protein interactions, which include electrostatics, desolvation, entropy, hydrophobicity, van der Waals and hydrogen bonding (Hendrickson *et al.*, 1987; Stites, 1997). Any particular complex can be dominated by any combination and relative weighting of these driving forces (Shoichet and Kuntz, 1991; Jones and Thornton, 1996). Therefore, a robust energy function should include as many potentials as is feasible.

When the binding site is unknown, a comprehensive search between two proteins is required to find the native complex in what has been termed the 'protein docking problem.' Unfortunately, a complete search of all possible complexes of two large flexible proteins is impossible because the number of configurations is truly vast. The docking problem can be simplified by treating the proteins as rigid bodies and searching over three translational and three rotational degrees of freedom. Impressive results were obtained by rigid docking methods to predict the binding of a β-lactamase inhibitory protein to TEM-1 β-lactamase (Strynadka *et al.*, 1996). This result is even more remarkable considering that the inhibitory protein undergoes a conformational change upon binding. Even for two modestly sized proteins, the computational cost of a rigid-body search over all space can be prohibitive. A key development to solve the problem was the formulation of a simplified energy function that evaluated geometric fit in terms of a correlation function, which is a special case of convolution (Katchalski-Katzir *et al.*, 1992). This formulation permits a rapid translational search between two molecules with their properties mapped on to grids and allows a thorough, systematic evaluation of many orientations between two proteins. Further work by Vakser's group (Vakser and Aflalo, 1994) concentrated on hydrophobic docking and on low-resolution representations of the molecular surfaces (Vakser, 1995, 1996; Vakser *et al.*, 1999).

Many previous reports of protein–protein docking algorithms employing convolution techniques have used geometric fit as the primary scoring function. It should also be noted that computer vision techniques, which can be faster than convolution methods, have also been used to dock proteins based upon geometric complementarity (Fischer *et al.*, 1993; Norel *et al.*, 1994). Although often yielding favorable results, this sole criterion is an oversimplification of the biophysics governing binding and is not expected to be sufficient for interactions with large electrostatic energy components. Harrison *et al.* (1994) employed a composite energy term consisting of both an electrostatic and Lennard-Jones term evaluated with convolutions, but used a relatively simple Coulombic electrostatic model that did not account for the difference in dielectric between the solvent and protein. Coulombic electrostatic energies have successfully been used as a secondary filter to discard geometric-fit predictions that have unfavorable charge interactions (Gabb *et al.*, 1997). In this filter-based method, all favorable (negative) electrostatic energies are treated equally regardless of magnitude, which is a large approximation. Previous work by one of us, embodied in the program TURNIP (Roberts *et al.*, 1991), performed a search that maintained a constant distance between two molecular surfaces and evaluated the Coulombic electrostatic potential energy between them, but did not include geometric fit explicitly and did not use convolution methods. We have previously reported some features of our program DOT (Daughter of TURNIP) (Ten Eyck *et al.*, 1995), which at that time did not account for van der Waals attractive energies, but used convolution methods both to calculate a more realistic continuum electrostatic energy term and to detect collisions.

To date, there has been no report of a convolution-based protein–protein docking program that incorporates both geometric fit and solvent continuum electrostatics into a single energy term. Unlike Coulombic models, solvent continuum electrostatic models capture the effects of the different dielectric constants of water, protein and lipid phases of a system and further account for shielding of charges by counter-ions in the solvent. We have included solvent continuum electrostatic interactions in our energy function, implemented through the program DOT, with the goal of creating a more accurate energy term. The solvent continuum electrostatic model is provided by solving the Poisson–Boltzmann equation (Gilson and Honig, 1988; Davis *et al.*, 1991; Honig and Nicholls, 1995; McCoy *et al.*, 1997). The Poisson equation is a partial differential equation that describes the variation of electrostatic potential in space due to a distribution of charges when the dielectric constant varies with position (the protein interior and surrounding solvent have different dielectric constants). The Poisson–Boltzmann equation is obtained when the charge distribution of counter-ions in the solution is added to the charge distribution of the macromolecule, assuming a Boltzmann distribution based on the electrostatic potential. The Poisson–Boltzmann equation thus allows us to solve for the electrostatic potential as a function of the dielectric constant and the charge density (charges from the macromolecule and from dissolved ions) throughout space. For systems that do not involve high charge densities, a simplified, linearized Poisson–Boltzmann equation can be more rapidly evaluated.

We have examined the benefit of using a composite energy function consisting of the sum of a Poisson–Boltzmann electrostatic energy and a van der Waals energy (implemented by geometric fit). Our results demonstrate that this composite energy term provides larger clusters of correct answers than either geometric fit or electrostatic energy alone. We also show that clusters of answers at the binding site can be found by analyzing the free energies of interaction. A major objective of the DOT program is to provide a method that is fast enough for routine use, cheap enough to be used in highly speculative modes and useful enough to guide the design of experiments to test the suggested interactions.

Materials and methods

Convolution functions

DOT models both the electrostatic and van der Waals energy terms of a protein–protein interaction. Hydrogen bonds were modeled electrostatically and hydrophobic interactions were modeled through van der Waals contacts. One molecule (‘stationary’) was held in a fixed position and the other molecule (‘moving’) was rotated and translated about the first. Two convolutions were performed, one to evaluate the electrostatic energy and the other to evaluate the van der Waals energy and simultaneously to count steric clashes. The implementations used here are similar to those previously described (Katchalski-Katzir *et al.*, 1992; Harrison *et al.*, 1994; Vakser and Aflalo, 1994; Ten Eyck *et al.*, 1995; Gabb *et al.*, 1997) with minor modifications. Both the electrostatic and van der Waals functions are expressed as integrals that are correlation functions, as described in detail below. Correlation functions are a special case of convolution products and thus can be calculated very efficiently using the Convolution Theorem. The Convolution Theorem states that the convolution product of the two functions is equal to the inverse Fourier transform of the scalar product of the Fourier transforms of the functions. Evaluating the correlation directly from the definition (given below) has a computational cost of N^2 multiplications. Evaluating the correlation using the Convolution Theorem and fast Fourier Transforms (FFTs) costs three FFTs, each proportional to $N \log N$ and N multiplications, so the computational cost is proportional to $N\ (3 \log N + 1)$. The convolutions in DOT were computed using a fast Fourier transform algorithm that was optimized for three-dimensional real transforms (Ten Eyck, 1973). Since multiple correlation functions are required and only the moving-molecule function changes, the stationary molecule FFTs can be omitted from all but the first calculation. This reduces the cost multiplier from 3 to 2 and the cost function becomes $N\ (2 \log N + 1)$. This process gives the values of either electrostatic energy or van der Waals and steric contacts throughout all space for a given orientation of the moving molecule.

Electrostatic energy convolution

The electrostatic energy is the product of electric charge and electrostatic potential, summed over the whole system. The stationary molecule was the source of the potential field and the moving molecule was described as a collection of partial charges centered at its atomic coordinates. If $V(\mathbf{x})$ is the electrostatic potential at point $\mathbf{x}$ and $Q(\mathbf{x})$ is the charge density at point $\mathbf{x}$, then the electrostatic energy of the system is given by

$$E = \int V\,(\mathbf{x})Q(\mathbf{x})\mathrm{d}\mathbf{x}$$

If the moving molecule is rotated through angle θ and translated to a position $\mathbf{x}_0$, the electrostatic energy of the system is the product of the rotated and translated charge distribution with the potential field and is given by

$$E_\theta(\mathrm{x}_0) = \int V\,(\mathrm{x})Q_\theta(\mathrm{x}–\mathrm{x}_0)\mathrm{dx}$$

This integral is a correlation function that can be evaluated efficiently through use of the Convolution Theorem as described above.

Partial charges and potential grid generation

Partial charges for both the stationary and moving proteins were assigned according to an AMBER parameter set that includes polar hydrogen atoms only (Weiner *et al.*, 1984, 1986). The partial charges of the moving molecule were placed on the grid using trilinear interpolation relative to the atomic centers. In the case of the stationary molecule, the potential grid was generated by solving the linearized Poisson–Boltzmann equation with the program University of Houston Brownian Dynamics (UHBD) (Davis *et al.*, 1991; Madura *et al.*, 1995). The potential was evaluated on a 128×128×128 grid with 1 Å spacing, a solvent dielectric of 80.0, a protein-interior dielectric of 3.0, a temperature of 300 K, an ionic radius of 1.4 Å and a solvent radius of 1.4 Å. A solvent ionic strength of 50 mM was used for cytochrome *c* peroxidase and acetylcholinesterase, 150 mM for hemoglobin, 145 mM for PKA and 100 mM for UDG. Approximately 2 min were required for each potential grid calculation on a Compaq DS20.

Van der Waals energy convolution

The van der Waals potential for the stationary molecule, $G(\mathbf{x})$, was defined as

$$G(\mathbf{x}) = \begin{cases} M \text{ if } \mathbf{x} \text{ is inside the stationary molecule} \\ 1 \text{ if } \mathbf{x} \text{ is in the surface layer of the stationary molecule} \\ 0 \text{ otherwise} \end{cases}$$

where M is an integer greater than the number of atoms in the moving molecule, ‘inside’ is within the van der Waals surface of the stationary molecule ($\mathbf{x} < 1.5$ Å from an atomic center) and the 'surface layer' consists of grid points in a 3 Å layer surrounding the van der Waals surface (1.5 Å $< r <$ 4.5 Å from an atomic center). The steric interaction was computed by evaluating the function

$$F(\mathbf{x}) = \int G(\mathbf{x})A(\mathbf{x})\mathrm{d}\mathbf{x}$$

where $A(\mathbf{x})$ is the set of delta functions at atomic centers of the moving molecule. If the moving molecule is translated by $\mathbf{x}_0$ and rotated by angle θ, then the steric overlap function for the rotated and translated system is given by the correlation function

$$\begin{aligned} F_\theta(\mathbf{x}_0) &= \int G(\mathbf{x})A_\theta(\mathbf{x}\text{–}\mathbf{x}_0)\mathrm{d}\mathbf{x} \\ &= j(\mathbf{x}_0)M + k(\mathbf{x}_0) \end{aligned}$$

The integer functions $j(\mathbf{x}_0)$ and $k(\mathbf{x}_0)$ count the atoms in the moving molecule that collide with atoms of the stationary molecule and that lie within its surface layer, respectively. The value $k(\mathbf{x}_0)$, proportional to the van der Waals attractive energy, is scaled by the depth of the van der Waals well. We chose a well depth of –0.1 kcal/mol for all interactions. This value was determined by plotting the Lennard-Jones 6–12 potentials for the interactions between carbon–carbon, carbon–nitrogen and carbon–oxygen pairs using parameters from the AMBER force field (Weiner *et al.*, 1984, 1986). The minimum well depth in all cases was close to –0.1 kcal/mol.

The stereochemical energy term was evaluated by first eliminating all grid points at which $j(\mathbf{x}_0)$ (the collision count) was greater than a threshold, typically zero. Implementations of this geometric fit algorithm by others (Katchalski-Katzir *et al.*, 1992; Gabb *et al.*, 1997) were usually formulated so as not to count the number of atomic collisions, *j*, but instead to assign a small penalty for each (around –15 units). If the sum of all penalties was large, the score was poor. We found the performance of DOT to be relatively insensitive to the value of this penalty parameter in the range 0 to –15 units. Instead, we chose to count the number of collisions (and to limit them) since evaluating the electrostatic energy inside the stationary molecule can introduce large errors as a result of singularities at atomic centers. Clamping the electrostatic potential grid can alleviate these artifacts (as described below). In some cases, it was useful to permit $j(\mathbf{x}_0)$ to be some small integral value such as five or ten to accommodate side-chain reorientation upon binding, as reported in Table II.

Clamping the electrostatic potential

The steric energy calculation eliminated configurations in which atoms of the moving molecule penetrated the van der Waals volume of the stationary molecule [when the tolerance for $j(\mathbf{x}_0)$ was 0], but atoms could approach to within 1.5 Å, closer than is physically realistic. This treatment allowed for small conformational changes caused by induced fit and for rounding of moving molecule atom positions to the closest grid point. Unfortunately, too close an approach can result in a few unrealistically large electrostatic energy terms. To alleviate this problem, all values of the electrostatic potential grid of the stationary molecule were clamped to the maximum positive and negative potentials found at its solvent-accessible surface, typically in the range –4 to +4 kcal/(mol.e) (Table I). The solvent-accessible surface, which is 1.4 Å out from the molecular surface, represents the closest approach of the center of a water molecule.

Calculation of the total energy

The total energy of the system, *U*, was found by summing the electrostatic energies (kcal/mol) and the scaled van der Waals energies (kcal/mol).

Calculation of the partition sum

To compute the partition sum, the Boltzmann factor was summed over all orientations at each grid point (excluding those orientations rejected for collisions as described above) as

$$Q_j = \sum_{i=1}^{R} \mathrm{e}^{-(U_i/k_\mathrm{B}T)}$$

where *j* is a grid point, *R* is the number of angles through which the moving molecule is rotated, *T* is the temperature in kelvin and k_B is the Boltzmann constant. Q_j was then converted to the Helmholtz free energy as

$$A_j = -k_\mathrm{B}\,T \log Q_j$$

Both Harrison *et al.* (1994) and Blom and Sygusch (1997) used Boltzmann-weighted probability distributions; Harrison *et al.* used these distributions to estimate free energies.

Rotations

The calculation time scaled linearly with the number of rotations, described with Eulerian angles, for the moving molecule. To test how fine a rotational spacing was required for the molecular systems examined, we used two rotation sets. One set had a mean resolution of about 6° and contained 54 000 rotations and the other set had a mean resolution of about 9° and contained 17 374 rotations. In general, larger moving molecules and interactions involving complex geometric fit required the finer rotation set. The rotation that generated the crystallographic answer was removed from the rotation list to eliminate bias. The finer rotation set (6° resolution) resulted in the evaluation of over 113 billion configurations between the two proteins. This calculation required ~65 h on a Compaq DS20 with two processors.

Program execution and implementation

DOT is a parallel program implemented in the C programming language and uses the Message Passing Interface (MPI) (http://www-unix.mcs.anl.gov/mpi/index.html) for inter-process communication. Various computer systems were used, including a network of 12 SGI Indigo 175 MHz R10000 processors with 128 Mbyte RAM each; 2 DEC 4100s each with 2GB RAM and four processors; a Compaq DS20 with two 500 MHz Alpha 21264 processors and 640 Mbyte RAM; 12 to 20 Sun SPARCstations with at least 64 Mbyte RAM each; and the Cray T3E and IBM SP2 machines, using up to 64 processors, at the San Diego Supercomputer Center. The parallel implementation was achieved by distributing the list of rotations amongst all processors. The energy functions were evaluated on grids containing 128 points in each dimension with 1 Å grid spacing. Each processor accumulated results until all rotations were processed. Dynamic load balancing was employed to ensure efficient partitioning of work. At the end of the calculation, results from all processors were merged. Merging was efficiently accomplished through an *N* log *N* algorithm in which pairs of processors independently merged their results until only the parent processor remained with the collective answers. The criteria for merging the minimum-energy grids was that the best energy at a given grid point was saved. Merging of the partition-sum grids was accomplished through addition of each processor's partition sum grid.

Table I. Properties of systems studied

Stationary protein + [moving protein]	No. of residues	No. of atoms[a]	Net charge (e)	Mean change in ASA (Å^2)[b]	Potential grid clamp[c] [kcal/(mol.e)]	(U)nbound or (B)ound coords.[d]
Hb $\alpha_1\beta_1$ +	287	2742[e]	−3.0		−3.5 to +3.5	B
[Hb $\alpha_2\beta_2$]	287	2742	−3.0	2388		B
PKA +	336	3419	+6.0		−4 to +4	B
[PKI (5−24)]	20	201	+2.0	1093		B
AChE +	543	4194	−10.0		−7 to +4	U and B
[Fas]	61	583	+4.0	1407		U and B
CCP +	294	2407	−12.0		−6 to +4	U and B
[YCC]	103	1100	+6.0	934		U and B
UDG +	223	2205	+5.0		−3.5 to +3.5	U and B
[UGI]	83	785	−11.0	1243		B

[a] Non-polar hydrogen atoms omitted in all cases.

[b] The mean change in solvent-accessible surface area (ASA) that occurs upon complexation in the crystallographically determined solution.

[c] See Materials and methods for the significance of this parameter.

[d] Coordinates used included those extracted from the crystal complex (termed ‘bound’) as well as those individually crystallized (termed ‘unbound’).

[e] CO was omitted as no partial charge data were available. Since this group is deeply buried and very small it does not affect these calculations.

Minimum-energy lists were merged by saving the best *n* values from lists each of size *n*.

Change in solvent-accessible surface area

GRASP (Nicholls *et al.*, 1991) was used to calculate the solvent-accessible surface area for the individual free proteins and the complex. The mean solvent-accessible surface area was defined as half the sum of the total change in solvent-accessible surface area for both proteins in the complex (Jones and Thornton, 1996). For this calculation, a probe radius of 1.4 Å was used on structures with polar hydrogen atoms only.

Protein coordinate files

Protein coordinates were obtained from the Protein Data Bank (PDB) (Berman *et al.*, 2000; http://www.rcsb.org/pdb) and were determined by X-ray diffraction methods. Coordinates have the following PDB codes: carbonmonoxyhemoglobin (1BBB) (Silva *et al.*, 1992); mouse acetylcholinesterase (1MAA) (Bourne *et al.*, 1999); fasciculin 2 (1FSC) (Le Du *et al.*, 1996); complex of acetylcholinesterase with fasciculin 2 (1MAH) (Bourne *et al.*, 1995); cytochrome *c* peroxidase (1CCP) (Wang *et al.*, 1990); yeast cytochrome *c* (1YCC) (Louie and Brayer, 1990); complex of cytochrome *c* peroxidase with yeast cytochrome *c* (2PCC) (Pelletier and Kraut, 1992); complex of the catalytic subunit of cAMP-dependent protein kinase with PKI (5–24), ATP and Mn^{2+} (1ATP) (Zheng *et al.*, 1993); free uracil-DNA glycosylase (1AKZ) (Mol *et al.*, 1995a); and the complex of uracil-DNA glycosylase with uracil-DNA glycosylase inhibitor (1UGH) (Mol *et al.*, 1995b). Water molecules were removed from the files. Polar hydrogen atoms were added to all protein structures with the computer graphics program InsightII (MSI, San Diego, CA) assuming a pH of 7.0. Histidine side chains were protonated only on Nε unless there was a compelling reason (metal ligation or hydrogen bonding) to protonate on Nδ or at both positions.

For cytochrome *c* peroxidase (unbound form), missing atoms of side chains were built with InsightII. To relieve steric interactions created by the inserted side chains, the structure was subjected to 100 iterations of steepest-descents minimization using Discover (MSI) with the *cvff* forcefield. The root-mean-square deviation (r.m.s.d.) of the protein backbone between the minimized and original structures was <0.18 Å.

Results

Throughout the docking calculations, three minimum-energy lists, a minimum-energy grid and a partition-sum grid were maintained. The minimum-energy lists stored the 500 most favorable solutions computed over all rotations and grid points and were scored by either the electrostatic energy, the van der Waals energy or the composite sum of both energies. It is possible that a single grid point could appear multiple times in a list if it had several low-energy solutions for different rotations. The minimum-energy grid stored the rotation with the single most favorable internal energy, *U*, computed at each grid point. The partition-sum grid, used to obtain the free-energy grid, stored the accumulated Boltzmann factor for all energies evaluated at each grid point and was then converted to the Helmholtz free energy, providing a free-energy landscape.

Systems studied

To test interactions spanning the range from those dominated by shape and hydrophobicity to those governed by electrostatics, we selected protein systems that differed considerably in size, charge and amount of surface area buried upon complexation (Table I). For example, dimerization of the hemoglobin (Hb) αβ subunits buries over 2300 Å^2, yet each subunit carries a net charge of only –3. On the other hand, the yeast cytochrome *c* (YCC)/cytochrome *c* peroxidase (CCP) interface buries only 934 Å^2 and involves proteins with net charges of +6 and –12, respectively. The other three complexes studied, acetylcholinesterase (AChE) with fasciculin (Fas), uracil-DNA glycosylase (UDG) with UDG inhibitor (UGI) and camp-dependent protein kinase (PKA) with protein kinase inhibitor (5–24) [PKI (5–24)], have interfaces with intermediate extents of buried surface area. Coordinates used included those extracted from the crystal complex (termed ‘bound’) as well as those individually crystallized (termed ‘unbound’).

Evaluation of DOT solutions

For all selected protein systems, the crystallographic complex has been published. To analyze the DOT results, the r.m.s.d. between the C_α atoms of the 500 best ranked DOT solutions and the X-ray structure was calculated (Table II). DOT solutions within specified r.m.s.d. cutoff values of the crystallographic position were deemed ‘correct’. Dockings involving unbound molecules often required a slightly larger r.m.s.d. criterion and sometimes a small number of allowed collisions for a satisfactory docking. The best 500 answers scored by the composite energy (electrostatic + van der Waals), the van der Waals energy alone and the electrostatic energy alone were examined.

Table II. The number of correct solutions scored by different energy functions

System[a]	Rot. res. (°)[b]	Collisions allowed	R.m.s.d. cutoff	Composite energy			Van der Waals energy			Electrostatic energy		
				No.[c]	Best rank/ r.m.s.d.[d]	Av. r.m.s.d.[e]	No.[c]	Best rank/ r.m.s.d.[d]	Av. r.m.s.d.[e]	No.[c]	Best rank/ r.m.s.d.[d]	Av. r.m.s.d.[e]
PKA$_{\text{bound}}$/PKI$_{\text{bound}}$	9	0	3	37	1/1.48	2.23	9	1/1.48	1.80	32	1/1.48	2.17
UDG$_{\text{bound}}$/UGI$_{\text{bound}}$	9	0	3	17	3/2.16	2.27	11	1/2.16	2.35	0		
UDG$_{\text{unbound}}$/UGI$_{\text{bound}}$	9	10	4	15	13/2.13	2.82	12	5/3.62	3.33	0		
Hb $\alpha_1\beta_{1\text{bound}}$/Hb $\alpha_2\beta_{2\text{bound}}$	6	0	3	13	1/1.85	2.00	6	4/1.85	1.72	0		
CCP$_{\text{bound}}$/YCC$_{\text{bound}}$	6	0	4	8	4/2.48	2.66	1	469/2.48	2.48	0		
CCP$_{\text{unbound}}$/YCC$_{\text{unbound}}$	6	0	4	1	266/3.50	3.50	0			0		
AChE$_{\text{bound}}$/Fas$_{\text{bound}}$	6	0	3	30	1/1.34	2.27	5	17/2.35	2.24	2	122/1.34	2.14
AChE$_{\text{bound}}$/Fas$_{\text{bound}}$	9	0	3	9	1/1.44	2.59	2	55/1.44	2.10	2	127/1.44	2.10
AChE$_{\text{unbound}}$/Fas$_{\text{unbound}}$	6	0	3	1	82/2.24	2.24	0			0		
AChE$_{\text{unbound}}$/Fas$_{\text{bound}}$	6	0	3	5	25/2.30	2.07	0			0		
AChE$_{\text{unbound}}$/Fas$_{\text{bound}}$	6	5	3	13	99/1.34	1.82	0			0		

[a] Coordinates are from the crystallographic complex (bound) or the individually determined (unbound) structures.
[b] The resolution of the rotation set for the moving molecule.
[c] Number of the top 500 DOT solutions within the r.m.s.d. cutoff of the crystallographic position.
[d] Highest ranked DOT solution within the r.m.s.d. cutoff and the r.m.s.d. of this solution from the crystallographic position.
[e] Average r.m.s.d. of all solutions within the r.m.s.d. cutoff of the crystallographic position.

Significantly, the composite-energy term yielded a larger number of correct solutions than either the van der Waals or the electrostatic energy terms alone. In four instances involving unbound coordinates (CCP$_{\text{unbound}}$/YCC$_{\text{unbound}}$, AChE$_{\text{unbound}}$/Fas$_{\text{unbound}}$ and AChE$_{\text{unbound}}$/Fas$_{\text{bound}}$), correct solutions were found by the composite energy, but not by the individual energy terms. These results strongly support the inclusion of electrostatic energy for predicting intermolecular docking.

Table II also shows the rank of the first correct solution in the energy list along with its r.m.s.d. With the composite-energy function, a correct solution was found within the best 25 answers for all but three cases and within the best 266 answers for all systems. For most dockings the van der Waals energy alone finds correct solutions with favorable ranks, consistent with the results others have obtained using geometric fit as the scoring function. The notable exceptions are CCP/YCC and AChE/Fas, which both have a large electrostatic dependence. For bound PKA/PKI, the number of solutions found by electrostatic energy (32) is much larger than the number found by van der Waals energy (nine). This is consistent with the highly electrostatic nature of this enzyme–inhibitor interaction (Grant *et al.*, 1996; Tsigelny *et al.*, 1996). Of all the dockings using bound coordinates, CCP/YCC gave the poorest results using van der Waals energy alone. Only one solution (rank 469) was within the cutoff criteria. This is consistent with the small number of contacts in the crystal structure of the complex (Pelletier and Kraut, 1992). However, eight correct solutions were found with the composite energy function, consistent with the strong ionic strength dependence of the interaction.

All of the dockings performed with coordinates obtained from crystallographic complexes showed a single cluster of correct solutions in the top 30 answers (except CCP/YCC), making the identification of correct solutions using bound coordinates straightforward.

Favorable energy clusters as binding site indicators

Given that DOT calculates the free-energy landscape, we investigated whether this information allows identification of the binding site. Determination of the binding site is particularly useful when using unbound coordinates for which the shape fit is not optimal. Favorable free-energy clustering was seen for all systems and conditions studied, even when the number of correct solutions meeting r.m.s.d. criteria in the top 500 predictions was small (Table II). For example, in the docking of bound UGI to unbound UDG, the free-energy grid shows a large ‘hot spot,’ or cluster, of favorable free energies surrounding the crystallographic solution (Figure 1A).

This effect is particularly pronounced in the CCP/YCC system, which displays long-range electrostatic guidance. Even though the docking of CCP$_{\text{unbound}}$/YCC$_{\text{unbound}}$ shows only one solution in the top 500 with an r.m.s.d. value within 4 Å of the crystallographic position (Table II), the free-energy landscape reveals that the largest concentration of solutions is clustered about the crystallographic position of YCC (Figure 1B). This cluster was not present, however, using the van der Waals energy term alone (Figure 1C), thus emphasizing the ability of the composite-energy term to produce larger clusters of correct answers. This information is useful for identifying the binding site and can in turn be used to select solutions for more detailed examination to identify the correct binding geometry. In the case of AChE/Fas, there were two free-energy clusters, only one of which corresponded to the known binding site. Clusters at unknown binding sites have also been seen by others (Vakser *et al.*, 1999), especially in dockings involving unbound coordinates. Filtering solutions with biochemical information would be useful in such instances (Gabb *et al.*, 1997).

Comparison of the free-energy grid with the minimum-energy grid gives information about the orientational effects on the interaction energies. If the energies at corresponding points between the grids are very close, the interaction is dominated by one or a few favorable orientations. For Hb, the energies of the first five correct answers at corresponding grid points in the minimum-energy and the free-energy grids were identical to six significant digits, indicating that only one orientation contributed significantly to the partition sums. In these cases, there is not only a tight binding configuration for the system, but also a free-energy trap in the region that acts to steer the ligand into the correct position. Other systems such as PKA/PKI had some grid points where correct answers

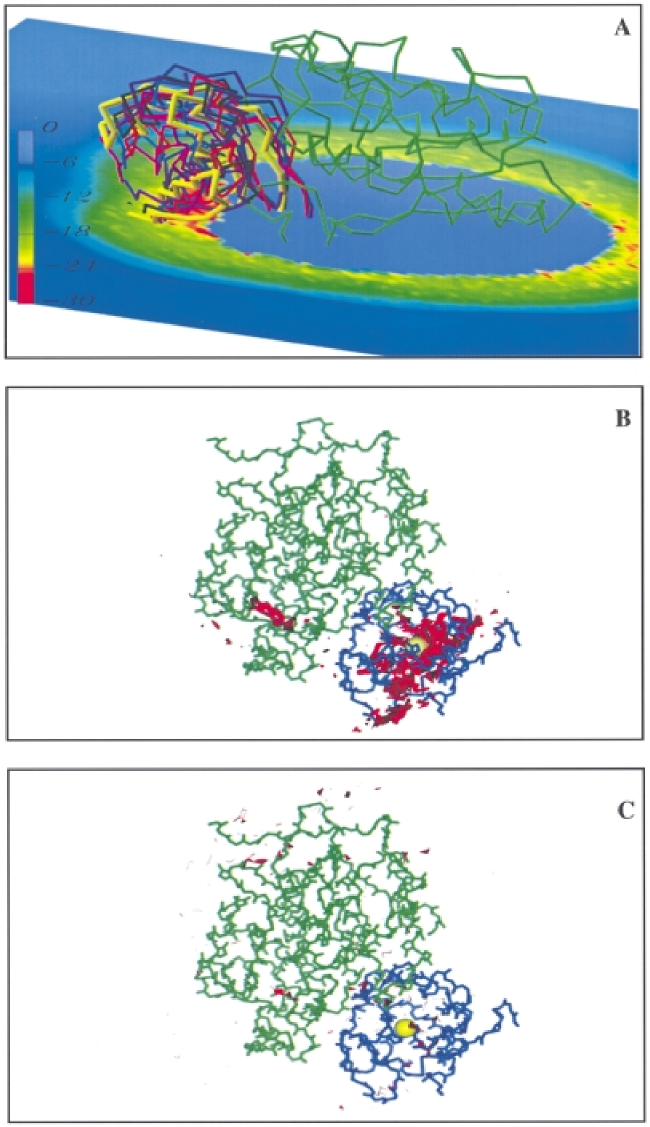

Fig. 1. Free energy maps. (**A**) Docking using the unbound coordinates of UDG (right, green C_α backbone) with the bound coordinates of UGI (left, thick yellow C_α backone) and the composite energy term shown in the crystallographic orientation. A slice through the grid of free energies (kcal/mol) shows a large 'hot spot' of favorable energies (left, red) about the crystallographic position. Other smaller red patches (right) not corresponding to a known binding site are also seen. A single cluster of five UGI solutions (left, shown by the C_α backbone traces) is found in the top 30 most favorable configurations. These solutions, which have r.m.s.d. values of 2.5–6.4 Å with the crystallographic coordinates for all non-hydrogen atoms, all have the appropriate β-strand centered in the protein–protein interface. This central β-strand is best aligned, with larger deviations away from the interface. (**B**) The unbound coordinates of CCP (green C_α backbone) and YCC (blue C_α backbone) shown in the crystallographic orientation of the complex with an isosurface (red) constructed at a level corresponding to the 500th best free energy. The free energy was computed using the composite-energy term. The largest free-energy cluster is clearly visible at the binding site. The sphere (yellow) marks the center of geometry of YCC. (**C**) The CCP/YCC system with free energies consisting of only the van der Waals energy term. Note that the free energy cluster present at the binding site in (B) has disappeared, demonstrating the importance of the composite-energy term.

were found with lower free-energy values than minimum-energy values. For example, at one grid point, the single most favorable energy from the minimum-energy grid was –22.4776, but the free energy for this grid point was –22.8993. Although

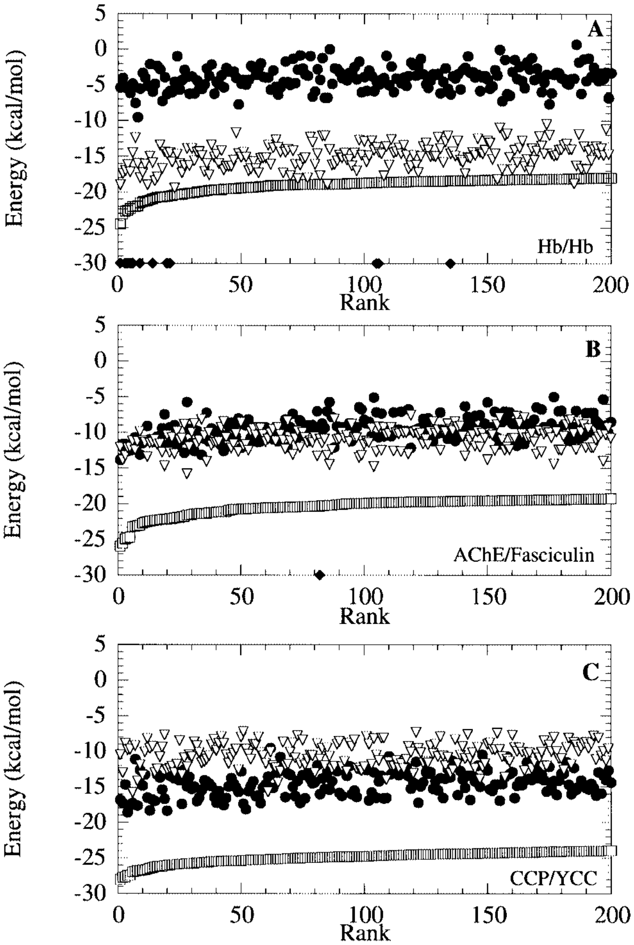

Fig. 2. Contributions of the electrostatic (●) and van der Waals attractive (▽) energies to the composite energy (□) for the top 200 solutions from the composite-energy list. The composite energy is the sum of the electrostatic and van der Waals energies. Solutions identified as correct (see Table II) are denoted on the abscissa (◆). (**A**) Docking of the Hb subunit $\alpha_1\beta_1$ to $\alpha_2\beta_2$. (**B**) $(\text{AChE})_{\text{unbound}}$ to $(\text{Fas})_{\text{unbound}}$. (**C**) $(\text{CCP})_{\text{unbound}}$ to $(\text{YCC})_{\text{unbound}}$.

this difference may appear insignificant, it actually represents several favorable solutions summed together since we are inspecting the logarithm of the partition sum.

Electrostatic and van der Waals contributions to the total energy function

Since the systems we examined vary considerably in total charge and size of the interface (Table I), we investigated the average contributions of the electrostatic and van der Waals energies to the composite energy for each complex (Figure 2). The energy term for Hb (Figure 2A) is clearly dominated by the attractive van der Waals term, while the energy term for CCP/YCC is dominated by the electrostatic term (Figure 2C). On the other hand, the AChE/Fas interaction has almost equal energy contributions (Figure 2B). Table III lists the average contributions of both the van der Waals and electrostatic energies to the composite energy for each system studied. The systems fall into three categories: those with large changes in mean solvent-accessible surface area (ASA) and small net charges (Hb), those with small changes in mean ASA and large net charges (CCP/YCC) and those with moderate net charges and moderate change in mean ASA (AChE/Fas, UDG/UGI and PKA/PKI). The first category is dominated by the van der Waals energy term, the second category by the

Table III. Average contributions to the composite-energy term

Stationary protein +	Av. total energy	Av. elec. energy	Av. vdW energy	Net charge (e)	Mean change in
[moving protein]	(kcal/mol)	(kcal/mol)	(kcal/mol)		ASA (Å^2)[a]
Hb $\alpha_1\beta_1$ +	−19.03	−4.02	−15.02	−3.0	
[Hb $\alpha_2\beta_2$]				−3.0	2388
PKA +	−16.76	−7.93	−8.83	+6.0	
[PKI (5−24)]				+2.0	1093
AChE$_{\text{unbound}}$[b] +	−20.39	−9.49	−10.90	−10.0	
Fas$_{\text{unbound}}$				+4.0	1407
CCP$_{\text{unbound}}$ +	−24.94	−14.41	−10.53	−12.0	
YCC$_{\text{unbound}}$				+6.0	934
UDG$_{\text{bound}}$ +	−17.42	−6.23	−11.20	+5.0	
UGI$_{\text{bound}}$				−11.0	1243

[a] The mean change in solvent-accessible surface area (ASA) that occurs upon complexation in the crystallographically determined solution.

[b] The subscripts 'bound' and 'unbound' refer to whether the coordinates were taken from the crystal complex or from individually solved structures, respectively.

electrostatic energy term and the third category has roughly equal contributions from both terms, the exception being UDG/UGI. For this complex the interaction energy is dominated by the van der Waals term, which may seem surprising given the large negative charge on UGI and the idea that UGI acts as a DNA mimic, but it is consistent with the large area of UGI buried within the interface (25% of the solvent-accessible surface).

Test of the required rotational resolution

Since the computational cost of a DOT run is linearly related to the number of rotations, it is important to determine the fineness of the rotation set required for a successful docking. A rigorous test is the crystallographic complex of AChE/Fas. Three β-sheet 'fingers' of Fas penetrate deeply into the active-site gorge of AChE and must be precisely aligned. A docking using 9° resolution produced significantly fewer correct answers than a docking using 6° resolution (Table II). This test illustrates one advantage of assigning the smaller molecule in the complex as the moving molecule; the smaller molecule is more finely sampled over its surface for a given rotational set. It also demonstrates that very fine sampling is required in cases with convoluted surface topology. For a less convoluted interface such as that in the PKA/PKI system, a 9° resolution rotation set found the largest number of correct solutions of all the systems studied.

Discussion

Inclusion of solvent continuum electrostatics improves docking results

We have found that a composite scoring function consisting of the sum of Poisson–Boltzmann electrostatic and van der Waals energies yields larger numbers of correct solutions than scoring by either energy component alone. Increases in the number of correct solutions were observed with both unbound and bound coordinates. The composite energy term, but not either component alone, was able to find solutions with a geometry very close to the crystallographic orientation within the best 266 minimum energies for all systems studied. Although inclusion of the electrostatic term in the scoring function clearly increased the number of correct solutions, it did not necessarily improve the rank of the best solution. Generally, geometric fit alone was sufficient to find well-ranked solutions except for systems that have a large electrostatic dependence.

Larger numbers of correct solutions aid identification of the binding site

A method capable of generating larger numbers of correct solutions has important advantages. The appearance of many similar low-energy solutions in a list of possibilities is a strong indicator of a correct solution. Even when currently available methods for predicting protein–protein interactions find the correct solution, it is not necessarily the most favorably scored. Clusters of low-energy solutions can aid identification of the native configuration. Larger numbers of correct solutions can also aid biochemical filtering procedures. In studies on a variety of systems using a geometric-fit algorithm (Gabb *et al.*, 1997), use of unbound coordinates usually resulted in no correct solutions with a rank better than 100. Correct solutions could be identified with stringent filtering using biochemical information, but some correct solutions were typically lost at each filtering step. In such cases, filters applied to a large number of close-to-correct solutions are likely to be more successful than application to a few close-to-correct solutions.

Free-energy clusters identify the binding site

A cluster of favorable free energies, such as that shown in Figure 1, implies that a significant volume of parameter space forms a free-energy trap or funnel, which increases the probability of productive binding. The results presented here and related results by others (Harrison *et al.*, 1994; King *et al.*, 1996; Weng *et al.*, 1996; Camacho *et al.*, 1999) demonstrate that clusters of favorable free energies tend to be found at the binding site, so methods that can potentially find large clusters of correct solutions, such as that described here, have a significant value. Analysis of the free-energy grid is especially useful when using unbound coordinates for which correct solutions are not well ranked. The solutions contained in a favorable free-energy cluster can be examined further by more rigorous energy refinement methods that permit limited conformational searching and eliminate grid artifacts (Mitchell *et al.*, 1999; Camacho *et al.*, 2000).

Implementation of the solvent continuum electrostatic model

DOT casts the evaluation of electrostatic energy as the atomic charges of one protein placed in the potential field of another, as opposed to an explicit evaluation of charge–charge interactions. Charged side chains on the protein surface of an independently determined structure are not necessarily oriented as they would be in a complex. Mismatch of two side chains within or near the interface can result in a large unfavorable energy, especially if the charge of each group is localized to a few points. In the solvent continuum electrostatic model used by DOT to describe the potential field of the stationary molecule, the electrostatic potential surrounding each group is modified by the surrounding charged environment, the greater dielectric of the solvent compared with the protein interior and the ionic strength. The net effect is to smooth the charge distribution of the stationary molecule. Inclusion of the continuum electrostatic energy term to the scoring function significantly improved the size of the clusters of correct solutions. Larger numbers of correct answers were found because those solutions with strongly favorable electrostatic energies made substantial contributions to the total energy. Although the computational cost of a solvent continuum electrostatic model is higher than that of a simple Coulombic model, the electrostatic potential is computed only once, requiring only a few minutes on a typical workstation. The cost of computing the electrostatic energy as a convolution product is the same for all models of the potential.

Because the electrostatic potential of the stationary molecule is computed in the absence of the moving molecule, the effective dielectric of the interior of the moving molecule is the same as the surrounding medium (~80). This approximation effectively overdamps the potential field and, if significant, would underestimate electrostatic contributions. Currently, using continuum methods to describe the moving molecule is infeasible. The electrostatic calculation would need to be done for each orientation of the moving molecule and this calculation takes at least ten times as long as performing the complete grid search for a given orientation. The effects of differing dielectrics of protein and solvent may be accounted for by adapting the partial charges of the moving molecule to mimic the potential calculated by continuum methods (Gabdoulline and Wade, 1996). These methods may be particularly useful for transient interactions dominated by electrostatic forces.

DOT successfully docked a variety of complexes

We applied DOT to five very different protein complexes chosen to represent distinct classes of protein–protein interactions (Jones and Thornton, 1996). The potential functions describing shape and electrostatics used by DOT were robust enough to reflect the varied energy contributions to protein association (Table III and Figure 2). For each complex, the van der Waals and electrostatic contributions to the total energy were consistent with experimental studies of the nature of the interaction (Pelletier and Kraut, 1992; Silva *et al.*, 1992; Zheng *et al.*, 1993; Bourne *et al.*, 1995; Mol *et al.*, 1995b; Radic *et al.*, 1997). This was true even though the complexes chosen vary from those strongly dominated by electrostatics to those strongly dominated by shape complementarity.

As others have found (Gabb *et al.*, 1997), using unbound coordinates resulted in poorer ranked correct answers than using bound coordinates. Bound coordinates represent a best-case scenario for rigid-body docking methods since major conformational changes are accounted for and were used here to test the advantages of including the Poisson–Boltzmann electrostatic term in the scoring function. For the UDG–UGI system, coordinates of both bound and unbound UDG were used. Although the rank and r.m.s.d. of the best solution were worse with the unbound coordinates, several solutions close to the crystallographic complex were found (Figure 1A). This demonstrates that our method of 'soft' docking, similar to that used by others but without the small penalty, can accommodate conformational change.

CCP/YCC differs from the other systems studied here in that the electrostatic term dominates the interaction energy. Indeed, free energy clusters could only be obtained for the CCP/YCC system by inclusion of the electrostatic energy term in the scoring function. Both the bound and unbound coordinates gave a free-energy grid that showed strong spatial focusing of YCC to the CCP binding site. This suggests that there may be a shallow energy well governing the interaction that may be important for creating a transient interaction, which is essential for biological efficiency.

Conclusion

This study was undertaken to validate a new approach to docking that incorporates an improved electrostatic treatment and rapid computational techniques for predicting diverse protein–protein interactions. We have found that larger numbers of correct solutions are found with a scoring function consisting of the sum of Poisson–Boltzmann and van der Waals energies rather than of either component term alone. We also found that examination of the free-energy grids allowed identification of the binding site. The fundamental docking problem distinguishes 'false positives' from thermodynamically significant binding sites. The statistical mechanical view hinted at by the DOT results may be significantly more useful than simply examining a few most favorably ranked solutions. Because DOT successfully predicts known complexes, we have begun to apply it to interacting proteins for which the complex structure has not yet been determined. For example, application of DOT to the electron-transfer partners cytochrome *c* oxidase and cytochrome *c* provided a docked complex (Roberts and Pique, 1999) consistent with concurrent mutagenesis, binding and time-resolved kinetic studies (Wang *et al.*, 1999; Zhen *et al.*, 1999). Predicted complexes can be used to direct experimental studies that, in turn, test the predictions and lead to refinement of the computational methods.

Acknowledgements

We thank Drs Elizabeth A.Komives, Adrian Elcock and J.Andrew McCammon for valuable discussions, Dr Garrett M.Morris for critical reading of the manuscript and the McCammon laboratory for AChE partial charge assignments. We thank Drs Thomas Macke and Michel Sanner for AVS modules used in visualization. J.G.M. acknowledges support from an NIH training grant in Growth Regulation and Oncogenesis (NIH/NCI T32 CA09523) and from the La Jolla Interfaces in Science Program funded by the Burroughs Wellcome Fund. The development and distribution of DOT were supported by grants DBI 9616114, DBI 9616115 and DBI 9631436 from the National Science Foundation and grant DE-FG03-96ER62262 from the Department of Energy. DOT may be obtained by sending an E-mail message to ccms-help@sdsc.edu.

References

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

Blom,N.S. and Sygusch,J. (1997) *Proteins: Struct. Funct. Genet.*, **27**, 493–506.

Bourne,Y., Taylor,P. and Marchot,P. (1995) *Cell*, **83**, 503–512.

Bourne,Y., Taylor,P., Bougis,P.E. and Marchot,P. (1999) *J. Biol. Chem.*, **274**, 2963–2970.

Camacho,C.J., Weng,Z., Vajda,S. and DeLisi,C. (1999) *Biophys. J.*, **76**, 1166–1178.

Camacho,C.J., Gatchell,D.W., Kimura,S.R. and Vajda,S. (2000) *Proteins: Struct. Funct. Genet.*, **40**, 525–537.

Davis,M.E., Madura,J.D., Luty,B.A. and McCammon,J.A. (1991) *Comput. Phys. Commun.*, **62**, 187–197.

Fischer,D., Norel,R., Wolfson,H. and Nussinov,R. (1993) *Proteins: Struct. Funct. Genet.*, **16**, 278–292.

Gabb,H.A., Jackson,R.M. and Sternberg,M.J. (1997) *J. Mol. Biol.*, **272**, 106–120.

Gabdoulline,R.R. and Wade,R.C. (1996) *J. Phys. Chem.*, **100**, 3868–3878.

Gilson,M.K. and Honig,B. (1988) *Proteins: Struct. Funct. Genet.*, **4**, 7–18.

Grant,B.D., TsigelnyI., Adams,J.A. and Taylor,S.S. (1996) *Protein Sci.*, **5**, 1316–1324.

Harrison,R.W., Kourinov,I.V. and Andrews,L.C. (1994) *Protein Eng.*, **7**, 359–369.

Hendrickson,W.A., Smith,S.L. and Royer,W.E. (1987) In Burnett,R.M. and Vogel,H.J. (eds), *Biological Organization: Macromolecular Interactions at High Resolution.* Academic Press, New York, pp. 235–244.

Honig,B. and Nicholls,A. (1995) *Science*, **268**, 1144–1149.

Jones,S. and Thornton,J.M. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 13–20.

Katchalski-Katzir,E., Shariv,I., Eisenstein,M., Friesem,A.A., Aflalo,C. and Vakser,I.A. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 2195–2199.

King,B.L., Vajda,S. and DeLisi,C. (1996) *FEBS Lett.*, **384**, 87–91.

Le Du,M.H., Housset,D., Marchot,P., Bougis,P.E., Navaza,J. and Fontecilla-Camps,J.C. (1996) *Acta Crystallogr.*, **D52**, 87–92.

Louie,G.V. and Brayer,G.D. (1990) *J. Mol. Biol.*, **214**, 527–555.

Madura,J.D. *et al.* (1995) *Comput. Phys. Commun.*, **91**, 57–95.

McCoy,A.J., Epa,C. and Colman,P.M. (1997) *J. Mol. Biol.*, **268**, 570–584.

Mitchell,J.C., Phillips,A.T., Rosen,B.J. and Ten Eyck,L.F. (1999) In Istrail,S., Pevzner,P. and Waterman,M. (eds), *Proceedings of the Third International Conference on Computational Molecular Biology (RECOMB 99).* ACM Press, New York, pp. 280–284.

Mol,C.D., Arvai,A.S., Slupphaug,G., Kavli,B., Alseth,I., Krokan,H.E. and Tainer,J.A. (1995a) *Cell*, **80**, 869–878.

Mol,C.D., Arvai,A.S., Sanderson,R.J., Slupphaug,G., Kavli,B., Krokan,H.E., Mosbaugh,D.W. and Tainer,J.A. (1995b) *Cell*, **82**, 701–708.

Nicholls,A., Sharp,K. and Honig,B. (1991) *Proteins: Struct. Funct. Genet.*, **11**, 281–296.

Norel,R., Fischer,D., Wolfson,H.J. and Nussinov,R. (1994) *Protein Eng.*, **7**, 39–46.

Pelletier,H. and Kraut,J. (1992) *Science*, **258**, 1748–1755.

Radic,Z., Kirchhoff,P.D., Quinn,D.M., McCammon,J.A. and Taylor,P. (1997) *J. Biol. Chem.*, **272**, 23265–23277.

Roberts,V.A. and Pique,M.E. (1999) *J. Biol. Chem.*, **274**, 38051–38060.

Roberts,V.A., Freeman,H.C., Olson,A.J., Tainer,J.A. and Getzoff,E.D. (1991) *J. Biol. Chem.*, **266**, 13431–13441.

Shoichet,B.K. and Kuntz,I.D. (1991) *J. Mol. Biol.*, **221**, 327–346.

Silva,M.M., Rogers,P.H. and Arnone,A. (1992) *J. Biol. Chem.*, **267**, 17248–17256.

Stites,W. (1997) *Chem. Rev.*, **97**, 1233–1250.

Strynadka,N.C. *et al.* (1996) *Nature Struct. Biol.*, **3**, 233–239.

Ten Eyck,L.F. (1973) *Acta Crystallogr.*, **A29**, 183–191.

Ten Eyck,L.F., Mandell,J., Roberts,V.A. and Pique,M.E. (1995) In Hayes,A. and Simmons,M. (eds), *Proceedings of the 1995 ACM/IEEE Supercomputing Conference*. ACM Press, New York.

Tsigelny,I., Grant,B.D., Taylor,S.S. and Ten Eyck,L.F. (1996) *Biopolymers*, **39**, 353–365.

Vakser,I.A. (1995) *Protein Eng.*, **8**, 371–377.

Vakser,I.A. (1996) *Biopolymers*, **39,** 455–464.

Vakser,I.A. and Aflalo,C. (1994) *Proteins: Struct. Funct. Genet.*, **20**, 320–329.

Vakser,I.A., Matar,O.G. and Lam,C.F. (1999) *Proc. Natl Acad. Sci. USA*, **96,** 8477–8482.

Wang,J.M., Mauro,M., Edwards,S.L., Oatley,S.J., Fishel,L.A., Ashford,V.A., Xuong,N.H. and Kraut,J. (1990) *Biochemistry*, **29**, 7160–7173.

Wang,K., Zhen,Y., Sadoski,R., Grinnell,S., Geren,L., Ferguson-Miller,S., Durham,B. and Millett,F. (1999) *J. Biol. Chem.*, **274**, 38042–38050.

Weiner,S.J., Kollman,P.A., Case,D.A., Singh,U.C., Ghio,C., Alagona,S., Profeta,S.,Jr and Weiner,J. (1984) *J. Am. Chem. Soc.*, **106**, 765–784.

Weiner,S.J., Kollman,P.A. and Nguyen,D.T. (1986) *J. Comput. Chem.*, **7**, 230–252.

Weng,Z., Vajda,S. and DeLisi,C. (1996) *Protein Sci.*, **5**, 614–626.

Zhen,Y., Hoganson,C.W., Babcock,G.T. and Ferguson-Miller,S. (1999) *J. Biol. Chem.*, **274**, 38032–38041.

Zheng,J., Knighton,D., Ten Eyck,L., Karlsson,R., Xuong,N.H., Taylor,S.S. and Sowadski,J. (1993) *Biochemistry*, **32**, 2154–2161.

Received June 10, 2000; revised October 13, 2000; accepted November 9, 2000