PROTEINS: Structure, Function, and Genetics 39:178–194 (2000)

Protein Docking Using Spherical Polar Fourier Correlations

David W. Ritchie[1,2*] **and Graham J.L. Kemp**[1]
[1] *Department of Computing Science, King's College, University of Aberdeen, Aberdeen, United Kingdom*
[2] *Department of Molecular and Cell Biology, University of Aberdeen, Aberdeen, United Kingdom*

ABSTRACT **We present a new computational method of docking pairs of proteins by using spherical polar Fourier correlations to accelerate the search for candidate low-energy conformations. Interaction energies are estimated using a hydrophobic excluded volume model derived from the notion of "overlapping surface skins," augmented by a rigorous but "soft" model of electrostatic complementarity. This approach has several advantages over former three-dimensional grid-based fast Fourier transform (FFT) docking correlation methods even though there is no analogue to the FFT in a spherical polar representation. For example, a complete search over all six rigid-body degrees of freedom can be performed by rotating and translating only the initial expansion coefficients, many infeasible orientations may be eliminated rapidly using only low-resolution terms, and the correlations are easily localized around known binding epitopes when this knowledge is available. Typical execution times on a single processor workstation range from 2 hours for a global search (5×10^8 trial orientations) to a few minutes for a local search (over 6×10^7 orientations). The method is illustrated with several domain dimer and enzyme–inhibitor complexes and 20 large antibody–antigen complexes, using both the bound and (when available) unbound subunits. The correct conformation of the complex is frequently identified when docking bound subunits, and a good docking orientation is ranked within the top 20 in 11 out of 18 cases when starting from unbound subunits. Proteins 2000;39:178–194.**

Key words: shape complementarity; macromolecular electrostatics; Laguerre polynomials; spherical harmonics

INTRODUCTION

A characteristic feature of biochemical protein reactions is that the protein must first bind to its ligand; either permanently, e.g., during transportation or inhibition, or temporarily, e.g., in catalysis. One of the current challenges in computational biology is to predict reliably whether and how a pair of proteins might associate. This is often referred to as "the docking problem."[1,2] Several crystallographic structures of protein complexes have now been determined, and these frequently exhibit high degrees of steric and chemical complementarity at the protein–protein interface.[3–8] Often, docking algorithms are developed and refined according to their ability to reproduce these known structures, although a more stringent test is to predict the structure of a complex when only the unbound structures of the constituent proteins are known in advance.[9–12]

Protein–protein docking is a hard problem to address computationally, even when one neglects the crucial presence of solvent, principally owing to the large number of atoms and hence degrees of freedom involved. It is currently feasible to use molecular mechanics to refine hypothesized docking orientations,[13] but the computational load is too great for this approach to be used to locate the binding epitopes *ab initio*.[14] For this reason most macromolecular docking algorithms assume rigid bodies, hence limiting the problem to a six-dimensional (6D) search space. However, it is not uncommon for docking methods to test hundreds of thousands[15,16] or many millions[17,18] of distinct rigid-body relative orientations. The most common way to reduce the amount of computation is to perform an initial geometric analysis of each surface to locate significant geometric features such as cavities,[2] local knobs and holes[19] and their associated surface normals.[16] These approaches reduce the complexity of the problem to a combinatorial search over a relatively small number of complementary surface features, which can be performed quite quickly using, e.g., geometric hashing.[15,20] However, such geometric methods generally require a post-processing step to remove sterically prohibited orientations.[16,21] In contrast, the Fourier correlation approach[22] uses a simple Cartesian grid model of protein topology which favors close contacts and automatically penalizes steric clashes. Calculating the degree of overlap of a pair of grids at successive translational increments can be performed very efficiently using a fast Fourier transform (FFT) technique.[22] However, high-resolution Fourier correla-

Abbreviations: CPU, central processor unit; FFT, fast Fourier transform; PDB, protein data bank; RMS, root mean squared; Mb, megabyte; VH, antibody variable heavy chain; VL, antibody variable light chain; 3D, three-dimensional; 6D, six-dimensional.

Grant sponsor: BBSRC; Grant number: 1/B10454.

*Correspondence to: David W. Ritchie, Department of Computing Science, King's College, University of Aberdeen, Aberdeen AB24 3UE, United Kingdom. E-mail: dritchie@csd.abdn.ac.uk

Received 8 July 1999; Accepted 4 November 1999

tions can still take many CPU-hours,[22] or even CPU-days.[18]

Nonetheless, the Fourier approach is attractive because in addition to correlating surface shapes, it may also be used to correlate other surface properties such as hydrophobicity[23] or to accelerate the calculation of electrostatic and van der Waals force field models of protein–small-ligand systems.[24,25] Another attractive feature of this method is that one can perform fast low-resolution searches, for example, to estimate main-chain complementarity[26] and to locate a ligand near a receptor-binding site.[27] In other words, there is an inherent degree of "softness" in a Fourier correlation which can be controlled by the number of terms employed. Even in high-resolution correlations this softness may be useful for accommodating small conformational changes that might be expected when docking unbound protein structures.[18]

The main disadvantages of existing Fourier docking methods are that a large grid is required to accommodate translations of one molecule about the second stationary molecule and that a new FFT must be calculated for each rotational increment of the stationary molecule. Hence, when docking large proteins such as antibody–antigen systems, the sampling resolution is limited to around 1 Å cubes for the problem to fit into main memory,[18] and accurate coverage of rotational space adds a significant overhead to a 6D docking calculation. Some progress has been made to alleviate the rotational search problem by using hydrogen bond filters[28] and surface segmentation techniques.[29] However, we feel that a much more substantial revision of the method is required to overcome these fundamental problems.

Here, we attempt to exploit the known computational advantages of Fourier-based approaches and to address the limitations described above. In our method, each protein surface shape is represented by a "double skin" model that describes thin regions of space exterior and interior to the molecular surface. The use of analytical surface skins in relation to molecular superposition has been described previously,[30] but this approach would be extremely expensive if applied to the protein–protein docking problem. Hence in order to avoid the geometrical complexities of manipulating explicit surface shapes, we represent each skin as a Fourier series expansion of real orthogonal radial and spherical harmonic basis functions. These functions are similar to those found in the solution to Schrödinger's equation for the hydrogen atom,[31] but are specially scaled to the dimensions of typical protein domains. In this surface skin model, the shape complementarity of a given docking orientation is scored by evaluating the degree of overlap between opposing pairs of interior and exterior molecular skins.

Using spherical harmonic functions to represent protein surface shape is not in itself new.[32–35] However, with the exception of Duncan and Olson's novel surface mapping approach,[11,36] it seems that the usefulness of this type of representation in the docking problem has been limited by the absence of a radial component. The introduction of special purpose radial functions, described here, allows arbitrary (e.g., re-entrant) shapes to be modeled and is crucial to the development of a 6D Fourier correlation model for macromolecular docking. The main reason for choosing to use spherical harmonic functions is that they transform among themselves under rotation.[37] From this, it follows that a rotation transforms the coefficients of a spherical polar parametrization in a predictable manner. In order to exploit this property, the 6 degrees of freedom in the rigid-body search space are divided into 5 Euler rotation angles and an intermolecular distance. Thus, apart from varying the intermolecular separation, the two molecules remain at fixed positions in space and are rotated about their own centroids. Consequently, this approach resolves the main disadvantages of existing FFT methods because the expansion coefficients need only be determined once, a large Cartesian grid is not required, and, during a docking search, both the search step size and the resolution of the correlation may be varied independently. Currently, a drawback with this approach is that it involves the calculation of many two-center overlap integrals.[38] However, these are independent of the protein identities and so may be calculated just once and stored on disc for subsequent use.

In this article, we describe the construction of parametric surface skins using real spherical polar basis functions. As the use of such functions for protein shape representation is novel, a brief summary of their properties is also provided. We then give a description of the algebraic manipulations necessary to develop an efficient search for docking orientations by incrementally rotating and translating the parametric representations. We also show that this spherical polar approach provides a natural way to model macromolecular electrostatic complementarity.

Although our docking algorithm is presented in terms of a blind search, often the location of one (or both) of the binding sites is known in advance. This knowledge can be used, e.g., in the form of distance constraints,[18] to help reduce the number of "false-positive" solutions obtained, particularly when attempting to predict a docking orientation using unbound subunits. When our spherical polar correlation is used predictively we also find that it helps to constrain the search space, but here specific distance constraints are not used. Instead, we attempt to let good solutions emerge relatively unaided by restricting the search to the region(s) of interest using only a simple constraint on one (or two) of the angular degrees of freedom. These angular constraints are natural ones if the task is to investigate a hypothesized binding orientation. For example, using our program, *Hex,* it is straightforward to maneuvre a pair of proteins into an approximate docking orientation and to search rapidly millions of trial docking modes local to the given starting position. This is possible because our spherical polar docking expression can be evaluated selectively (in contrast to Cartesian FFT methods which produce 3D arrays of translational correlation scores[22]), using only those orientations that fall within the region of interest. Additionally, because most trial orientations are trivially infeasible, many of them may be eliminated by performing a low-resolution scan of

the search space: Only a relatively small number of surviving orientations need be evaluated at high resolution. Hence, execution times can often be reduced to a matter of minutes.

The docking algorithm is demonstrated using the structures of 30 protein complexes taken from the Protein Data Bank (PDB).[39] These include two domain dimers, eight enzyme–inhibitor systems, and 20 antibody complexes. In 24 cases the known structure is ranked within the top ten orientations (18 are ranked first) in a global 6D search of some 5×10^8 trial orientations. For 18 of the complexes, one or both of the protein structures have been determined crystallographically in the unbound conformation, and in these cases we attempt to predict the orientation of the complex starting from randomly oriented unbound subunits. The rank obtained depends largely on the degree of conformational change induced by binding: By constraining the ligand to tumble over the receptor binding site, a good solution is often found within the top few hundred orientations. By further focusing the search around the ligand epitope, the rank of the correct orientation can be improved to within the top 20 in 11 of the 18 cases.

METHODS

Theory

In this article protein shape and electrostatic properties are represented using series expansions of orthonormal spherical polar basis functions. For example, letting $A(\underline{r})$ represent an arbitrary function in 3D space we have

$$A(\underline{r}) = \sum_{nlm}^{N} a_{nlm} R_{nl}(r) y_{lm}(\theta, \phi); N \geq n > l \geq |m| \geq 0 \quad (1)$$

where a_{nlm} are the expansion coefficients, to be determined, and N is the order of the expansion: for docking, we use $N \leq 25$. The functions $y_{lm}(\theta, \phi) = \vartheta_{l|m|}(\cos\theta)\varphi_m(\phi)$ are real spherical harmonics; these functions have been described extensively elsewhere.[37,40,41] The radial functions, $R_{nl}(r)$, are based on the generalized Laguerre polynomials, $L_q^{(\alpha)}(\rho)$.[42] For surface shape representations we use shape-scaled radial functions, $S_{nl}(r)$, of the form

$$S_{nl}(r) = \left[\left(\frac{2}{k^{3/2}}\right)\frac{(n-l-1)!}{\Gamma(n+1/2)}\right]^{1/2} e^{-\rho/2}\rho^{l/2}L_{n-l-1}^{(l+1/2)}(\rho) \quad (2)$$

where the square root term is a normalization factor, and ρ is a scaled distance, $\rho = r^2/k$, with scaling parameter $k = 20$. This choice of radial scaling ensures that most of the zeros of the radial function fall within about 30Å of the origin. Effectively, these functions are scaled to the dimensions of typical globular protein domains. For half-integral arguments, the gamma function may be evaluated using the identity $\Gamma(n + 1/2) = \sqrt{\pi}(1/2)_n$ where $(x)_n = x.(x + 1)\ldots(x + n - 1)$ is a rising factorial and $(x)_0 \equiv 1$. Electrostatic potentials are represented using unscaled radial functions, $V_{nl}(r)$:

$$V_{nl}(r) = \left[(2\Lambda)^3\frac{(n-l-1)!}{(n+l+1)!}\right]^{1/2} e^{-\rho/2}\rho^{l}L_{n-l-1}^{(2l+2)}(\rho) \quad (3)$$

where $\rho = 2\Lambda r$, with scale factor $\Lambda = 1/2$. In quantum mechanics the functions $S_{nl}(r)$ and $V_{nl}(r)$ correspond to the radial eigenfunctions of the harmonic oscillator and certain Coulomb potential problems, respectively.[43]

Expansions such as equation 1 are useful in the rigid body docking problem for several reasons. First, given some function, $A(\underline{r})$, which might be available only as discrete samples, the orthonormality property allows the expansion coefficients to be determined easily. Multiplying both sides of equation 1 by $R_{n'l'}(r)y_{l'm'}(\theta, \phi)$ and integrating gives an expression for each coefficient:

$$a_{n'l'm'} = \int A(\underline{r})R_{n'l'}(r)y_{l'm'}(\theta, \phi)\, dV. \quad (4)$$

An expansion to order $N = 25$ involves the calculation of $N(N + 1)(2N + 1)/6 = 5{,}525$ such integrals.

The second property we wish to exploit is that the series representation of $A(\underline{r})$ may be rotated by transforming only the expansion coefficients. For example, applying a rotation to each side of equation 1 gives

$$\hat{R}(\alpha, \beta, \gamma)A(\underline{r}) = A'(\underline{r}) = \sum_{nlm}^{N} a'_{nlm}R_{nl}(r)y_{lm}(\theta, \phi) \quad (5)$$

and it can be shown that the rotated coefficients, a'_{nlm}, are related to the unrotated coefficients, a_{nlm}, by

$$a'_{nlm} = \sum_{m'=-l}^{l} a_{nlm'}R^{(l)}_{mm'}(\alpha, \beta, \gamma). \quad (6)$$

Each rotation matrix, $R^{(l)}$, is a real $(2l + 1) \times (2l + 1)$ matrix whose elements are functions of the Euler rotation angles, (α, β, γ). Equation 6 follows directly from the special rotational properties of the spherical harmonic functions.[37,41] The real rotation matrix elements, $R^{(l)}_{mm'}$, may be derived from the complex Wigner D-matrices as described previously.[35] Because there are N distinct values of l, it can be seen that a series expansion to order N can be rotated using just N rotation matrices, with each rotation matrix, $R^{(l)}$, being used $N - l$ times. For rotations about only the z-axis ($\beta = \gamma = 0$), equation 6 reduces to

$$a'_{nlm} = a_{nlm}\cos m\alpha + a_{nl\bar{m}}\sin \bar{m}\alpha. \quad (7)$$

It can also be shown that translating $A(\underline{r})$ by Δz along the positive z-axis may be represented as

$$\hat{T}(\Delta z)A(\underline{r}) = A''(\underline{r}) = \sum_{nlm}^{N} a''_{nlm}R_{nl}(r)y_{lm}(\theta, \phi) \quad (8)$$

where the translated coefficients, a''_{nlm}, are given by an expression of the form

$$a''_{nlm} = \sum_{n'l'm'}^{N} a_{n'l'm'}T^{(|m|)}_{nl,n'l'}(\Delta z)\delta_{mm'} \quad (9)$$

in which each $T^{(|m|)}_{nl,n'l'}(\Delta z)$ represents a translation matrix element appropriate for the chosen radial basis. It may be noted that taking the transpose of the translation matrix gives

$$a''_{nlm} = \sum_{n'l'm'}^{N} a_{n'l'm'} T^{(|m|)}_{n'l',nl}(\Delta z)\delta_{mm'} \tag{10}$$

which corresponds to a translation of $A(\underline{r})$ along the negative z-axis. However, since a direct calculation of the translation matrix elements is somewhat involved,[44,45] we currently estimate them numerically, as outlined below.

Finally, representing 3D functions as orthonormal expansions provides a straightforward way of calculating the correlation, or degree of overlap, between pairs of functions. Even after rotating and translating the original functions, the correlation has the form

$$\int A'(\underline{r})B''(\underline{r})dV = \sum_{nlm}^{N} a'_{nlm}b''_{nlm} = \underline{a}'.\underline{b}''. \tag{11}$$

Thus, in the present case, our aim is to reduce the task of evaluating candidate docking orientations in a rigid-body search to one of calculating scalar products of suitably rotated and translated coefficient vectors.

Correlating Surface Skins

In order to exploit the above properties, protein surfaces are represented by a "double skin" shape model. This is derived from a dot representation of the familiar molecular surface,[46] defined by rolling a probe sphere over the van der Waals surface of the molecule. The first, exterior, skin is defined as the volume bounded by the molecular surface and the solvent-accessible surface (which is offset from the molecular surface by the radius of the probe sphere). Here, these surfaces are estimated using our own algorithm[47] which is an adaptation of Shrake and Rupley's[48] method of calculating solvent-accessible dot surfaces. It is convenient to define the second, interior, skin as the union of the van der Waals volumes of all atoms just inside the molecular surface. Both skins are represented as density functions, $\sigma(\underline{r})$ and $\tau(\underline{r})$, respectively, defined to have a value of unity inside the skin and zero everywhere else:

$$\sigma(\underline{r}) = \begin{cases} 1; & \underline{r} \in \text{exterior skin} \\ 0; & \text{otherwise,} \end{cases} \tag{12}$$

$$\tau(\underline{r}) = \begin{cases} 1; & \underline{r} \in \text{surface atom} \\ 0; & \text{otherwise.} \end{cases} \tag{13}$$

Approximating these density functions as series expansions to order N gives, for example,

$$\sigma(\underline{r}) = \sum_{nlm}^{N} a^{\sigma}_{nlm} S_{nl}(r) y_{lm}(\theta, \phi). \tag{14}$$

The skin density functions for each protein are estimated by projecting skin sample points onto a 3D Cartesian grid, represented as bits in a 3D byte array. A grid spacing of 0.75Å requires an array of about 10^6 elements (1Mb) to sample the skins of typical protein domains. Each protein's coordinate origin is taken as the "center of mass" of all heavy atoms. Skin sample points are generated by centering a small test sphere, with a diameter of the desired skin thickness, on each molecular dot surface normal so that at each position the sphere just touches the molecular surface. A local Cartesian grid of $(0.2\text{Å})^3$ cells is centered on the test sphere, and any local grid cell whose center lies inside the test sphere is taken as a sample point to be projected onto the main grid. Many sample points may map to the same main grid cell, although only non-empty grid cells are considered in the following integration step. The coefficient integrals (equation 4) are estimated using

$$a^{\sigma}_{nlm} \simeq \sum_{c} S_{nl}(r_c) y_{lm}(\theta_c, \phi_c)\Delta V \tag{15}$$

where the summation is over all non-zero bits in the grid, ΔV is the cell volume and (r_c, θ_c, ϕ_c) are the spherical polar coordinates of the centre of the c^{th} cell. The interior skin is sampled in a similar manner, this time by centering local 0.2Å grids on each surface atom. The centre of each local grid cell is considered within the interior skin if that point falls within the van der Waals volume of the atom. The basis functions are evaluated at each non-zero grid cell using standard recursion formulae.[35,42] Figure 1 shows the resolution with which the skin density functions are encoded when using various expansion orders, N. This figure shows that atom-scale features (i.e., 3Å features) begin to be resolved at around $N = 16$, although significantly higher-order expansions are required to properly distinguish the mutually exclusive exterior and interior skin volumes. As there is relatively little visible improvement on going from $N = 25$ to $N = 30$, an upper limit of $N = 25$ was chosen to give a reasonable compromise between speed and accuracy in the following docking calculations.

The use of a surface skin representation to model protein shape complementarity is justified by considering Figure 2. This figure suggests that a good strategy for finding complementary orientations between a pair of proteins is to maximize the overlap between the interior skin of one protein with the exterior skin of the other. Steric clashes may be penalized with an interior–interior skin overlap penalty term. Using these ideas, the shape complementarity score, S (in Å^3 units), for proteins A and B is written as

$$S = \int \sigma_A \tau_B dV + \int \tau_A \sigma_B dV - Q \int \tau_A \tau_B dV \tag{16}$$

where $\sigma_A = \sigma_A(\underline{r}_A)$ and $\tau_B = \tau_B(\underline{r}_B)$, etc., and Q is a steric penalty factor: here, $Q = 12$. With a skin thickness of 1.4Å, the first two terms give an expression for the volume of solvent expelled from the protein surfaces upon association (see Fig. 2). With a suitable scale factor, this expelled volume can be used as a first-order approximation to the hydrophobic free energy of association.[49]

Substituting the appropriate series expansion for each

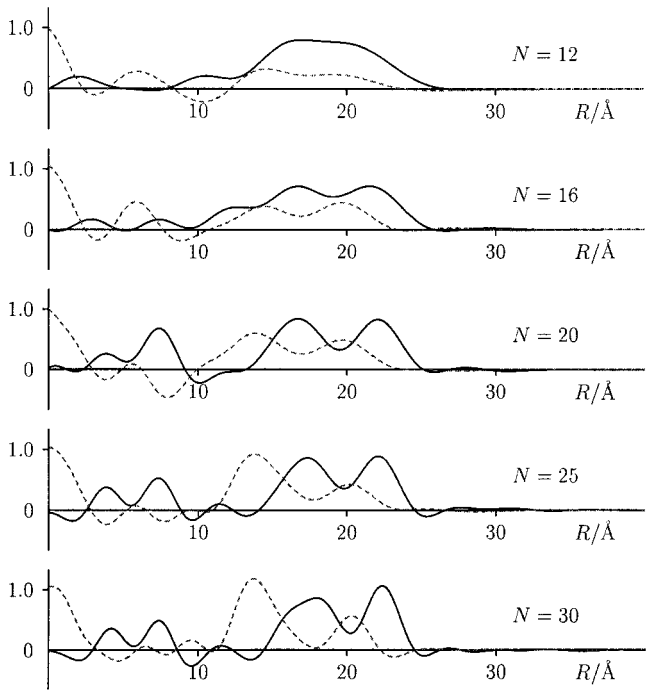

Fig. 1. Plots of the exterior and interior skin density functions σ(R) and τ(R) (solid and dashed lines, respectively) of lysozyme in the antibody HyHel-5–lysozyme complex (3HFL) as a function of distance along the intermolecular axis, calculated at several expansion orders, N. In the sequence N = 12, 16, 20, 25, 30, the total number of shape coefficients for each density function approximately doubles at each step (with 650, 1,496, 2,870, 5,525, and 9,455 coefficients, respectively). Here, lysozyme is centered on the origin, and HyHel-5 (not shown) is located on the positive R axis. The lysozyme center of mass (R = 0) lies between LEU56:C_β and LEU56:$C_{\delta 2}$, and the intermolecular axis passes through ARG45:$C_{\eta 1}$ near R = 20 causing the double peak around R = 17 and R = 23 (i.e., the axis cuts a reentrant region of the molecular surface in these plots). At the origin, the initial surface sampling algorithm finds LEU56:C_β and LEU56:$C_{\delta 2}$ to be accessible to a solvent probe, giving an interior skin density of nearly unity and an exterior skin density of zero, as observed.

skin density function into the above overlap expression (equation 16) gives

$$S = \sum_{nlm}^{N} \sum_{n'l'm'}^{N} (a^{\sigma}_{nlm} b^{\tau}_{n'l'm'} + a^{\tau}_{nlm} b^{\sigma}_{n'l'm'} - Q a^{\tau}_{nlm} b^{\tau}_{n'l'm'}) I_{nn'll'mm'} \quad (17)$$

and on writing $b^{Q}_{nlm} = b^{\sigma}_{nlm} - Q b^{\tau}_{nlm}$ this can be simplified to

$$S = \sum_{nlm}^{N} \sum_{n'l'm'}^{N} (a^{\sigma}_{nlm} b^{\tau}_{n'l'm'} + a^{\tau}_{nlm} b^{Q}_{n'l'm'}) I_{nn'll'mm'}. \quad (18)$$

The factor $I_{nn'll'mm'}$ on the right-hand side arises from the residual overlap of pairs of basis functions centered on local coordinate systems, $\underline{r}_A$ and $\underline{r}_B$:

$$I_{nn'll'mm'} = \int S_{nl}(r_A) \vartheta_{l|m|}(\cos\theta_A) \varphi_m(\phi_A) S_{n'l'}(r_B) \vartheta_{l'|m|'}(\cos\theta_B) \times \varphi_{m'}(\phi_B) dV. \quad (19)$$

These integrals may be simplified by aligning the intermolecular axis with the global z-axis and by changing variables to a prolate spheroidal coordinate system.[38] This gives $\phi_A = \phi_B = \phi$ so that the circular functions, $\varphi_m(\phi)$, can be integrated out. It can then be shown that the remaining terms may all be expressed as functions of the intermolecular separation, R. Hence

$$I_{nn'll'mm'} = K_{nn'll'|m|}(R) \delta_{mm'}. \quad (20)$$

However, owing to the large numbers of terms in a prolate spheroidal expansion, calculating the $K(R)$ integrals analytically is not feasible; hence, they are estimated by numerical integration in the (r, θ) plane and stored on disc for subsequent use.[47] Storing all $K(R)$ integrals to $N = 25$ for R between 1Å and 50Å in 1Å increments requires around 55 Mb of disc space.

Composing Transformations

Although equation 18 could be used directly to evaluate the complementarity score for candidate docking orientations, it is much more efficient to compose orientations by rotating and translating each protein incrementally. Initially, both proteins are assumed to share a common coordinate system, and the docking search is performed as a nested sequence of rotation and translation operations. The first four rotational degrees of freedom are taken as Euler rotation angles, (β_1, γ_1) and (β_2, γ_2), calculated from the angular coordinates of the vertices of a pair of tessellated icosahedra. The remaining axial rotation, α_2, is assigned to the ligand. Conceptually, at each intermolecular distance all pairs of vertices are rotated in turn onto the intermolecular axis, followed by a search over the twist angle, α_2. This is illustrated in Figure 3. Translations are calculated using the $K(R)$ overlap integrals. For example, expanding the first integral in equation 16 and collecting terms with unprimed subscripts gives

$$\int \sigma_A \tau_B dV = \sum_{n'l'm'}^{N} \Big(\sum_{nlm}^{N} a^{\sigma}_{nlm} K_{nn'll'|m|}(R) \delta_{mm'} \Big) b^{\tau}_{n'l'm'} = \sum_{n'l'm'}^{N} a''^{\sigma}_{n'l'm'} b^{\tau}_{n'l'm'}. \quad (21)$$

But this could equally be calculated by first collecting primed subscripts:

$$\int \sigma_A \tau_B dV = \sum_{nlm}^{N} a^{\sigma}_{nlm} \Big(\sum_{n'l'm'}^{N} K_{nn'll'|m|}(R) \delta_{mm'} b^{\tau}_{n'l'm'} \Big) = \sum_{nlm}^{N} a^{\sigma}_{nlm} b''^{\tau}_{nlm}. \quad (22)$$

Comparing these expressions with equations 9 and 10 shows that the $K(R)$ overlap integrals are formally equivalent to the translation matrices, $T(\Delta z)$. Thus, the coefficient transformations are as follows:

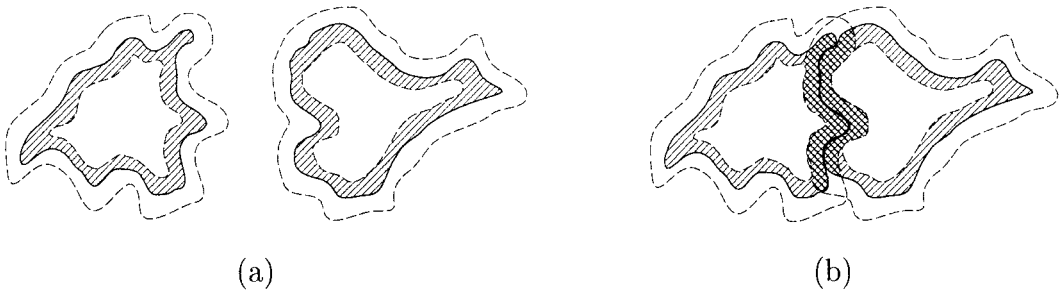

Fig. 2. A cartoon illustration of protein shape complementarity with the "double skin" model. The exterior skin is the volume bounded by the solvent-accessible surface (dashed lines) and the molecular surface (solid lines). Shaded regions represent interior skins. On moving from the orientation in (**a**) to that of (**b**) so as to maximize the degree of overlap between respective pairs of interior and exterior skins, an inevitable consequence is to bring the two surfaces into very close contact. The central hatched area shows the solvent-accessible volume occluded on association.

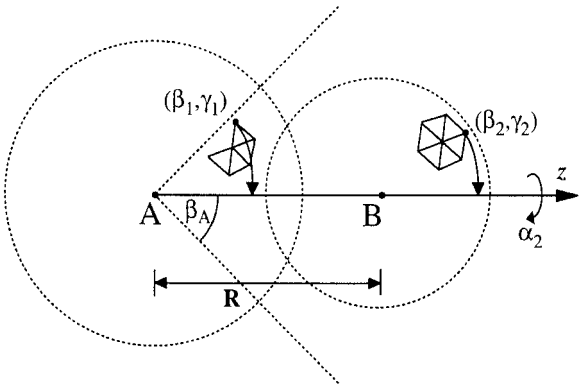

Fig. 3. Schematic illustration of a six-dimensional docking search using tessellated icosahedra. A and B label the receptor and ligand local coordinate origins, respectively. The angular coordinates of each tessellation vertex provide molecular rotational increments (β, γ). If desired, the search may be localized to a known binding site on protein A, assumed to be centered on the z-axis, by calculating the correlation only for orientations where $\beta_1 \leq \beta_A$. A similar constraint may also be applied to β_2 if the ligand epitope is known.

1. Rotate the receptor (protein A):

$$a'_{nlm}\sigma = \sum_{m'=-l}^{l} a^{\sigma}_{nlm'} R^{(l)}_{mm'}(0, \beta_1, \gamma_1). \tag{23}$$

2. Translate the receptor along the negative z-axis:

$$a''_{nlm}\sigma = \sum_{n'l'}^{N} a'^{\sigma}_{n'l'm} K_{n'nl'l|m|}(R). \tag{24}$$

3. Rotate the ligand (protein B):

$$b'^{Q}_{nlm} = \sum_{m'=-l}^{l} b^{Q}_{nlm'} R^{(l)}_{mm'}(0, \beta_2, \gamma_2). \tag{25}$$

4. Twist the ligand about the z-axis:

$$b''^{Q}_{nlm} = b'^{Q}_{nlm} \cos m\alpha_2 + b'^{Q}_{nl\bar{m}} \sin \bar{m}\alpha_2. \tag{26}$$

Similar operations are applied to the remaining coefficient vectors, $\underline{a}^{\tau}$ and $\underline{b}^{\tau}$.

Rather than calculating coefficient scalar products explicitly, the amount of computation can be reduced by substituting the above-transformed coefficients into equation 16 and collecting coefficients of α_2 to give

$$S = \sum_{m=-L}^{L} Q^{+}_{m} \cos m\alpha_2 + Q^{-}_{m} \sin \bar{m}\alpha_2; L = N - 1 \tag{27}$$

where $S \equiv S(R, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2)$ and

$$Q^{+}_{m} \equiv Q^{+}_{m}(R, \beta_1, \gamma_1, \beta_2, \gamma_2) = \sum_{nl}^{N} (a''^{\sigma}_{nlm} b'^{\tau}_{nlm} + a''^{\tau}_{nlm} b'^{Q}_{nlm}) \tag{28}$$

$$Q^{-}_{m} \equiv Q^{-}_{m}(R, \beta_1, \gamma_1, \beta_2, \gamma_2) = \sum_{nl}^{N} (a''^{\sigma}_{nlm} b'^{\tau}_{nl\bar{m}} + a''^{\tau}_{nlm} b'^{Q}_{nl\bar{m}}). \tag{29}$$

Finally, using the identities $\cos \bar{m}\phi = \cos m\phi$ and $\sin \bar{m}\phi = -\sin m\phi$, equation 27 may be simplified to obtain a real Fourier series in α_2:

$$S = Q^{++}_{0} + \sum_{m=1}^{L} Q^{++}_{m} \cos m\alpha_2 + Q^{--}_{m} \sin m\alpha_2 \tag{30}$$

for which the coefficients are given by

$$Q^{++}_{m} = Q^{+}_{m} + Q^{+}_{\bar{m}}(1 - \delta_{m0}) \tag{31}$$

and

$$Q^{--}_{m} = Q^{-}_{\bar{m}} - Q^{-}_{m}. \tag{32}$$

In contrast to the discrete correlations of the FFT method, the above Fourier series (equation 30) gives the shape complementarity score as a continuous function of all six rigid-body degrees of freedom (although discrete steps in R must be taken when using precalculated overlap integrals). For a given partial orientation, (R, β_1, γ_1, β_2, γ_2), the docking score at successive twist increments, α_2, can be calculated extremely rapidly using equation 30. Indeed, on contemporary workstations, it is feasible to store in memory precalculated vectors of rotated coeffi-

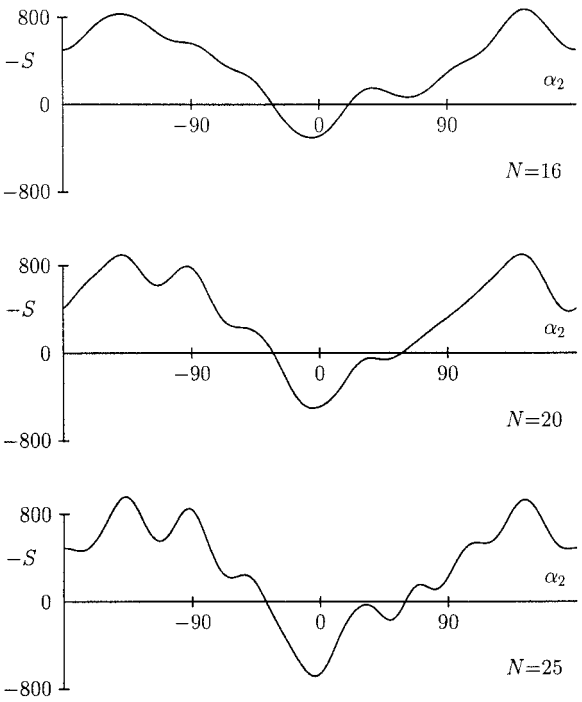

Fig. 4. The shape complementarity score, S (equation 30, in KJ/mol units), as a function of the twist angle, α_2, for the HyHel-5–lysozyme complex shown at increasing expansion orders, N. The minimum near the center of the plots ($\alpha_2 = 0$) corresponds to the crystallographic orientation.

cients for each rotational increment of the ligand, along with arrays of cos $m\alpha_2$ and sin $m\alpha_2$. Thus, the innermost two cycles of a docking search require only $O(N^2)$ operations to update the Q_m coefficients (equations 28 and 29) and just $O(N)$ operations to evaluate the correlation at successive twist increments (equation 30). Plots of the shape complementarity score (equation 30, multiplied by a negative constant described below) for the HyHel-5–lysozyme complex are shown in Figure 4. These plots show that the shape correlation function produces a strong signal for the correct crystallographic orientation even when only low-resolution (e.g., $N = 16$) terms are used.

Correlating Electrostatics

As suggested in the Introduction, spherical polar correlations provide a natural way to model electrostatic complementarity. Proper treatment of solvent effects is difficult (for reviews, see e.g., refs. 50 and 51), so in this initial development an in vacuo electrostatic model is used. Because we wish to evaluate electrostatic interactions across (often extensive) protein–protein interfaces, assuming an isotropic medium is, at least in part, justified. Classically, the electrostatic energy of a charge distribution, $\rho(\underline{r})$, under the influence of a potential, $\phi(\underline{r})$, is given by[52]

$$E = \frac{1}{2}\int \rho(\underline{r})\phi(\underline{r})dV. \quad (33)$$

Writing $\rho(\underline{r}) = \rho_A(\underline{r}_A) + \rho_B(\underline{r}_B)$ and $\phi(\underline{r}) = \phi_A(\underline{r}_A) + \phi_B(\underline{r}_B)$, and representing each function as a spherical polar expansion (using the $V(r)$ radial functions, equation 3) immediately gives an expression for the electrostatic interaction energy of a pair of proteins:

$$E(R, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2) = \frac{1}{2}\sum_{nlm}^{N}\sum_{n'l'm'}^{N}(a'^{\rho}_{nlm}b'^{\phi}_{n'l'm'} + a'^{\phi}_{nlm}b'^{\rho}_{n'l'm'}) \times J_{nn'll'|m|}(R)\delta_{mm'} \quad (34)$$

where a'^{ρ}_{nlm} and a'^{ϕ}_{nlm} denote rotated coefficients of the charge density and electrostatic potential expansions of protein A, etc., and where $J(R)$ is the matrix of overlap integrals calculated in the $V(r)$ basis. With an infinite number of terms equation 34 would be *exact* although, as before, the expansion is truncated at $N \leq 25$ to give a "soft" electrostatic correlation.

The charge density coefficients for each protein are calculated by equating a series expansion to the classical expression[52] for the charge density due to a distribution of point charges, q_i, located at positions $\underline{x}_i \equiv \underline{r}_i$:

$$\rho(\underline{r}) = \sum_i q_i\delta(\underline{x} - \underline{x}_i) = \sum_{n'l'm'}^{N} a^{\rho}_{n'l'm'}V_{n'l'}(r)y_{l'm'}(\theta, \phi) \quad (35)$$

where $\delta(\underline{x})$ is the Dirac delta function in three dimensions. Multiplying both sides of equation 35 by $V_{nl}(r)y_{lm}(\theta, \phi)$ and integrating gives the remarkably simple result:

$$a^{\rho}_{nlm} = \sum_i q_iV_{nl}(r_i)y_{lm}(\theta_i, \phi_i). \quad (36)$$

The expansion coefficients for the potential are calculated from the charge density coefficients by solving Poisson's equation:

$$\nabla^2\phi(\underline{r}) = -4\pi\rho(\underline{r}). \quad (37)$$

Substituting the series expansion for each side, applying ∇^2 to the basis functions, multiplying both sides of the result by $V_{n'l'}(r)y_{l'm'}(\theta, \phi)$, and integrating gives

$$\sum_{n=l+1}^{N} a^{\phi}_{nlm}\int_0^{\infty}(V''_{nl}(r) + 2V'_{nl}(r)/r - l(l+1)V_{nl}(r)/r^2)V_{n'l}(r)r^2dr = -4\pi a^{\rho}_{n'lm} \quad (38)$$

where V' denotes $\partial V/\partial r$, etc. Integrating by parts the term in $V'_{nl}(r)$ gives

$$\sum_{n=l+1}^{N} a^{\phi}_{nlm}G^{(l)}_{nn'} = -4\pi a^{\rho}_{n'lm} \quad (39)$$

where the elements of $G^{(l)}$ have the symmetric form

$$G^{(l)}_{nn'} = -\int_0^{\infty}(V'_{nl}(r)V'_{n'l}(r)r^2 + l(l+1)V_{nl}(r)V_{n'l}(r))\,dr. \quad (40)$$

It can be seen that for each l (and m), equation 39 represents a set of simultaneous equations in the coeffi-

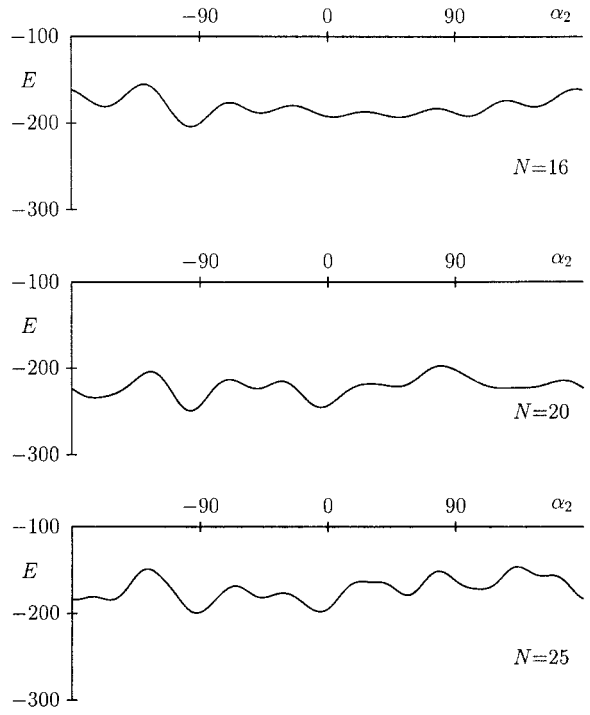

Fig. 5. Electrostatic interaction energy, E (equation 34, in KJ/mol units), as a function of the twist angle, α_2, for the HyHel-5–lysozyme complex shown at increasing expansion orders, N. The minimum near the center of the plots ($\alpha_2 = 0$) corresponds to the crystallographic orientation.

cients, $a^{\Phi}_{n'lm}$, which can be determined by inverting each $G^{(l)}$ matrix. The elements of $G^{(l)}$ may be calculated by direct manipulation of the series expansion for $V_{nl}(r)$. Using

$$L_q^{(\alpha)}(\rho) = \sum_{k=0}^{q} \binom{q+\alpha}{q-k} \frac{(-\rho)^k}{k!} \tag{41}$$

and denoting by C_{nlk} the coefficients of ρ in the expansion of $L^{(2l+2)}_{n-l-1}(\rho)$, one obtains after some working

$$G^{(l)}_{nn'} = -\Lambda^2 \left[\frac{(n-l-1)!}{(n+l+1)!}\frac{(n'-l-1)!}{(n'+l+1)!}\right]^{1/2} \times \sum_{k=0}^{n-l-1}\sum_{k'=0}^{n'-l-1} C_{nlk}C_{n'lk'}(2l+k+k')! \times (2(2l+1)(l+1)+k+k'-(k-k')^2). \tag{42}$$

It should be noted that with large numbers of charges the above method of calculating the potential as an expansion about a single point is not especially accurate. Greater accuracy can be obtained using so-called tree code fast multipole methods.[53,54] Nevertheless, since equation 34 can also be rearranged into a Fourier series in α_2 (cf. equation 30), this approach provides an extremely efficient way to estimate macromolecular electrostatic complementarity at arbitrary orientations. Figure 5 shows the variation of the electrostatic interaction energy (equation 34) as a function of α_2 for the HyHel-5–lysozyme complex, calculated using all polar atoms. Although the minima are less marked than in the steric case (Fig. 4), this figure clearly shows that our electrostatic correlation can help to identify the crystallographic orientation.

The Docking Correlation

The electrostatic energy, E, and the steric complementarity score, S, may be combined to give a pseudo energy for the complex (in KJ/mol units) using

$$E_{\text{total}} = \left(\frac{1391.4}{K_R}\right) E + K_H S. \tag{43}$$

Here, the relative permittivity, K_R, and hydrophobic free energy factor, K_H, are treated as adjustable parameters: $K_R = 8$ is used to approximate the electrostatic energy calculated explicitly for the HyHel-5–lysozyme complex,[55] and $K_H = -0.8$ KJ/mol/Å^3 was chosen empirically to produce a reasonable weighting of the two contributions to E_{total}. These values tend to overestimate the absolute binding energy for most complexes, although only relative energies are needed to distinguish different docking orientations.

As it is time-consuming to evaluate equation 43 at high resolution for every orientation in a docking search, the calculations described below were accelerated by performing an initial low-resolution ($N = 16$) filtering scan of the search space. This involves generating partial orientations (R, β_1, γ_1, β_2, γ_2) for the complex and recording the lowest energy found after searching over the twist angle, α_2. Any partial orientation that gives a positive energy ≥ 100 KJ/mol is immediately discarded. The top 25% of the surviving orientations (subject to a maximum of 20,000) are then passed to a final high-resolution scoring stage which uses a simple "peak picking" algorithm to locate all local minima in each twist angle search. Because the initial scan is considered primarily as a proximity test, the electrostatic contribution is calculated only in the final stage at $N = 25$.

RESULTS

Our docking algorithm was applied to the known structures of two protein domain dimers, eight enzyme–inhibitor complexes, and 20 antibody complexes, taken from the PDB. These are listed in Table I. In crystal structures where the unit cell contains multiple copies, the first instance of each structure was used. All water molecules were removed prior to docking, as were any alternate (e.g., "B") atom positions. Polar hydrogens were added using standard geometries, as necessary. Atom charges and united atom radii were assigned from the AMBER[56] parameter set. No attempt was made to model missing heavy atoms (e.g., antibody F9.13.7). All antibody calculations used only the F_v fragments, with the exception of the F_{ab}–protein G complex (IGC), in which only the F_c was used. The idiotype–anti-idiotype antibody complexes (IAI, DVF and KB5) are the largest complexes we have attempted to model using the current method.

Molecular and solvent-accessible surfaces were calcu-

TABLE I. Protein Complexes Used in the Current Docking Study†

Case	Receptor	Ligand	Å	PDB	Ref.
DHB	Hemoglobin α_1	Hemoglobin β_1	2.8	2DHB	65
CCY	Cytochrome C′ A	Cytochrome C′ B	1.7	2CCY	66
CSE	Subtilisin Carlsberg	Eglin C	1.2	1CSE	67
SNI	Subtilisin BPN′	CI-2	2.1	2SNI	68
SIC	Subtilisin BPN′	SSI	1.8	2SIC	69
KAI	Kallikrein A	BPTI	2.5	2KAI	70
PTC	Trypsin	BPTI	1.9	2PTC	71
CGI	Chymotrypsinogen A	HPTI	2.3	1CGI	72
CHO	α-chymotrypsin	OMTKY3	1.8	1CHO	73
BGS	Barnase	Barstar	2.2	1BGS	74
GGI	50.1	Peptide	2.8	1GGI	75
TET	TE33	Peptide	2.3	1TET	76
FPT	C3	Peptide	3.0	1FPT	77
IGF	B13I2	Peptide	2.8	2IGF	78
JEL	Jel42	HPr	2.8	1JEL	79
BQL	HyHel-5	Quail lysozyme	2.6	1BQL	80
HFL	HyHel-5	Chicken lysozyme	2.7	3HFL	81
HFM	HyHel-10	Chicken lysozyme	3.0	3HFM	82
VFB	D1.3	Chicken lysozyme	1.8	1VFB	60
MLC	D44.1	Chicken lysozyme	2.1	1MLC	61
MEL	cAb	Chicken lysozyme	2.5	1MEL	83
JHL	D11.15	Pheasant lysozyme	2.4	1JHL	84
FBI	F9.13.7	Guineafowl lysozyme	3.0	1FBI	85
NCA	NC41	N9 neuraminidase	2.5	1NCA	86
NMB	NC10	N9 neuraminidase	2.5	1NMB	87
NSN	N10	Staph. nuclease	2.9	1NSN	88
IAI	730.1.4	409.5.3	2.9	1IAI	89
DVF	D1.3	E5.2	1.9	1DVF	90
KB5	Desire-1	KB5-C20	2.5	1KB5	91
IGC	MOPC21	Protein G	2.6	1IGC	92

†These are listed by PDB code and crystallographic resolution and are grouped as 2 domain-dimers, 8 enzyme/inhibitors, and 20 antibody complexes.

lated from all heavy atoms using a 1.4Å probe sphere and a dot surface density of approximately 4 dots/Å^2. Surface skins were sampled onto a (0.75Å)^3 grid as described in Methods. With this grid size, it takes from 1 to 2 minutes to determine all skin coefficients up to $N = 25$ for each pair of proteins (all times are given as elapsed times on a Silicon Graphics R5000 processor). Calculating the charge density coefficients takes about 2 seconds per protein, and solving Poisson's equation for the electrostatic potential coefficients takes a further 0.2 seconds.

Recognizing Known Complexes

Although our ultimate goal is to predict the association of unbound protein subunits, as we are describing a new algorithm it is important to investigate its performance in cases in which the expected result is known. Hence, we first give some results for the recognition of known protein complexes. This exercise also gives an indication of how well a rigid-body docking algorithm might be expected to perform, without expert intervention, in those cases (usually idealized) where conformational changes are negligible.

Table II shows how well our algorithm recognizes the correct orientation of several complexes following a global search in all 6 degrees of freedom, consisting of approximately 5.4×10^8 distinct trial orientations. When using correlations to $N = 25$, this table shows that 18 of the 30 structures are correctly recognized and that 24 out of 30 are ranked within the top 10. Here, successful recognition (a "match") is assumed when the lowest energy orientation found is within 3Å RMS of the correct structure. RMS deviations are calculated as the RMS distance between all ligand C_α atoms of the docked orientation and those of the complex, following a least-squares superposition[57] of the docked structure onto the complex using only receptor C_α atoms (this is the same method of computing RMS deviations as used by Fischer et al.[15]). In this test, trial orientations were generated using 492 vertices of a pair of tessellated icosahedra to give angular increments of about 10° for each of the first four rotational degrees of freedom, and by using 72 twist increments of 5° for the final twist angle search (a smaller twist increment costs little and helps to locate minima accurately). The angular search was repeated at each of 31 intermolecular distances in steps of ±1Å from the starting conformation of the complex, with all distances being rounded to the nearest whole number (thus the orientations of the "correct" structures in Table II may differ from those of the complex by up to ±0.5Å in R). All calculations were preceded by an initial filtering scan at $N = 16$, as described in Methods. Each

TABLE II. Recognition of Known Complexes in a Global 6D Search over 5.4×10^8 Alternative Test Orientations†

	$N = 16$		$N = 20$		$N = 25$	
Case	Top	RMS	Top	RMS	Top	RMS
DHB	2	0.00	2	0.00	1	1.55
CCY	1	0.04	1	0.04	1	1.59
CSE	37	0.73	1	0.08	1	0.92
SNI	15	0.58	1	0.42	1	0.42
SIC	3,407	0.00	2	0.22	1	0.82
KAI	17	0.41	3	0.69	7	0.81
PTC	132	0.52	2	0.48	1	0.48
CGI	1	0.38	1	0.38	1	0.38
CHO	1	0.45	1	0.55	1	0.55
BGS	1	0.82	1	0.82	1	0.88
GGI	1	2.47	1	0.90	1	0.90
TET	5	1.48	1	1.16	1	1.03
FPT	102	1.04	1	0.42	1	0.42
IGF	3	0.71	1	0.77	1	0.77
JEL	4,867	0.81	1,060	0.81	2	0.81
BQL	524	1.85	12	0.96	1	0.39
HFL	318	1.01	5	1.00	1	1.00
HFM	7	2.19	27	1.09	10	1.09
VFB	8,344	1.49	216	0.20	9	0.20
MLC	1,401	0.00	116	0.00	187	0.84
MEL	9,898	1.03	27	1.03	3	1.03
JHL	385	0.62	8	0.38	1	1.08
FBI	14	1.09	1	1.09	1	0.38
NCA	68	1.53	1	0.32	1	0.32
NMB	160	2.43	1,630	1.39	1,009	1.39
NSN	19,992	1.11	716	0.75	1,130	2.29
IAI	1,381	1.48	111	0.37	20	1.39
DVF	11,145	0.00	88	1.38	49	0.44
KB5	140	0.34	1	0.34	78	1.38
IGC	1,328	1.74	269	0.81	1	0.34

†Listed are the rank and C_α RMS deviations of the lowest energy (top scoring) orientation found within 3Å RMS of the complex, evaluated at increasing expansion orders, N. Calculation times are around 2 hours per complex.

global scan takes about 2 hours, with final scoring at $N = 25$ adding a further 10 minutes per complex.

As might be expected, Table II shows that a higher-order expansion generally gives a better rank for the complex, although this trend is not necessarily monotonic. However, it is worth noting that even low-order correlations score the correct docking orientation remarkably favorably. Correlations at $N = 16$ place the correct solution well within the top 1,500 orientations in all but six cases. The most difficult complexes to recognize are the antibody complexes with larger antigens, particularly the NC10–neuraminidase complex (NMB) and the N10–staphylococcal nuclease complex (NSN). However, even the large idiotype–anti-idiotype antibody complexes (tabulated as IAI, DVF, and KB5) are still ranked remarkably favorably.

Localizing the Search

In order to reduce the number of false-positive orientations for the large antibody complexes, the above calculations were repeated at $N = 25$ but with the search constrained to the receptor binding site by excluding rotational samples for which $\beta_1 > 45°$. Table III shows that this simple constraint is sufficient to improve the rank of the best solutions to within the top ten in all but one case and that 23 of the 30 complexes are now ranked first by the algorithm. Also tabulated are the calculated steric and electrostatic contributions to the total energy (E_{shape} and E_{elec}, respectively). E_{elec} may be compared with the exact electrostatic interaction energy, E_{coul}, given in the next column, calculated from point atom charges using Coulomb's law with $K_R = 8$. Considering the very large numbers of polar atoms involved (up to $\simeq$ 3,000 per complex), it is seen that the electrostatic correlation often gives a remarkably good estimate of E_{coul}. Table III shows that E_{elec} is unfavorable for only four of the 20 antibody complexes (VFB, NMB, KB5, IGC), although two of these (KB5, IGC) are still ranked first. Unfavorable electrostatics are also observed for the cytochrome C' domain dimer (CCY) and for two of the enzyme–inhibitor complexes (SIC and PTC) although E_{shape} dominates and all three complexes are still correctly recognized.

Compared to a global search, using a receptor cut-off angle of $\beta_1 \leq 45°$, reduces the number of trial orientations by a factor of about 7. This does not affect the high rank of the enzyme–inhibitor complexes, but it does help the recognition of many of the antibody complexes, improving their average rank by about a factor of 5. This suggests that a good proportion of the false-positive antibody–antigen orientations are located away from the binding site, whereas the signal for the correct docking orientations of the other protein–protein complexes is much clearer. This supports the proposition[58] that antibody–antigen interfaces exhibit poorer shape complementarity than other protein–protein interfaces. The relatively low rank obtained for NMB appears to be due to the largely planar interface in this complex: Different translations in the plane, and rotations perpendicular to it, are not easily distinguished. Nonetheless, Table III shows that localizing the search with a single simple constraint is sufficient to bring the ranking of antibody complexes significantly closer to the high levels observed for the enzyme–inhibitor complexes.

Predictive Docking

Following the encouraging results of the above tests, we used our algorithm to attempt to predict the orientations of 18 complexes using wherever possible the unbound structures of the constituent proteins. The structures used are listed in Table IV. These include the same examples investigated by Gabb et al.,[18] except that we used more recent structures 1VFA/1VFB and 3SSI for the D1.3 antibody–lysozyme (VFB) and subtilisin–SSI (SIC) complexes, respectively. The barnase–barstar complex used here (1BGS) has a mutation (C40A) relative to the unbound barstar (1A19), as not all wild-type structures are available. In each case, the unbound structures were separately superposed onto the structure of the complex by C_α least-squares fitting, to give a consistent "reference orientation" for each calculation: The residual RMS deviation between the complex and the reference structure

TABLE III. Recognition of Known Complexes Using Correlations to $N = 25$ With $\beta_1 \leq 45°$ (Approximately 8×10^7 Trial Orientations)†

Case	Rank	E_{total}	E_{shape}	E_{elec}	E_{coul}	Top	E_{top}	RMS	Hits
DHB	3	−743.3	−685.1	−58.3	−36.4	1	−814.7	1.55	4
CCY	2	−611.8	−686.1	+74.3	+94.6	1	−748.0	1.59	5
CSE	2	−721.7	−651.9	−69.8	−72.9	1	−732.7	0.92	4
SNI	1	−773.2	−712.4	−60.8	−36.1	1	−773.2	0.42	6
SIC	1	−590.1	−659.1	+69.0	+59.9	1	−631.0	0.82	1
KAI	23	−1,436.8	−538.9	−897.8	−788.4	7	−1,535.4	0.81	4
PTC	1	−556.8	−704.1	+147.3	+137.9	1	−556.8	0.48	3
CGI	1	−1,106.4	−1,026.9	−79.5	−69.6	1	−1,106.4	0.38	14
CHO	23	−555.4	−490.4	−65.0	−61.9	1	−732.9	0.55	4
BGS	3	−859.8	−554.4	−305.4	−242.0	1	−1,061.3	0.88	7
GGI	1,131	−430.5	−250.0	−180.4	−175.6	1	−718.2	0.90	33
TET	3	−512.6	−504.3	−8.3	−21.4	1	−574.0	1.03	21
FPT	1	−605.6	−461.2	−144.4	−127.6	1	−605.6	0.42	30
IGF	7	−499.4	−385.0	−114.4	−106.0	1	−591.9	0.77	56
JEL	6	−568.1	−541.3	−26.8	−17.1	2	−640.0	0.81	2
BQL	1	−845.7	−647.8	−197.9	−183.6	1	−845.7	0.39	1
HFL	1	−852.1	−657.5	−194.6	−183.3	1	−877.2	1.00	2
HFM	4	−815.1	−588.3	−226.8	−196.8	8	−890.6	1.09	2
VFB	1	−231.1	−571.0	+340.0	+209.5	1	−231.1	0.20	2
MLC	23	−639.4	−475.9	−163.5	−147.6	8	−663.8	0.84	1
MEL	1	−804.6	−627.2	−177.4	−147.4	1	−869.0	1.03	4
JHL	1	−835.9	−594.6	−241.2	−182.3	1	−842.7	1.08	2
FBI	1	−1,008.4	−647.5	−360.9	−308.7	1	−1,008.4	0.38	3
NCA	1	−974.7	−828.3	−146.4	−88.9	1	−974.7	0.32	2
NMB	502	−306.0	−363.1	+57.1	−0.4	421	−314.4	1.39	0
NSN	8	−565.6	−535.0	−30.6	−16.4	4	−603.1	1.11	4
IAI	15	−612.8	−611.1	−1.7	−3.4	5	−655.3	1.39	2
DVF	5	−600.1	−553.5	−46.6	−62.1	1	−830.5	1.42	2
KB5	55	−469.4	−496.1	+26.7	+8.4	1	−692.1	1.49	2
IGC	1	−623.2	−657.5	+34.3	+40.2	1	−623.2	0.34	1

†The first four columns of figures give the rank and energies calculated for each complex. The following column, E_{coul}, is the exact electrostatic interaction energy of the complex, calculated using Coulomb's law. The next three columns give the rank, energy, and RMS deviation of the top-scoring docking orientation found by the algorithm. The number of orientations found within 3Å of the complex and within the top-scoring 100 orientations (considered as "hits") is given in the final column. All energies, E, are in KJ/mol. Calculation times are around 45 minutes per complex.

represents a good estimate of the best orientation attainable within the rigid-body assumption. As these deviations can be quite large, we relaxed the threshold for a "match" (3Å RMS) between the reference and predicted docking orientations by an amount equal to this deviation.

Table V shows the rank obtained for the reference orientation of each complex, along with the best-matching orientations found by the algorithm in a search localized to the receptor-binding site using $\beta_1 \leq 45°$, as in Table III. Comparing the rank of the reference orientations with those of the bound subunits in Table III shows that the algorithm often detects a significant conformational change between the bound and unbound protein structures. However, in most cases much higher ranking orientations are found within 3Å RMS of the reference orientation, indicating that small rigid-body motions can, to a certain extent, compensate for poorly fitting starting orientations. In order to assess how the electrostatic correlation contributes to the overall docking score, each docking calculation was repeated using just the surface skin correlation. The rank of the best solution found in these shape-only docking calculations (indicated in parentheses in Table V) are seen to be significantly worse in the majority of cases. This shows that our simple *in vacuo* electrostatic correlation model can play a useful role in helping to identify favorable docking orientations.

It should be noted, however, that the above calculations are biased toward finding good solutions because the search space always includes the reference orientation. Thus, in order to simulate genuinely blind docking predictions, each calculation was repeated using a starting orientation in which the ligand was shifted away from the reference orientation (columns marked with an asterisk in Table V). The shift was defined by a random step of ± 0.5Å in R and by a small rotation (β_2, γ_2) using angular coordinates taken from the midpoint of a randomly selected edge near the "north pole" of the icosahedral tesselation. Thus, the search space now systematically *excludes* the sought solution and, therefore, the rank obtained from these pseudo-random starting orientations corresponds to the worst rank that might be expected following a large number of random trials. The relatively large range observed between the rank obtained from the shifted and unshifted starting orientations in Table V indicates that

TABLE IV. Protein Structures Used for the Predictive Docking Calculations, Listed by PDB Code and Crystallographic Resolution†

	Receptor			Ligand		
Case	PDB	Å	Ref.	PDB	Å	Ref.
SNI	1SUP	1.6	93	2CI2	2.0	59
SIC	1SUP	1.6	93	3SSI	2.3	94
KAI	2PKA	2.05	95	1BPI	1.1	96
PTC	2PTN	1.55	97	4PTI	1.5	71
CGI	1CHG	2.5	98	1HPT	2.3	99
CHO	5CHA	1.67	100	2OVO	1.5	101
BGS	1A2P	1.50	102	1A19	2.76	103
JEL	1JEL	2.8	79	1POH	2.0	104
BQL	1BQL	2.6	80	1DKJ	2.0	80
HFL	3HFL	2.65	81	1LZA	1.6	105
HFM	3HFM	3.0	82	1LZA	1.6	105
VFB	1VFA	1.8	60	1LZA	1.6	105
MLC	1MLB	2.1	61	1LZA	1.6	105
MEL	1MEL	2.5	83	1LZA	1.6	105
JHL	1JHL	2.8	84	1GHL	2.1	106
FBI	1FBI	3.0	85	1HHL	1.9	106
IAI	1IAI	2.9	89	1AIF	2.9	62
IGC	1IGC	2.6	92	1IGD	1.1	92

†The unbound structures of most of the antibodies have not been determined, and so the conformation from the complex is used in these cases (1JEL, 1BQL, 3HFL, 3HFM, 1MEL, 1JHL, 1FBI, 1IAI, 1IGC).

10° angular search increments are too crude to give good coverage of the search space.

Although the above calculations often yield good docking orientations within the top few hundred solutions, even when the reference orientation is excluded from the search, it was felt that visual inspection (for example) of this number of orientations would be impractical. Thus, we investigated the effect of constraining the search to the ligand epitope by restricting the allowed range of β_2. Because reducing the search space reduces execution times proportionately, it is now feasible to use finer angular search increments. Hence, we used icosahedral tessellations with 720 faces to generate receptor (β_1, γ_1) and ligand (β_2, γ_2) rotational steps of about 6.7° each (the twist angle, α_2, was held at 5°). As before, pseudo-random shifts were applied to the ligand prior to each calculation. The results of these high-resolution docking calculations are summarized in Table VI. Inspection of this table shows that many high-ranking docking orientations are found at each level of constraint with just a few exceptional cases, following a similar trend to Table V. However, despite the much denser coverage of the search space than in Table V, the absolute rankings are now often dramatically improved. For example, even with fairly weak constraints ($\beta_1 \leq 45°$, $\beta_2 \leq 45°$), five of the 11 antibody complexes are ranked within the top 30 and three of the seven enzyme–inhibitor complexes are ranked within the top 40. With the strongest constraint level ($\beta_1 \leq 30°$, $\beta_2 \leq 30°$), the majority of the complexes (seven out of 11 antibody complexes and four of the seven enzyme–inhibitor complexes) are ranked within the top 20 orientations. These results show that despite sometimes significant conformational changes, our algorithm can often find good docking orientations given only a very loose specification of the binding epitopes.

DISCUSSION

The spherical polar Fourier docking method presented here has been shown to provide a fast and accurate way to find feasible protein–protein docking orientations. Although this is conceptually similar to former grid-based FFT docking methods, we began by constructing explicit spherical polar series expansions of surface shape and electrostatic representations of pairs of proteins. This allowed expressions to be developed for the overlap of appropriate pairs of functions to give a full six-dimensional Fourier docking correlation. Because the steric and electrostatic expansion coefficients need only be calculated once for each protein, the remaining computational cost is largely one of rotating and multiplying pairs of coefficient vectors. In order to use the correlation most effectively, an icosahedral tessellation of the sphere was used to sample rotational space evenly and fairly. Despite the additional programming effort required, this approach allows a 6D docking search to be reduced to just four nested loops. The innermost loop over the twist angle, α_2, involves finding the local minima of a one-dimensional real Fourier series, and this can be performed very rapidly. However, to make this approach tractable it is currently necessary to store many megabytes of pre-calculated overlap integrals, but this is not a significant overhead on modern workstations. Nonetheless, having established that the general approach is viable, we are developing an improved method of calculating the overlap/translation matrices using Fourier-Bessel transform theory, which should help address this drawback.

The results presented here, for a single processor workstation, show that the performance of the spherical polar correlation compares favorably with the Cartesian grid-based FFT approaches, particularly for large protein docking problems such as antibody–antigen systems. This is primarily because evaluation of the spherical polar correlations is completely de-coupled from the choice of the original sampling grid, and so a fast low-order scan ($N = 16$) could be used to eliminate rapidly many infeasible orientations. Consequently we could use finer rotational increments (10° or less) than is feasible in a global FFT docking search,[18] while still keeping execution times down to a reasonable level (around 2 hours per global docking search). Clearly the algorithm is highly vectorizable, and significant performance improvements could be expected on suitable hardware. Indeed, on multi-processor platforms, which are becoming increasingly common, our docking calculations can be distributed over the available processors using standard Unix multi-tasking facilities. This reduces execution times almost linearly with the number of processors used.

However, the reported performance improvement has not been at the expense of reduced accuracy. In a global search, the spherical polar correlation correctly identifies the orientation of the complex in 18 out of 30 cases, and 24 out of 30 are ranked within the top ten out of 5×10^8 trial

TABLE V. Docking Unbound Subunits Using Correlations to $N = 25$ With the Search Constrained to the Receptor Binding Site Using $\beta_1 \leq 45°$[†]

Case	Rank[a]	RMS[b]	Top[c]	(Top)[d]	Δ[e]	RMS[f]	Top*	(Top)*	Δ*	RMS*
SNI	—	0.46	—	(−)	=	—	—	(−)	=	—
SIC	380	0.64	240	(125)	−	1.03	—	(−)	=	—
KAI	261	0.54	136	(843)	+	3.08	334	(1,928)	+	2.94
PTC	133	1.10	130	(192)	+	1.26	587	(865)	+	1.34
CGI	5,150	1.78	48	(12)	−	2.19	215	(71)	−	2.29
CHO	4,402	1.16	6	(13)	+	1.88	5	(19)	+	2.06
BGS	—	0.51	394	(1,665)	+	2.29	593	(236)	−	1.90
JEL	5,083	1.38	87	(87)	=	3.74	149	(544)	+	3.68
BQL	3,643	0.84	144	(191)	+	0.99	60	(75)	+	1.69
HFL	575	0.53	203	(316)	+	1.16	486	(399)	−	2.40
HFM	—	0.58	1,636	(1,402)	−	2.24	—	(1,579)	−	—
VFB	2,327	1.07	158	(185)	+	3.07	211	(231)	+	2.18
MLC	—	0.63	—	(1,696)	−	2.82	904	(262)	−	2.94
MEL	3	0.67	1	(1)	=	1.39	1	(1)	=	1.37
JHL	63	0.51	48	(466)	+	1.23	456	(1,408)	+	2.54
FBI	—	0.68	2,079	(−)	+	—	—	(−)	=	—
IAI	2,182	1.10	778	(291)	−	1.78	—	(−)	=	—
IGC	2,235	1.05	691	(79)	−	2.69	636	(84)	−	1.55

[†]Shape-only calculations are given in parentheses. Columns marked with an asterisk are for calculations starting with randomized ligand orientations (see main text for details). A dash indicates no solution found within the top 10,000 orientations. Calculation times are from 8 to 12 minutes per complex.

[a]The rank of the reference structure, calculated using unbound subunits fitted to the main-chain conformation of the complex.

[b]Residual C_α RMS deviation of the reference structure. Calculated docking orientations will always have RMS errors that exceed this value.

[c]The rank of the top-scoring solution found within 3Å RMS of the bound complex (allowing for the residual RMS deviation) using the combined steric and electrostatic correlation.

[d]The rank of the best solution found using only the steric correlation.

[e]The effect of the electrostatic correlation on the result: +, improves; −, worsens; =, no change.

[f]RMS deviation of the top-scoring orientation found using the combined steric and electrostatic correlation.

orientations (Table II). Such figures are comparable to the excellent results obtained by Meyer et al.[28] using an FFT correlation in conjunction with hydrogen bond filters and an accurate angular search algorithm. However, the majority of the complexes investigated here are large antibody–antigen systems compared to only three such complexes in the former study. Similarly, our results compare favorably to those of the geometric hashing method (Table 2 of Fischer et al.[15]) and also to the combined steric/electrostatic FFT correlation method of Gabb et al.[18] (their Table 4), despite our steric penalty term ($Q = 12$) being "softer" than the equivalent quantity in the FFT method (assigning interior grid cells a value of −15). Unfortunately, Meyer et al. do not report results for docking unbound subunits, and so it is uncertain how robust the hydrogen bond approach would be when the starting conformations are poorly fitting. Nonetheless, it is clear that docking bound subunits of protein complexes presents few difficulties to current algorithms. Tables II and III show that our approach easily achieves this high standard.

When the algorithm is used predictively, to dock unbound structures, the quality of the results depends largely on the degree of conformational change induced by binding. Of the examples studied here, probably the most dramatic conformational changes are to be found in the main-chain and side-chains (particularly MET:59) of the CI-2 inhibitory loop when binding to subtilisin BPN′[59] (tabulated as SNI). This presumably accounts for the relatively poor rank obtained when attempting to dock this complex. Nonetheless, despite the induced-fitting nature of enzyme–inhibitor interfaces, our constrained docking calculation ($\beta_1 \leq 30°$, $\beta_2 \leq 30°$) was able to find a good docking orientation that ranked within the top 20 solutions for four of the seven enzyme–inhibitor complexes. The two antibody–lysozyme complexes D1.3 and D44.1 (tabulated as VFB and MLC, respectively) are also of interest, being the only antibody–antigen complexes for which both unbound subunits were available. On binding, both antibodies exhibit conformational changes in the hypervariable loops, and both complexes have several buried interfacial waters.[60,61] Hence, it is somewhat surprising that the best solution found for VFB is ranked so highly, particularly since the electrostatic component is strongly unfavorable. Considering the relatively weak nature of the search constraints used here, it is noteworthy how well our correlation performs with the remaining antibody complexes, with good docking predictions being placed within the top 20 solutions in seven out of 11 cases and with many more "hits" falling within the top 100 orientations. Even the large idiotype–anti-idiotype complex (IAI) was predicted relatively well despite H3 loop and VH/VL domain motions in the anti-idiotype.[62] Thus, our spherical polar docking correlation is seen to be remarkably robust with respect to conformational changes induced by binding.

Like other docking methods, knowledge of at least one of

TABLE VI. High-Resolution Predictive Docking With Angular Search Constraints[†]

	RMS		$\beta_1 \leq 45°$[a] $\beta_2 \leq 45°$		$\beta_1 \leq 30°$[b] $\beta_2 \leq 45°$		$\beta_1 \leq 30°$ $\beta_2 \leq 30°$[c]	
Case	Ref. struct.[d]	Docked struct.*[e]	Top*	Hits*	Top*	Hits*	Top*	Hits*
SNI	0.46	1.30	1,257	0	401	0	143	0
SIC	0.64	2.24	32	1	30	1	16	2
KAI	0.54	3.17	68	1	62	1	31	2
PTC	1.10	3.80	108	0	9	3	3	5
CGI	1.78	2.08	20	2	20	2	13	6
CHO	1.16	2.26	1	14	1	14	1	17
BGS	0.51	2.06	162	0	139	0	73	1
JEL	1.38	3.98	4	3	4	3	3	3
BQL	0.84	1.32	6	3	7	4	2	6
HFL	0.53	1.30	45	4	42	4	27	5
HFM	0.58	3.04	147	0	72	1	50	1
VFB	1.07	2.81	54	1	46	1	19	2
MLC	0.63	2.84	467	0	308	0	91	1
MEL	0.67	1.64	1	3	1	6	1	5
JHL	0.51	1.77	4	3	4	3	3	4
FBI	0.78	3.69	28	1	25	3	19	4
IAI	1.10	3.64	508	0	384	0	173	0
IGC	1.05	1.97	54	1	15	2	6	3

[†]Listed are the rank and RMS deviation of the best orientation ("top") obtained using the search constraints given in the column headings, along with the number of good orientations found within the top 100 solutions ("hits"). Here, 720 icosahedral increments of about 6.7° are used to generate receptor (β_1, γ_1) and ligand (β_2, γ_2) rotations, subject to the given constraints on β_1 and β_2. All calculations used randomized ligand starting orientations ("*").
[a]2.7×10^7 trial orientations: about 20 minutes docking time.
[b]1.3×10^7 trial orientations: about 11 minutes docking time.
[c]6.0×10^6 trial orientations: about 8 minutes docking time.
[d]Residual C_α RMS deviation of the reference structure, carried over from Table V.
[e]RMS deviation of the first orientation found within 3Å RMS of the complex, allowing for the residual deviation of the reference structure. With pseudo-random starting orientations, each calculation finds the same top-ranking orientation; hence these deviations are listed only once.

the binding sites is necessary to reduce the number of false-positive solutions found. For example, when the ligand is constrained (rather loosely) to tumble over the receptor binding site (Table V), the correct solution is often ranked within the top few hundred orientations. This compares favorably to the results of similar docking calculations using the geometric hashing algorithm (in particular, cf. CHO, PTC, and HFL with Table 4.C of Fischer et al.[15]). However, direct comparison with the predictive FFT calculations of Gabb et al.[18] is difficult because even their "loose" filtering constraint (which calls for specific residue or hypervariable loop contacts) reduces the solution set to just a few hundred candidate orientations. In comparison, our most tightly constrained calculations involve evaluating several million possible docking orientations. Hence, despite using a sampling strategy which systematically avoided the sought solution, it is most encouraging to see that the spherical polar approach still generates so many high-ranking solutions.

Given the practical necessity of constraining predictive docking calculations to known (or hypothesized) binding sites, it is clear that a spherical polar formulation provides a convenient way to apply the constraints before, rather than after, the correlation is evaluated. This can reduce docking calculation times to a matter of minutes, even when using fine angular search steps (e.g., 6.7° or less). Using the interactive graphics features of our program, *Hex,* it is straightforward to place a ligand near the antigen-binding site of an antibody, for example, and to perform constrained docking calculations which search around the given starting orientation. We believe that constrained docking calculations of this type could help experimentalists gain useful insights when considering possible binding modes of pairs of proteins. We find that even the 30° constraints are quite generous for large antibody–antigen systems, although β_2 must be allowed to vary freely for small peptide ligands. Clearly, we could have used tighter filtering constraints to further improve the rankings presented here, but forcing the desired solution in this manner is a rather unsatisfactory way to deal with the problem of macromolecular conformational flexibility.

One reason for the relative success of the spherical polar correlation approach is that the basis functions are "tuned" to the dimensions of typical protein domains (or domain dimers such as antibody F_v fragments), whereas the distance scale in the FFT method is much larger because the FFT grid must be big enough to accommodate translations of one molecule about the other. However, the price to be paid for this "tuning" is that very large macromolecules, e.g., the large trimeric hemagglutinin–antibody complex presented in the CASP2 docking challenge,[12] would be represented very poorly because of the exponential decay of the radial basis functions beyond about $R = 30Å$.

Nonetheless, docking an antibody to hemagglutinin also poses problems for the FFT approach: In this case, Vasker obtained the best single docking prediction by splitting the hemagglutinin moiety into smaller fragments prior to performing low-resolution FFT correlations on each fragment and by using symmetry to eliminate those parts of its surface which are buried in the trimer and therefore presumed to be non-antigenic.[63] To dock macromolecular complexes of this size, we expect similar measures would also be required with the spherical polar approach. For such cases, we plan to investigate using several local coordinate origins within the larger molecule so that each part of its surface is represented accurately at least once; hence, accurate localized searches could be performed over each surface patch. On the other hand, with a suitable choice of scale factor, one would expect that the surface shape and electrostatic properties of small organic molecules should be captured quite well using relatively low-order spherical polar expansions, and this could provide an efficient way of searching small-molecule databases for lead drug molecules that are similar to a given template.

Certainly, in the protein docking case, we have shown that spherical polar representations can encode atom-scale protein surface properties relatively compactly while still being sufficiently soft to absorb the effect of moderate conformational change when docking unbound subunits. Furthermore, we believe that the spherical polar approach could be extended to model limited side chain flexibility in a tractable way. For example, since each atom contributes to the charge density coefficients additively (equation 35), one could model the electrostatic effect of side chain motion by subtracting the contributions from the old atom positions and by adding similar contributions for some new conformation (the cost of re-solving Poisson's equation for the potential coefficients is almost negligible). Updating the surface shape coefficients in a similar manner is more problematic, because our surface skins are defined by Richards' rolling probe construction,[46] which is a global operation. However, an alternative and extremely promising avenue we are currently investigating is to model shape complementarity using Gaussian expansions of Lennard-Jones potentials.[64] Because our shape-scaled radial basis functions (equation 2) are effectively modified Gaussian functions, calculating the expansion coefficients of Lennard-Jones potentials should be no more expensive than the present electrostatic calculations.

CONCLUSIONS

We have described a new protein docking algorithm based on spherical polar correlations of protein surface shape and electrostatic representations. We have shown that these correlations provide a fast and accurate way to find feasible docking orientations of protein complexes and that this approach is highly competitive compared to former grid-based FFT docking methods. Starting from unbound subunits, we can often get close to the desired conformation of the complex. However, knowledge of one or both binding sites is still necessary to reduce the number of false-positive solutions. In the spherical polar approach, this information is given as a simple constraint in just one or two of the angular degrees of freedom. Execution times can be reduced to a matter of minutes by applying these constraints before, rather than after, the correlation is evaluated. Hence our interactive docking program, *Hex,* could provide a useful and practical tool for experimentalists. However, we argue that constraining the search to known binding sites is currently a necessary, but unsatisfactory, way to deal with conformational flexibility. We have discussed some of the ways in which spherical polar Fourier correlations might help address this problem, which will be vital if we are ever to throw off the shackles of the rigid-body assumption.

ACKNOWLEDGMENTS

We thank Prof. J.E. Fothergill for useful discussions and encouragement.

ADDENDUM

The program described (*Hex* 2.0) is available on the Internet at "http://www.biochem.abdn.ac.uk/hex/".

REFERENCES

1. Wodak SJ, Janin J. Computer analysis of protein–protein interaction. J Mol Biol 1978;124:323–342.
2. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982;161(2):269–288.
3. Chothia C, Novotný J, Bruccoleri R, Karplus M. Domain association in immunoglobulin molecules: the packing of variable domains. J Mol Biol 1985;186:651–663.
4. Davies DR, Padlan EA, Sheriff S. Antibody-antigen complexes. Annu Rev Biochem 1990;59:439–473.
5. Janin J, Chothia C. The structure of protein-protein recognition sites. J Biol Chem 1990;265(27):16027–16030.
6. Novotny J, Sharp K. Electrostatic fields in antibodies and antibody/antigen complexes. Prog Biophys Mol Biol 1992;58:203–224.
7. Jones S, Thornton JM. Principles of protein–protein interactions. Proc Natl Acad Sci 1996;93(1):13–20.
8. Davies DR, Cohen GH. Interactions of protein antigens with antibodies. Proc Natl Acad Sci 1996;93(1):7–12.
9. Janin J. Protein–protein recognition. Prog Biophys Mol Biol 1995;64(2–3):145–166.
10. Shoichet BK, Kuntz ID. Predicting the structure of protein complexes: a step in the right direction. Chem Biol 1996;3:151–156.
11. Strynadka NCJ, Eisenstein M, Katchalski-Katzir E, Shoichet BK, Kuntz ID, Abagyan R, Totrov M, Janin J, Cherfils J, Zimmerman F, Olson A, Duncan B, Rao M, Jackson R, Sternberg M, James MNG. Molecular docking programs successfully predict binding of a β-lactamase inhibitory protein to TEM-1 β-lactamase. Nature Struct Biol 1996;3(3):233–239.
12. Dunbrack RL, Jr, Gerloff DL, Bower M, Chen X, Lichtarge O, Cohen FE. Meeting review: the second meeting on the critical assessment of techniques for protein structure prediction (CASP2), Asilomar, California, December 13–16, 1996. Folding Design 1997;1:R27–R42.
13. Cherfils J, Bizebard T, Knossow M, Janin J. Rigid-body docking with mutant constraints of influenza hemagglutinin with antibody HC19. Proteins 1994;18:8–18.
14. Totrov M, Abagyan R. Detailed *ab initio* prediction of lysozyme-antibody complex with 1.6Å accuracy. Struct Biol 1994;1(4):259–263.
15. Fischer D, Lin SL, Wolfson HL, Nussinov R. A geometry based suite of molecular docking processes. J Mol Biol 1995;248:459–477.
16. Norel R, Lin SL, Wolfson HJ, Nussinov R. Molecular surface

complementarity at protein–protein interfaces: the critical role played by surface normals at well placed, sparse, points in docking. J Mol Biol 1995;252:263–273.
17. Walls PH, Sternberg MJE. New algorithm to model protein–protein recognition based on surface complementarity. J Mol Biol 1992;228(1):277–297.
18. Gabb HA, Jackson RM, Sternberg MJE. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 1997;272(1):106–120.
19. Connolly ML. Shape complementarity at the hemoglobin $\alpha_1\beta_1$ subunit interface. Biopolymers 1986;25:1229–1247.
20. Fischer D, Norel R, Wolfson H, Nussinov R. Surface motifs by a computer vision technique: Searches, detection, and implications for protein–ligand recognition. Proteins 1993;16:278–292.
21. Ausiello G, Cesareni G, Helmer-Citterich M. ESCHER: a new docking procedure applied to the reconstruction of protein tertiary structures. Proteins 1997;28:556–567.
22. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. Proc Natl Acad Sci USA 1992;89:2195–2199.
23. Vasker IA, Aflalo C. Hydrophobic docking: A proposed enhancement to molecular recognition techniques. Proteins 1994;20:320–329.
24. Harrison RW, Kourinov IV, Andrews LC. The Fourier-Greens function and the rapid evaluation of molecular potentials. Protein Eng 1994;7(3):359–369.
25. Blom NS, Sygusch J. High resolution fast quantitative docking using Fourier domain correlation techniques. Proteins 1997;27:493–506.
26. Vasker IA. Protein docking for low-resolution structures. Protein Eng 1995;8(4):371–377.
27. Vasker IA. Long-distance potentials: an approach to the multiple-minima problem in ligand-receptor interaction. Protein Eng 1996;9:37–41.
28. Meyer M, Wilson P, Schomburg D. Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking. J Mol Biol 1996;264(1):199–210.
29. Ackermann F, Herrmann G, Posch S, Sagerer G. Estimation and filtering of potential protein–protein docking positions. Bioinformatics 1998;14(2):196–205.
30. Masek BB, Merchant A, Matthew JB. Molecular skins—a new concept for quantitative shape—matching of a protein with its small-molecule mimics. Proteins 1993;17(2):193–202.
31. Pauling L, Wilson EB. Introduction to quantum mechanics. New York: McGraw-Hill; 1935.
32. Max NL, Getzoff ED. Spherical harmonic molecular surfaces. IEEE Comput Graphics Appl 1988;8(4):42–50.
33. Duncan BS, Olson AJ. Approximation and characterization of molecular surfaces. Biopolymers 1993;33:219–229.
34. Leicester S, Finney J, Bywater R. A quantitative representation of molecular-surface shape. 1. Theory and development of the method. J Math Chem 1994;16(3–4):315–341.
35. Ritchie DW, Kemp GJL. Fast computation, rotation and comparison of low resolution spherical harmonic molecular surfaces. J Comp Chem 1999;20(4):383–395.
36. Ruf W, Shobe J, Rao M, Dickinson CD, Olson A, Edgington TS. Importance of factor VIIa gla-domain residue arg-36 for recognition of the macromolecular substrate factor X gla-domain. Biochemistry 1999;38:1957–1966.
37. Biedenharn LC, Louck JC. Angular momentum in quantum physics. Reading, MA: Addison-Wesley; 1981.
38. Pople JA, Beveridge DL. Approximate molecular orbital theory. New York: McGraw-Hill; 1970.
39. Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J, Ritter O, Abola EE. Protein data bank (PDB): database of three-dimensional structural information of biological macromolecules. Acta Cryst 1998;D54:1078–1084.
40. Hobson EW. The theory of spherical and ellipsoidal harmonics. London: Cambridge University Press; 1931.
41. Rose ME. Elementary theory of angular momentum. New York: John Wiley & Sons; 1957.
42. Erdélyi A, Magnus W, Oberhettinger F, Tricomi FG. Higher transcendental functions, Vol 2. New York: McGraw-Hill; 1953.
43. Biedenharn LC, Louck JC. The Racah-Wigner algebra in quantum theory. Reading, MA: Addison-Wesley; 1981.
44. Danos M, Maximon LC. Multipole matrix elements of the translation operator. J Math Phys 1965;6(1):766–778.
45. Talman JD. Special functions: a group theoretical approach. New York: W. A. Benjamin Inc.; 1968.
46. Richards FM. Areas, volumes, packing, and protein structure. Annu Rev Biophys Bioeng 1977;6:151–176.
47. Ritchie DW. Parametric protein shape recognition. PhD thesis, University of Aberdeen, U.K., 1998.
48. Shrake A, Rupley JA. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. J Mol Biol 1973;79:351–371.
49. Richmond TJ. Solvent accessible surface area and excluded volume in proteins. J Mol Biol 1984;178:63–89.
50. Sharp K, Honig B. Electrostatic interactions in macromolecules: Theory and applications. Annu Rev Biophys Biophys Chem 1990;19:301–332.
51. Warshel A, Åqvist J. Electrostatic energy and macromolecular function. Annu Rev Biophys Biophys Chem 1991;20:267–298.
52. Jackson JD. Classical electrodynamics. New York: Wiley; 1975.
53. Bharadwaj R, Windemuth A, Sridharan S, Honig B, Nicholls A. The fast multipole boundary element method for molecular electrostatics—an optimal approach for large systems. J Comp Chem 1995;16(7):898–913.
54. Wang HY, LeSar R. An efficient fast-multipole algorithm based on an expansion in the spherical harmonics. J Chem Phys 1996;104(11):4173–4179.
55. Novotny J, Bruccoleri RE, Saul FA. On the attribution of binding energy in antigen-antibody complexes. McP603, D1.3 and Hy-Hel-5. Biochemistry 1989;28:4735–4749.
56. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Jr, Weiner P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 1984;106:765–784.
57. Kabsch W. A solution for the best rotation to relate two sets of vectors. Acta Cryst 1976;A32:922–923.
58. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. J Mol Biol 1993;234:946–950.
59. McPhalen CA, James MNG. Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. Biochemistry 1987;26(1):261–269.
60. Bhat TN, Bentley GA, Boulot G, Greene MI, Tello D, Dall'Acqua W, Souchon H, Schwarz FP, Mariuzza RA, Poljak RJ. Bound water molecules and conformational stabilization help mediate an antigen–antibody association. Proc Natl Acad Sci USA 1994;91(3):1089–1093.
61. Braden BC, Souchon H, Eiselé J-L, Bentley GA, Bhat TN, Navaza J, Poljak RJ. Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. J Mol Biol 1994;243(4):767–781.
62. Ban N, Escobar C, Hasel KW, Day J, Greenwood A, McPherson A. Structure of an anti-idiotypic Fab against feline peritonitis virus-neutralizing antibody and a comparison with the complexed Fab. Faseb J 1995;9(1):107–114.
63. Vasker I. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. Proteins 1997;1:226–230.
64. Kostrowicki J, Piela L, Cherayil BJ, Scheraga HA. Performance of the diffusion equation method in searches for optimum structures of clusters of Lennard-Jones atoms. J Phys Chem 1991;95:4113–4119.
65. Bolton W, Perutz MF. Three dimensional Fourier synthesis of horse deoxy-haemoglobin at 2.8 Å resolution. Nature 1970;228:551–552.
66. Finzel BC, Weber PC, Hardman KD, Salemme FR. Structure of ferricytochrome c′ from *rhodospirillum molischianum* at 1.67 Å resolution. J Mol Biol 1985;186(3):627–643.
67. Bode W, Papamokos E, Musil D. The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin C, an elastase inhibitor from the leech *Hirudo medicinalis*. Structural analysis, subtilisin structure and interface geometry. Eur J Biochem 1987;166(3):673–692.
68. McPhalen CA, James MNG. Structural comparison of two serine proteinase–protein inhibitor complexes: Eglin-C-subtilisin carlsberg and CI-2-subtilisin novo. Biochemistry 1988;27:6582–6598.
69. Takeuchi Y, Satow Y, Nakamura NT, Mitsui Y. Refined crystal structure of the complex of subtilisin BPN′ and *streptomyces*

subtilisin inhibitor at 1.8 Å resolution. J Mol Biol 1991;221:309–325.
70. Chen Z, Bode W. Refined 2.5 Å X-ray crystal structure of the complex formed by porcine kallikrein A and the bovine pancreatic trypsin inhibitor. Crystallization, Patterson search, structure determination, refinement, structure and comparison with its components and with the bovine trypsin-pancreatic trypsin inhibitor complex. J Mol Biol 1983;164:283–311.
71. Marquart M, Walter J, Deisenhofer J, Bode W, Huber R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta Cryst 1983; B39:480–490.
72. Hecht HJ, Szardenings M, Collins J, Schomburg D. Three-dimensional structure of the complexes between bovine chymotrypsinogen A and two recombinant variants of human pancreatic secretory trypsin inhibitor (Kazal-type). J Mol Biol 1991;220: 711–722.
73. Fujinaga M, Sielecki AR, Read RJ, Ardelt W, Laskowski M, Jr, James MNG. 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 1987;195:397–418.
74. Guillet V, Lapthorn A, Hartley RW, Mauguen Y. Recognition between a bacterial ribonuclease, barnase, and its natural inhibitor, barstar. Structure 1993;1:165–177.
75. Rini JM, Stanfield RL, Stura EA, Salinas PA, Profy AT, Wilson IA. Crystal structure of a human immunodeficiency virus type 1 neutralizing antibody, 50.1, in complex with its V3 loop peptide antigen. Proc Natl Acad Sci USA 1993;90(13):6325–6329.
76. Shoham M. Crystal structure of an anticholera toxin peptide complex at 2.3 Å. J Mol Biol 1993;232(4):1169–1175.
77. Wien MW, Filman DJ, Stura EA, Guillot S, Delpeyroux F, Crainic R, Hogle JM. Structure of the complex between the FAB fragment of a neutralizing antibody for the type-1 poliovirus and its viral epitope. Nat Struct Biol 1995;2:232–243.
78. Stanfield RL, Fieser TM, Lerner RA, Wilson IA. Crystal structures of an antibody to a peptide and its complex with peptide antigen at 2.8 Angstroms. Science 1990;248:712–719.
79. Prasad L, Sharma S, Vandonselaar M, Quail JW, Lee JS, Waygood EB, Wilson KS, Dauter Z, Delbaere LTJ. Evaluation of mutagenesis for epitope mapping: Structure of an antibody–protein complex. J Biol Chem 1993;268(15):10705–10708.
80. Chacko S, Silverton EW, Smith-Gill SJ, Davies DR, Shick KA, Xavier KA, Willson RC, Jeffrey PD, Chang CYY, Sieker LC, Sheriff S. Refined structures of bobwhite quail lysozyme uncomplexed and complexed with the HyHel-5 Fab fragment. Proteins 1996;26(1):55–65.
81. Cohen GH, Sheriff S, Davies DR. Refined structure of the monoclonal antibody Hy/Hel-5 with its antigen hen egg-white lysozyme. Acta Cryst 1996;D52:315–326.
82. Padlan EA, Silverton EW, Sheriff S, Cohen GH, Smith-Gill SJ, Davies DR. Structure of an antibody–antigen complex: crystal structure of the HyHel-10 Fab–lysozyme complex. Proc Natl Acad Sci USA 1989;86:5938–5942.
83. Desmyter A, Transue TR, Ghahroudi MA, Thi MHD, Poortmans F, Hamers R, Muyldermans S, Wyns L. Crystal structure of a camel single-domain V_H antibody fragment in complex with lysozyme. Nature Struct Biol 1996;3(9):803–811.
84. Chitarra V, Alzari PM, Bentley GA, Bhat TN, Eiselé J-L, Houdusse A, Lescar J, Souchon H, Poljak RJ. Three-dimensional structure of a heteroclitic antigen–antibody cross-reaction complex. Proc Natl Acad Sci USA 1993;90:7711–7715.
85. Lescar J, Pellegrini M, Souchon H, Tello D, Poljak RJ, Peterson N, Greene M, Alzari PM. Crystal structure of a cross-reaction complex between Fab F9.13.7 and guinea fowl lysozyme. J Biol Chem 1995;270(30):18067–18076.
86. Tulip WR, Varghese JN, Laver WG, Webster RG, Colman PM. Refined crystal structure of the influenza virus N9 neuraminidase–NC41 Fab complex. J Mol Biol 1992;227(1):122–148.
87. Malby RL, Tulip WR, Harley VR, McKimm-Breschkin JL, Laver WG, Webster RG, Colman PM. The structure of a complex between the NC10 antibody and influenza virus neuraminidase and comparison with the overlapping binding site of the NC41 antibody. Structure 1994;2(8):733–746.
88. Bossart-Whitacker P, Chang CY, Novotny J, Benjamin DC, Sheriff S. The crystal structure of antibody N10–staphylococcal nuclease complex at 2.9 Angstrom resolution. J Mol Biol 1995;253: 559–575.
89. Ban N, Escobar C, Garcia R, Hasel K, Day J, Greenwood A, McPherson A. Crystal structure of an idiotype–anti-idiotype Fab complex. Proc Natl Acad Sci USA 1994;91(5):1604–1608.
90. Fields BA, Goldbaum FA, Ysern X, Poljak RJ, Mariuzza RA. Molecular basis of antigen mimicry by an anti-idiotype. Nature 1995;374:739–742.
91. Housset D, Mazza G, Grégoire C, Piras C, Malissen B, Fontecilla-Camps JC. The three-dimensional structure of a T-cell antigen receptor $V_\alpha V_\beta$ heterodimer reveals a novel arrangement of the V_β domain. EMBO J 1997;16:4205–4216.
92. Derrick JP, Wigley DB. The third IgG binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab. J Mol Biol 1994;243(5):906–918.
93. Gallagher T, Oliver J, Bott R, Betzel C, Gilliland GL. Subtilisin BPN′ at 1.6 Å resolution: Analysis for discrete disorder and comparison of crystal forms. Acta Cryst 1996;D52(6):1125–1135.
94. Takeuchi Y, Nonaka T, Nakamura KT, Kojima S, Miura K, Mitsui Y. Crystal structure of an engineered subtilisin inhibitor complexed with bovine trypsin. Proc Natl Acad Sci USA 1992; 89(10):4407–4411.
95. Bode W, Chen Z, Bartels K, Kutzbach C, Schmidt-Kastner G, Bartunik H. Refined 2 Å X-ray crystal structure of porcine pancreatic kallikrein A, a specific trypsin-like serine proteinase. Crystallization, structure determination, crystallographic refinement, structure and its comparison with bovine trypsin. J Mol Biol 1983;164(4):237–282.
96. Parkin S, Rupp B, Hope H. Structure of bovine pancreatic trypsin inhibitor at 125 K: Definition of carboxyl-terminal residues Gly57 and Ala58. Acta Cryst 1996;D52(1):18–29.
97. Walter J, Steigemann W, Singh TP, Bartunik H, Bode W, Huber R. On the disordered activation domain in trypsinogen: chemical labelling and low-temperature crystallography. Acta Cryst 1982; B38:1462–1472.
98. Freer ST, Kraut J, Robertus JD, Wright HT, Xuong NH. Chymotrypsinogen: 2.5-Å crystal structure, comparison with α-chymotrypsin, and implications for zymogen activation. Biochemistry 1970;9(9):1997–2009.
99. Hecht HJ, Szardenings M, Collins J, Schomburg D. Three-dimensional structure of a recombinant variant of human pancreatic secretory trypsin inhibitor (Kazal Type). J Mol Biol 1992; 225(4):1095–1103.
100. Blevins RA, Tulinsky A. The refinement and the structure of the dimer of α-chymotrypsin at 1.67-Å resolution. J Biol Chem 1985;260(7):4264–4275.
101. Bode W, Epp O, Huber R, Laskowski M, Ardelt W. The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3). Eur J Biochem 1985;147(2):387–395.
102. Mauguen Y, Hartley RW, Dodson EJ, Bricogne GG, Chothia C, Jack A. Molecular structure of a new family of ribonucleases. Nature 1982;297:162–164.
103. Ratnaparkhi GS, Ramachandran S, Udgaonkar JB, Varadarajan R. Discrepancies between the NMR and X-ray structures of uncomplexed barstar: analysis suggests that packing densities of protein structures determined by NMR are unreliable. Biochemistry 1998;37:6958–6966.
104. Jia Z, Quail JW, Waygood EB, Delbaere LTJ. The 2.0-Å resolution structure of *escherichia coli* histidine-containing phosphocarrier protein HPr: A redetermination. J Biol Chem 1993;268(30): 22490–22501.
105. Maenaka K, Matsushima M, Song H, Sunada F, Watanabe K, Kumagai I. Dissection of protein–carbohydrate interactions in mutant hen egg-white lysozyme complexes and their hydrolytic activity. J Mol Biol 1995;247(2):281–293.
106. Lescar J, Souchon H, Alzari PM. Crystal structures of pheasant and guinea fowl egg-white lysozymes. Protein Sci 1994;3(5):788–798.