and the form of π given in Table 3 then show that

$$P_1 = [2\pi_{2311}\sigma_{12} + 2\pi_{2323}\sigma_{31}]H_3$$
$$P_2 = [\pi_{2311}(\sigma_{11} - \sigma_{22}) + 2\pi_{2323}\sigma_{23}]H_3$$
$$P_3 = [\alpha_{33} + \pi_{3311}(\sigma_{11} + \sigma_{22}) + \pi_{3333}\sigma_{33}]H_3,$$

i.e. only the piezomagnetoelectric effect contributes to P_1 and P_2. Also $U_2N_2P_2$, U_2N_2S, $Nb_2Co_4O_9$ and $Nb_2Mn_4O_9$ have symmetry $\bar{3}'m'$ according to Oleś *et al.* (1976).

The author is indebted to Professor H. Schmid for his suggestion to investigate the form of the piezomagnetoelectric effect and for stimulating discussions.

References

AIZU, K. (1970). *Phys. Rev. B*, **2**, 754-772.
GRIMMER, H. (1991*a*). *Acta Cryst.* A**47**, 226-232.
GRIMMER, H. (1991*b*). *Helv. Phys. Acta*, **64**, 187-188.
KOPSKÝ, V. (1979). *Acta Cryst.* A**35**, 95-101.
LYUBIMOV, V. N. (1965). *Kristallografiya*, **10**, 520-524; *Sov. Phys. Crystallogr.* **10**, 433-436.
MARKELOV, V. A., NOVIKOV, M. A. & TURKIN, A. A. (1977). *Pis'ma Zh. Eksp. Teor. Fiz.* **25**, 404-407; *JETP Lett.* **25**, 378-380.
OLEŚ, A., KAJZAR, F., KUCAB, M. & SIKORA, W. (1976). *Magnetic Structures Determined by Neutron Diffraction.* Warsaw: Państwowe Wydawnictwo Naukowe.
RADO, G. T. (1962). *Phys. Rev.* **128**, 2546-2556.
SCHMID, H. (1973). *Int. J. Magn.* **4**, 337-361.
SIROTINE, Y. & CHASKOLSKAÏA, M. (1984). *Fondements de la Physique des Cristaux.* Moscow: Mir.
STEFANAKOS, E. K., TINDER, R. F. & THAPLIYAL, H. V. (1979). *J. Phys. C*, **12**, 4921-4925.

Acta Cryst. (1992). A**48**, 271-276

A Resolution-Sensitive Procedure for Comparing Protein Surfaces and its Application to the Comparison of Antigen-Combining Sites

BY MARK GERSTEIN

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England, and University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

(*Received* 9 *July* 1991; *accepted* 22 *October* 1991)

Abstract

Resolution is a crucial parameter to consider in making surface comparisons. A method is presented here for the rapid, objective and automatic comparison of selected parts of protein surfaces as a function of resolution using differences and correlations of Fourier coefficients. A test-case application of this procedure to the surfaces of five immunoglobulin antigen-combining sites allowed them to be partitioned into two categories.

1. Introduction

Knowledge of the topography of protein surfaces is essential to understanding molecular recognition. A solved X-ray crystal structure contains a wealth of information about these surfaces and using modern graphics technology it is relatively easy to examine and compare them in all their detail (*e.g.* Max, 1984; Connolly, 1983*a*). However, comparisons based on human observation are subjective, qualitative and time-consuming. The number of solved structures is continually increasing and so are the methods for generating new conformations and surface descriptions from these structures - *e.g. CONGEN* (Brucoleri & Karplus, 1987) and the molecular surface (Richards, 1977). The time necessary to look through a comprehensive sample of surfaces can be prohibitive. Consequently, it is expedient to develop an objective and quantitative procedure for comparing protein surfaces rapidly and automatically.

An essential parameter to consider in any surface-shape comparison method is resolution. Surfaces - such as those of two human faces - that are different in medium-resolution detail may have similar low-resolution features and high-resolution texture. By comparing surfaces in terms of Fourier coefficients, one naturally obtains information ordered in terms of increasing resolution.

We are specifically interested in immunoglobulin recognition and have tested our approach by comparing the surfaces of antigen-combining sites. The problem of surface comparisons is particularly evident for these molecules, since recombination of a small number of genes can produce an estimated 10^8 different antigen-combining sites, each with distinctly different surface-recognition properties (Milstein, 1990). However, before the results of the procedure on this specific case is described, it is presented in a more general context.

0108-7673/92/030271-06$03.00

2. Method

The procedure consists of four steps:

1. generating molecular envelopes;
2. superimposing the envelopes;
3. Fourier transforming them;
4. comparing the transforms at various resolutions.

The surface of a protein is represented here by a selected section of its molecular envelope, which is prepared in such a way so as to remove as much detail as possible that is not directly related to the protein surface. An initial density map $\rho_0(\mathbf{r})$ is generated using a Gaussian approximation for atomic shapes (Diamond, 1971). It is assumed to be represented on a finely spaced grid and the following operations are to be performed simultaneously at every point on this grid. First the map is thresholded at a suitable value τ. The technique of simple region growing with expansion and contraction operations is used to fill cavities inside the envelope [see Rosenfeld & Kak (1982) for a general discussion of this technique]. The expansion operation, denoted by **E**, fills an empty grid point (value 0) if it is adjacent to a filled one (value 1) and the contraction operation, **Q**, empties a filled grid point if it is next to an empty one. The operations are constructed so that m expansions, denoted $\mathbf{E}^m$, followed by the same number of contractions, $\mathbf{Q}^m$, returns the original shape unless one or more holes or cavities are completely filled during the expansion. Holes and cavities of arbitrary size can be filled by choosing large enough values of m. To arrive at a final envelope $\rho(\mathbf{r})$, the filled envelope is clipped by a cubic boundary $B(\mathbf{r}_0, l)$ of side l centered at $\mathbf{r}_0$ on the area of interest.

$$\rho(\mathbf{r}) = B(\mathbf{r}_0, l)\mathbf{Q}^m\mathbf{E}^m\mathbf{T}[\rho_0(\mathbf{r})] \qquad (1)$$

where

$$\mathbf{T}(\rho) = \begin{cases} 0 & \text{if } \rho < \tau \\ 1 & \text{otherwise,} \end{cases}$$

$$\mathbf{E}(\rho) = \begin{cases} 1 & \text{if } \rho = 1 \text{ for a grid point in } G(\mathbf{r}) \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathbf{Q}(\rho) = \begin{cases} 0 & \text{if } \rho = 0 \text{ for a grid point in } G(\mathbf{r}) \\ 1 & \text{otherwise} \end{cases}$$

and $G(\mathbf{r})$ is a set of 27 grid points, which comprises a central point nearest $\mathbf{r}$ and the 26 points neighboring it.

Note that the final 'envelope' ρ is a binary function on a three-dimensional volume and not on a two-dimensional surface. It, nevertheless, only contains information about the protein surface. It could have been generated by a number of other procedures – in particular, simply filling in the region enclosed by a Connolly surface (Connolly, 1983*b*). A simple binary envelope and clipping boundary provide enough surface detail for the comparisons discussed here. It is, however, straightforward to generalize the envelope to a potential that takes on a continuum of values outside the protein surface, *e.g.* to represent fully the electrostatic potential from a program such as *DELPHI* (Sharp & Honig, 1990).

In making a number of surface comparisons, it was found best to compare the envelope of a particular surface ρ with a common reference ρ_R. These envelopes must first be aligned. Superimposing density maps involves a time- and memory-consuming multidimensional search (*e.g.* Bricogne, 1976; Muirhead, Cox, Mazzarella & Perutz, 1967). To sidestep this issue, it is convenient to assume that envelope volumes will be optimally aligned when the protein coordinates from which they are derived are intelligently superimposed. Superimposing coordinate sets is a straightforward and relatively rapid calculation (McLachlan 1972, 1979; Diamond, 1976, 1989; Lesk, 1986; Kearsley, 1989). To perform the superposition, one set of coordinates was chosen as the reference and the others were fit to it. A number of recently developed procedures (Diamond, 1988, 1992; Kearsley, 1990) may make it possible to perform this coordinate superposition without choosing a reference, but it will still be necessary to decide which atoms are equivalent for structures without sequence identity.

The superimposed envelopes are then convolved with a three-dimensional Gaussian of half-width σ to smooth the sharp edges introduced by the thresholding and clipping. They are embedded in a suitable cell and Fourier transformed to get the complex Fourier coefficients $\mathbf{F}(\mathbf{s})$ for the surface and $\mathbf{R}(\mathbf{s})$ for the reference. In practice the convolution is performed by reciprocal-space multiplication,

$$\mathbf{F}(\mathbf{s}) = \exp[-2(\pi s\sigma)^2] \int_{\text{cell}} \rho(\mathbf{r}) \exp(2\pi i\mathbf{r}\cdot\mathbf{s})\mathrm{d}^3\mathbf{r}. \qquad (2)$$

Of the many possible statistics for comparing Fourier coefficients, two were found to be particularly useful. The first is the mean square of the difference in the coefficients as a function of resolution $S = |\mathbf{s}|$

$$D(S) = \langle(\mathbf{F}-\mathbf{R})(\mathbf{F}-\mathbf{R})^*\rangle. \qquad (3)$$

Note this is a vector difference involving both amplitude and phase. It is defined at a constant magnitude of $\mathbf{s}$, at which averaging is performed over all reciprocal-lattice directions. In practice, one constructs a spherical shell of radius S and thickness Δs centered on the origin in reciprocal space and averages over the $N(S)$ Fourier coefficients contained in the shell. The difference D differs from the conventional χ^2 statistic [§ 14.1 in Press, Flannery, Teukolsky & Vetterling (1988)] that would be computed to fit the molecular envelopes, *i.e.* $\chi^2(S) = N(S)D(S)$. Since $N(S) \propto S^2$, D gives more weight to the low-resolution terms, which are of particular interest. Notice that, if the envelope ρ is dimensionless, the

Fourier coefficients **F** and **R** have the dimension of volume with the first coefficient equal to the volume of the envelope: $\mathbf{F}(0) = \int \rho \, d^3\mathbf{r} = V$. Consequently, the difference D is measured in units of volume squared. It incorporates the most straightforward measure of shape difference, the difference in volume with respect to the reference: $D(0) = (V - V_R)^2$.

The second statistic is a linear correlation coefficient on both amplitude and phase, which can be generalized from Pearson's scalar correlation coefficient,

$$C(S) = \langle \delta\mathbf{F}\delta\mathbf{R}^* + \delta\mathbf{F}^*\delta\mathbf{R} \rangle / 2\sigma_F\sigma_R, \quad (4)$$

where the deviation from the mean $\delta\mathbf{F}(\mathbf{s}) = \mathbf{F} - \langle\mathbf{F}\rangle$, the standard deviation $\sigma_F = \langle \delta\mathbf{F}\delta\mathbf{F}^* \rangle^{1/2}$, and $\delta\mathbf{R}$ and σ_R are defined analogously. The correlation coefficient is a dimensionless number normalized to lie between -1 and 1. It is invariant under scaling, *i.e.* replacing **F** by $\alpha\mathbf{F} + \beta$ for constant α and β.

This implicit scaling inherent in using the correlation coefficient has advantages in some circumstances, but it can often obscure large-scale differences in shapes. This fact is manifest by explicitly relating the correlation and the difference,

$$D = (\langle\mathbf{F}\rangle - \langle\mathbf{R}\rangle)^2 + (\sigma_F - \sigma_R)^2 + 2(1 - C)\sigma_F\sigma_R. \quad (5)$$

The first and second terms on the right-hand side of the equation are differences in the means and standard deviations of coefficient distributions. D is more sensitive to these differences than C. That is, at a given resolution, it is more sensitive to the gross characteristics of the two surfaces. It is better suited for comparing surfaces different in size. The correlation C, in contrast, is best for comparisons of coefficient distributions at resolutions where their first and second moments are similar but their higher moments are different. That is, it is best for comparing surfaces at resolutions where they are similar in size but different in shape. In this case, its normalization makes it more readily interpretable than D. In terms of the analogy to facial recognition made earlier, D is better for comparing a bird's face with a human face, while C is better suited to comparing two human faces.

3. Implementation and results

The above procedure was implemented* and used to compare antigen-combining sites. Five crystal structures were used: REI (Epp, Latham, Schiffer, Huber & Palm, 1975), D1.3 (Amit, Mariuzza, Phillips & Poljak, 1986), HyHEL-10 (Padlan *et al.*, 1989), McPC603 (Satow, Cohen, Padlan & Davies, 1986) and 17/9 (J. Rini & I. Wilson, personal communication). Complete antibody molecules have six hypervariable loops that form a binding site for an antigen. However, the comparisons here were restricted to the surface formed by the three loops that are part of the light chain. The main-chain conformations of the structures fall into two classes. Corresponding hypervariable loops in REI, D1.3 and HyHEL-10 have the same canonical structure and roughly the same main-chain conformation (Chothia *et al.*, 1989). In contrast, 17/9 (Schulze-Gahmen *et al.*, 1988) and McPC603 have an eight-residue insertion in the L1 loop that gives them a very different main-chain conformation.

REI was chosen as the reference and the other four structures were superposed onto it using the conserved atoms of the central β-sheet framework in the V_L domain (Chothia & Lesk, 1987). Density maps were generated on a grid with spacings of $\frac{1}{3}$ Å, thresholded at $\tau = 0.002$ electrons Å^{-3} and filled by region growing. A cube of side 29 Å centered on the three loops was used to clip the molecular envelope ρ, as shown in Fig. 1(*a*). Note that most of the atoms

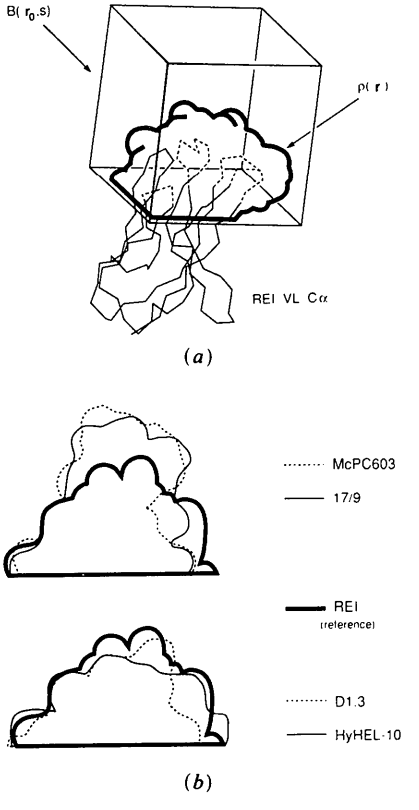

Fig. 1. (*a*) A view of principal constructions used in the surface comparisons. C^α trace of V_L domain of REI. The solid part of the trace was used for superpositions, while the dotted part was not. The box shows the clipping boundary $B(\mathbf{r}_0, l)$ and the final envelope volume $\rho(\mathbf{r})$. (*b*) A selected cross section through molecular envelopes $\rho(\mathbf{r})$ of the five immunoglobins shows that McPC603 and 17/9 have significantly different surface shapes from REI, HyHEL-10 and D1.3.

* The programs to perform the surface comparisons were written in ANSI C and Fortran77 and run under Unix on a Silicon Graphics Iris. Typically, comparing a new surface to the reference took about 220 s of CPU time, half of which was devoted to calculating the Fourier transform.

used in the superposition were not contained within the clipping boundary. As a result of their different main-chain conformations, 17/9 and McPC603 have substantially larger differences in envelope shape from REI than have HyHEL-10 and D1.3 (Fig. 1*b*).

The envelope volume ρ of each surface was embedded in the center of a cell, convolved with a Gaussian of half-width 3 Å, and Fourier transformed in space group $P1$ by standard crystallographic programs (*CCP*4, 1991). The differences D and the correlations C were computed as discussed above. Ideally, these statistics should not depend on the exact details of how the Fourier transform is made, *i.e.* on the size of the cell or its orientation and origin with respect to the envelope. On the one hand, they were found to vary slightly with different types of sampling. To eliminate this problem, a standard orientation was used. The envelope was centered in a much bigger cell (of side 60 Å), which had its coordinate axes parallel to the axes of the clipping boundary $B(\mathbf{r}_0, l)$. On the other hand, the difference and correlation were much less affected by the details of the transform than the standard crystallographic statistics on amplitudes.

In practice, the computation of C using (4) is awkward. It is necessary to sum twice over reciprocal space and these sums must explicitly include a coefficient and its complex conjugate. Taking advantage of the fact that all the averages must be real, it can be rearranged into a more useful form:

$$\begin{aligned} C(S) = {} & [n\sum(ac+bd)-\sum a\sum c] \\ & \times[n\sum(a^2+b^2) - (\sum a)^2]^{-1/2} \\ & \times[n\sum(c^2+d^2) - (\sum c)^2]^{-1/2} \end{aligned} \quad (6)$$

where $\mathbf{F}(\mathbf{s}) = a(\mathbf{s}) + ib(\mathbf{s})$ and $\mathbf{R}(\mathbf{s}) = c(\mathbf{s}) + id(\mathbf{s})$. In this simplified expression, the summations need only include one of a conjugate pair and are made over the $n = N(S)/2$ unique coefficients in each shell.

As is evident in Fig. 2(*a*), the difference D distinguishes between two classes of surfaces at low resolution. In fact, the classification is clearer in reciprocal space than in real space from visual inspection of Fig. 1(*b*). The analysis can be further simplified by just listing the overall low-resolution (500–12 Å) D for each surface (in units of 10 000 Å^6). This is 26 for McPC603, 24 for 17/9, 5.1 for HyHEL-10 and 3.9 for D1.3. On the basis of these four numbers a person (or an algorithm) could classify 17/9 and McPC603 differently from D1.3 and HyHEL-10. With regard to immunoglobulin structure, this sort of classification is useful insofar as it demonstrates that hypervariable regions with the same main-chain structure can also be classified as having similar overall surface shape.

As shown in Fig. 2(*b*), the correlation C provides more detail at higher resolutions. McPC603 and 17/9, for instance, which are similar at low resolution, become progressively more different at resolutions higher than 10 Å. However, the magnitude of this higher-resolution difference is much less than that of the low-resolution difference between the surface classes. This is why, referring to equation (5) relating D and C, it is better to use the correlation at the higher resolution and the difference at the lower one.

4. Discussion

A method for comparing protein surfaces in a resolution-sensitive fashion has been presented. Other approaches to surface comparison are possible. They differ from the method presented either in the way the surface is represented or in how the comparison statistic is determined.

The conventional three-dimensional Fourier series in Cartesian space, $\exp[2\pi i(hx+ky+lz)]$, was used here to represent surface shape. There are, however, other resolution-dependent basis sets that could have been used. Representations in terms of spherical harmonics $Y_m^l(\theta, \varphi)$, which are Fourier series in spherical coordinates, have been attempted (Leicester, Finney & Bywater, 1988; Max & Getzoff, 1988). These do not require the choice of an origin, *i.e.* the surfaces to be compared need not be superposed on their main chains. However, the representation chosen in these papers is two-dimensional. Since real protein surfaces have a fractal dimension between 2 and 3 (Lewis & Rees, 1985), this representation cannot describe all aspects of a protein surface - in particular, re-entrant surfaces, such as cavities and crevices, which are often most interesting. Nor can it be used on surface descriptions more complex than simple envelopes, such as complete electrostatic potentials. Furthermore, expansion in terms of spherical harmonics is much slower than a fast-Fourier transform (FFT) at present.

Spherical harmonics and trigonometric functions are ultimately geared toward the representation of continuous data, whereas the envelope ρ used here is discrete and, in fact, simply two-valued. Consequently, a number of schemes specially designed for resolution-dependent digital-image compression are especially appealing (Wintz, 1969). Hadamard transforms, which can be completely expressed in terms of binary operators, would be significantly faster than an FFT, and Karhunen-Loève transforms, which have coefficients with statistically minimal correlation, would achieve better compression.

All the surface representations discussed so far are tied to a choice of basis and consequently a coordinate system. An optimum representation scheme would represent the protein surface in a coordinate-free manner. Connolly (1986) has proposed such a measure based on determining the solid angle Ω when the molecular surface is intersected with a probe sphere. While one can vary the radius of the probe

sphere, the solid-angle method is not resolution dependent in the sense of the Fourier-series method presented here. It provides an intrinsic measure of protein shape rather than a hierarchically ordered set of numbers that can be used in a metric for surface comparison.

Turning now to alternate comparison statistics, one finds that they fall into three categories. First are the real-space measures that only refer to the density ρ. These include a residual (Jones, Zou, Cowan & Kjeldgaard, 1991) and a variety of differently normalized correlations (*e.g.* Zhang & Main, 1990; Carbo, Leyda & Arnau, 1980; Hodgkin & Richards, 1987). These measures are straightforward to implement and interpret, but it is difficult to see how they vary with resolution. A second class of measures comprises the standard statistics used during crystallographic structure determination: residuals and correlations on amplitudes and average changes in phases. These suffer (in the context of surface comparisons) because they discard much important information by not including both amplitude and phase on the same

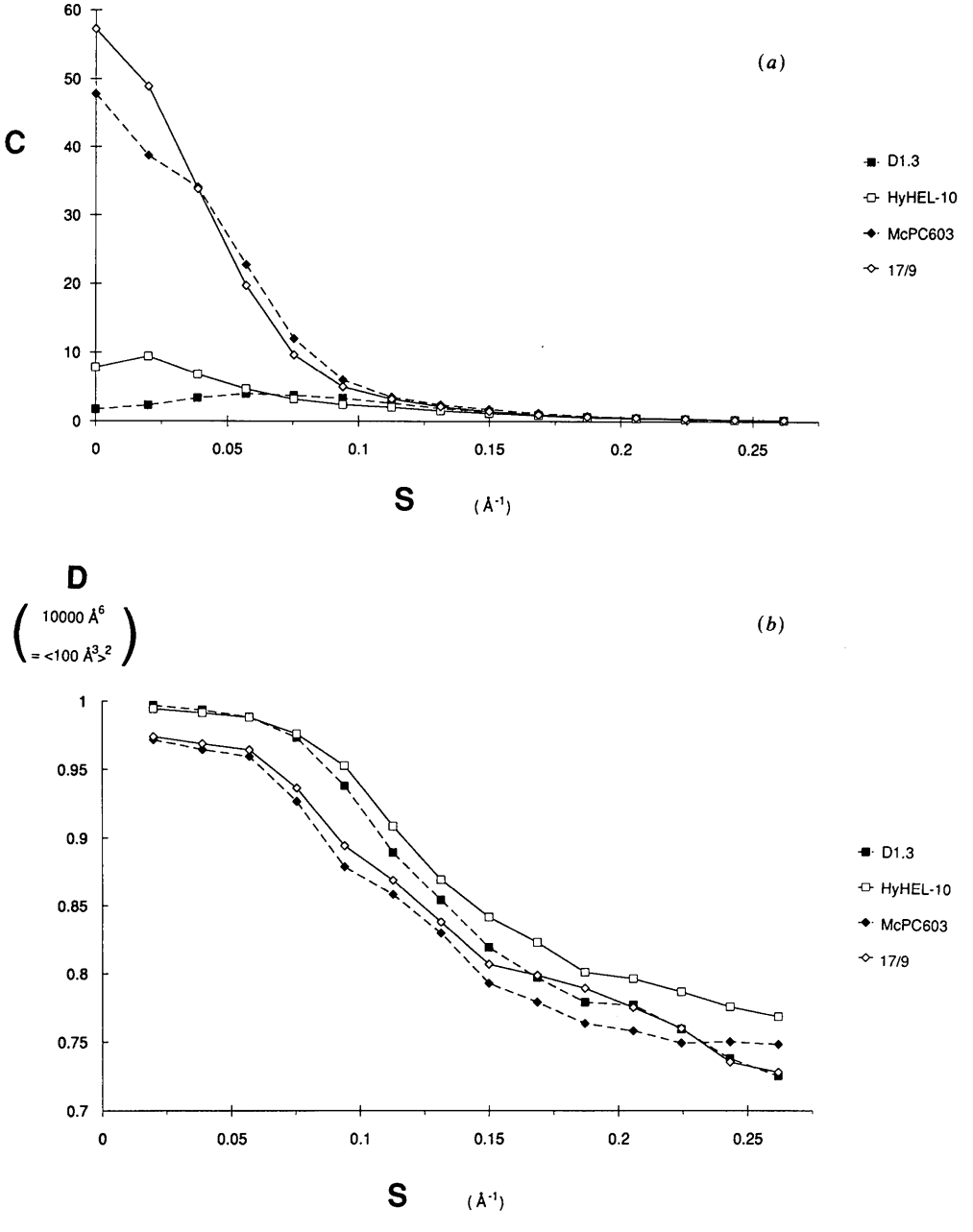

Fig. 2. Plots of (*a*) the difference *D* and (*b*) the correlation *C versus* shell radius *S*. All differences are with respect to the REI reference surface.

footing. The third and final category includes non-parametric statistics such as rank correlations (Namasivayam & Dean, 1986). For the price of using less information from the initial distributions, these provide greater reliability in associating a significance with their particular value. However, this feature is not particularly relevant here, since the significance or insignificance of one comparison is determined by judging it in the context of the other comparisons with the reference structure.

Irrespective of the details of the representation or comparison statistic chosen, there are many applications for comparing surfaces in a resolution-sensitive fashion. As an illustration, consider some applications to immunoglobulin structure. At present there are roughly 15 structures containing the three hypervariable loops compared here. For each structure it is possible to generate surfaces representing Lennard-Jones and electrostatic potentials. Do two surfaces that have the same shape classification with respect to one potential have the same classification with respect to the other? The procedure presented could be used automatically to classify these 30 surfaces on the basis of their low-resolution differences D. A second low-resolution-classification application is assessing the hypothetical loop conformations generated by a program such as *CONGEN* (Bruccoleri & Karplus, 1987), which has already been applied with some success in the prediction of hypervariable loop structure (Bruccoleri, Haber & Novotny, 1988; Martin, Cheetham & Rees, 1989). Surface comparisons could be used to screen automatically generated conformations for matching certain key surface features, such as a large central cavity. By simply inverting the envelopes ρ it would also be possible to make comparisons with the complement of a combining-site surface, an antigen.

The surface-comparison methods presented here could also be useful for categorizing surfaces at higher resolutions. There has been considerable interest in measuring the local roughness of a protein surface since it may be related to substrate capture and antigen binding. Measurements and comparisons based on the fractal dimensionality have been tried (*e.g.* Pfeifer, Welz & Wipperman, 1985; Lewis & Rees, 1985; Aqvist & Tapia, 1987). However, comparisons of surface texture could also be easily made by assessing the relative mix of low- and high-frequency Fourier coefficients in the methods presented here.

Thanks are given to J. Rini and I. Wilson for providing the 17/9 coordinates prior to publication. Comments on the manuscript from R. Diamond, A. M. Lesk, A. D. McLachlan and especially R. M. Lynden-Bell and C. Chothia are very much appreciated. Computing assistance from P. R. Evans, helpful suggestions from G. Bricogne and support from a Herchel-Smith Fellowship are acknowledged.

References

AMIT, A. G., MARIUZZA, R. A., PHILLIPS, S. E. V. & POLJAK, R. J. (1986). *Science*, **233**, 747-758.

AQVIST, J. & TAPIA, O. (1987). *J. Mol. Graph.* **5**, 31-34.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832-847.

BRUCCOLERI, R. E., HABER, E. & NOVOTNY, J. (1988). *Nature (London)*, **335**, 564-568.

BRUCCOLERI, R. E. & KARPLUS, M. (1987). *Biopolymers*, **26**, 137-168.

CARBO, R., LEYDA, L. & ARNAU, M. (1980). *Int. J. Quantum Chem.* **17**, 1185.

*CCP*4 (1991). The Cooperative Computing Project in Crystallography. SERC Daresbury Laboratory, Warrington, England.

CHOTHIA, C. & LESK, A. M. (1987). *J. Mol. Biol.* **196**, 901-917.

CHOTHIA, C., LESK, A. M., TRAMONTANO, A., LEVITT, M., SMITH-GILL, S. J., AIR, G., SHERIFF, S. PADLAN, E. A., DAVIES, D., TULIP, W. R., COLMAN, P. M., SPINELLI, S., ALZARI, P. M. & POLJAK, R. J. (1989). *Nature (London)*, **342**, 877-883.

CONNOLLY, M. (1983*a*). *Science*, **221**, 709-713.

CONNOLLY, M. (1983*b*) *J. Appl. Cryst.* **16**, 548-558.

CONNOLLY, M. (1986). *J. Mol. Graph.* **4**, 3-6.

DIAMOND, R. (1971). *Acta Cryst.* A**27**, 436-52.

DIAMOND, R. (1976). *Acta Cryst.* A**32**, 1-10.

DIAMOND, R. (1988). *Acta Cryst.* A**44**, 211-216.

DIAMOND, R. (1989). *Acta Cryst.* A**45**, 657.

DIAMOND, R. (1992). In preparation.

EPP, O., LATHAM, E., SCHIFFER, M., HUBER, R. & PALM, W. (1975). *Biochemistry*, **14**, 4934-4952.

HODGKIN, E. E. & RICHARDS, W. G. (1987). *Int. J. Quantum Chem. Symp.* **14**, 105.

JONES, T. A., ZOU, J. Y., COWAN, S. W. & KJELDGAARD, M. (1991). *Acta Cryst.* A**47**, 110-119.

KEARSLEY, S. K. (1989). *Acta Cryst.* A**45**, 208-210.

KEARSLEY, S. K. (1990). *J. Comput. Chem.* **11**, 1187-1192.

LEICESTER, S. E., FINNEY, J. L. & BYWATER, R. P. (1988). *J. Mol. Graph.* **6**, 104-108.

LESK, A. M. (1986). *Acta Cryst.* A**42**, 110-113.

LEWIS, M. & REES, D. C. (1985). *Science*, **230**, 1163-1165.

MCLACHLAN, A. D. (1972). *Acta Cryst.* A**28**, 656-657.

MCLACHLAN, A. D. (1979). *Acta Cryst.* A**38**, 871-873.

MARTIN, A. C. R., CHEETHAM, J. C. & REES, A. R. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 9268-9272.

MAX, N. L. (1984). *J. Mol. Graph.* **2**, 8-13.

MAX, N. L. & GETZOFF, E. D. (1988). *IEEE Comput. Graph. Appl.* **8**, 42-50.

MILSTEIN, C. (1990). *Proc. R. Soc. London Ser B*, **239**, 1-16.

MUIRHEAD, H., COX, J. M., MAZZARELLA, C. & PERUTZ, M. F. (1967). *J. Mol. Biol.* **28**, 117-156.

NAMASIVAYAM, S. & DEAN, P. M. (1986). *J. Mol. Graph.* **4**, 46-50.

PADLAN, E. A., SILVERTON, E. W., SHERIFF, S., COHEN, G. H., SMITH-GILL, G. S. & DAVIES, D. R. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 5938-5942.

PFEIFER, P., WELZ, U. & WIPPERMAN, H. (1985). *Chem. Phys. Lett.* **113**, 535-540.

PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1988). *Numerical Receipes in C.* Cambridge Univ. Press.

RICHARDS, F. M. (1977). *Annu. Rev. Biophys. Bioeng.* **6**, 151-176.

ROSENFELD, A. & KAK, A. C. (1982). *Digital Picture Processing*, Vol. 2, 2nd ed. New York: Academic Press.

SATOW, Y., COHEN, G. H., PADLAN, E. R. & DAVIES, D. R. (1986). *J. Mol. Biol.* **190**, 593-604.

SCHULZE-GAHMEN, U., RINI, J. M., AREVALO, J., STURA, E. A., KENTEN, J. H. & WILSON, I. A. (1988). *J. Biol. Chem.* **263**, 17100-17105.

SHARP, K. A. & HONIG, B. (1990). *Annu. Rev. Biophys. Biophys. Chem.* **19**, 301-332.

WINTZ, P. A. (1969). *Proc. IEEE*, **60**, 809-820.

ZHANG, K. Y. J. & MAIN, P. (1990). *Acta Cryst.* A**46**, 41-46.