A new tool for the qualitative and quantitative analysis of protein surfaces using B-spline and density of surface neighborhood

Nathalie Colloc'h and Jean-Paul Mornon

Laboratoire de Minéralogie-Cristallographie, Universités de Paris VI et VII, CNRS URA09, Paris, France

To improve the qualitative and quantitative analysis of surfaces of protein, two new methods are proposed: one that smoothes the MS surface of Connolly with B-spline smoothing functions to highlight the significant features of the surface, and one that computes the density of surface neighborhood to allow quantitative comparison.

Keywords: proteins, protein surfaces, B-spline smoothing function, graphic display, concavity, convexity

INTRODUCTION

The shape of a protein is an important factor in interactions with other components of living cells. The high specificity of molecular recognition depends on several interactions. Shape complementarity is a convenient representation of these effects. Chothia and Janin[1] have shown that while hydrophobicity is the major factor stabilizing protein–protein complexes, the specificity of recognition requires a close-packed interface with hydrogen bonds and van der Waals contacts. During automatic docking procedures, shape complementarity has been used as a first test to choose between different modes of association[2] or as a check of the goodness of fit.[3] It has also been used alone without evaluation of chemical interactions to search among a data base of small molecules for ligands to a known receptor binding site.[4] Shape complementarity also helps to predict affinity of analogs, the best analog being the one with the best fit at the interface.[5] In bioorganic chemistry, molecular recognition requires complementarity in size, shape, and functionalities.[6] We therefore decided to analyze and compare a number of protein surfaces, qualitatively and quantitatively, in an attempt to find new clues for molecular recognition and also structural homologies between various proteins not related by primary nor secondary structure. For example, the relative orientation of the two pockets and the cavity of uteroglobin[7] is similar to the relative orientation of the two active site pockets and one of the two cavities of phospholipase A_2.[8] This could explain why uteroglobin inhibits phospholipase A_2 activity.[9,10] This observation would suggest that the shape homology might explain some functional properties of proteins.

The surface of a protein can be defined by a probe rolling over the molecule whose atoms are described by van der Waals spheres.[11] With an infinitely small probe sphere, a rough surface equivalent to the van der Waals surface is defined. This is not especially useful for macromolecules. With a radius of 1.4 Å, the probe sphere approximates a molecule of water. The solvent-accessible surface[12] is defined by the locus of the center of the probe rolling on the atoms. The molecular surface[13] is defined by the combination of the contact surface (the part of the van der Waals surface that is accessible to the probe sphere) and the reentrant surface (the part of the probe sphere looking toward the molecule when it is in contact with more than one atom). A larger probe has been used to help locate antigenic epitopes,[14,15] with antigens being approximated by a probe sphere of around 10 Å in radius.

The numerical calculation of the molecular surface is done with the program MS[16] of M.L. Connolly. MS describes the surface with regions of different shapes, convex, saddle-shaped, or concave, according to the translational degree of freedom of the probe sphere. This representation of molecular surfaces is very effective for detailed surface examination but it is not so useful for an overall view. The few significant features are often buried in many atomic size bumps. Computing a molecular surface with a larger probe sphere smoothes the surface but removes the pocket.

Address reprint requests to Dr. Mornon at the Laboratoire de Minéralogie-Cristallographie, Universités de Paris VI et VII, CNRS URA09, Tour 16, 4 place Jussieu, 75252 Paris Cedex 05, France.

Dr. Colloc'h's present address is the Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, USA.
Received 11 April 1990; accepted 1 May 1990

Therefore, we decided to create a new tool for describing protein surfaces that removes the small atomic bumps by smoothing the surface without eliminating pockets or pits.

A new method for automatically highlighting the significant features of a surface is described that facilitates quantitative comparisons between protein surfaces. We have not chosen a cartographic method where the surface is projected onto an ellipsoid[17,18] because this always adds some distortion to the quantitative data, making comparison more difficult. Instead, we used an approach that determines local concavity and convexity by defining the density of surface neighborhood.

METHODS

Qualitative description

Description within a cubic grid A cubic grid was used to describe the molecular surface. This representation facilitates the application of one-dimensional smoothing functions. This approach was used by Stouch and Jurs[19] for the rapid comparison of the surface and the volume of a small molecule.

In the first step, calculation of the molecular surface is performed with Connolly's MS algorithm,[16] with a density of 4 or 5 points per square angstrom. Then, the surface is cut by a succession of parallel equidistant planes, orthogonal to one of the basic axes X, Y, or Z of the protein. All the points included between two parallel planes, typically 1 Å apart, are projected onto the lower plane. For each plane, the projected points are sorted into a square grid. For the small molecules, of less than 300 atoms, we used a grid of 0.5 Å. The projected points are surrounded with a curve that passes through the outermost points. In the case of an internal cavity or two bumps perpendicular to the planes, two or more closed curves could lay on the same plane.

Such a description is quite precise, but it is better to describe the surface with curves in three orthogonal directions, so the whole process is repeated for parallel planes orthogonal to the X, Y, then Z axis. In this way, no information about the shape of the surface is lost. This kind of description is not very helpful for the display on a graphics screen, but is very convenient for the two next steps: smoothing the surfaces (qualitative applications) and counting the structural features (quantitative applications).

The program that describes Connolly's surface within a cubic grid is called SECTION. The number of dots for describing the surface is approximately half the number required for an MS representation, assuming a density of 4 or 5 points per square angstrom, and the calculation of SECTION with a distance between two planes of 1 Å, in the three orthogonal directions. With the conditions described above, the CPU time is about a quarter that of the MS surface calculation.

Smoothing the surface When the surface is described within a cubic grid, it is very easy to smooth each curve with a one-dimensional smoothing function. We use the B-spline smoothing function,[20,21] with the Bernstein basis, to smooth the surface and highlight the significant structural features. It is similar to the ribbon description used to smooth the carbon alpha chain and accentuate the secondary structure.[22]

The relationship between a set of points in a plane that describes the vertices of a polygon and the smooth curve generated by these points is called the basis or the weighting function. The B-spline basis has been chosen because the order of the resulting curve is not linked to the number of vertices that define the curve. Since the number of vertices in each polygon is important, it would have been impossible to use a basis such as the Bezier basis,[23,24] where the order of the resulting curve is equal to the number of vertices of the polygon.

A curve generated with the use of the B-spline basis is given by equation 1 (see Table 1) where $P(t)$ describes the smooth curve as a function of the parameter t, P_i is each of the $n + 1$ points defining n polygon vertices, and $N_{i,k}(t)$ is each of the $n + 1$ weighting functions for a resulting curve of order k. The weighting functions are defined by the recursive formula 2, with x_i an element of a knot vector defined by formula 3. Each x_i is an integer such that $x_i \leq x_i + 1$. The range of the parameter t used to generate a B-spline curve is $0 \leq t \leq t_{max}$, t_{max} being given by equation 4 if there are no duplicate vectors, which means that each vertex is of equal weight. The order of the curve is also reflected in the knot vector with knots of multiplicity k at both the beginning and the end of the knot set.

The smoothness is based on the continuity of higher-order derivatives, which means that the order of the curve determines how smooth the curve is. But, as the order increases, the resulting curve becomes tighter between the beginning and the end points. It is better to choose a low order of smoothing and to reiterate the smoothing function a great number of times than to choose a higher order of smoothing that distorts the shape of the curves too much.

A fourth-order B-spline basis has been chosen to get a curve that is smooth but sufficiently similar to the shape of the defining polygon. To get a smoother curve, the fourth-order smoothing function is iteratively applied to the set of points. After several trials, we found seven iterations gives the best results (see an example with crambin[25] in Color

Table 1. Equations for a B-spline curve[a]

(1) $$P(t) = \sum_{i=0}^{n} P_i N_{i,j}(t)$$

(2) $$N_{i,1}(t) = 1 \quad \text{if } x_i \leq t \leq x_i + 1$$
$$= 0 \quad \text{elsewhere}$$
$$N_{i,j}(t) = \frac{(t - x_i)N_{i,j}(t)}{x_{i+j-1} - x_i} + \frac{(x_{i+j} - t)N_{i+1,j-1}(t)}{x_{i+j} - x_{i+1}}$$

(3) $$x_i = 0 \quad \text{if } i < j$$
$$x_i = i - j + 1 \quad \text{if } j \leq i \leq n$$
$$x_i = n - j + 2 \quad \text{if } i > n$$

(4) $$t_{max} = n - j + 2$$

[a](1) Vectorial equation of a B-spline curve of order k, defined with $n + 1$ points P_i (polygon with n vertices). (2) Weighting functions, defined by a recursive equation $N_{i,k}(t)$ and a initial equation $N_{i,1}(t)$. (3) Equations to determine the knot vector, which is linked to the order of the curve. (4) The parameter t of the curve can vary between 0 and t_{max}.

Plate 1). Since the resulting curve always goes through the first and the last points, these points would become definite points, so it is better to move the beginning and the end points forward between two iterations. Since all the polygons are closed, the first and the last points are in fact identical.

Each polygon of the surface described within a cubic grid is smoothed seven times. The resulting surface is described by a set of smooth curves in the three orthogonal directions. The program that smooths the surface from the one described within a cubic grid is called SURSPLIN. The CPU time is about the same that of the MS surface calculation, under the conditions described above.

Quantitative description

To improve the qualitative analysis and to reinforce it with quantitative values, a program called SURSCOP was developed, based on a simple method for determining the concavity and the convexity on each part of the surface. The general idea is analogous to the van der Waals signature[26]: A probe is moved around the molecule at a given distance from the surface and particular properties are calculated from the whole molecule toward the probe point.

It is computationally more efficient to work with the surface described within a cubic grid because it is easier to perform computer calculations with integral values rather than real values. To get more accurate results, it is better to use a surface described within a cubic grid of 0.5 Å side length.

The probe dots are located above and below the surface at 1 Å from each surface dot to determine concave and convex features. For each probe dot or explored position, the number of surface points within a cube of 7 Å side length centered on the probe dot are counted. The resulting number is called the density of surface neighborhood. So, for each explored position, a density of surface neighborhood is determined and a color is attributed. For example, the probe dots that lie above a very concave area (a pit or a pocket) or below a very convex area (a bump) are both associated with a high-density value.

The probe points that have more than a certain density value are displayed on a graphics screen, superimposed on the smoothed surface. The color of the dots changes with the density value. The external probe points that have a density of surface neighborhood between 90 and 100 surface points per cube are displayed in red, those between 100 and 110 are displayed in yellow, and those above 110 in green. Therefore, red dots are located above large valleys, and yellow and green dots above pockets and pits. The internal dots that have a density of surface neighborhood between 100 and 110 are displayed in pale blue, those between 110 and 120 are displayed in blue, and those above 120 in dark blue. Therefore, blue points are located under bumps; the darker the blue, the more convex the bump. Thus, all the main significant features are highlighted and their comparisons become more objective.

It would have been possible just to use probe dots lying above the surface and keep the dots with the higher density values for the concave region, and similarly keep those with the lower density values for the convex region. However, it was deemed more convenient to draft both the concave and the convex areas with a similar approach.

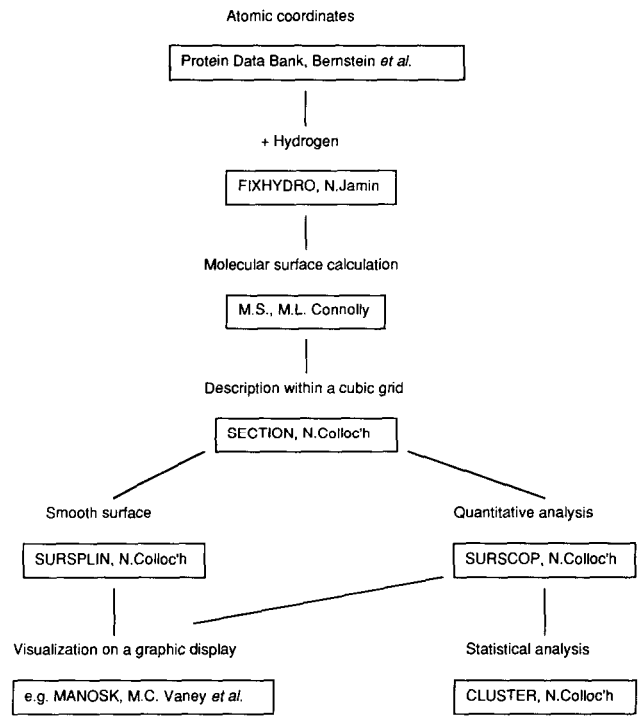

Figure 1. Flow diagram of program linkage

A program called CLUSTER reads the output of SURSCOP, and counts the number of dots for a range of density of surface neighborhood. It performs statistical analysis and quantitative comparisons of the different features on the surface.

The four programs SECTION, SURSPLIN, SURSCOP, and CLUSTER are written in Fortran 77 without extension. The output of the program SECTION, a molecular surface described within a cubic grid, is read by both SURSPLIN and SURSCOP. Both the output of SURSPLIN, a smooth surface described by a set of curves, and the output of SURSCOP, the dots showing the concavity and the convexity of the surface, are then loaded onto a graphic display. In our case, it is loaded on an Evans & Sutherland PS390 with the help of an option of our general purpose modeling software MANOSK.[27,28] The programs are independent of the graphics device or the graphic screen; only a few sets of lines are used for the interface between the output of SURSPLIN or SURSCOP and a graphics software. Figure 1 is a flow diagram of program linkage.

RESULTS

Qualitative results

We obtained atomic coordinates for the proteins we studied from the Protein Data Bank[29,30] at Brookhaven National Laboratory or directly from the crystallographers.

The display of a smooth surface on a graphics screen facilitates extensive studies of the protein surface. The advantages of smooth surfaces are several. To obtain a smooth surface, Connolly's surface has to be calculated using the

MS program, then the surface is described within a cubic grid with the program SECTION and smoothed with the program SURSPLIN. On a Vax Station 3500, the two programs SECTION and SURSPLIN take 84 min for phosphoglycerate kinase[31] (3192 atoms). Although the calculation time is lengthy, the resulting surface is described with about two-thirds less points than the MS surface under the conditions described above (MS surface density: 4 to 5 points per square angstrom, curves in three orthogonal directions spaced by 1 Å, and seven smoothing iterations of fourth-order B-spline). The smooth surface occupies less memory, facilitating multiple surface comparisons. It also takes up less space in the graphics memory, thereby allowing greater interaction on the graphics screen. The surfaces can be rotated or translated faster than if they are described with more than twice as many points. Moreover, the significant features are more obvious with the smoothed representation because all the atomic scale features have disappeared and a display with crossing lines rather than dots highlights the relief. For a variety of proteins, it is easier to determine the significant features with this representation. It also allows us to study the surface of very large macromolecules or complexes that would have been difficult to study with other kinds of representation.

Moon and Howe[32] have developed a method to more rapidly compute the molecular surface. Unfortunately, the resulting surface is rougher than Connolly's surface. We find this to be an unsatisfactory trade-off. For example, the surface of an icosahedral subunit of the rhinovirus 14,[33] which contains more than 800 residues, has been displayed (see Color Plate 2). The smooth representation highlights the existence of three antigenic zones and a deep canyon at the center of the triangular subunit.[34] The surfaces of complexes between an antigen (lysozyme) and the Fab fragment from a monoclonal antibody against lysozyme[35,36] have also been studied to compare the interactions of two kinds of monoclonal antibodies with lysozyme.

Despite the fact that a smooth surface is less precise than a molecular surface performed by Connolly's MS program, it retains enough information for interactions to be studied between two proteins, as in the case of the antigen–antibody complex of Sheriff et al.[36] (see Color Plate 3). It even highlights the surfaces that are most strongly interacting.

Since the smoothed protein surface has less points than the MS surface, the display of several surfaces of proteins simultaneously on the screen is possible. This helps us to search for characteristic global shapes. For example, the existence of some bilobed proteins, with two lobes and a central valley where the active site is often located, has been noted. The bilobed shape was noticed by Anderson et al. for the kinase family.[37] This shape is often encountered in proteins that are coded for by duplication genes such as penicillopepsin[38] (see Color Plate 4). However, this characteristic shape has also been noticed for some proteins that do not, to our knowledge, arise from gene duplication activity, as is the case for dihydrofolate reductase,[39] papain,[40] and lysozyme.[41] These three proteins, which have different structure and function, possess the same global shape and the same size regardless of the direction from which they are observed.

During our study of more than 50 protein surfaces among a variety of families, we have noticed that the active site is often the largest feature on the surface. With the smooth representation where all the small bumps are removed, the active site becomes very obvious. We have, however, focused our study on all features, other than the active site, that exist on the surface of many proteins as little pits or bumps of about 7 Å of depth and diameter. The results of this study will be reported elsewhere.

Quantitative results

The probe dots are located at 1 Å from the surface and the density of the surface neighborhood calculation is performed in a cube of 7 Å length centered on each probe dot. With these values, the surface is observed from far enough away so as to overlook atomic scale features, but close enough to concentrate on a reasonably small number of features at any given time and not to take into account distant features. The value has been adjusted so that the density is determined for about one significant feature at a time.

If we keep all the exterior dots that have more than a certain density of surface neighborhood, the dots that are close together could be grouped into a cluster. This allows us to draw a graph of the number of clusters that exists for each density level. For a low value of the limiting density, we observe only one cluster of external probe dots surrounding the surface. This cluster will be broken up into a number of clusters by selecting a higher limiting density, typically about 80 surface points per cube. Then, as the value of the minimum density of surface neighborhood increases, the number of remaining clusters decreases since the clusters above wide features vanish (see Figures 2 and 3).

The graph for the external probe dots shows that a value of 90 for the minimum density displayed on a graphics screen is convenient and corresponds to a large number of clusters. For the same graph with the internal probe dots, we notice that the graph has the same shape but is shifted to the right. The limiting density for a large number of

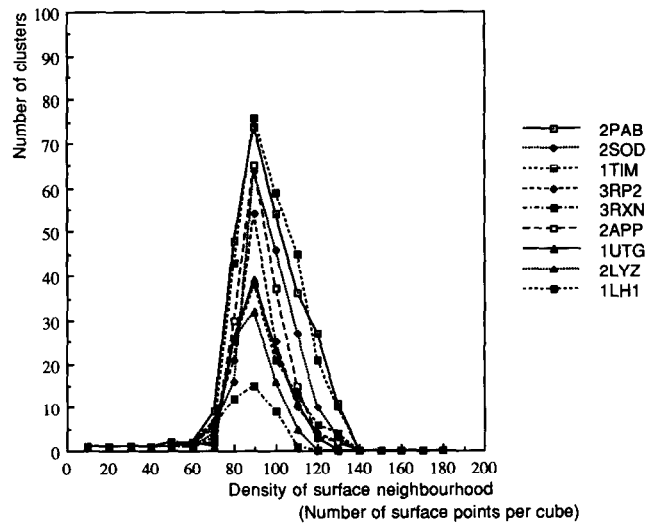

Figure 2. Graph of the number of clusters that exists for each density level, the density of surface neighborhood being performed for the exterior dots

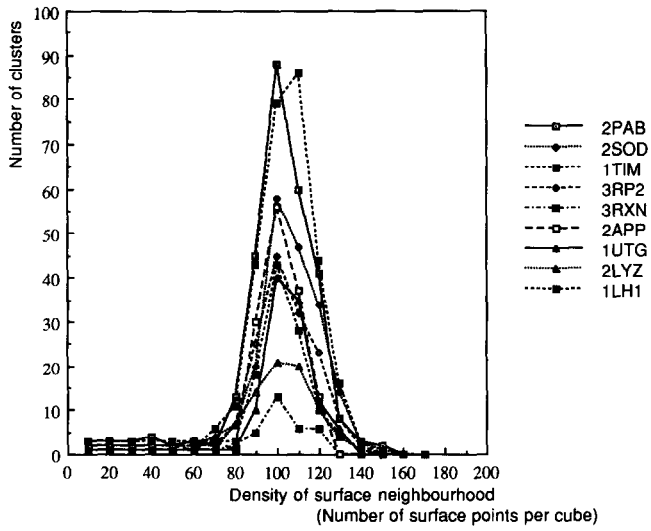

Figure 3. Graph of the number of clusters that exists for each density level, the density of surface neighborhood being performed for the interior dots

clusters is around 100 instead of 90. This could be because the internal probe points are below the surface, and consequently have more neighbors in close proximity. This shift explains why we have chosen to display graphically the internal probe points having a density greater than 100.

On a graphics screen, the probe dots are above or below all the significant features. Thus, each feature has a characteristic quantitative measurement: the number of points that have a density between 90 and 100, between 100 and 110, etc. For example, a wide valley would have many points between 90 and 100 but few points above. And a narrow and deep pocket would have less points between 90 and 100 than a wide valley, but would have also many points between 100 and 110, between 110 and 120, etc.

The distribution of the number of points for each density level is sharply related to the shape of the feature. It allows a quantitative description of the shape of each significant feature. These quantitative data have not yet been explored, but with a graphics screen, the display of the colored dots allowed a precise visual comparison, for example, between Asp tRNA[42] and Phe tRNA.[43] The two structures are very similar, and the colored dots help to highlight the differences (see Color Plate 5).

The visual study of the colored dots superimposed on the smooth surface reveals that the same kind of features have the same color (and thus the same value of concavity or convexity) for all the proteins studied. For example, an active site is rather yellow with a few green dots (concave to very concave area), small pits are red with a few yellow dots (medium concave area), and the bumps with lysine or arginine are blue with a few dark blue dots (convex to very convex areas).

On the other hand, the average color of a protein determines its degree of roughness. A smooth protein such as azurin[44] is mainly red and pale blue because there are few areas that are very concave or very convex on smooth features. A rougher protein, such as rhodanese,[45] has a large range of colors with all kinds of concavity and convexity.

Another application of the colored dots is the study of the distribution of the major features on the surface. By looking at the colored dots alone without the surface on a graphics screen, we observe that for a large number of proteins, the detected concave and convex features are not uniformly distributed (to be presented elsewhere).

In conclusion, the colored probe dots are very helpful for graphic inspection, but could also have other applications, such as quantitative comparison or distribution studies.

DISCUSSION

The descriptions of volume within a cubic grid have already been done for small molecules, primarily to compute and compare molecular volumes. Stouch and Jurs[19] have used a grid with a density of points of 1–6 per linear angstrom, and so distances between planes of 1–$\frac{1}{6}$ Å. Bohacek and Guida[46] have improved this method to quickly compute the volume of molecules using a grid of 0.1 Å for molecules of less than 100 atoms and 0.2 Å otherwise. Both groups started from the van der Waals surface. We have preferred to start with Connolly's surface described with dots, because all the work of overlapping the van der Waals sphere for each atoms has already been done and, thus, it was the quickest way to describe the surface of macromolecules within a cubic grid. We have used a grid of 1 Å for proteins since the significant features of the surface we are interested in are larger than 1 Å. For molecules of about 100 atoms, we have used a grid of 0.5 Å for a better visualization of molecules whose maximum length is of the order of less than 20 Å. With a grid of 1 Å, the bottom of deep pockets may not be well defined even with curves in the three orthogonal directions. This is a minor disadvantage for visual inspection since the eyes create the missing curves, but it is a big problem for the automatic analysis of the surface with the program SURSCOP. We therefore were forced to use a surface described within a cubic grid of 0.5 Å. This requires a lot more CPU time to compute. Consequently, until now the quantitative analysis has been possible only for molecules less than 500 residues.

The description with smoothed surfaces could not have been used for the quantitative analysis because the smooth curves in the three directions do not cross each other. Since the curves in each direction are smoothed independently with a B-spline smoothing function, nothing forced the curves to cross each other. This again is not a problem for visual inspection since the relief is still very visible. However, it was a problem for the quantitative analysis. It would have been impossible to make an automatic search for particular features with this kind of description since the exterior and the interior are not well defined without crossing curves. That was one of the reasons why we used the surface described within a cubic grid for quantitative analysis. The other reason was that it is easier and faster to perform calculations with integers than with reals. The fact that the curves do not cross gave rise to another, as yet unresolved, problem: how to display smooth surfaces on a raster screen. This would have improved the classical display of space-filling models, where the relief is often not easy to highlight. But we need to improve the surfaces with curves that cross before hoping to see a raster smooth surface.

The quantitative count of the significant features repre-

sents a large imprecision in the case of a protein with an internal cavity. To distinguish exterior probe dots from interior probe dots, we look only for external dots, the remaining dots being internal. So the count of interior probe dots includes the ones surrounding an eventual internal cavity. Visually, these dots are separated, but on a graph of the number of clusters versus the limiting density, the dots near a cavity will be counted with the dots below the outer surface, so the numbers we get are higher in this case than they should be. This method has only a local view of the surface, and a global view is needed to detect and separate internal cavities.

The proposed methods for displaying protein molecular surfaces and for the comparison of significant features have allowed us to glean new information about protein surfaces. Our method of obtaining a quantitative description of these features, using a surface described in a cubic grid, can be compared to other approaches.

Connolly proposed a method based on the calculation of solid angles.[47] For a circle centered on a planar contour, if the contour is convex, only a small part of the circle is inside the contour, and vice versa for a concave contour. Connolly extended this argument to the 3D case, with a sphere instead of a circle, and solid angle instead of planar angle calculations. In this way he was able to obtain quantitative data about the concavity and convexity of the surface, independent of the global shape of the protein and also of the coordinate system. This method has been used to automatically dock the two subunits α_1 and β_1 of hemoglobin.[48] However, it fails to predict any other association, because the number of parameters is too great to adjust simultaneously to obtain a manageable number of solutions.

Another way to evaluate the shape of globular proteins was proposed by Lewis and Rees[49] who use a fractal approach. The mathematical notion of fractals introduced by Mandelbrot[50] is used to describe a structure that is invariant with scale change. This means that each part, whatever its dimensions, is similar to the whole object. Quantitatively, the degree of roughness is given by the fractal dimension. This property is useful for the study of molecular surface. Avnir et al.[51] noted that most matter, at the atomic scale, has a fractal surface. The fractal dimension of a surface is ranged from 2 to 3. The rougher is the surface, the closer to 3 is the fractal dimension. Lewis and Rees[49] applied the notion of fractal dimensions to proteins. They observed that the regions of proteins that interact with other proteins to form stable complexes, such as solvent interfaces, possess a high fractal dimension and therefore a more irregular surface; regions that temporarily interact with ligands possess a low fractal dimension and therefore a smoother surface. With similar methods, Åqvist and Tapia[52] used the values of the fractal dimension to predict the binding site between the prealbumin and the retinol binding protein.

Without any hypothesis concerning the fractal nature of the protein surface, and with only a rough calculation of concavity and convexity, we have found similar results. We observed that the interface areas are very rough, and that the active sites, although very rough, are often surrounded by large smooth regions.

We have compared the prealbumin tetramer[53] displayed with our color-coded dots representing concavity and convexity with Åqvist and Tapia color dots representing the fractal dimension (see Color Plate 1b from Ref. 52): areas with high fractal dimension are located at about the same places as regions with a high density of dots.

There is, however, a difference between the two calculations: We calculated the density of surface neighborhood in the same unit (cube of 7 Å length) whatever the size of the protein, whereas the fractal dimension calculations are performed in cones including the same fraction of surfaces and thus are dependent upon the size of the protein. We think it is preferable not to take into account the protein size since the size and shape of the various features do not depend on the size of the protein.

Another method has been proposed to perform quantitative calculations of molecular surface shape using Fourier descriptors.[54] This is a powerful technique that allows quantitative comparison of surface shape. It is based on representation of a molecular surface in terms of spherical harmonics. But, because it is very CPU-time intensive, it is, for the moment, used only for small molecules. It is interesting to note that the qualitative representation is quite similar. Smoothing with a fourth-order B-spline with seven iterations gives visually the same results as taking a convergent series of spherical harmonics with 20 terms (see Color Plate 6 from Ref. 54). Thus, qualitatively, the two methods give about the same results.

Compared to these three methods, our approaches of calculating the concave and convex parts of the surface afford a different view and new information. Our method is simpler and certainly less time-consuming than the fractal and the Fourier descriptor approach. Our method allows us to calculate the concavity and the convexity routinely for proteins up to 500 residues. We do lose some accuracy, since the starting point is the approximate surface described within a cubic grid, and all calculations are done with integral values. For the moment, the quantitative data have been used to great effect for visual analysis, affording more accurate comparison, and also for the extensive study of the distribution of the significant features at the surface of the proteins. This method has not yet been used for an automatic docking process but could perhaps help in obtaining a rough approach to docking with subsequent use of other methods to obtain a more precise value of the fit.

The display of smooth surfaces is very helpful for qualitative analysis. It is as if we observe the mean surface position of all the atoms over time. This is why the smoothing does not erase important topological information. After studying more than 50 proteins surfaces among various families (results shown elsewhere), it appears that the chosen strength of smoothing efficiently removes the small bumps on the molecular surface keeping intact the relatively small number of significant features.

NOTE

The programs SECTION, SURSPLIN, and SURSCOP are written in Fortran 77 without extension, so are easily portable. They are already implanted in several laboratories with different hardware or software equipments. The codes are available upon request from N. Colloc'h.

ACKNOWLEDGMENTS

The authors thank P. Andrew, G. E. Schulz, L. Sawyer, E. M. Westbrook, R. J. Poljak, D. Moras, and S. Brunie for their atomic coordinates, A. A. Shaw for language correction, and F. E. Cohen and L. M. Gregoret for their very helpful advice and a critical reading of the manuscript. This work was supported in part by a grant of the Institut Scientifique Roussel.

REFERENCES

1 Chothia, C. and Janin, J. Principles of protein–protein recognition. *Nature* 1975, **256**, 705–708
2 Wodak, S. J. and Janin, J. Computer analysis of protein–protein interaction. *J. Mol. Biol.* 1978, **124**, 323–342
3 Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. and Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* 1982, **161**, 269–288
4 Desjarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D. and Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* 1988, **31**, 722–729
5 Blaney, J. M., Jorgensen, E. C., Connolly, M. L., Ferrin, T. E., Langridge, R., Oatley, S. J., Burridge, J. M. and Blake, C. C. F. Computer graphics in drug design: molecular modeling of thyroid hormone-prealbumin interactions. *J. Med. Chem.* 1982, **25**, 785–790
6 Rebek, J., Jr. Model studies in molecular recognition. *Science* 1987, **235**, 1478–1484
7 Morize, I., Surcouf, E., Vaney, M. C., Epelboin, Y., Buehner, M., Fridlansky, F., Milgrom, E. and Mornon, J. P. Refinement of the $C222_1$ crystal form of oxidized uteroglobin at 1.34 Å resolution. *J. Mol. Biol.* 1987, **194**, 725–739 (Entry 1UTG from PDB)
8 Brunie, S., Bolin, J., Gewirth, D. and Sigler, P. B. The refined crystal structure of dimeric phospholipase A_2 at 2.5 Å. Access to a shielded catalytic center. *J. Biol. Chem.* 1985, **260**, 9742–9749 (Entry 1PP2 from PDB)
9 Levin, S. W., Butler, J. B., Schumacher, U. K., Wightman, P. D. and Mukherjee, A. B. Uteroglobin inhibits phospholipase A_2 activity. *Life Sci.* 1986, **38**, 1813–1819
10 Miele, L., Cordella-Miele, E. and Mukherjee, B. Uteroglobin: structure, molecular biology, and new perspectives on its function as a phospholipase A_2 inhibitor. *Endocr. Rev.* 1987, **8**, 474–490
11 Bondi, A. Van der Waals volumes and radii. *J. Phys. Chem.* 1964, **68**, 441–451
12 Lee, B. and Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 1971, **55**, 379–400
13 Richards, F. M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 1977, **6**, 151–176
14 Novotny, J., Handschumacher, M., Haber, E., Bruccoleri, R. E., Carlson, W. B., Fanning, D. W., Smith, J. A. and Rose, G. D. Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). *Proc. Natl. Acad. Sci. USA* 1986, **83**, 226–230
15 Novotny, J., Handschumacher, M. and Haber, M. Location of antigenic epitopes on antibody molecules. *J. Mol. Biol.* 1986, **189**, 715–721
16 Connolly, M. L. *Quantum Chemistry Program Exchange Bulletin* 1981, **1**, 75
17 Fanning, D. W., Smith, J. A. and Rose, G. D. Molecular cartography of globular proteins with application to antigenic sites. *Biopolymers* 1986, **25**, 863–883
18 Chirgadze, Y., Kurochkina, N. and Nikonov, S. Molecular cartography of proteins: surface relief analysis of the calf eye lens protein gamma-crystallin. *Protein Eng.* 1989, **3**, 105–110
19 Stouch, T. R. and Jurs, P. C. A simple method for the representation, quantification, and comparison of the volumes and shapes of chemical compounds. *J. Chem. Inf. Comput. Sci.* 1986, **26**, 4–12
20 De Boor, C. On calculating with B-splines. *J. Approx. Theory* 1972, **6**, 50–62
21 Riesenfeld, R. F. *Bernstein-Bezier methods for the computer-aided design of free form curves and surfaces.* Ph.D. thesis, Syracuse University, Syracuse, NY, 1973
22 Carson, M. Ribbon models of macromolecules. *J. Mol. Graphics* 1987, **5**, 103–106
23 Bezier, P. E. Example of an existing system in the motor industry: the Unisurf System. *Proc. R. Soc., London Ser. A* 1971, **321**, 207–218
24 Bezier, P. E. *Numerical Control. Mathematics and Applications.* Wiley, London, 1972 (translated by Forrest, D. R. and Pankhurst, A. F. from *Emploi des machines a commande numerique,* Masson, Paris, 1970)
25 Hendrickson, W. A. and Teeter, M. M. Private communication, 1981 (Entry 1CRN from PDB)
26 Surcouf, E. and Mornon, J. P. Signature de van der Waals: outil d'etude des associations moleculaires. *C. R. Acad. Sci. Paris* 1982, **295(II)**, 923–926
27 Vaney, M. C., Surcouf, E., Cherfils, J., Morize, I. and Mornon, J. P. MANOSK, a new program designed for macromolecular modelling. *J. Mol. Graphics* 1985, **3**, 123–124
28 Cherfils, J., Vaney, M. C., Morize, I., Surcouf, E., Colloc'h, N. and Mornon, J. P. MANOSK: a graphics program for analyzing and modeling molecular structure and functions. *J. Mol. Graphics* 1988, **6**, 155–160
29 Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, **112**, 535–542
30 Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. and Weng, J. *Crystallographic Databases—Information Content, Software Systems, Scientific Applications* (F. H. Allen, G. Bergerhoff and R. Seivers, Eds.). Data Commission of the Int'l Union of Crystallography, Bonn/Cambridge/Chester, 1987, 107–132
31 Watson, H. C., Walker, N. P. C., Shaw, P. J., Bryant, T. N., Wendell, P. L., Fothergill, L. A., Perkins, R. E., Conroy, S. C., Dobson, M. J., Tuite, M. F.,

Kingsman, A. J. and Kingsman, S. M. Sequence and structure of yeast phosphoglycerate kinase. *EMBO* 1982, **1**, 1635–1640 (Entry 3PGK from PDB)
32 Moon, J. B. and Howe, W. J. A fast algorithm for generating smooth molecular dot surface representations. *J. Mol. Graphics* 1989, **7**, 109–112
33 Arnold, E. and Rossmann, M. G. The use of molecular-replacement phases for the refinement of the human rhinovirus 14 structure. *Acta Crystallogr. Sect. A* 1988, **44**, 270–282 (Entry 4RHV from PDB)
34 Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H.-J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Rueckert, R. R., Sherry, B. and Vriend, G. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 1985, **317**, 145–152
35 Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. and Poljak, R. J. Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. *Science* 1986, **233**, 747–753
36 Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C. and Davies, D. R. Three-dimensional structure of an antibody–antigen complex. *Proc. Natl. Acad. Sci. USA* 1987, **84**, 8075–8079 (Entry 2HFL from PDB)
37 Anderson, C. M., Zucker, F. H. and Steitz, T. A. Space-filling models of kinase clefts and conformation changes. *Science* 1979, **204**, 375–380
38 James, M. N. G. and Sielecki, A. R. Structure and refinement of penicillopepsin at 1.8 Å resolution. *J. Mol. Biol.* 1983, **163**, 299–361 (Entry 2APP from PDB)
39 Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C. and Kraut, J. Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution. I. General features and binding of methotrexate. *J. Biol. Chem.* 1982, **257**, 13650–13662 (Entry 3DFR from PDB)
40 Jansonius, J. N. Private communication, 1984 (Entry 1PPD from PDB)
41 Diamond, R. Real-space refinement of the structure of hen egg-white lysozyme. *J. Mol. Biol.* 1974, **82**, 371–391 (Entry 2LYZ from PDB)
42 Westhof, E., Dumas, P. and Moras, D. Restrained refinement of two crystalline forms of yeast aspartic and phenylalanine transfer RNA crystals. *Acta Crystallogr. Sect. A* 1988, **44**, 112–123 (Entry 2TRA from PDB)
43 Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M. and Kim, S.-H. Crystal structure of yeast phenylalanine transfer RNA. I. Crystallographic refinement. *J. Mol. Biol.* 1978, **123**, 607–630 (Entry 6TNA from PDB)
44 Baker, E. N. Structure of azurin from *Alcaligenes denitrificans*. Refinement at 1.8 Å and comparison of the two crystallographically independent molecules. *J. Mol. Biol.* 1988, **203**, 1071–1095 (Entry 2AZA from PDB)
45 Ploegman, J. H., Drent, G., Kalk, K. H. and Hol, W. G. J. Structure of bovine liver rhodanese. I. Structure determination at 2.5 Å resolution and a comparison of the conformation and sequence of its two domains. *J. Mol. Biol.* 1978, **123**, 557–594 (Entry 1RHD from PDB)
46 Bohacek, R. S. and Guida, W. C. A rapid method for the computation, comparison and display of molecular volumes. *J. Mol. Graphics* 1989, **7**, 113–117
47 Connolly, M. L. Measurement of proteins surface shape by solid angles. *J. Mol. Graphics* 1986, **4**, 3–6
48 Connolly, M. L. Shape complementarity at the hemoglobin $\alpha_1\beta_1$ subunit interface. *Biopolymers* 1986, **25**, 1229–1247
49 Lewis, M. and Rees, D. C. Fractal surfaces of proteins. *Science* 1985, **230**, 1163–1165
50 Mandelbrot, B. B. *The Fractal Geometry of Nature*. Freeman & Cie, New York, 1983
51 Avnir, D., Farin, D. and Pfeifer, P. Molecular fractal surfaces. *Nature* 1984, **308**, 261–263
52 Åqvist, J. and Tapia, O. Surface fractality as a guide for studying protein–protein interactions. *J. Mol. Graphics* 1987, **5**, 30–34
53 Blake, C. C. F., Geisow, M. J., Oatley, S. J., Rerat, B. and Rerat, C. Structure of prealbumin: secondary, teritary and quaternary interactions determined by Fourier refinement at 1.8 Å. *J. Mol. Biol.* 1978, **121**, 339–356. (Entry 2PAB from PDB)
54 Leicester, S. E., Finney, J. L. and Bywater, R. P. Description of molecular surface shape using Fourier descriptors. *J. Mol. Graphics* 1988, **6**, 104–108