19 August 1983, Volume 221, Number 4612 SCIENCE

Solvent-Accessible Surfaces of Proteins and Nucleic Acids

Michael L. Connolly

Summary. A method is presented for analytically calculating a smooth, three-dimensional contour about a molecule. The molecular surface envelope may be drawn on either color raster computer displays or real-time vector computer graphics systems. Molecular areas and volumes may be computed analytically from this surface representation. Unlike most previous computer graphics representations of molecules, which imitate wire models or space-filling plastic spheres, this surface shows only the atoms that are accessible to solvent. This analytical method extends the earlier dot surface numerical algorithm, which has been applied in enzymology, rational drug design, immunology, and understanding DNA base sequence recognition.

Computer graphics has made the results of x-ray crystallographic studies of proteins and nucleic acids more accessible to biochemists and molecular biologists. Traditionally, computer-generated images of molecular structures have consisted of lines for the chemical bonds (*1–3*) or spheres (*4–7*) and ellipsoids (*8*) for the atoms. I present an alternative representation, that of a smooth, continuous envelope in contact with the atoms that are accessible to solvent.

Applications of this surface representation include enzymology, rational drug design, the elucidation of molecular diseases such as sickle cell anemia, recognition of specific DNA base sequences by proteins and drugs, and the location of possible antigenic determinants on viruses.

The historical basis for the smooth surface envelope method is the work of Richards (*9*) and colleagues on solvent-accessible area. Their emphasis was on chemical calculations measuring quantities of hydrophobic and hydrophilic area, while the methods described below were developed primarily for the purpose of visualizing molecular structure and interactions. Nevertheless, these new methods also permit the measurement of area and volume in conjunction with the graphical display.

Solvent-Accessible Area

Solvent-accessible area was originally defined and computed by Lee and Richards (*10*) as the area traced out by the center of a probe sphere representing a solvent molecule as it is rolled over the surface of the molecule of interest. These computational methods were invented as a tool for attacking the protein folding problem (*9*). The problem is that of predicting the three-dimensional structure of a protein given only its primary sequence of amino acids.

Simply measuring a quantity of area is insufficient for the study of many aspects of protein and nucleic acid function, such as substrate binding and catalysis, drug–nucleic acid interaction, and recognition by the immune system. A method for visualizing solvent-accessible surfaces is needed. For this purpose, an alternative solvent-accessible surface definition, proposed by Richards (*9*), is appropriate. Unlike the original surface of Lee and Richards (*10*), this alternative molecular surface is not displaced from the van der Waals surface. Instead, it consists of the part of the van der Waals surface of the atoms that are accessible to the probe sphere (contact surface), connected by a network of concave and saddle-shaped surfaces (reentrant surface) that smooths over the crevices and pits between the atoms. This surface is the boundary of the volume from which a probe sphere is excluded if it is not to experience van der Waals overlap with the atoms.

Improving on the algorithms of Greer and Bush (*11*) for calculating contact and reentrant surfaces and of Shrake and Rupley (*12*) for calculating solvent-accessible area, I developed a numerical computer algorithm for placing dots over the solvent-accessible molecular surface of a protein (*13, 14*). Below, I briefly review the dot surface algorithm and present a new, analytical surface method.

Dot Surfaces

The basic approach in this method is to place a probe sphere, representing a solvent molecule, tangent to the atoms of the protein at several thousand different positions. For each probe position that does not experience van der Waals overlap with the atoms of the protein, points lying on the inward-facing surface of the probe sphere become part of the protein's solvent-accessible surface. The probe may be placed tangent to (i) single atoms, creating a dot at the point of tangency, (ii) pairs of atoms, creating a concave arc of dots connecting the two points of tangency, and (iii) triples of atoms, creating a concave triangle of dots between the three points of tangency. For each surface point generated, the numerical algorithm produces not only its coordinates but also an approximate solvent-accessible area associated with the point and an outward-pointing unit vector perpendicular to the surface at that point. The pancreatic trypsin–trypsin inhibitor complex (*15*) is shown in Fig. 1, with a dot surface for the enzyme only.

The author is a Helen Hay Whitney postdoctoral fellow in the Molecular Biology Department, Research Institute of Scripps Clinic, La Jolla, California 92037.

This method has proved useful in enzymology (*16–20*), immunology (*21, 22*), virology (*23*), molecular pathology (*24*), and the study of protein-ligand (*25*) and protein-protein (*26, 27*) interactions.

Despite the many applications of the dot surface numerical algorithm, it was necessary to invent an analytical surface algorithm in order to generate high-resolution color raster display images and to compute more accurate molecular areas and volumes.

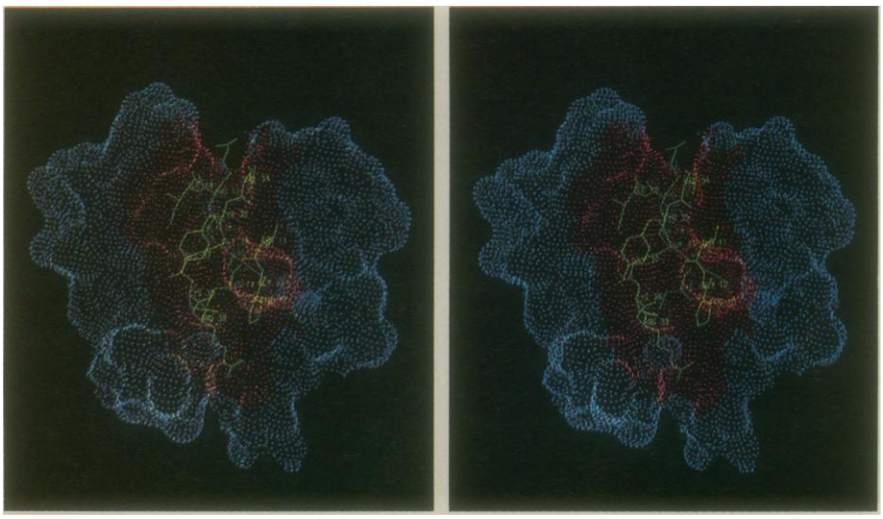


Fig. 1. Stereo pair of the pancreatic trypsin–trypsin inhibitor complex. The enzyme is represented by a dot surface. The residues of the inhibitor in contact with the enzyme are represented by bonds. The part of the trypsin surface that is kept from contact with the solvent by the presence of the inhibitor is colored red.

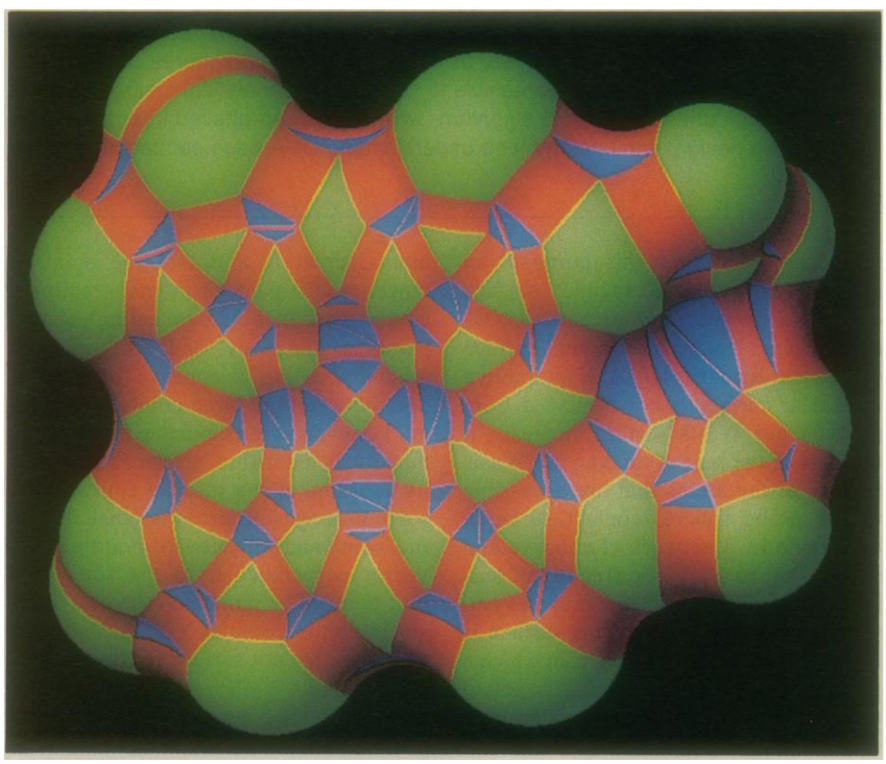


Fig. 2. Heme molecule drawn on a color raster graphics system. Green, convex surface; red, saddle surface; blue, concave surface. Surface pieces join at circular arcs.

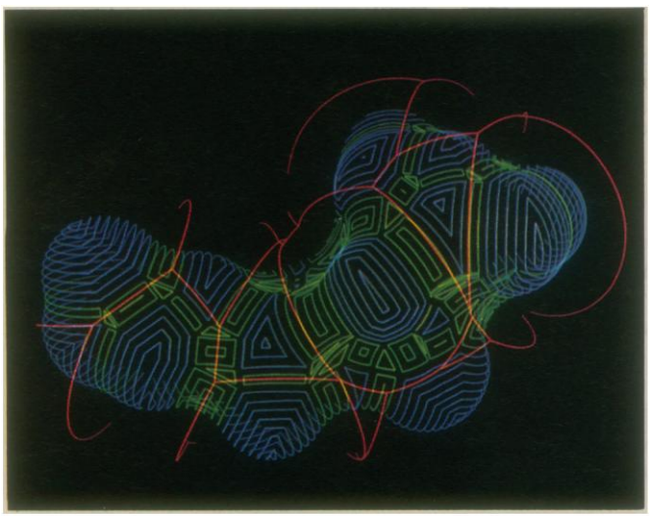
Analytical Molecular Surface

A continuous molecular surface contour is defined as the union of pieces of spheres and tori joining smoothly at circular arcs. There are three kinds of pieces: concave spherical triangles, saddle-shaped rectangles, and convex spherical regions (Fig. 2).

The computer algorithm proceeds in three steps, one for each shape of surface. First, a probe sphere is placed tangent to every set of three neighboring atoms, and a concave triangle is generated whenever the probe sphere experiences no collisions with any other atoms of the molecule. Each concave triangle has three concave arcs as edges. Next, the saddle rectangles are formed by connecting adjacent concave arcs along the inner surfaces of tori (Fig. 3). The edges of each saddle rectangle consist of a pair of concave arcs and a pair of convex arcs. In the final step, the convex arcs on each atom are grouped to form closed circuits, or cycles, and the boundary of each convex face is defined by zero, one, or more cycles. The equations defining the surface and the details of the computer algorithm will be presented elsewhere (*28*).

Since molecular areas and volumes are important physical chemical properties of molecules, efforts have been made to calculate them from x-ray coordinates. The areas of the convex faces are referred to in the literature as contact areas, and approximate numerical methods for their measurement have been developed (*10, 29, 30*). With the surface defined in an analytical fashion, it is now possible to calculate these contact areas exactly. This is done by using the Gauss-Bonnet theorem (*31*) from differential geometry. This theorem is traditionally used to study the relation of surface topology to integrals of curvature, but since the curvature of a convex spherical face is constant, the integrals simplify and the contact area may be expressed as a function of the atomic radius and the geometry and topology of the boundary cycles. The area of a concave face is calculated in a similar fashion. The area of a saddle face may be calculated by using integral calculus, since it is part of a surface of revolution, the torus.

Molecular volumes have been calculated from protein x-ray crystallographic coordinates, using a polyhedral definition of the protein surface, and these calculated volumes have been compared to experimentally measured partial specific volumes of proteins in solution (*32*). A smoothly curved definition of the protein surface, such as the analytically defined solvent-accessible surface, should help provide a more accurate measurement of molecular volume. The volume enclosed by the solvent-accessible surface may be calculated by partitioning this volume into simpler shapes whose volumes may be easily calculated by solid geometry and integral calculus. Most of the molecular volume is contained within an interior polyhedron whose vertices are the centers of the solvent-accessible atoms. Coating this



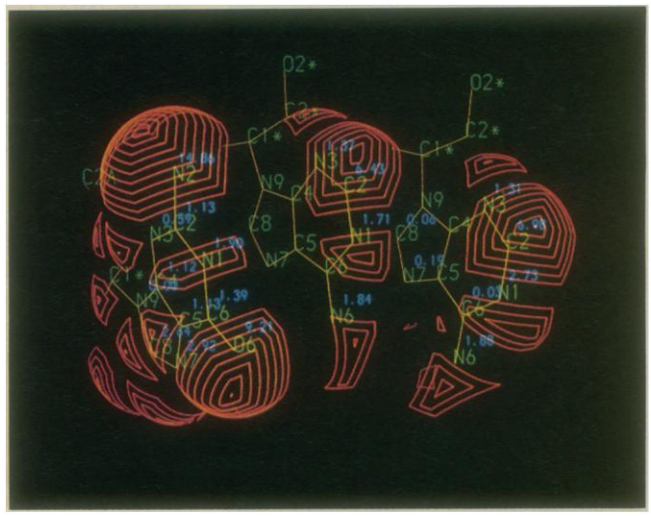


Fig. 3 (left). Trajectory of probe rolling over a molecular surface. The trajectory arcs (red) connect positions where the probe is simultaneously tangent to three atoms. In a corresponding manner, saddle rectangles connect concave triangles. These reentrant surfaces (green) then define the boundaries of the convex surfaces (magenta). Fig. 4 (right). Yeast phenylalanyl transfer RNA anticodon (GAA). The contact surface of the three anticodon bases is shown. The contact areas in square angstroms are displayed next to the atom labels.

polyhedron is a surface layer that is made up of one piece for each curved face of the analytical surface. This surface layer has an average thickness of about an atomic radius.

As an example of an application of the area method, the contact areas of the atoms of the transfer RNA anticodon (*33*) have been computed and are shown in Fig. 4. The conjunction of both graphical and area measurement methods makes it possible to see not only how much of an atom is accessible but also where the accessible regions are. For this anticodon and the DNA structures presented below, van der Waals radii with implicit hydrogens have been taken from (*30*) and a probe with a radius of 1.5 Å has been used.

To illustrate the ability of the analytical method to measure small changes in area and volume, the room-temperature (*34*) and low-temperature (*35*) DNA dodecamer structures are compared. The molecular areas and volumes are 3631 $Å^2$ and 6534 $Å^3$ (290 K, 1.9-Å resolution) and 3623 $Å^2$ and 6514 $Å^3$ (16 K, 2.7-Å resolution), respectively. The low-temperature structure is 0.3 percent smaller in volume. The room-temperature structure is shown in Fig. 5.

Computer Graphics

The analytically defined surface, being continuous rather than discrete, is well suited to raster display. The input to a raster graphics system consists of a two-dimensional array of picture elements, or pixels, each of which has a color and shade value (*36*). In order to produce this pixel array from a three-dimensional curved surface, a hidden-surface elimination algorithm is required (*37*). The analytical molecular surface representation is substantially different from previous curved surface representations, such as polygon mesh, parametric bicubic patches, and solid modeling (*38*, p. 506), so it was necessary to invent a hidden-surface algorithm for it, which will be published elsewhere (*39*).

The use of stereo, in conjuction with hidden-surface elimination and shading, gives a vivid demonstration of protein topography (Fig. 6). The copper atom is seen to lie in a deep pit at the active site of Cu,Zn superoxide dismutase (*40*).

One is not restricted to using spheres to represent individual atoms. For large molecular complexes, it is useful to model a group of atoms with a single sphere. The 2.8-Å structure of aspartate carbamoyltransferase (*41*) has been modeled with each amino acid residue represented by a sphere centered at the alpha carbon (Fig. 7).

A method for smoothing the junctures between atoms by summing Gaussian densities for each atom and drawing surfaces at various density contour levels was developed by Blinn (*42*). While the probe sphere method does a similar smoothing, its main effect is not the smoothing of crevices and pits, but rather the complete removal of the van der Waals surface of interior atoms. This interior surface removal is important, because most of a protein's van der Waals surface is in the interior and not directly involved in molecular interactions.

Display of dot surfaces on a real-time color vector system is, in general, more useful than the raster surface representation because (i) the dot surfaces are transparent, enabling chemical bonds and atom labels to be seen, and (ii) the image may be rotated and sectioned in real time. However, the raster system does have the advantage that it can show a larger region of surface at high resolution. This is because raster systems typically display a quarter of a million pixels, while real-time vector systems can handle only 10,000 to 20,000 vectors.

In addition to the shaded-surface raster representation, an analytical surface has a real-time vector representation, where each face of the surface is represented by a set of concentric curved polygons (Fig. 3). These polygons may be calculated in a straightforward manner for concave and saddle faces, but convex faces have less regular shapes. For convex faces, concentric cycles bordering a shrinking contact area are generated by progressively incrementing the radii of neighboring atoms.

An interactive display program is required to manipulate molecular surfaces on a vector graphics system. For the Evans and Sutherland Multi Picture System, this need is satisfied by the general-purpose graphics program GRAMPS (GRAphics for the Multi Picture System), developed by O'Donnell and Olson (*43*). GRAMPS may simultaneously display any combination of curved polygonal surfaces, chemical bonds, atom labels, dot surfaces, and arbitrary geometric figures, such as an icosahedron representing a virus capsid (*44*). The various graphical objects are organized into a hierarchical tree structure and each ob-

Fig. 5. DNA dodecamer with sequence: CGCGAATTCGCG. The part of the van der Waals surface of each atom that is accessible to solvent is colored by atom type (red, oxygen; green, carbon; blue, nitrogen). Reentrant surface (white) smooths out the crevices and pits between the atoms.

ject may be independently transformed and colored in real time.

A primary value of the graphical display of solvent-accessible surfaces is that it provides immediately comprehensible information about steric complementarity. This is illustrated by the work of Blaney *et al.* (*45*), who used real-time color dot surface graphics to study the fit of the thyroid hormone thyroxin into the binding site of a blood transport protein, prealbumin. They noticed an empty pocket adjacent to one of the phenyl rings of thyroxin. Computer graphics modeling showed that naphthyl analogs of thyroxin would fit into the binding site and the larger naphthyl ring would fill this pocket. When a wide variety of thyroid hormone analogs were tested, those with a naphthyl ring filling this pocket were found to bind better than those which left the pocket empty.

Another use of this surface representation has been to paint chemical information onto it. Weiner *et al.* (*46*) did this by coloring the surface dots of proteins and nucleic acids according to electrostatic potential. Interfacing surfaces in protein-protein, protein-ligand, and drug–nucleic acid interactions were seen to have not only topographic but also electrostatic complementarity. The electrostatic surface potential of DNA was seen to be strongly sequence-dependent. This method has also been used to study the binding of the negatively charged superoxide radical to the enzyme superoxide dismutase (*47*). In the electrostatic method, the potential is evaluated at the center of each probe sphere position that generates a surface point. That is, in addition to being a canvas for displaying chemical information, the surface can play a fundamental role in calculating that information.

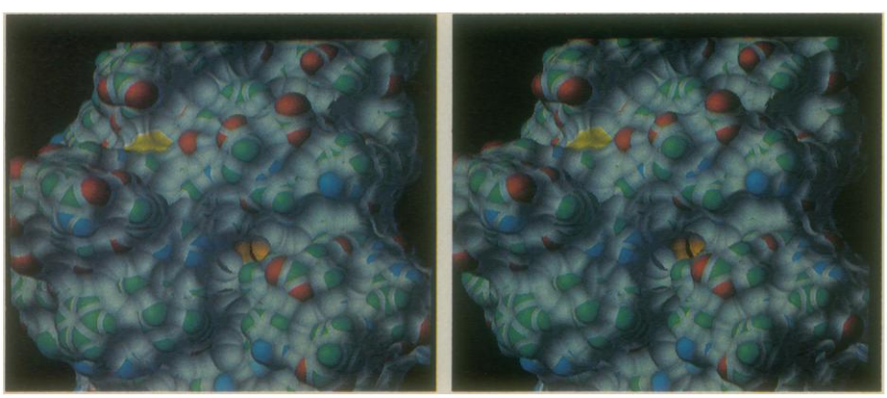


Fig. 6. Stereo pair of Cu,Zn superoxide dismutase. Same color-coding as in Fig. 5, but with the contact and reentrant surfaces of sulfur and copper colored yellow and copper. The copper atom is part of the active site and interacts with the superoxide radical. Hydrogen atoms are given the color of the heavy atom they are bonded to. Self-intersecting surfaces create point and edge cusps and other artifacts in deep grooves.

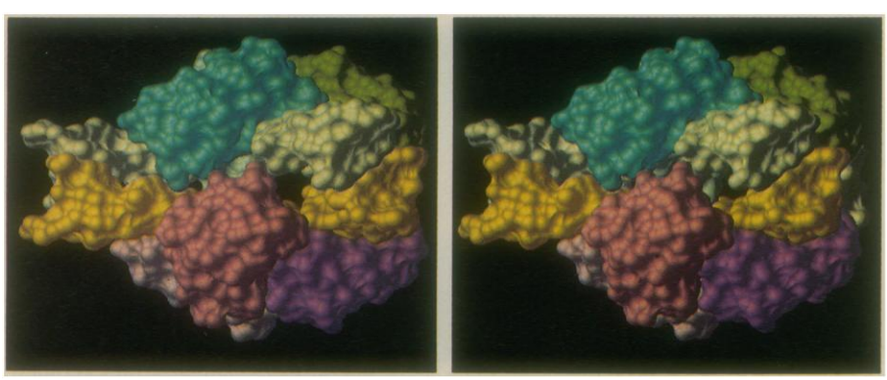


Fig. 7. Stereo pair of aspartate carbamoyltransferase. The top catalytic trimer is colored green, light green, and cyan. The bottom catalytic trimer is colored red, pink, and magenta. The regulatory dimers are colored yellow and white. Each amino acid residue is represented by one sphere 3 Å in radius positioned at the alpha carbon, and a probe sphere 3 Å in radius was used to calculate the surface of each subunit.

Conclusions

The principal use of computer graphics by macromolecular x-ray crystallographers has been in fitting the model to the electron density and in refining the structure (*1–3*). The methods described above will help crystallographers in the succeeding step of interpreting the solved structure. Scientists in related disciplines will also benefit, since the structures of more than 100 proteins, nucleic acids, and virus capsids have been deposited for general distribution at the Protein Data Bank at Brookhaven National Laboratory (*48*).

While the display of solvent-accessible surfaces on real-time vector graphics systems is preferred for interactively exploring a macromolecular structure, the color raster display of solvent-accessible surfaces made possible by the analytical algorithm is better able to communicate structural discoveries because of its higher resolution and greater visual realism.

In time, the raster solvent-accessible surface display will acquire more of the capabilities of the vector display. For example, raster graphics methods for displaying transparent surfaces exist (*49*) and can be adapted to this system. Also, it should be possible to section away the front surface of a protein to display interior pockets and cavities, since the hidden-surface algorithm (*39*) uses a depth buffer (*38*, pp. 560–561), where the height of each pixel is stored.

Although these graphical methods were developed to study the protein surface, they should also be useful in visualizing the packing of alpha helices and beta sheets in the protein interior, simply by giving these structural elements individual surface contours. This will bring solvent-accessibility studies back full-circle to their original scientific problem, the understanding of the folding of the polypeptide chain to form protein tertiary structure.

References and Notes

1. R. Diamond, in *Computational Crystallography*, D. Sayre, Ed. (Oxford Univ. Press, Oxford, 1982), p. 318; T. A. Jones, in *ibid.*, p. 303.
2. F. P. Brooks, Jr., in *Proceedings of the 1977 International Federation of Information Processing*, B. Gilchrist, Ed. (North-Holland, Amsterdam, 1977), p. 625.
3. C. D. Barry, C. E. Molnar, F. U. Rosenberger, *Technical Memo 229* (Computer Systems Laboratory, Washington University, St. Louis, Mo., January 1976); C. D. Barry, H. E. Bosshard, R. A. Ellis, G. R. Marshall, in *Computers in Life Science Research*, W. Siler and D. A. B. Lindberg, Eds. (Plenum, New York, 1975), p. 137.
4. T. Porter, *Comput. Graphics* **12**, 282 (1978); *ibid.* **13**, 234 (1979).
5. R. J. Feldmann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5409 (1978).
6. K. Knowlton and L. Cherry, *Comput. Chem.* **1**, 161 (1977).
7. N. L. Max, *Comput. Graphics* **13**, 165 (1979).
8. C. K. Johnson, *Oak Ridge Natl. Lab. Tech. Rep. 5138* (1976).
9. F. M. Richards, *Annu. Rev. Biophys. Bioeng.* **6**, 151 (1977).
10. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).
11. J. Greer and B. L. Bush, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 303 (1978).
12. A. Shrake and J. A. Rupley, *J. Mol. Biol.* **79**, 351 (1973).
13. M. L. Connolly, thesis, University of California, Berkeley (1981).
14. ———, *QCPE Bull. 1* (1981), p. 75. The dot molecular surface program (MS) is written in Fortran and may be obtained by writing to Quantum Chemistry Program Exchange, Department of Chemistry, Indiana University, Bloomington 47405.
15. R. Huber, D. Kukla, W. Bode, P. Schwager, K. Bartels, J. Deisenhofer, W. Steigemann, *J. Mol. Biol.* **89**, 73 (1974).
16. R. N. Smith, C. Hansch, F. H. Kim, B. Omiya, G. Fukumura, C. D. Selassie, P. Y. C. Jow, J. M. Blaney, R. Langridge, *Arch. Biochem. Biophys.* **215**, 319 (1982).
17. C. Hansch, R. Li, J. M. Blaney, R. Langridge, *J. Med. Chem.* **25**, 777 (1982).
18. S. Sprang, R. Fletterick, M. Stern, D. Yang, N. Madsen, J. Sturtevant, *Biochemistry* **21**, 2036 (1982).
19. E. Goldsmith, S. Sprang, R. Fletterick, *J. Mol. Biol.* **156**, 411 (1982).
20. S. R. Sprang, E. J. Goldsmith, R. J. Fletterick, S. G. Withers, N. B. Madsen, *Biochemistry* **21**, 5364 (1982).
21. R. A. Lerner, *Nature (London)* **299**, 592 (1982).
22. A. J. Olson, G. Cohen, D. Davies, *Antibody Structure,* film available from Byron Motion Pictures, 65 K Street, NE, Washington, D.C. 20002.
23. A. J. Olson, *Virus Wars,* computer-generated film presented at the International School of Crystallography, Conference on Crystallography in Molecular Biology, Italy, June 1982.
24. R. E. Dickerson and I. Geis, *Hemoglobin: Structure, Function, Evolution, and Pathology* (Benjamin/Cummings, Menlo Park, Calif., 1983), p. 136.
25. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin, *J. Mol. Biol.* **161**, 269 (1982).
26. R. Langridge, T. E. Ferrin, I. D. Kuntz, M. L. Connolly, *Science* **211**, 661 (1981).
27. R. O. Fox, Jr., and F. M. Richards, *Nature (London)* **300**, 325 (1982).
28. M. L. Connolly, *J. Appl. Crystallogr.*, in press.
29. T. J. Richmond and F. M. Richards, *J. Mol. Biol.* **119**, 537 (1978).
30. C. J. Alden and S.-H. Kim, *ibid.* **132**, 411 (1979).
31. M. P. do Carmo, *Differential Geometry of Curves and Surfaces* (Prentice-Hall, Englewood Cliffs, N.J., 1976), pp. 274–276.
32. F. M. Richards, *J. Mol. Biol.* **82**, 1 (1974).
33. J. L. Sussman, S. R. Holbrook, R. W. Warrant, G. M. Church, S.-H. Kim, *ibid.* **123**, 607 (1978).
34. H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2179 (1981).
35. H. R. Drew, S. Samson, R. E. Dickerson, *ibid.* **79**, 4040 (1982).
36. T. Whitted, *Science* **215**, 767 (1982).
37. W. N. Newman and R. F. Sproull, *Principles of Interactive Computer Graphics* (McGraw-Hill, New York, 1979), pp. 367–388.
38. J. D. Foley and A. Van Dam, *Fundamentals of Interactive Computer Graphics* (Addison-Wesley, Reading, Mass., 1982).
39. M. L. Connolly, in preparation.
40. J. A. Tainer, E. D. Getzoff, K. M. Beem, J. S. Richardson, D. C. Richardson, *J. Mol. Biol.* **160**, 181 (1982).
41. H. L. Monaco, J. L. Crawford, W. N. Lipscomb, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5276 (1978).
42. J. F. Blinn, *ACM Trans. Graphics* **1**, 235 (1982).
43. T. J. O'Donnell and A. J. Olson, *Comput. Graphics* **15**, 133 (1981).
44. A. J. Olson, *Tomato Bushy Stunt Virus,* film available from Palmer Film Services, 611 Howard Street, San Francisco, Calif. 94105.
45. J. M. Blaney *et al.*, *J. Med. Chem.* **25**, 785 (1982).
46. P. K. Weiner, R. Langridge, J. M. Blaney, R. Schaefer, P. A. Kollman, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3754 (1982).
47. E. D. Getzoff, thesis, Duke University, Durham, N.C. (1982).
48. F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
49. T. Whitted, *Commun. ACM* **23**, 343 (1980).

50. I thank T. J. O'Donnell and A. J. Olson for the use of their program GRAMPS, the Brookhaven National Laboratory Protein Data Bank for x-ray coordinates (*48*). M. Pique for assistance at the Computer Graphics Laboratory, University of North Carolina at Chapel Hill (NIH RR 00898, F. P. Brooks, Jr., principal investigator), J.-P. Dumas for assistance at the Salk Institute, and the Helen Hay Whitney Foundation for a postdoctoral fellowship.

Ground Water Contamination in the United States

Veronica I. Pye and Ruth Patrick

Ground water that is used by humans consists of subsurface water which occurs in fully saturated soils and geological formations. Nearly half the population of the United States use ground water from wells or springs as their primary source of drinking water (*1, 2*); 36 percent of the municipal public drinking water supply comes from ground water (*1*); and 75 percent of major U.S. cities depend on ground water for most of their supply (*3*). Total fresh ground water withdrawals in 1980 were estimated as 88.5 billion gallons per day, of which 65 percent were used for irrigated agriculture (*4*). Although ground water contamination has occurred for centuries, increased industrialization, population density, and agricultural activities have greatly exacerbated the problem in some areas. As our dependence on ground water increases, its quality becomes an ever more important issue.

Ground water is not only important to man, it is also an integral part of the hydrologic cycle of the earth—the circulation of water between the oceans, atmosphere, and land. It constitutes approximately 4 percent of the water in the hydrologic cycle, second only to the oceans and seas, which account for about 94 percent (*5*). The volume of ground water in storage exceeds the volume of fresh surface water in lakes, streams, and rivers. Approximately 30 percent of the streamflow of the United States is supplied by ground water emerging as natural springs or other seepage areas (*2*). Ground water forms most, if not all, of the low water flow of streams during dry periods. The interrelation between surface water and ground water is further indicated by the fact that, under certain conditions, surface water may recharge ground water aquifers.

Aquifers may be composed of permeable or porous geological material, either unconsolidated sand and gravel or consolidated material such as carbonate

Veronica I. Pye is Research Director of the Environmental Assessment Council, Academy of Natural Sciences, Philadelphia, Pennsylvania 19103. Ruth Patrick is Chairman of the Environmental Assessment Council, is Senior Curator of Limnology and occupies the Francis Boyer Research Chair at the Academy of Natural Sciences, Philadelphia, and is Adjunct Professor at the University of Pennsylvania.

The report on ground water on which this article is based was prepared by the Environmental Assessment Council. Council members were Robert G. Dunlop, Caryl Haskins, Richard E. Heckert, Lane Kirkland, George Lamb, Charles F. Luce, Ruth Patrick, Glen Paulson, William Reilly, Laurance S. Rockefeller, Abel Wolman, and George Wills.