A new approach to illustrating electrostatic molecular surfaces

Haruki Nakamura, Masam Kusunoki* and Noritake Yasuoka*

Department of Applied Physics, Faculty of Engineering, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan
*Institute for Protein Research, Oosaka University, Suita, Oosaka 55, Japan

A computer program system is used to illustrate electrostatic molecular surfaces of proteins by a polyhedron model. The whole solvent-accessible surface of the molecule is approximated by a polyhedron, the surface colour of which is able to represent the electrostatic potential at the position of each polygon. The program written in Fortran is composed of two stages. During the first stage, it calculates the positions of all the surface polygons and the potentials. In the second, the polyhedron is illustrated as a picture viewed from any desired direction on a raster computer graphic system. Optionally, the types of atoms or the hydrophobicity can be represented instead of the electrostatic potential.

Keywords: surface approximation, raster graphics, electrostatic potential

received 1 December 1983

Fairly significant anisotropy of electrostatic potential has been found in many globular proteins, due to the asymmetric distribution of both macroscopic electric dipole moments produced by α-helices[1,2] and ionic charges of polar amino acid residues[3-11]. Although there are other fundamental shortrange interactions, the electrostatic potential involved not only in protein–protein interactions but also in protein–DNA interactions has been recognized to have important roles to play. These include stabilizing tertiary structures[2,3,12], binding substrates or enhancing enzymatic reaction rates[1,9-11].

Several methods of 2D[8,10,13,14] and 3D[9-11] representation have been developed in order to exhibit electrostatic potential maps. Among them, a recent method illustrating the electrostatic potential molecular surfaces[11] may be the most powerful. It is one of the applications of interactive computer graphic displays of solvent-accessible surfaces developed by M L Connolly[15-18] where colour-coded electrostatic potential is displayed on the solvent-accessible surfaces by dots. This method has also made it possible to assign the protein interactive region which may be associated with other molecules, and to predict the conformation of a molecular complex[11,19]. However, it is a little difficult to understand the 3D features from those interactive illustrations, when they are represented or printed as pictures in repose.

We propose here a new approach to illustrating an electrostatic potential molecular surface, which is approximated by a polyhedron with coloured polygons coded as the electrostatic potential values at the positions. This illustration is suited especially for a raster graphic system, and one can easily grasp a tertiary image of the potential distribution from it without any shading or stereo-viewing procedures.

METHOD

Following F M Richards' definition[20], a solvent-accessible surface of a molecule is yielded as an envelope surface, which is the locus of the centre of a probe sphere rolling on the van der Waal's surface of the molecule.

The basic procedure for approximating the accessible surface of a molecule by a polyhedron is as follows:

- 1 A polyhedron is inscribed in a sphere, the centre of which is the ith atom position. The radius of the sphere is the probe radius (a_p) plus the van der Waal's radius of the ith atom (a_{vi})[20]. The sphere is called the accessible sphere. The surface of the accessible sphere is represented as (a_p+a_{vi}, θ, φ) in polar coordinates, the origin of which coincides with the atom position and the vertices of the polyhedron are determined as

$$\theta = \theta_j \equiv \pi\,(j-1)/m \qquad (1 \leq j \leq m+1) \qquad (1)$$
$$\phi = \phi_k \equiv 2\pi\,(k-1)/n \qquad (1 \leq k \leq n+1) \qquad (2)$$

  They produce $m \times n$ tetragons, the vertices of which are assigned as (θ_j, ϕ_k), (θ_{j+1}, ϕ_k), (θ_{j+1}, ϕ_{k+1}) and (θ_j, ϕ_{k+1}) where ($1 \leq j \leq m$, $1 \leq k \leq n$).
- 2 Those tetragons are selected which do not completely overlap the accessible spheres originating from other atoms. In order to check quickly whether each vertex or the polyhedron overlaps or not, the candidate accessible spheres, which possibly overlap the polyhedron, are selected first. Then a table of (θ_j, ϕ_k) is made indicating whether each vertex overlaps or not, and non-overlapping tetragons are selected.
- 3 Tetragons that overlap partially other accessible spheres have their non-overlapping parts arranged into new polygons. When two or more polyhedra produced in step (2) overlap each other, the intersecting tetragons are cut and deformed into

generally complex polygons. They are treated simply as follows. First, when only one vertex overlaps other accessible spheres, two new non-overlapping vertices are nominated where the tetragon's sides intersect the sphere(s). With these two new vertices and the other three original vertices, a new pentagon is generated. In the same manner, when two or three vertices of the tetragon overlap, a triangle, a tetragon, a pentagon or a hexagon is generated as a non-overlapping polygon. Finally, when all four vertices overlap accessible spheres, and all of them do not overlap any one particular sphere, there is a possibility that a certain region of the tetragon may not overlap and indeed it may appear on the surface. In this case, the tetragon is further divided into small tetragons, from which non-overlapping ones are selected again.

Lots of polygons generated by following procedures (1)–(3) compose the polyhedron model of the solvent-accessible surface of the molecule. When the radius of the probe sphere a_p is set to zero, the surface represents the van der Waal's surface of the molecule, which resembles the molecular surface with curved cross-hatching[21].

Colour plate 1 (see p C4) shows an example of the polyhedron model of the amino acid residue Phe-6 of glucagon, where a_p was zero. The colour code of each polygon corresponds to the atom type as indicated on the right of the screen.

Calculation of the electrostatic potential Ψ at a point $\vec{r}_0$ for a set of charges $\{q_i\}$ at the positions $\{\vec{r}_i\}$ in a medium with dielectric constant ε, follows the classical formula:

$$\Psi(\vec{r}_0) = \Sigma \frac{q_i}{i\,\varepsilon\,|\vec{r}_i - \vec{r}_0|} \qquad (3)$$

The charges $\{q_i\}$ used in the present study for proteins are the atomic partial charges, which are determined as Mulliken's gross atomic population[22] from quantum chemical calculations. Several tables of the partial charges are available for amino acid residues. Among them, two data sets are used for two individual programs; one obtained from *ab initio* calculation[14] and the other using the CNDO/2 method[23]. The former values have an advantage of the non-empirical procedure in the calculation. The latter values have been widely used for the internal energy calculations of proteins[24]. These partial charges are assigned at the positions of all the atoms in the protein molecule including ionized residues. All the coordinate data $\{\vec{r}_i\}$ available from X-ray studies are taken from the Brookhaven data bank[25]. When the position of a hydrogen atom can be determined geometrically, its atomic partial charge is set at the calculated position. Otherwise, for $-CH_3$, $-NH_2$, $-OH$ and $-SH$, the sum of the partial charges of the groups are assigned to C, N, O and S, respectively. In these cases, the van der Waal's radii are also enlarged to the group values by the united atom approximation[24]. For ionic residues, two different procedures are available. In the first procedure, none of the residues are ionized. In the second procedure, under a neutral pH value, almost all the ionic residues including N– and C–termini except *His* residues are ionized. For each *His* residue, an interpolated value between ionized and non-ionized partial charge is calculated according to the pK value assigned. The electrostatic potential to be assigned to each tetragon or rearranged polygon is now calculated from the position that is on the accessible sphere surface using the polar angles of the middle point of the original tetragon before rearrangement.

In the next step, the colour-coded polygons are displayed on the computer graphic display. Since modern raster graphics hardware can paint a polygon quickly with an assigned colour and pattern, it is not necessary to develop a new hidden surface algorithm. After a proper rotation of the whole polyhedron model, the polygons are sorted and painted in order of the z coordinates of their centres. Here, the z-axis is defined as the axis perpendicular to the screen and coming forward. During this procedure, polygons, with normals to their surfaces which point into the screen, are eliminated. When the polygons are displayed, their outlines are marked in white so that a tertiary image is grasped easily. However, for small tetragons, which were generated in step (3), no outlines are painted. Otherwise, there might be too many lines to understand the whole structure.

As an option, pattern-coded polygons can be displayed instead of colour coded polygons for a monochromic graphic display or a hard copy system.

PROGRAM AND GRAPHIC DISPLAY SYSTEM

The computer program TERAS (tertiary representation of accessible surface) written in Fortran is in two stages, since the software of a graphic display depends upon the hardware system. During the first stage, it calculates the positions of the vertices of all the accessible surface polygons and the electrostatic potentials assigned to the polygons following the method mentioned above. The resulting values are then stored temporarily in an intermediate datafile. Optionally, the hydrophobicity of the amino acid residues or the types of atoms can be assigned as colour codes instead of electrostatic potential in this stage. During the second stage, the polyhedron model is illustrated as an image viewed from any desired direction by a computer graphic system. The stereo image may also be drawn. In this stage, minor changes to the program make it possible to use it with different graphics hardware. Two options are also available for illustrations of the hydrophobicity and the space-filling model, respectively.

The authors have used two raster computer graphic systems:

- HITAC-M200H/M280H computer (host) and colour graphic video terminal (HT-5425-E82), which belong to the Computer Centre, University of Tokyo
- ACOS-series-77-system-900 computer (host) and colour graphic video-terminal (N6960A) which belong to the Institute for Protein Research, Oosaka University

The former uses the 'CGDM' software graphic package[26] and the latter 'SPLOT'[27].

RESULTS AND DISCUSSION

Colour plates 2 (a) and (b) show examples of the electrostatic molecular surfaces of triose phosphate isomerase which has a typical parallel α/β type structure called a β-barrel[26]. The charges from CNDO/2 method were used with dielectric constant 4 and no ionisable residues were assumed to be dissociated. The radius of the probe sphere was 3.0 Å, a slightly large value to help grasp the whole trend of the potential distribution. The mainchains were simultaneously displayed in the illustrations. They include no ionic contribution, but they help to emphasize the effects of parallel α-helices and β-barrel. As expected from its tertiary structure, it is clearly seen that this protein has a typical dipolar potential distribution without the ionic contribution[26]. Moreover, the position where the phosphate group of the co-enzyme is attached has the highest positive electrostatic potential, just as predicted by Hol[1].

Although illustrations like Colour plates 2 (a) and (b) may be useful to classify the protein according to the type of its electrostatic multipole moment[26], they underestimate both the contributions from the ionic residues and the shielding effects produced by the large water dielectric and by the small ions around the protein. When charges due to the ionic residues are added, the shielding effects should be included in order to calculate the electrostatic potential especially outside the protein. As far as the formula of equation (3) is used, the shielding effects should be included in the apparent dielectric constant. There have been several proposals such as $\varepsilon \propto |\vec{r}_i - \vec{r}_0|$[11,30,31] and $\varepsilon \propto \exp(\kappa |\vec{r}_i - \vec{r}_0|)$ (κ; Debye shielding parameter)[8,9] and so on. Another approach for the calculation uses a much more realistic model of polarized atoms with mobile water dipoles, which was first examined by Warshel and Levitt[31]. This microscopic model is certainly superior to the continuous model of equation (3), but may take more time because of its iterative calculations. The present program uses equation (3) and so the resulting potential map includes inevitable over- and underestimation of the shielding effects. Moreover, Mulliken's gross atomic population has some ambiguities. In a somewhat arbitrary manner it assigns continuous electron density as separate point charges[26]. Care should be taken when calculating the potential very near to atoms.

The representation of a macromolecule fragment shown in Colour plate 1 and the overall representation in Colour plates 2 (a) and (b) were made by the same program TERAS in principle. However, for the latter colour plate, the rearranging procedure to make a polyhedron (described in step 3) was omitted, as otherwise too many polygons and outlines might appear to obtain a whole image of the potential surface. The time to calculate and display the surface can be reduced considerably by this rather rough simplification.

CONCLUSIONS

In studies of macromolecules it may be a good idea to illustrate the accessible surface of the molecules with colours coded by several physical chemical properties at the position, as not only the conformations of the molecules but also their properties can be seen at a glance. The present approach, which is to approximate the surface by a polyhedron, is one method of display for a raster graphic system. This polyhedron model may be developed in many ways, such as:

- making the polygon meshes quickly with 3D spline curves which cover the surface completely
- using half-transparent colours for the polygons so that it is possible to see inside the surface
- illustrating the polyhedron in realtime with rotation and parallel movement using interactive graphic systems

ACKNOWLEDGMENTS

This work was partly supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Japan. The authors wish to thank Professor A Wada, University of Tokyo, for his helpful comments and discussions.

REFERENCES

1 **Hol, W G J, van Duijnen, P T and Berendsen, H J C** *Nature* No 273 (1978) pp 443–446
2 **Hol, W G J, Halie, L M and Sander, C** *Nature* No 294 (1981) pp 532–536
3 **Wada, A and Nakamura, H** *Nature* No 293 (1981) pp 757–758
4 **Barlow, D J and Thornton, J M** *J. Mol. Biol.* No 168 (1983) pp 867–885
5 **Friend, S H and Gurd, F R N** *Biochemistry* No 18 (1979) pp 4620–4630
6 **Matthew, J B, Friend, S H and Gurd, F R N** *Biochemistry* No 20 (1981) pp 571–580
7 **Friend, S H, Matthew, J B and Gurd, F R N** *Biochemistry* No 20 (1981) pp 580–586
8 **Matthew, J B and Richards, F M** *Biochemistry* No 21 (1982) pp 4989–4999
9 **Matthew, J B, Weber, P C, Salemme, F R and Richards, F M** *Nature* No 301 (1983) pp 169–171
10 **Koppenol, W H and Margoliash, E** *J. Biol. Chem.* No 257 (1982) pp 4426–4437
11 **Weiner, P K, Langridge, R, Blaney, J M, Schaefer, R and Kollman, P A** *Proc. Natl. Acad. Sci. USA* No 79 (1982) pp 3754–3758
12 **Sheridan, R P, Levy, R M and Salemme, F R** *Proc. Natl. Acad. Sci. USA* No 79 (1982) pp 4545–4549
13 **Sheridan, R P and Allen, L C** *Biophys. Chem.* No 11 (1980) pp 133–136
14 **Sheridan, R P and Allen, L C** *J. Am. Chem. Soc.* No 103 (1981) pp 1544–1550 and its supplementary material
15 **Connolly, M L in Sayre, D (ed)** *Computational Crystallogr.* Clarendon Press (1982) p 273
16 **Langridge, R, Ferrin, T E, Kuntz, I D and Connolly, M L** *Science* No 211 (1981) pp 661–666
17 **Connolly, M L** *Science* No 221 (1983) pp 709–713
18 **Kuntz, I D, Blaney, J M, Oatley, S J, Langridge, R and Ferrin, T E** *J. Mol. Biol.* No 161 (1982) pp 269–288
19 **Richards, F M** *Ann. Rev. Biophys. Bioeng.* No 6 (1977) pp 151–176
20 **Bondi, A** *J. Phys. Chem.* No 68 (1964) pp 441–451

21 **Smith, G M and Gund, P** *J. Chem. Inf. Comput. Sci.* No 18 (1978) pp 207–210
22 **Mulliken, R S** *J. Chem. Phys.* No 23 (1955) 1833–1840
23 **Momany, F A, McGuire, R F, Burgess, A W and Scheraga, H A** *J. Phys. Chem.* No 79 (1975) pp 2361–2381
24 **Dunfield, L G, Burgess, A W and Scheraga, H A** *J. Phys. Chem.* No 82 (1978) pp 2609–2616
25 **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** *J. Mol. Biol.* No 112 (1977) pp 535–542
26 *HITAC character/graphic display manager manual* Second version No 8080–7–140, Hitachi (1982)
27 *ACOS N6960 graphic display library manual* FXG 07–1, NEC (1983)
28 **Banner, D W, Bloomer, A C, Petsko, G A, Phillips, D C, Pogson, C I, Wilson, I A, Corran, P H, Furth, A J, Milman, J D, Offord, R E, Priddle, J D and Waley, S G** *Nature* No 255 (1975) pp 609–614
29 **Nakamura, H and Wada, A** to be published
30 **Gelin, B R and Karplus, M** *Biochemistry* No 18 (1979) pp 1256–1268
31 **Warshel, A and Levitt, M** *J. Mol. Biol.* No 103 (1976) pp 227–249
32 **Kern, C W and Karplus, M** *J. Chem. Phys.* No 40 (1964) pp 1374–1389