ENHANCEMENT AND SIMPLIFICATION OF MACROMOLECULAR IMAGES

KEIICHI NAMBA,*‡ D. L. D. CASPAR,‡ AND GERALD STUBBS*
*Department of Molecular Biology, Vanderbilt University, Nashville, Tennessee 37235; and ‡Structural Biology Laboratory, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254

ABSTRACT Computer graphics programs have been devised to display selected atomic features and to simplify images of complex macromolecular structures. By using boundary outlines, adjustment of size and shape of the molecular components, color coding, shading, and selective omission of obscuring detail, attention can be focused on specific interactions which determine higher levels of organization. A balanced color table has been constructed in which different hues have equal steps in brightness; this table has facilitated distinction of atom types and sequence coding together with representation of an optimum range of depth cueing and surface shading. The graphics system has been used with the atomic coordinates of the tobacco mosaic virus structure to simplify images of the protein subunit, to illustrate intermolecular interactions, and to relate subunit packing arrangements in different assemblies to the underlying atomic structure. The system has also been used to construct a schematic representation of the polyomavirus capsid, based on low resolution data. Application of artistic methods contributes to the effective presentation and interpretation of detailed scientific information about complex macromolecular structures.

INTRODUCTION

Computer graphics methods have been applied to macromolecular structures to represent overall shapes, molecular surfaces, skeletal connections, and details of atomic arrangements (Feldman, 1976; Langridge et al., 1981; Bash et al., 1983). Increasingly complex biological structures are being solved to atomic resolution by x-ray diffraction analysis. Interpretation of this wealth of biological structural data requires graphic enhancement and simplification of atomic models to illustrate ordered conformations, stabilizing interactions, and subunit packing arrangements, which correspond to aspects of macromolecular organization categorized as secondary, tertiary, and quaternary structures.

We have illustrated distinctive features in the chain-folding, subunit packing, and overall morphology of tobacco mosaic virus (TMV) (Namba et al., 1985) by constructing sharply outlined pictures of the parts of this complex macromolecular assembly, with the image resolution reduced. Simplification by reduction in resolution however, tends to obscure the relationship between subunit conformation and higher levels of organization. We describe here another approach to simplifying macromolecular graphics starting with atomic coordinates. By enhancing key atomic features, using boundary outlines, size and shape adjustment, color coding, shading, and selective omission of obscuring detail, attention can be focused on essential spatial relationships at different levels of organization. The procedures used with atomic models can also be applied to construct more schematic representations of macromolecular assemblies.

Programming and Display System

The system consists of two programs which are written in FORTRAN. The first program reads a file which contains coordinates. They are generally atomic coordinates, but they can be any kind of model coordinates, which are treated as atomic coordinates. Coordinate points can be represented as solid spheres and vectors between coordinate points can be represented as solid cylinders. Sizes of spheres and cylinders, viewing direction, lighting angle and image size are all adjustable by input parameters. Color codes are assigned to types of atoms, chemical compositions and backbone sequence positions. The program makes a file of two-dimensional arrays, which stores shading light intensities, z coordinates, color codes, and a subunit identification code of surface points.

The second program reads this output file and analyzes z coordinates to draw external and internal edge lines, as previously described for illustrating the surface morphology of TMV protein assemblies (Namba et al., 1985). Subunit boundaries are drawn thicker than the lines marking relief within the subunit. These two types of edges are distinguished by comparing the subunit identification code of each pixel with those of the neighboring pixels. The balance of shading light and depth cueing may be adjusted by input parameters in this program. Color codes, shading

Dr. Namba's present address is ERATO 5-9-5 Tokodai, Tsukuba, Ibaragi 300-26, Japan.

light intensities, and levels of depth cueing are combined to determine numbers corresponding to colors from a table. These numbers are sent to the memory of the graphics terminal and the image is displayed on the screen. The cpu running time required to generate the pictures shown here ranged from ~1 min for a single protein subunit (Fig. 4), 3 min for a symmetric view of a helix turn composed of 17 subunits (Fig. 10), and 20 min for an oblique view of seven turns of the TMV helix (Fig. 16).

The devices used to produce the pictures were a VAX 11/780 computer (Digital Equipment Corp., Maynard, MA), a 512 raster graphics terminal (Advanced Electronic Design Inc., Sunnyvale, CA), and a Matrix color graphics recorder (Matrix Promotions, Morris Plains, NJ). The AED 512 has 512 × 512 picture elements and can display 256 colors at a time. The 256 colors can be chosen from combinations of red, green, and blue with 256 levels of brightness. The Matrix recorder was adjusted to display grey on polaroid color prints and on color negative and positive film for combinations of red, green, and blue of equal numerical magnitude.

Color Table

The dimensions of the color table are hue and brightness. The table in Fig. 1 has been set up with columns of different colors and rows of different brightness. We have used color for two purposes: to identify different types of atoms, and to code sequence position. For distinguishing atom categories, the colors should be as distinct as possible, while for marking gradation in sequence, the changes in hue should be as smooth as possible. Brightness is used to indicate shape and depth. The bright end of each column should have recognizably the same hue as the dark end and the steps in intensity should appear uniform.

There does not appear to be any simple analytical way to construct a color table to meet the requirements of maintaining the same hue in each column and the same brightness in each row. There are $(256)^3$ possible numbers which can define a three-color combination for the AED raster graphics terminal, of which 256 were to be selected for the 16 × 16 color table. The hue and brightness displayed on the terminal for any three-color combination is not the same as those seen in a color print or transparency generated with the Matrix recorder. Polaroid color prints were used to guide the construction of the color table by trial and error, starting with initial values for the grid based on the crude approximation that brightness should be proportional to the sum of the numbers representing the three component colors. Brightness was augmented by adding white as the hue saturated. The proportion of white required to match the brightness of different light hues was greater in the blue range than in the complementary yellow range of the color table. The numerical steps of the component primary colors between adjacent rows of different brightness were much larger at the dark and light ends of each column than in the middle, since there is a nonlinear relation between perceived brightness and the numerical intensity scale of the graphics terminal. Reproduction of the color table (Fig. 1) tends to distort the color balance and brightness scale. As reproduced here, the magenta and crimson have been darkened and the cyan and green lightened compared with the initial polaroid version. When feasible, the color table should be refined against the final form of reproduction.

The first column at the left of the color table (Fig. 1) contains black for edge lines, and grey, blue, beige, and white to choose for background. The second to eighth columns contain distinct colors for coding atom types: green for heavy atoms and ribose carbon; cyan for ionic nitrogen; blue for nonionic nitrogen; purple for phosphorus; magenta for aromatic carbon; crimson for ionic oxygen; and red for nonionic oxygen. The ninth to sixteenth columns contain colors ranging from orange to yellow for sequence coding of the peptide backbone and aliphatic side chain carbons. These eight columns of graded colors (which are set brighter than the distinct colors) were used with the TMV protein sequence to distinguish blocks of 20 residues in the polypeptide backbone, starting with yellow at the amino terminus and ending with orange at the carboxy terminus. The colors in the table can be reassigned for other purposes. Modified versions of this color table, with altered hues and brightness, can be constructed by interpolation and further trial and error refinement.

Simplifying Features in Atomic Models

A picture of a space filling model of the TMV protein subunit, viewed from the direction corresponding to the top of the virus helix (Namba et al., 1985; Namba and Stubbs, 1986), is shown in Fig. 2. This is a conventional representation of the molecular structure with all nonhydrogen atoms drawn as 3 Å diameter van der Waals spheres, using the atom-type color coding designated in the description of the color table but without sequence coding for the orange backbone atoms. The perimeter of the molecule is delineated by a dark outline, and thinner lines are used to mark boundaries of atoms which are not in contact with the atoms seen below them in projection. Edge lines, as used in this figure, enhance the three-dimensional appearance of the molecular surface, but it is almost impossible to visualize the polypeptide conformation in such a space filling representation of the structure.

A ball-and-stick illustration of the protein subunit backbone is shown in Fig. 3. The α-carbon atoms are represented by spheres of 2 Å diameter and the peptide groups by 1 Å-diameter rods. Color coding the backbone, and marking boundary lines facilitate recognition of the chain continuity. The fine detail used to represent the backbone, however, impedes perception of the chain topology.

A simplified illustration of the polypeptide chain is shown in Fig. 4. This picture is similar to a conventional line drawing of the α-carbon backbone, but with the chain segments represented by 2 Å-diameter rods. Blisters on the

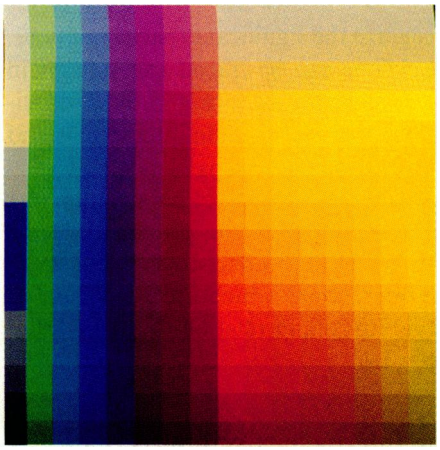

FIGURE 1 Color table constructed to maintain equal brightness in horizontal rows with columns of distinct colors for atom type and graded colors for sequence coding.

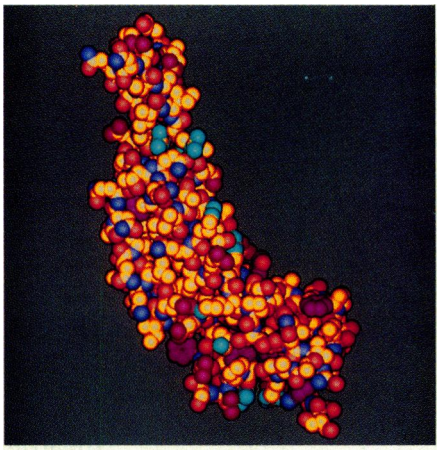

FIGURE 2 TMV protein subunit (*top view*) with all nonhydrogen atoms represented by 3 Å diameter spheres.

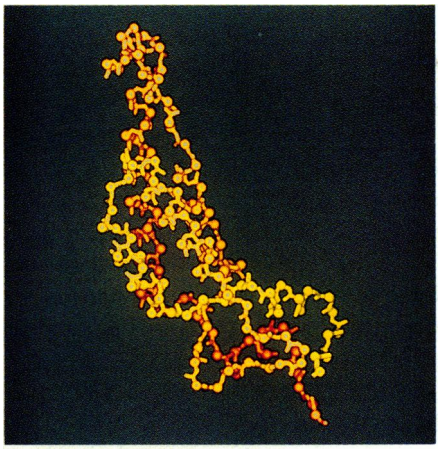

FIGURE 3 TMV protein subunit backbone represented by 2 Å diameter balls for C-α atoms and 1 Å-wide rods for bonds between peptide group atoms.

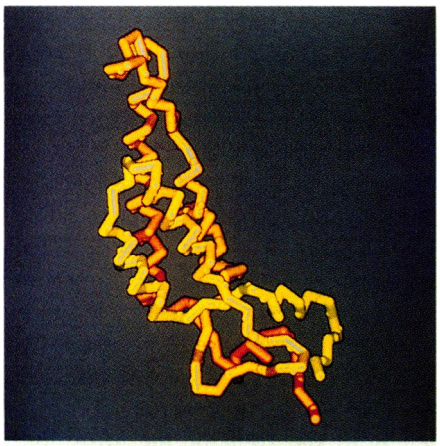

FIGURE 4 TMV protein subunit α-carbon backbone represented by rod segments of 2 Å diameter with the carbonyl carbons marked as blisters.

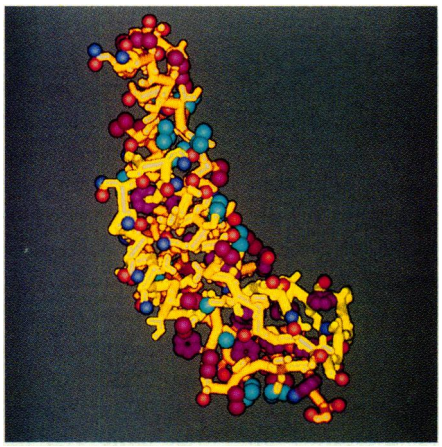

FIGURE 5 TMV protein subunit showing the amino acid side chains attached to the α-carbon backbone, with ionic oxygen and nitrogen atoms accentuated.

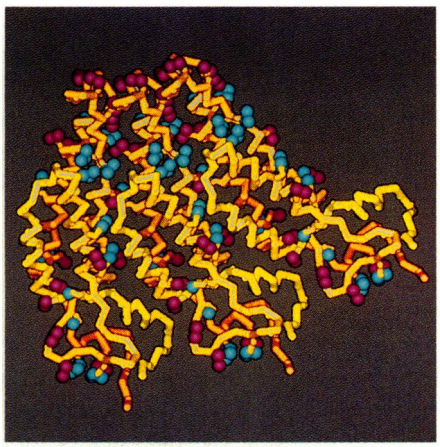

FIGURE 6 Three TMV protein subunits with only ionizable side chains marked, assembled side-to-side in the helical arrangement.

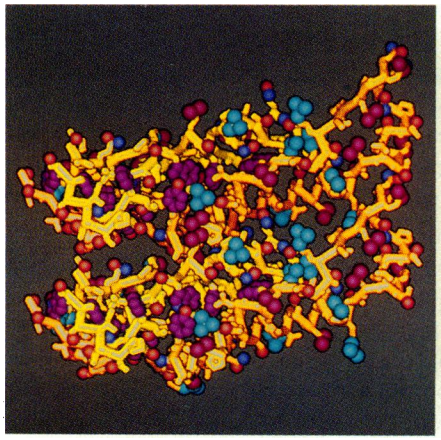

FIGURE 7 Two TMV protein subunits in edge view, with all side chains, illustrating the axial helical packing. (Helix axis is 20 Å to the right of the 70 Å-long subunits.)

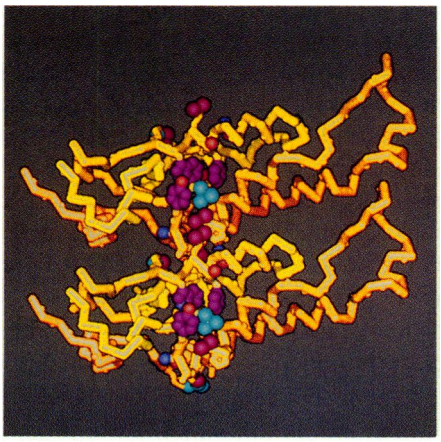

FIGURE 8 Two subunits (as in Fig. 7), with only the side chains involved in the intersubunit contacts between radii of 55 and 65 Å marked.

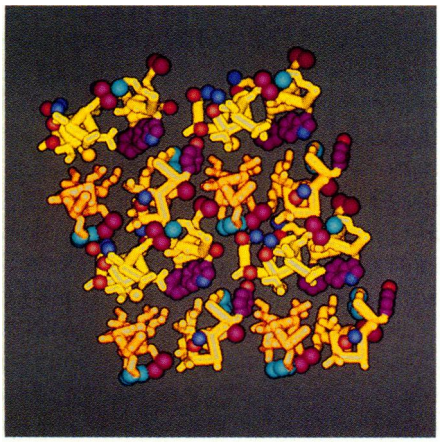

FIGURE 9 Cross-section of four subunits viewed from outside the helix illustrating the intersubunit contacts made by the side chains marked in Fig. 8.

rods, made by marking the position of carbonyl carbons with a 1.5 Å-diameter sphere, indicate the polarity of the polypeptide chain without introducing distracting detail.

Color coding the chain sequence, delineating the edge lines, simulating illumination of the cylindrical rod segments, and depth cueing all contribute to enhancing the perception of the three-dimensional shape of the backbone. Some portions of the chain are obscured by overlaps, but for most of the length, residue numbers can be counted from the kinks in the chain. For a molecule as simple as the TMV protein subunit, two orthogonal views about the long axis are sufficient to define the three-dimensional backbone structure (cf., Fig. 8). Stereo views of these pretzel stick drawings (Namba and Stubbs, 1986) give a vivid impression of the backbone conformations, but even in the two-dimensional view, the shape is evident.

Adjustment of each parameter used for constructing the backbone illustration in Fig. 4 involves a compromise among conflicting goals. A narrower chain would have less obscuring overlaps, but would also have less distinguishable shape and substance. A more extended carbonyl group would emphasize the chain polarity and α-helical hydrogen bonding but would introduce distracting detail. A thicker outline would more clearly delineate the chain crossovers, but would diminish the photographic illusion. A greater range of colors for the sequence coding would enhance the discrimination among segments, but would produce a motley appearance. A greater range of shading for depth cueing would reduce that available for surface shading. There are ranges of satisfactory combinations of parameters for representing the backbone, depending on the emphasis aimed for in the illustration. The compromises made in choosing the parameters for Fig. 4 produce an informative and aesthetically pleasing picture. Simplicity has been attained by omitting side chains and schematically representing the peptide links. Reattaching side chains to the backbone restores important information but complicates the image.

Fig. 5 shows all the side chains attached to the simplified backbone of Fig. 4. The information content in Fig. 5 is essentially the same as that in Fig. 2; there is however, considerable difference in the ease with which the information in these two pictures can be comprehended. In Fig. 5 there is no problem in distinguishing side chains from backbone; furthermore, the different types of side chains are readily recognized, the oxygens and nitrogen atoms of ionic residues are larger than those which are nonionic as well as being differently colored; aliphatic portions of side chains are composed of rod segments which are narrower than the backbone and have the same sequence specific color; and aromatic rings are distinguished by shape and color. Other color coding, size and shape discrimination schemes could be used to emphasize different features in the structure. Whatever scheme is used, if there is a great deal of information in the picture, as in Fig. 5, patient examination and comparison with other pictures are necessary to be able to recognize the significant spatial relations. For example, the backbone can be traced and side chain sequence positions can be identified in Fig. 5 by comparison with Fig. 4; and backbone-side chain interactions and overall shape can be perceived by comparison with Fig. 2. There is no way to display a great deal of information in a single picture without complexity. The graphics scheme used in these illustrations has been devised to allow selective focusing on the interactions which are significant at different levels of organization, in relation to the underlying atomic structure.

Illustrating Intermolecular Interactions

Assembly of three protein subunits, packed side to side as in the helical virus particle viewed from the top, is illustrated in Fig. 6. Only the acidic and basic residues are marked, leaving enough open space to see the conformation of the backbone. This picture shows how comfortably the subunits pack together sideways and indicates the nature of some of the electrostatic interactions among the charged amino acid residues which are significant in assembly. To visualize the intermolecular interactions in three dimensions requires more views with more detail.

Fig. 7 shows two subunits in edge view with all the side chains marked, illustrating the axial interactions in the virus helix. As in Fig. 5, there is a confusion of detail. Careful scrutiny and reference to simpler pictures are needed to assimilate the structurally significant information. A simpler view from the same direction of the two subunits is shown in Fig. 8. In this picture, the only side chains marked are those between radii of 55 and 65 Å from the virus axis (which is to the right of the 70-Å long subunits). This selective side chain omission focuses attention on the pair of carboxyl groups, at the center, which are critically important in controlling the virus assembly (Caspar, 1963; Namba and Stubbs, 1986). Another, more comprehensive representation of the intermolecular interactions between radii of 55 and 65 Å is shown in Fig. 9, which is the view from outside the virus particle of this section cut through four subunits. Each subunit has four α-helical segments in cross section, and the boundaries between subunits can be identified by the symmetry relationships. The function of the critical carboxyl pair of the center of Fig. 9 could be illustrated further by selective highlighting and with views in other directions.

Figs. 6–9 demonstrate the capacity of our graphics system to focus on specific intermolecular interactions in the context of the macromolecular architecture. The three-dimensional illusion could be enhanced by use of other graphics systems to simulate perspective, cast shadows and illuminate with multiple light sources. Motion picture displays, which utilize a very large number of individual images with zoom, and rotation and clipping, have been effective for displaying spatial relations in molecular interactions (O'Donnell and Olsen, 1981). A goal of our

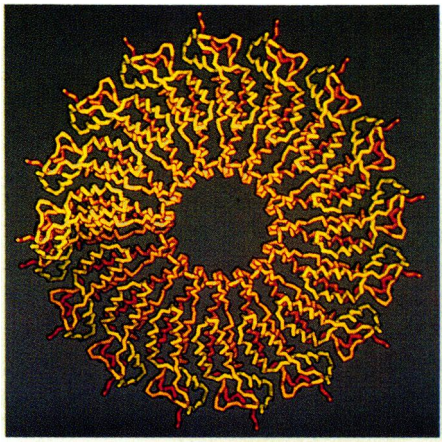

FIGURE 10 Seventeen TMV subunits represented by the α-carbon backbone of 2.5 Å diameter, arranged in one turn of the virus helix.

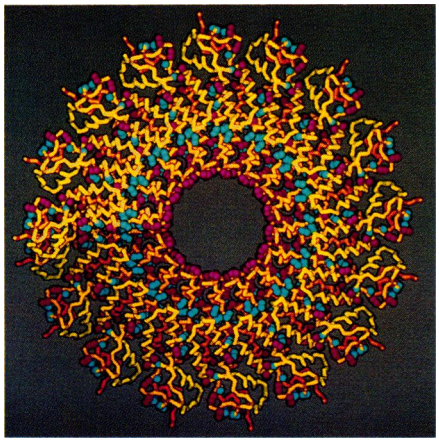

FIGURE 11 The subunit arrangement as in Fig. 10, with all ionizable residues marked, illustrating the concentrations of basic groups at 40 Å radius and acidic groups at 25 Å radius.

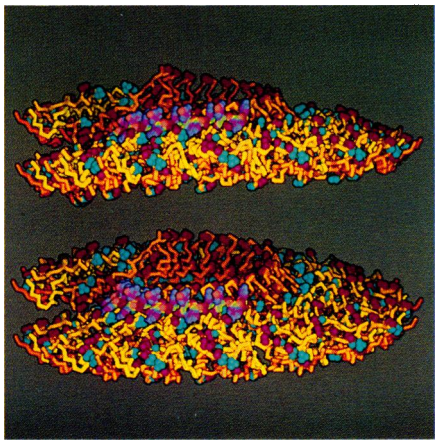

FIGURE 12 Edge view and oblique view of one helix turn, as in Fig. 11, with a highlighted quarter turn of the RNA chain interacting with the basic residues.

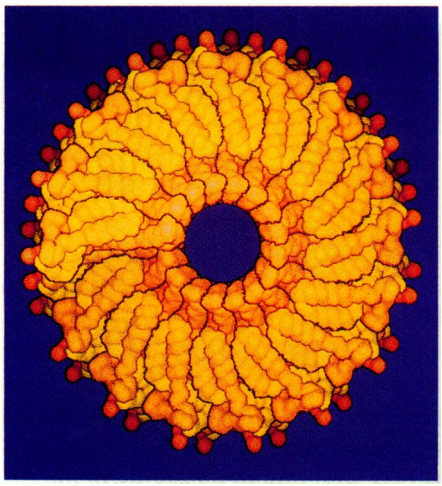

FIGURE 13 Low resolution image of two and a half turns of the virus helix with a space filling representation of the subunits illustrating the side-to-side packing.

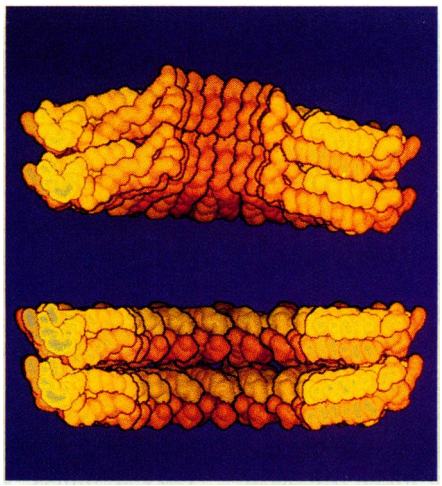

FIGURE 14 Cross-sectional edge views of two turns of the virus helix (*top*) and two layers of the four-layer protein disk (*bottom*) illustrating differences in the axial packing.

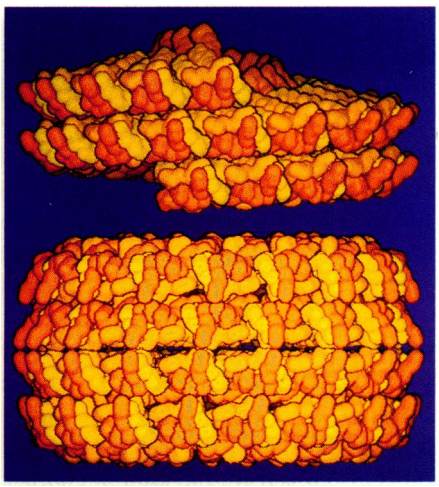

FIGURE 15 Edge views of the virus helix (*top*) and of the four-layer protein disk (*bottom*), illustrating differences in subunit arrangement seen from the surface.

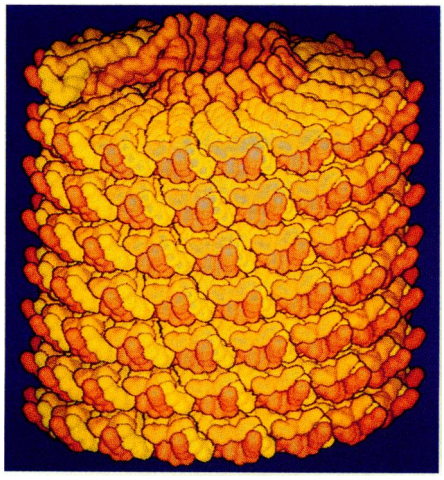

FIGURE 16 Oblique view of seven turns of the TMV helix.

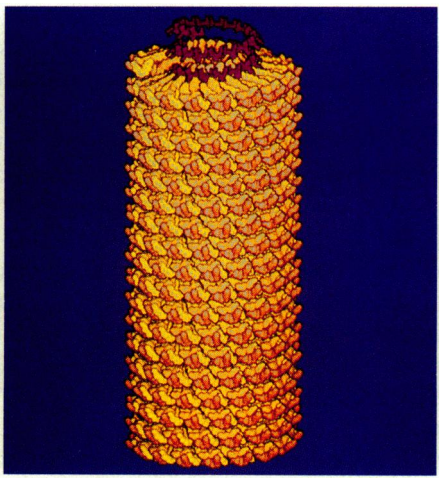

FIGURE 17 Oblique view of 18 turns of the TMV helix (415 Å length) with two turns of the RNA helix exposed at the top.

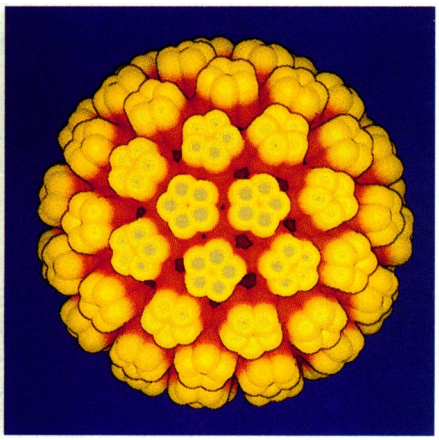

FIGURE 18 Schematic low resolution representation of the polyomavirus capsid of diameter 495 Å.

graphics system is to produce a small number of informative pictures, suitable for publication, which can simply and clearly illustrate critical interactions in complex macromolecular assemblies.

Representing Subunit Packing Arrangements

Proceeding to a higher level of organization, one turn of the helical assembly of TMV protein subunits is illustrated in Fig. 10, using the simplified drawing of the backbone. The depth cueing gives an impression of the pitch of helix, and distinguishes between the overlapped first and seventeenth subunits at the completed turn. This skeletal representation provides a framework in which to relate the subunit structure to the structure-determining interaction in the helical assembly.

Fig. 11 shows the acidic and basic residues attached to the seventeen subunits of Fig. 10. The information content in Fig. 11 is essentially the same as in Fig. 6, which shows only three helically arranged subunits with ionic residues attached. The redundancy in Fig. 11 focuses attention on the symmetric rationale of the higher level of organization. Basic residues, localized at ~40 Å radius, form a helical ribbon of positive charges designed to neutralize the negative charges of the RNA chain; and electrostatic repulsion among the carboxyl groups concentrated at the interior of the helix serves to regulate the virus assembly.

Attachment of a quarter turn of the RNA chain in the groove at 40 Å radius is illustrated in edge and oblique views of one helix turn in Fig. 12. Highlighting of the RNA by using a brighter portion of the color table, helps to distinguish it from the darker protein chains. These pictures also illustrate the kind of confusion which can occur in using the α-carbon skeletal drawing to represent the helical packing. Viewed from the top (Fig. 11), there is very little overlap among chain segments in one turn of the helix; but viewed from the edge (Fig. 12) there is a confusion of superposition which obscures the subunit boundaries. Visualizing overlapped arrangements of subunits in larger aggregates requires a simpler, more solid representation of the subunit.

Fig. 13 shows a top view of a two and a half turn segment of the virus helix, which has been built with solid, space-filling subunits. The information used to construct this picture is the same as that used for Fig. 10; the appearances are different because the rods forming the backbone have a diameter of 8 Å in Fig. 13 compared with 2.5 Å in Fig. 10. The 8-Å diameter backbone gives the subunit a volume similar to that of the conventional space-filling van der Waals representation (Fig. 2). Shading light intensity has been reduced to minimize surface detail, and the brightness scale for depth cueing has been expanded. In an earlier scheme for illustrating a smoothed subunit surface from a contoured low resolution map (Namba et al., 1985), the backbone continuity was obscured because of the interleaving of the smeared density from the side chains. In contrast, the chain continuity is still evident from the smoothed subunit surface generated from the swollen backbone as illustrated in Figs. 13–17. The essential features in these space-filling representations are the dark external edge lines delineating the subunit boundaries and the lighter internal edge lines which mark distinct segments of the folded polypeptide chain. The surface shading, depth cueing, and color coding provide additional information about the molecular shape and chain folding.

Representing the subunit as a sharply outlined, simple shape makes it possible to clearly illustrate the packing relationships in assemblies. The side-to-side interlocking of the subunits is displayed more distinctly in Fig. 13 than in Fig. 10, although these figures differ only in the diameter of the backbone model. Fig. 14 compares the axial interactions between two layers of the virus helix (top) and two layers of the four-layer crystalline disk aggregate (bottom) seen in cross section. In this picture of the disk, the inner loop of the subunit polypeptide chain is omitted, since it is disordered in the crystal (Bloomer et al., 1978) and its coordinates are indeterminate. Fig. 15 compares the axial interactions in the helix (top) and disk (bottom) as seen from the outer surface. The disk aggregate is illustrated as it occurs in the crystal (Bloomer et al., 1978), showing two-fold symmetric interactions as well as polar axial contacts in the four layer structure. In Fig. 15, the depth cueing has been set differently for the helix and disk to illustrate the use of relative brightness and color contrast to emphasize different aspects of the structure. A darker range of the color table was used for the picture of the helix segment, which enhances the color coded distinction between the yellow amino terminus and the orange carboxy terminus; whereas, the brighter hues used for the picture of the four-layer disk wash out the color difference for features nearest to view but intensify the appearance of depth.

The views of the helical aggregate in Figs. 15 and 16 correspond to the edge-on and oblique views of the skeletal models in Fig. 12. Compared with the skeletal models, the simplified subunit representations provide more comprehendible illustrations of the subunit packing. The view of a larger portion of the TMV particle in Fig. 17 emphasizes the exposed coils of the RNA chain with the conformation determined by its packing between the turns of the protein helix. Figs. 13–17 exemplify ways in which the structure of macromolecular assemblies can be related to more detailed images of the parts to explain higher levels of organization in terms of specific molecular interactions.

The graphics system developed for illustrating the shape and packing of TMV protein subunits can be readily adapted for constructing more schematic images. Fig. 18 illustrates the surface morphology of the polyomavirus capsid (Salunke et al., 1986), based on the electron density map solved by x-ray diffraction analysis at 22 Å resolution (Rayment et al., 1982). The pentameric capsomeres are

constructed from rods with rounded ends which are color-coded according to radial coordinate, from red at the inside to yellow at the outside. The size, shape, and packing arrangement of these pentamers have been adjusted to conform to the information from the low resolution structural analysis. Outlining of the simplified capsomeres emphasizes the differences in packing arrangement of the hexavalent and pentavalent capsomeres. Even though the actual subunit shape has not yet been determined, this schematic picture effectively illustrates the nonequivalent bonding relationships of the identical subunits.

CONCLUSION

Information content and informativeness are generally antithetical aspects in macromolecular graphics. The more detail displayed in a picture of a complex structure, the greater the possibilities for confusion. Simplification requires reduction in information content. The goal of controlled reduction of image detail is to facilitate recognition of selected information. This is the goal of artists. Presenting and interpreting the massive amounts of information now becoming available about macromolecular structures will require more concern with the methods of art. Aesthetic use of simplification, outlining, color balance, symmetry, and composition can all contribute to effective transmittal of quantitative scientific information.

We thank Dr. A. C. Bloomer for the coordinates of the TMV protein disk aggregate and Dr. D. M. Salunke for the parameters for the schematic polyomavirus model. This work has been supported by Public Health Service grants GM33265 (Gerald Stubbs) and CA15468 (D. L. D. Caspar).

Received for publication 26 June 1987 and in final form 2 November 1987.

REFERENCES

Bash, P. A., N. Pattabiraman, C. Huang, T. E. Ferrin, and R. Langridge. 1983. Van der Waals surfaces in molecular modeling: implementation with real-time computer graphics. *Science (Wash. DC).* 222:1325–1327.

Bloomer, A. C., J. N. Champness, G. Bricogne, R. Staden, and A. Klug. 1978. Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. *Nature (Lond.).* 276:362–368.

Caspar, D. L. D. 1963. Assembly and stability of the tobacco mosaic virus particle. *Adv. Protein Chem.* 18:37–121.

Feldmann, R. J. 1976. The design of computing systems for molecular modeling. *Annu. Rev. Biophys. Bioeng.* 5:477–510.

Langridge, R., T. E. Ferrin, I. D. Kuntz, and M. L. Connolly. 1981. Real-time color graphics in studies of molecular interactions. *Science (Wash. DC).* 211:661–666.

Namba, K., D. L. D. Caspar, and G. Stubbs. 1985. Computer graphics representation of levels of organization in tobacco mosaic virus structure. *Science (Wash. DC).* 227:773–776.

Namba, K., and G. Stubbs. 1986. Structure of tobacco mosaic virus at 3.6 Å resolution: implications for assembly. *Science (Wash. DC).* 231:1401–1406.

O'Donnell, T. J., and A. J. Olson. 1981. GRAMPS—A graphics language interpreter for real-time interactive three-dimensional picture editing and animation. *Computer Graphics.* 15:133–142.

Rayment, I., T. S. Baker, D. L. D. Caspar, and W. T. Murakami. 1982. Polyoma virus capsid structure at 22.5 Å resolution. *Nature (Lond.).* 295:110–115.

Salunke, D. M., D. L. D. Caspar, and R. L. Garcea. 1986. Self-assembly of purified polyomavirus capsid protein VP_1. *Cell.* 46:895–904.