ELSEVIER

MOLMOL: A program for display and analysis of macromolecular structures

Reto Koradi, Martin Billeter, and Kurt Wüthrich

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zurich, Switzerland

MOLMOL is a molecular graphics program for display, analysis, and manipulation of three-dimensional structures of biological macromolecules, with special emphasis on nuclear magnetic resonance (NMR) solution structures of proteins and nucleic acids. MOLMOL has a graphical user interface with menus, dialog boxes, and on-line help. The display possibilities include conventional presentation, as well as novel schematic drawings, with the option of combining different presentations in one view of a molecule. Covalent molecular structures can be modified by addition or removal of individual atoms and bonds, and three-dimensional structures can be manipulated by interactive rotation about individual bonds. Special efforts were made to allow for appropriate display and analysis of the sets of typically 20–40 conformers that are conventionally used to represent the result of an NMR structure determination, using functions for superimposing sets of conformers, calculation of root mean square distance (RMSD) values, identification of hydrogen bonds, checking and displaying violations of NMR constraints, and identification and listing of short distances between pairs of hydrogen atoms.

GENERAL CONCEPT

MOLMOL is a molecular graphics program with special emphasis on convenient display and analysis of sets of conformers used to represent the result of nuclear magnetic resonance (NMR) structure determinations.[1] To this end special features were implemented that are not readily available in other molecular graphics programs.[2–4] Additional potential applications of MOLMOL include the analysis of the trajectories of molecular dynamics simulations.[5]

MOLMOL applications can be grouped into three classes. First, different visualization possibilities were implemented for evaluating and comparing macromolecular structures. The different displays can readily be combined, and additional information such as interatomic distances, hydrogen bonds, etc., can be added to the displays. Special functions were implemented for presentation of the structural uncertainty given by the spatial spread among groups of NMR conformers, and for visualization of experimental NMR data in the structure. Second, different conformations of a given molecule, or chemically different structures such as homologous proteins, can be superimposed and quantitative comparisons by root mean square distances (RMSDs) are provided. Other quantitative analyses include searches for short interatomic contacts, evaluation of dihedral angles, identification and listing of hydrogen bonds, and the calculation of solvent-accessible surfaces. These data can be obtained in the form of graphs or tables. Third, MOLMOL is used for the generation of high-quality pictures for documentation or publication, which includes various ribbon presentations and other schematic, simplified views of complex macromolecular structures.

The graphical user interface of MOLMOL is based on the Motif toolkit,[6] using dialog boxes, menus, and buttons; the configuration of the latter two is up to the user. For all commands, on-line help is available. The main platforms for MOLMOL are Silicon Graphics workstations, or similar machines with powerful 3D graphics using the OpenGL library.[7] For less demanding applications, MOLMOL can be compiled and run on any UNIX machine, using only the X11 library.[8]

The next section describes the basic tools available in MOLMOL, which is followed by an account of advanced, novel features and their implementation in the program, and by a presentation of analysis tools that go beyond molecular displays.

BASIC TOOLS FOR STRUCTURE DISPLAY

Standard display possibilities implemented in MOLMOL include that atoms can be drawn as spheres of various sizes,

Color Plates for this article are on pages 29–32.

Address reprint requests to Dr. Wüthrich at the Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zurich, Switzerland.

Journal of Molecular Graphics 14:51–55, 1996

655 Avenue of the Americas, New York, NY 10010

0263-7855/96/$15.00
PII S0263-7855(96)00009-4

allowing both space-filling (CPK) and variable ball-and-stick representations. Bonds can be drawn either as lines or as cylinders with variable, user-selected radii. These standard displays can readily be combined with novel schematic presentations described in the next section. Display attributes such as color, shininess, and opacity can be set independently for each item. An example of a combination of various display styles within a single molecule is given in Color Plate 1. In addition to highlighting different structural elements, MOLMOL can also display nonbonding relations such as hydrogen bonds (Color Plate 2). NMR data can also be visualized, for example, as a complete set of distance constraints (Color Plate 2), or as a subset of all those constraints that are violated to an extent exceeding a predetermined limit.[9]

Independent choice of display styles for different parts of a molecular structure is made possible by a selection mechanisms with a powerful expression syntax. User-defined sets of atoms, bonds, distances, or primitives can thus be addressed, for example, to represent parts of a structure as ribbons, ellipsoids, or other simple geometric shapes (see below). Alternatively, conventional interactive selection by mouse clicks is also possible.

Display features supported by MOLMOL include complex annotations with text and lines (Color Plate 1), which can be properly placed in three-dimensional space for correct stereo images. The chemical structures can be modified by adding and deleting atoms and bonds. New three-dimensional structures can be generated by variation of dihedral angles about individual covalent bonds. The following stereo views may be displayed and plotted: single views for the left or right eye, stereo views as side-by-side or cross-eyed displays. For machines with the necessary hardware, stereo display with shutter glasses is also supported.

MOLMOL can read and write atom coordinates for three-dimensional macromolecular structures in the common data formats. High-quality plots from all displays can be generated as PostScript[10] or FrameMaker (MIF)[11] formatted files. These two formats yield files containing geometric primitives that will make use of the full resolution of the output device. Raster files at arbitrary resolution can be generated in the TIFF (Tagged Image File Format) standard.[12] In addition, the program can also produce input files for the public domain ray-tracing program POV-Ray. These files yield high-quality figures, with effects such as shadows, reflections, transparency, and texture mapping (see Color Plates 1 and 3–5).

The algorithm used for hidden surface elimination (which is needed for generating output in vector formats such as PostScript) is similar to the Weiler–Atherton algorithm.[13] Unlike the version described by Foley et al.,[14] it does not use a depth sorting step, but processes elements such as polygons and lines in the order they are generated, keeping track of and storing only the visible parts of the individual elements. Bounding rectangles are calculated for all elements, and the four coordinates of the rectangles (left, right, bottom, top) are used as keys for a four-dimensional tree,[15] which ensures that potential intersections can be determined rapidly. Care has been taken to avoid unnecessary intersection tests and splits. The algorithm accepts convex polygons as input, and produces convex polygons in the resulting output.

SPECIAL DISPLAY FEATURES

Polypeptide secondary structures

Schematic representations of regular secondary structures are indispensable for effective visualization of complex protein structures.[16] Such displays must be based on efficient identification of secondary structure elements in globular proteins. MOLMOL includes an algorithm for automatic identification of regular secondary structures.[17] As alternatives, the secondary structure elements may be read from a PDB file, or taken from publications on the structure determination. The exact start and end of each regular secondary structure element can be further modified interactively.

The drawing elements used by MOLMOL are ribbons for the presentation of helices, flat arrows for β strands, planes to emphasize β sheets, and tubes for loops and other nonregular structures (Color Plate 1). Size and shape of individual elements can be modified interactively, and neighboring elements can be smoothly connected by blending one shape into the other. Helices follow a natural cubic spline through given atoms, normally the α-carbon atoms. The orientation of the ribbon is determined by the axis of a cylinder fitted into the same atomic coordinates, and the deviation from the cylinder surface is minimized by a nonlinear least-squares algorithm. In contrast to other algorithms (for example, Ref. 18), no fixed radius or other "magic" constants are assumed, so that the fit results in nice displays for all types of helical polypeptide structures, such as α helices and 3_{10} helices (Color Plate 1). In β strands the α carbons follow a zig-zag pattern along the backbone. Before fitting to a cubic spline they are therefore first smoothed by replacing the coordinates $P(i)$ of atom i with $0.25 \times P(i-1) + 0.5 \times P(i) + 0.25 \times P(i+1)$. To further improve the display of the β strands within a given β sheet, the surfaces of the arrows are oriented so as to lie approximately in the plane defined by the neighboring strands. β sheets and β barrels can be emphasized as spline surfaces (Color Plate 1), in which the smoothed points used for the β strands are also used as control points for calculating a bicubic spline surface. In the two-dimensional parameter space each point is placed on a grid, where the parameter in the first dimension is given by the index of the individual strand within the sheet and the parameter in the second dimension by the index of the atom within the strand. The parameter is interpolated whenever data points are missing for certain grid positions, which happens typically in β bulges, or when the individual strands within a sheet have different lengths.

Schematic presentation of the backbone in nucleic acids

The same cubic spline function is used as for polypeptide helices. The spline is normally fitted to the phosphate atoms in the same way as the aforementioned fit to α-carbon atoms in polypeptides. An illustration of the resulting display is given in Color Plate 3.

Special presentations of discrete groups of atoms

Ring structures in aromatic amino acids and nucleic acids can be drawn as solid plates (Color Plate 3). An automatic

search of the atom coordinates locates and orients the rings without further input. The criterion used to identify a set of atoms forming a ring is that they are arranged in a cycle connected by bonds, and that the polygon defined by their coordinates is convex.

Discrete parts of a molecule, for example, individual amino acid side chains, structural units such as the helix–turn–helix motif, or entire protein domains can be emphasized with the use of geometric shapes such as spheres, ellipsoids, cylinders, cones, or rectangular boxes (Color Plate 3). The size of the bodies representing a set of user-selected atoms is defined by the smallest volume that contains a predetermined percentage of these atoms. This minimization problem, which has many local minima, is reliably solved by a simulated annealing algorithm[19] with the volume as the target function. Solutions that do not contain the required number of atoms are rejected. Part of the parameters can often readily be optimized once other parameters are given, e.g., the minimal cylinder radius can be calculated once the axis is given. Such parameters are calculated only when evaluating the target function and are not included in the annealing protocol.

Molecular surfaces

Three different kinds of molecular surfaces can be treated with MOLMOL, i.e., van der Waals surfaces, solvent-accessible surfaces,[20] and contact surfaces.[21,22] All of these can be determined and displayed either with dots or as shaded surfaces (Color Plate 4). As we use different algorithms for calculating dot surfaces or shaded surfaces, respectively, the approaches are described separately below. The solvent-accessible surface is defined by spheres centered about the positions of protein atoms for which the radius has been increased by addition of the radius of the sphere that represents the solvent molecule. Calculation of the solvent-accessible surface is thus otherwise identical to the van der Waals surface and will therefore not be discussed separately.

The dot representation of the van der Waals surface consists of the set of dots from all spheres centered on atom positions that are not inside another sphere. For the potentially costly testing for dots inside other spheres, a spatial grid is used that dramatically reduces the number of tests needed. The calculation with this technique can be done in linear time.[23]

Calculation of the dot representation of the contact surface is considerably more complex. Besides the points where the sphere representing the solvent touches a sphere of a protein atom, one also needs to generate points on surface parts where the solvent sphere rolls along the intersection of two spheres representing protein atoms, or where it touches three spheres representing protein atoms. The algorithm used in MOLMOL generates potential centers of the solvent sphere for all pairs and triples of protein atoms that are close enough, and tests whether the solvent sphere intersects another sphere. If not, the corresponding dots on the surface are generated. All the sphere centers are also stored, and in a second pass all points that were generated from pairs or triples of spheres are tested for inclusion in one of these solvent spheres. This eliminates self-intersections of the resulting surface. Again, for all sets of points that need to be stored, the aforementioned grid data structure is applied to make proximity queries more efficient.

The algorithms used for generating dot surfaces cannot readily be applied for generating shaded surfaces, as this would require a costly triangulation step. So we chose a different approach that directly produces a triangulated surface. We describe only the calculation of the contact surface, since van der Waals and solvent-accessible surfaces are evaluated accordingly. As a first step a dot representation (dots at points $\mathbf{s}_j$) of the solvent-accessible surface is generated, as explained in the previous section. In a second step, we build up a three-dimensional grid that encloses all spheres representing protein atoms i centered at $\mathbf{c}_i$ with radii a_i. For each grid point, $\mathbf{p}$, we calculate a distance D from the solvent-accessible surface. If r is the radius of the sphere representing the solvent and d the diagonal length of a cell in the grid, with all $a_i > d$, then

$$D = \begin{cases} r + d & \text{if } \exists_i\colon |\mathbf{p} - \mathbf{c}_i| < a_i - d \\ 0 & \text{if } \forall_i\colon |\mathbf{p} - \mathbf{c}_i| > a_i + r \\ \min\limits_j (|\mathbf{p} - \mathbf{s}_j|) & \text{otherwise} \end{cases}$$

Again, spatial data structures are used to minimize the number of tests. Finally, we calculate an isosurface at value r in this grid, using an algorithm similar to the marching cubes algorithm,[24] in which each cube is subdivided into five tetrahedra.[25]

Additional facilities

Among the additional features available in MOLMOL, we find the following three visualization routines particularly attractive: First, all display elements with solid surfaces, such as spheres, tubes, cylinders, ribbons, and ellipsoids, can be displayed with different surface properties, including shininess, opaqueness, and material character such as metallic or stonelike (Color Plates 1 and 3). Second, a series of snapshots resulting, for example, from the simulation of the dynamic behavior of a macromolecule may be displayed sequentially with a user-defined delay, thus providing an animated view of time-dependent processes. Third, variable precision of a structure determination may be displayed by a smooth tube with varying radius, so that local structural uncertainty is indicated by a wide tube, while a thin tube indicates high precision of the local structure determination. Combined with the use of color codes for other local parameters, such as the temperature factors of crystal structures, amide proton exchange rates, or NMR relaxation parameters, one can visualize multiple independently obtained measures in one display for direct comparison. Color Plate 5 shows the backbone of the NMR structure of the pheromone E*r*-2.[26] The thickness of the tube indicates the local precision with which this structure has been determined, and the colors visualize the experimentally determined amide hydrogen exchange rates.[1]

TOOLS FOR STRUCTURE ANALYSIS

Structure superpositions and RMSD calculations

MOLMOL is set up for simultaneous analysis of a large number of structures, and for quantitative comparisons be-

tween those. The structures can be different conformers of one molecule, as in typical presentations of NMR structures (Color Plate 2). Alternatively, the structures can also be chemically different; in this case the user selects fragments of corresponding lengths from the different molecules for superposition and RMSD calculation. For a group of structures a table of the global RMSD values[27] between all pairs of structures may be calculated, including the average, the standard deviation, and the minimum and maximum values. Plots of global or local displacements and local RMSD values versus the sequence can also be generated.[28] Corresponding entries from two or several groups of different structures can be compared, for example, sets of NMR conformers obtained with and without energy minimization. All these measures can be calculated either for the entire structures or for arbitrary selections of atoms. Based on the RMSD calculations, optimal superpositions may be obtained (Color Plate 2), and a mean structure may be determined by averaging the coordinates of equivalent atoms.

Analyses based on interatomic distances

Within a given set of atoms, interatomic distances that are shorter than a user-defined limit can be listed. If MOLMOL is further supplied with a file of chemical shifts, the program can generate a list of expected nuclear Overhauser enhancement (NOE) cross-peaks[1] for comparison with the experimental spectra. From user-defined criteria of interatomic distances and bond angles,[29] hydrogen bonds can be identified in a given 3D structure, listed in a table, and displayed in the molecular structure (Color Plate 2).

Constraints on interatomic distances and torsion angles that are used for the calculation of an NMR structure can be read into MOLMOL. Their residual violations can be determined and listed in a table. All constraints, the violated constraints, or some other selection of constraints can be visualized directly in the structure (Color Plate 2).

Additional options

MOLMOL can evaluate the torsion angles of the polypeptide backbone and draw a Ramachandran plot,[30] with "allowed regions" obtained by statistical analysis of a selection of structures from the data banks (Color Plate 6). Other useful plots include presentations of the distribution of dihedral angle values, such as ϕ, ψ, and χ^1, in the group of conformers used to represent the NMR solution structure versus the sequence, display of selected dihedral angles taken from snapshots of a molecular dynamics trajectory versus time (Fig. 1)[30a], or presentation of the solvent-accessible surface of the individual residues versus the sequence. MOLMOL also determines the principal axes of the global molecular shapes and calculates the corresponding radii of gyration.

AVAILABILITY

MOLMOL is available without charge from the following FTP servers: ftp.spectrospin.ch and ftp.mol.biol.ethz.ch. The licensing conditions are included with the program. Compiled versions are available for SGI, Sun, and IBM (AIX) machines. The source code can be compiled on any UNIX workstation, but the Motif library is required. More information is also available at the following URL: http://www.mol.biol.ethz.ch/wuthrich/software/molmol.

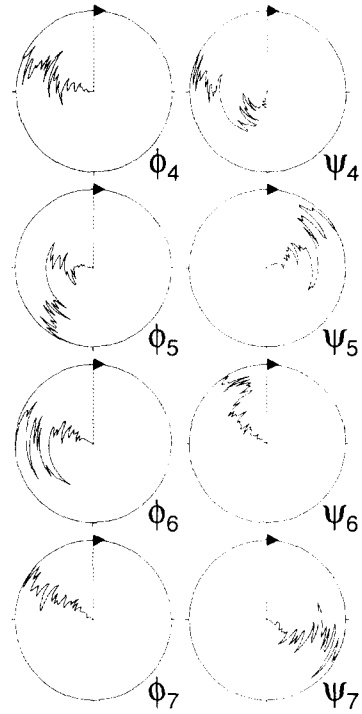

Figure 1. Visualization of a molecular dynamics trajectory by presentation of the real-time variations of the dihedral angles. MOLMOL is laid out for presentation of all angles ω_i, ϕ_i, ψ_i, and χ^j_i, although only the data for ϕ and ψ are shown here. The illustration was selected from a 2-nsec molecular dynamics simulation at ambient conditions of an Antennapedia homeodomain–DNA complex.[30a] Snapshots were taken between 1000 and 1050 psec at 1-psec intervals. For each angle the time axis runs from the center of the circle to its periphery; the angle values are given in a clockwise manner relative to the 0° position indicated at the top. Note the pairwise correlated variations of ψ_4 and ϕ_5, and ψ_5 and ϕ_6, respectively (see text).

ACKNOWLEDGMENTS

This work was financially supported by BRUKER/Spectrospin. The final version of Color Plates 1, 3, 4, and 5 was obtained with POV-Ray. POV-Ray is available without charge on CompuServe's GraphDev Forum or by anonymous FTP from ftp.povray.org. We thank the members of our research group for suggestions and feedback on the practical use of MOLMOL, and R. Marani for the careful processing of the manuscript.

REFERENCES

1 Wüthrich, K. *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York, 1986
2 Ferrin, T.E., Huang, C.C., Jarvis, L.E., and Langridge,

R. The MIDAS display system. *J. Mol. Graphics* 1988, **6,** 13–27
3 Jones, T.A., Zou, J.-Y., Cowan, S.W., and Kjelgaard, M. Improved methods for building protein models in electron density and location of errors in these models. *Acta Crystallogr.* 1991, **A47,** 110–119
4 Sayle, R.A., and Milner-White, E.J. RASMOL: Biomolecular graphics for all. *Trends Biochem. Sci.* 1995, **20,** 374–376
5 McCammon, J.A., and Harvey, S.C. *Dynamics of Proteins and Nucleic Acids.* Cambridge University Press, Cambridge, 1987
6 Heller, D., and Ferguson, P.M. *Motif Programming Manual.* O'Reilly & Associates, Sebastopol, CA, 1994
7 Silicon Graphic Incorporated. *The OpenGL Programming Guide.* Addison-Wesley, Reading, MA, 1993
8 Nye, A. *Xlib Programming Manual.* O'Reilly & Associates, Sebastopol, CA, 1992
9 Billeter, M., Engeli, M., and Wüthrich, K. Interactive program for investigation of protein structures based on ^{1}H NMR experiments. *J. Mol. Graphics* 1985, **3,** 79–83, 97–98
10 Adobe Systems Incorporated. *PostScript Language Reference Manual.* Addison-Wesley, Reading, MA, 1985
11 Frame Technology International Ltd. *MIF Reference.* San Jose, CA, 1991
12 Aldus Developers Desk. *TIFF Revision 6.0.* Aldus Corporation, Seattle, WA, 1992
13. Weiler, K., and Atherton, P. Hidden surface removal using polygon area sorting. SIGGRAPH '77 Proceedings (George, J., ed.), ACM, New York, 1977, pp. 214–222
14 Foley, J.D., van Dam, A., Feiner, S.K., and Hughes, J.F. *Computer Graphics: Principles and Practice.* Addison-Wesley, Reading, MA, 1990, pp. 689–693
15 Sedgewick, R. *Algorithms.* Addison-Wesley, Reading, MA, 1988, pp. 380–386
16 Richardson, J.S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 1981, **34,** 167–339
17. Kabsch, W., and Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983, **22,** 2577–2637
18 Kraulis, P.J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 1991, **24,** 946–950
19. Kirkpatrick, S., Gelatt, C.D., Jr., and Vecchi, M.P. Optimization by simulated annealing. *Science* 1983, **220,** 671–680
20 Lee, B., and Richards, F.M. The interpretation of protein structure: Estimation of static accessibility. *J. Mol. Biol.* 1971, **55,** 379–400
21 Richards, F.M. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* 1977, **6,** 151–176
22 Connolly, M.L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* 1983, **16,** 548–558
23 Halperin, D., and Overmars, M.H. *Spheres, Molecules, and Hidden Surface Removal.* Technical Report, Utrecht University-Computer Science, Utrecht, The Netherlands, 1994
24 Lorenson, W.E., and Cline, H.E. Marching cubes: a high resolution 3D surface construction algorithm. SIGGRAPH '87 Proceedings (Stone, M.C., ed.), ACM, New York, 1987, pp. 163–169
25 Ning, P., and Bloomenthal, J. An evaluation of implicit surface tilers. *IEEE Computer Graphics and Applications* 1993, **13(6),** 33–41
26 Ottiger, M., Szyperski, T., Luginbühl, P., Ortenzi, C., Luporini, P., Bradshaw, R.A., and Wüthrich, K. The NMR solution structure of the pheromone Er-2 from the ciliated protozoan *Euplotes raikovi. Protein Sci.* 1994, **3,** 1515–1526
27 McLachlan, A.D. Gene duplication in the structural evolution of chymotrypsin. Appendix: Least square fitting of two structures. *J. Mol. Biol.* 1979, **128,** 49–70
28 Billeter, M. Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals. *Q. Rev. Biophys.* 1992, **25,** 325–377
29 Widmer, H., Billeter, M., and Wüthrich, K. Three-dimensional structure of the neurotoxin ATX Ia from *Anemonia sulcata* in aqueous solution determined by nuclear magnetic resonance spectroscopy. *Proteins* 1989, **6,** 357–371
30 Ramachandran, G.N., and Sasisekharan, V. Conformation of polypeptides and proteins. *Adv. Protein Chem.* 1968, **23,** 283–437
30a Billeter, M., Güntert, P., Luginbühl, P., and Wüthrich, K. Hydration and DNA recognition by homeodomains. Submitted.
31 Bushweller, J.H., Billeter, M., Holmgren, A., and Wüthrich, K., The nuclear magnetic resonance solution structure of the mixed disulfide between *Escherichia coli* glutaredoxin(C14S) and glutathione. *J. Mol. Biol.* 1994, **235,** 1585–1597
32 Billeter, M., Qian, Y.Q., Otting, G., Müller, M., Gehring, J., and Wüthrich, K. Determination of the nuclear magnetic resonance solution structure of an *Antennapedia* homeodomain–DNA complex. *J. Mol. Biol.* 1993, **234,** 1084–1093
33 Berndt, K.D., Güntert, P., Orbons, L.P.M., and Wüthrich, K. Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. *J. Mol. Biol.* 1992, **227,** 757–775
34 Hobohm, U., Scharf, M., Schneider, R., and Sander, C. Selection of representative protein data sets. *Protein Sci.* 1992, **1,** 409–417