NEW PROGRAMS

The molecular surface package

Michael L. Connolly

San Francisco, CA, USA

The Molecular Surface Package is a reimplementation, in C, of a set of earlier FORTRAN programs for computing analytical molecular surfaces, areas, volumes, polyhedral molecular surfaces, and surface curvatures. The software does not do interactive molecular graphics, but it will produce pixel maps of smooth molecular surfaces. The polyhedral molecular surfaces are suited to display on graphics systems with real-time rendering of polyhedra.

Keywords: molecule, surface, analytical, area, volume, polyhedron, curvature

INTRODUCTION

The Molecular Surface Package (MSP) is a collection of five computer programs written in C that compute and analyze molecular surfaces. The programs compute surface envelopes for protein molecules in three formats: smoothly curved, polyhedral, and as an array of points. Areas and volumes are computed analytically. Internal cavities are identified. Protein surfaces can be rendered to produce a pixel map. The curvature (concavity/convexity) of a protein surface can be quantitated. The software is based on ideas of solvent accessibility developed by Fred Richards and his colleagues at Yale.[1–3] An early, widely used implementation of these ideas, was the program MS.[4,5] Other numerical surface algorithms have also been developed.[6–15]

Address reprint requests to Dr. M.L. Connolly at 2269 Chestnut St. #279, San Francisco, CA 94123, USA.

Received 15 September 1992; accepted 7 October 1992

ANALYTICAL ALGORITHMS

At the Research Institute of Scripps Clinic a series of computer algorithms was developed based on an analytical description of the molecular surface. This work was done in FORTRAN on a VAX-11/750 computer operating under VMS. The AMS,[16] LAMS,[16] VAM,[17] RAMS,[18] and MCS[18] programs were written while a Helen Hay Whitney postdoctoral fellow working under the supervision of Arthur J. Olson. The CT,[19] MST,[19] PC1,[20] PC2,[20] HEE,[20] shape measurement,[21] docking,[22] and chain-complex intersection[23] programs were written while an Assistant Member in the Department of Molecular Biology. This latter work was partially supported by NIH grant GM 34338.

REIMPLEMENTATION OF ALGORITHMS

Because of a dispute with Scripps Clinic over the ownership of the copyright to the source code of the aforementioned computer programs, it was decided to rewrite this software while self-employed in 1986, working under the fictitious business name Biohedron. The new programs were written in C, but they were not translations of the earlier FORTRAN programs (which would have made them derived works), but rather were entirely new implementations of the old algorithms. This work was performed on a MASSCOMP 5500 workstation with 2 MB of RAM and a floating-point board, operating under UNIX. After this computer broke down in 1987, I ported the software to the Macintosh. Further development occurred in 1988–90 on a Macintosh SE with 1 MB RAM and Aztec C, and during 1991–92 on a Macintosh IIsi with 5 MB of RAM, a MC68882 floating-point chip, Aztec C, and the Apple MPW Shell.

Table 1 describes the five programs of the Molecular Surface Package. It also shows the relationship between the earlier FORTRAN programs and the later C programs. There is not an exact one-to-one correspondence between the FORTRAN and C programs.

The PQMS program does the following five things:

(1) It rolls a sphere over a molecule, generating convex, concave, and saddle-shaped bits of surface.
(2) It removes bits of reentrant surface lying inside the solvent-excluded volume of the molecule (cusp trimming).
(3) It divides the surface generated into connected components (cavity identification).
(4) It computes the area of each surface face.
(5) It computes the volume of each surface component.

The program's actions are controlled by a command script. Areas and volumes can be written to disk in ASCII format, while the surface itself can be written to disk in binary format, for further use by SRF and TRB. The term *piecewise quartic* refers to the fact that the molecular surface is made up of pieces of surface defined by second (spherical) or fourth (toroidal) degree equations.

© 1993 Butterworth–Heinemann

Table 1. List of programs

Program	Description	Lines of code	Programs replaced
PQMS	Piecewise quartic molecular surface	9000	AMS, VAM, CT
SRF	Surface rendering by foliation	5000	RAMS
TRB	Triangulation by recursive bisection	7000	MST
Omega	Solid-angle measurement of shape	3000	
DS	Dot surface	6000	MS

The SRF program produces a pixel map of one or more molecular surfaces and one or more ball-and-stick models. The surfaces can be colored according to atom, surface component, or some chemical property defined for each point of a dot surface. The image is written to disk in sunraster format, which is displayable under X-Windows by the public-domain xloadimage utility. The word *foliation* means a family of parallel, adjacent arcs that form a surface. Each face of the surface is sliced into arcs one pixel wide, which are then rendered individually. The front surface of a protein molecule can be clipped to reveal internal cavities, which can be given distinct colors.

The TRB program reads the binary surface file written by PQMS and produces a polyhedral surface from it. Each polygon of the polyhedral surface is a triangle. The vertices, edges and triangles are written to disk. Each triangle is identified by the atom and surface component it belongs to. This information can be used to color code the display of the surface on a graphics system. The term *recursive bisection* refers to the fact that each curved face of the molecular surface is bisected to form two, smaller curved faces; each of these is, in turn, bisected, and so on, until only small triangles remain.

The Omega program reads the polyhedral molecular surface produced by PQMS and TRB, and then computes its curvature in the vicinity of each vertex. It rewrites the polyhedral molecular surface to disk, with a curvature (solid angle) associated with each vertex. The name *Omega* comes from the mathematical convention that the capital Greek letter Ω is used to denote a solid angle. A solid angle is the area of a region on a sphere divided by the sphere radius squared. A sphere centered at a vertex of the polyhedral surface lies partially inside and partially outside the solvent-excluded volume of the molecule. The solid angle of the part that lies inside measures the curvature of the surface lying inside the sphere. It is greater than 2π for concave regions, and less than 2π for convex regions.[21] The size of the sphere, of course, determines the scale of features identified.

The following programs have not yet been reimplemented: LAMS, MCS, PC1, PC2, and HEE. Some of the shape measurement and docking algorithms have been reimplemented in the VOID program.[24,25] The chain-complex intersection algorithm has also been reimplemented in the VOID program.

ADVANTAGES OF THE MOLECULAR SURFACE PACKAGE

One of the reasons that the MSP is better than the original FORTRAN programs is that it is coded in C. C has dynamic memory allocation, structures, and is more object-oriented than FORTRAN. However, the main advantage of the MSP is that it is more robust than the original FORTRAN programs. The AMS program failed fairly often due to probe spheres being simultaneously tangent to four atoms. This happens far less often with PQMS for two reasons: PQMS is in double precision, while AMS was in single-precision; and if the PQMS program fails for a molecule, it perturbs the radii of the atoms that caused the difficulty by 0.1 Å and recomputes the surface. (Obviously, AMS could have been made double precision, and some users have done so.)

The areas and volumes reported by PQMS are more accurate than those of VAM, because areas and volumes are computed after cusp trimming, rather than before. The cusp trimming in PQMS handles several difficult cases not handled by CT. In addition, PQMS reports volumes of individual cavities, which VAM did not do. Another enhancement is the reporting of areas broken down by atom and cavity. For example, if a phenyl ring lies between two internal cavities, PQMS will report the amount of surface area bordering each cavity.

The TRB program is more robust than the MST program, because its decision on which pair of vertices of a face's boundary to connect by a bisecting arc is based upon empirically determined optimal parameters. The improved cusp trimming in PQMS also helps to produce a smoother polyhedral surface. The TRB program also has the advantage that the user can specify the fineness of the tessellation on a per-atom basis, not just a global basis.

The DS program has two modes: complete and connected. The former is similar to the program MS, and includes surfaces for internal cavities. In the latter mode, the probe sphere traces a continuous path over the protein surface. This algorithm is similar to the algorithm of MSEED.[26,27] The continuous rolling algorithm is also similar to the convex hull sweep algorithm,[28] and the variable-radius-probe molecular interstitial skeleton algorithm.[29] The DS connected mode fails to identify internal cavities, but it is 2–3 times faster than the complete mode. The DS program has the advantage over the MS program that the dot surface calculation code is partitioned into a distinct set of subroutines that do no disk I/O and that can be linked with a molecular modeling program to provide interactive dot surface generation.

DISTRIBUTION

The software is distributed on Macintosh floppy disks. The documentation consists of several documents in MacWrite II format. A perpetual source

code site license costs $850 for North American academic end users, $1500 for noncommercial foreign end users, and $5000 for commercial end users. Interested parties should write to me at 2269 Chestnut St. #279, San Francisco, CA 94123, USA.

ACKNOWLEDGMENTS

I thank the following people for testing this software: Henry Dayringer of Monsanto; Fred Brooks, Helga Thorvaldsdottir and Doug Schiff of the University of North Carolina at Chapel Hill; Frank Prendergast, Brad Erickson, Chris Haydock, Andrew Myers, and Arthur Weaver at the Mayo Clinic; Rik Wierenga, Martin Noble, and Karlheinz Gross of EMBL, Heidelberg.

REFERENCES

1 Lee, B. and Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* 1971, **55,** 379–400
2 Richards, F.M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 1977, **6,** 151–176
3 Richards, F.M. Calculation of molecular volumes and areas for structures of known geometry. *Methods in Enzymology* 1985, **115,** 440–464
4 Connolly, M.L. PhD dissertation thesis, University of California at Berkeley, 1981.
5 Connolly, M.L. Molecular surface program. *QCPE Bull.* 1981, **1,** 75
6 Greer, J. and Bush, B. Macromolecular shape and surface maps by solvent exclusion. *Proc. Natl. Acad. Sci. USA* 1978, **75,** 303–307
7 Pearl, L.H. and Honegger, A. Generation of molecular surfaces for graphic display. *J. Mol. Graphics* 1983, **1,** 9–12
8 Lorensen, W. and Cline, H. Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics* 1987, **21,** 163–169
9 Colloc'h, N. and Mornon, J.P. Representation and analysis of protein surfaces using spline functions. *J. Mol. Graphics* 1988, **5,** 170
10 Moon, J.B. and Howe, W.J. A fast algorithm for generating smooth molecular dot surface representations. *J. Mol. Graphics* 1989, **7,** 109–112
11 Voorintholt, R., Kosters, M.T., Vegter, G., Vriend, G., and Hol, W.G.J. A very fast program for visualizing protein surfaces, channels and cavities. *J. Mol. Graphics* 1989, **7,** 243–245
12 Pascual-Ahuir, J.L. and Silla, E. GEPOL: An improved description of molecular surfaces. I. Building the spherical surface set. *J. Comp. Chem.* 1990, **11,** 1047–1060
13 Silla, E., Villar, F., Nilsson, O., Pascual-Ahuir, J.L., and Tapia, O. Molecular volumes and surfaces of biomacromolecules via GEPOL: A fast and efficient algorithm. *J. Mol. Graphics* 1990, **8,** 168–172
14 Silla, E., Tuñón, I., and Pascual-Ahuir, J.L. GEPOL: An improved description of molecular surfaces. II. Computing the molecular area and volume. *J. Comp. Chem.* 1991, **12,** 1077–1088
15 Agishtein, M.E. Fuzzy molecular surfaces. *J. Biomolec. Struct. Dynamics* 1992, **9,** 759–768
16 Connolly, M.L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* 1983, **16,** 548–558
17 Connolly, M.L. Computation of molecular volume. *J. Am. Chem. Soc.* 1985, **107,** 1118–1124
18 Connolly, M.L. Depth buffer algorithms for molecular modeling. *J. Mol. Graphics* 1985, **3,** 19–24
19 Connolly, M.L. Molecular surface triangulation. *J. Appl. Crystallogr.* 1985, **18,** 499–505
20 Connolly, M.L. Plotting protein surfaces. *J. Mol. Graphics* 1986, **4,** 93–96
21 Connolly, M.L. Measurement of protein surface shape by solid angles. *J. Mol. Graphics* 1986, **4,** 3–6
22 Connolly, M.L. Shape complementarity at the hemoglobin $\alpha_1\beta_1$ subunit interface. *Biopolymers* 1986, **25,** 1229–1247
23 Connolly, M.L. An application of algebraic topology to solid modeling in molecular biology. *Visual Computer* 1987, **3,** 72–81
24 Connolly, M.L. Volume Overlap Isotopy and Docking. *QCPE Bull.* 1992, **12,** 43–44
25 Connolly, M.L. Shape distributions of protein topography. *Biopolymers* 1992, **32,** 1215–1236
26 Perrot, G. and Maigret, B. New determinations and simplified representations of macromolecular surfaces. *J. Mol. Graphics* 1990, **8,** 141–144
27 Perrot, G., Cheng, B., Gibson, K.D., Vila, J., Palmer, K.A., Nayeem, A., Maigret, B., and Scheraga, H.A. MSEED: A program for the rapid analytical determination of accessible surface areas and their derivatives. *J. Comp. Chem.* 1992, **13,** 1–11
28 Sedgewick, R. *Algorithms* Addison-Wesley, Menlo Park, California, 1983
29 Connolly, M.L. Molecular interstitial skeleton. *Comp. Chem.* 1991, **15,** 37–45