Protein Science (1999), *8*:676–679. Cambridge University Press. Printed in the USA.

FOR THE RECORD

GRASS: A server for the graphical representation and analysis of structures

MURAD NAYAL, BENJAMIN C. HITZ, AND BARRY HONIG

Department of Biochemistry and Molecular Biophysics, Columbia University, 630 W. 168th Street, New York, New York 10032

(RECEIVED July 29, 1998; ACCEPTED November 17, 1998)

Abstract: GRASS (Graphical Representation and Analysis of Structures Server), a new web-based server, is described. GRASS exploits many of the features of the GRASP program and is designed to provide interactive molecular graphics and quantitative analysis tools with a simple interface over the World-Wide Web. Using GRASS, it is now possible to view many surface features of biological macromolecules on either standard workstations used in macromolecular analysis or personal computers. The result is a World-Wide Web-based, platform-independent, easily used tool for macromolecular visualization and structure analysis.

Keywords: chime; computational biology; continuum electrostatics; functional genomics; GRASP; molecular graphics; VRML; World-Wide Web.

The widespread use of the GRASP program (Nicholls et al., 1991) has demonstrated the utility of mapping various physical, chemical, and biological properties on the molecular surface. A limitation of the current version of GRASP is that it can only be run on Silicon Graphics (SGI) workstations. GRASS (Graphical Representation and Analysis of Structures Server) is an internet tool that uses GRASP as its internal "engine," and can be used on PCs and Macintoshes as well as Unix workstations, if the platform supports one of the three helper programs needed by the server for displaying molecular graphics. GRASS is designed to be intuitive, easy to use, and platform independent. Graphics processing and the calculation of various molecular properties are performed on the server and are then sent to the client for viewing. At present, GRASS can be accessed by pointing an internet browser to the URL: http://honiglab.cpmc.columbia.edu/cgi-bin/GRASS/surfserv_enter.cgi (Fig. 1) or to a link at the PDB web site at http://www.pdb.bnl.gov/pdb-bin/pdbmain.

Reprint requests to: Barry Honig, Department of Biochemistry and Molecular Biophysics, Columbia University, 630 W. 168th Street, New York, New York 10032; e-mail: honig@honiglab.cpmc.columbia.edu.

Abbreviations: HSSP, Homology-Derived Structures of Proteins; HTML, Hyper Text Markup Language; HTTP, Hyper Text Transfer Protocol; MIME, Multipart Internet Mail Extensions; PARSE, Parameters for Solvation Energy; PDB, Protein Data Bank; VRML, Virtual Reality Modeling Language; WWW, World-Wide Web.

GRASS is used in three steps: first, a macromolecular structure is selected for viewing and analysis. This is done by either uploading the user's own structure (in PDB format) to the server or searching a locally maintained database of macromolecular structures for entries that satisfy any number of search criteria. Two databases are available for searching: a local mirror of the Protein Data Bank (PDB) (Bernstein et al., 1977) updated weekly, and a database of PDB structures modified to enable full atomic charges assignments (see below) that is updated several times a year. Entries can be retrieved based on the chemical name of the macromolecule studied, the biological species where the macromolecule is found, the names of the scientists who determined the structure, the experimental method used for structure determination, the resolution at which the structure was determined, if applicable, or the date the structure was submitted to the PDB. The server will parse the selected structure, grouping the different molecules found in four different categories: peptide chains, nucleic acid chains, ions, and ligands (including waters) (Fig. 2). The user selects the molecules to be displayed by selecting one or more display styles for each. The available display styles are: CPK atoms, bonds, backbone tracing ribbons, and molecular surfaces (Richards, 1977). GRASS will also enable users to construct interaction surfaces defined as the portion of the molecular surface that is less than a certain interaction distance (specified by the user) from a bound ligand.

Second, the color-coding scheme for molecules is selected by choosing a molecular property to be calculated or fetched from a database. The available properties include atomic partial charges, electrostatic potential calculated using either formal charges or the PARSE charge set (Sitkoff et al., 1994), molecular surface curvature, hydrophobicity [inferred from the atom type, the residue type or the side-chain cyclohexane to water transfer free energy (Radzicka & Wolfenden, 1988)] amino acid sequence variability [obtained from the HSSP, Homology-Derived Structures of Proteins, database (Sander & Schneider, 1991)], distance to a ligand, and temperature factors. The molecule will be colored according to the molecular property chosen using a property value to RGB color correspondence.

Third, one of three programs is selected to display the graphics: A Virtual Reality Modeling Language Viewer (VRML) (Vollhardt et al., 1995; Bell et al., 1996), for example CosmoPlayer, available from Silicon Graphics Inc. (SGI) at http://cosmo.sgi.com/player/;

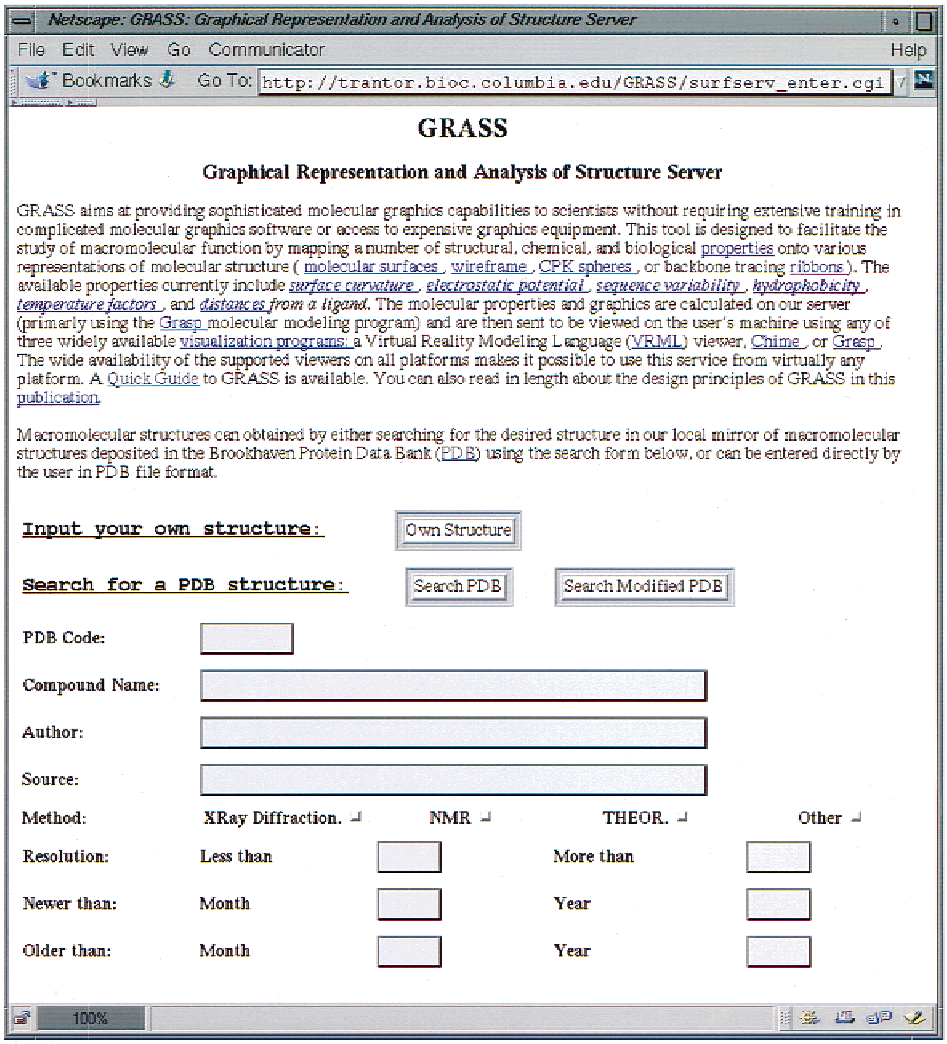

Fig. 1. The structure selection WWW interface of GRASS.

the Chime molecular graphics program (Chemscape Chime software, MDL information systems), available from Molecular Design Limited Inc. (MDL) at http://www.mdli.com/chemscape/chime/; or the GRASP (Nicholls et al., 1991) molecular modeling program available at http://honiglab.cpmc.columbia.edu/grasp/. All three programs allow the user to interactively manipulate the molecules to obtain the desired orientation. In addition, it is possible to request specific initial orientations for optimal view of bound ligands. The display can also be obtained in stereo mode when using GRASP as the local viewer program.

GRASS implements a client-server design concept where the needed molecular databases as well as the programs that carry out the computations are maintained on a central server computer. Currently, GRASS is mirrored at the PDB, but other sites are anticipated. (A license can also be obtained to install GRASS locally. Contact Katie Rosa at katie@honiglab.cpmc.columbia.edu.) GRASS takes full advantage of various internet technologies. The graphical user interface is implemented as interactive Hyper Text Markup Language HTML 3.2 pages (Raggett, 1997) and JavaScript 1.2 (Netscape Communication Corporation, 1997) that can be interpreted by most modern-day browsers. Displaying interactive molecular graphics on the client side is accomplished using widely available plugins and helper programs. Also, the client/server communication is implemented using well-established internet protocols.

The database of modified PDB structures: GRASS can use either raw data from the PDB as input, or a locally modified PDB in which hydrogen atom positions and coordinates of side-chain atoms that are not seen crystallographically have been added. Hydrogen and missing side-chain atoms were added using the WhatIf program (Vriend, 1990). Ambiguity in side-chain placement is arbitrarily resolved by selecting the first listed atomic position in the X-ray file. Also, a single structure is chosen from the ensemble of structures determined with NMR by selecting the structure of the ensemble that is closest to the average structure. If an unmodified PDB file is used, the calculated electrostatic potentials result only from ionizable groups in their charged state. The modified PDB is required if the effects of the partial charges on polar groups are to be included in the calculations.

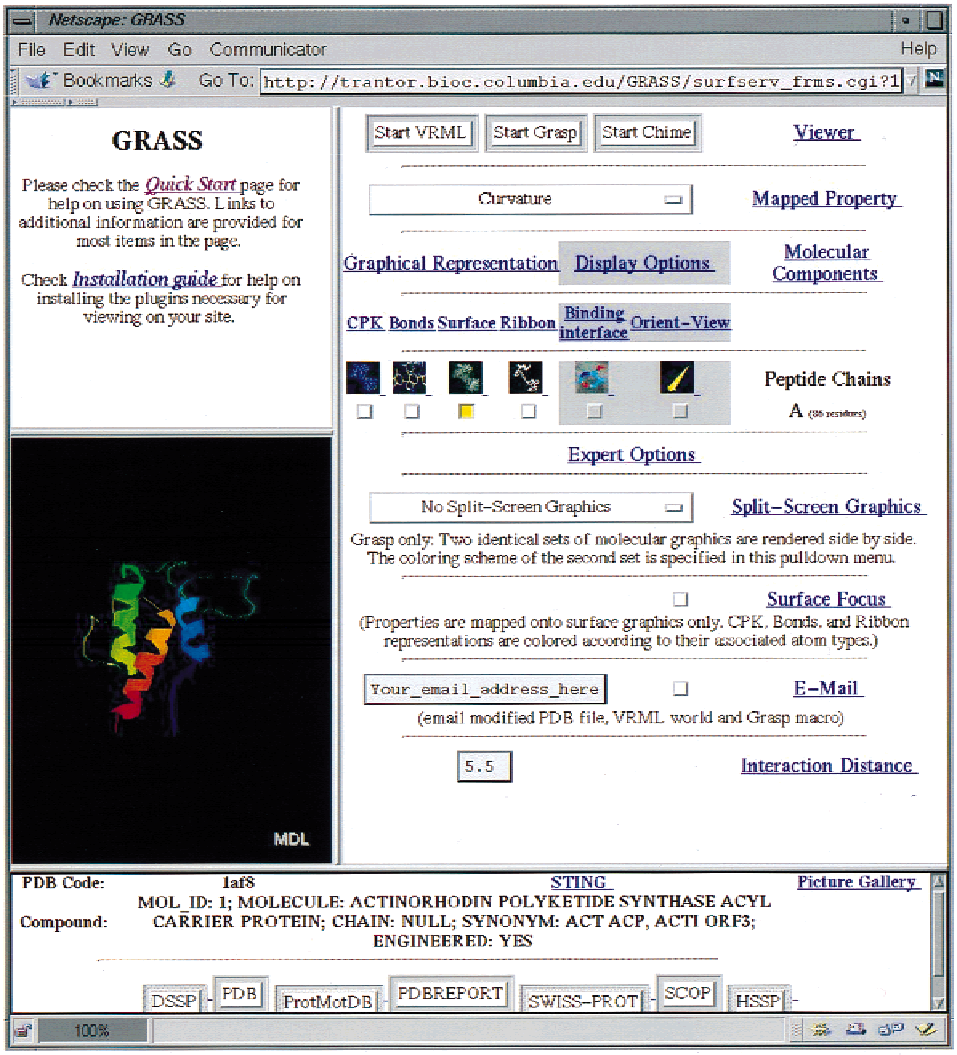

Fig. 2. The molecular graphics generator WWW interface of GRASS. The user begins by selecting a one or more graphical representation of the molecules of interest by checking the corresponding box on the page. The coloring scheme is selected next from a pull-down menu. Finally, the viewer to display the graphics is chosen by clicking the appropriate start button at the top of the page. Links to other databases available on the web are also provided.

Client–server design: GRASS is based on a client–server paradigm using three interacting software layers (Fig. 3). On the client side, the graphical user interface is provided by a Hyper Text Transfer Protocol HTTP (Fielding et al., 1997) client capable of interpreting Hyper Text Markup Language (HTML) 3.2 and JavaScript 1.2 (Netscape Communication Corporation). Most modern-day World-Wide Web (WWW) browsers provide this functionality. Interactive molecular graphics on the client side are rendered using one of three possible plug-ins/helper programs that are configured to interact with the WWW browser (VRML browser, Chime, and GRASP). The sequence of operations that take place when using GRASS is as follows: a web browser running on the user's computer provides the graphical user interface and allows the user to issue commands using a pointing device (mouse). On the server computer side, an HTTP server listens to commands transmitted over the internet from HTTP clients (WWW browsers). Commands issued by a GRASS user are routed via the HTTP server to an intermediate layer of programs using the CGI (Common Gateway Interface) protocol. Intermediate layer programs are responsible for interpreting the user's commands and invoking third layer programs that do the actual calculation of the necessary molecular properties and the construction of the graphics. The results of the calculations and the constructed molecular graphics are then sent back to the client to be displayed. The results, properly encoded, can also be sent to the user using e-mail.

Computations of molecular properties: The computational heart of GRASS is the molecular modeling program GRASP (Nicholls et al., 1991). A program, GraspScript, is written to generate GRASP macros based on the users' input. These macros are then used to cause GRASP on the server to calculate electrostatic potentials using either formal charges or the PARSE charge set (Sitkoff et al., 1994), molecular surface curvature, and distances to a ligand. The output of these calculations is placed in a PDB file with the calculated property replacing the field normally used for the temper-

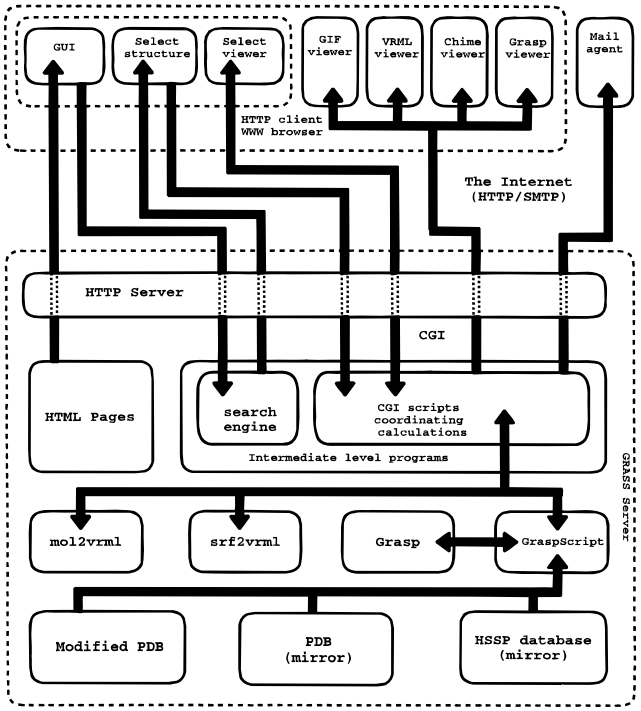

Fig. 3. A schematic demonstrating the logical organization of the GRASS. An HTTP server communicates with the user's HTTP client using standard Internet protocols. The client requests are relayed to an intermediate set of programs that coordinate molecular properties calculations and graphics generation performed by a third layer of programs. The requested information takes the same pathway in reverse order and is relayed back to the client.

ature factors. Other molecular properties that can be displayed include atomic charges; hydrophobicity, which is determined either based on the atom type, the residue type, or the experimentally measured cyclohexane to water transfer free energies (Radzicka & Wolfenden, 1988); temperature factors; and amino acid sequence variability within the protein family obtained from the HSSP database (Sander & Schneider, 1991). The resulting PDB file with the calculated molecular property values replacing the temperature factors is then used to generate the molecular display, and color coded according to the property of interest. This PDB file can also be sent encoded as an e-mail attachment to the user.

Construction of the graphical representations: GRASS allows the user to construct four graphical representations of molecular structures: CPK representation of the atoms, wireframe representation of the chemical bonds, ribbon representation of the backbone fold, and the molecular surface (Richards, 1977). The molecule is subsequently displayed on the client computer using one of the three helper programs. The method by which the graphics are constructed is different for each viewing program. For the programs that "understand" molecular structure (Chime and GRASP), the display is generated by providing the molecular structure file (the PDB file) and a program-specific macro (set of commands) that instructs the program to create the graphics and run the calculations requested by the user. The PDB file and the necessary macro are created by the server and packaged in a Multipart Internet Mail Extensions MIME (Borenstein & Freed, 1993) encoded internet message and sent to the client. GRASS can also create molecular graphics in Virtual Reality Modeling Language (VRML) format that can be displayed using any VRML viewer. VRML is a general-purpose three-dimensional scene description language that does not support special provisions for the description of chemical structures. For this reason the molecular structure has to be described in VRML using general graphics primitives such as spheres, rods, and polygons. GRASS creates VRML specified CPK, Wireframe, and Ribbons graphics while molecular surfaces are constructed using Grasp, and then converted to the VRML 2.0. The VRML image can also be sent to the user via e-mail for archiving and future examination.

Gallery of macromolecular surfaces: A database of static images of the molecular surface of macromolecules in the PDB has been constructed and is accessible from GRASS. The molecular surfaces are color coded with different molecular properties. No special plug-ins or helper applications are needed to use the database, as the images can be viewed using any WWW browser. This database is updated several times a year, and currently contains 64,362 images.

Acknowledgments: This work was supported in part by NSF Grant DB1-9601463. We thank Donald Petrey for providing us with the mol2vrml program. We also would like to thank Rick Fine and Boris Klebansky for useful discussions.

References

Bell G, Carey R, Marrin C. 1996. The Virtual Reality Modeling Language Specification, version 2.0 ISO/IEC CD 14772.

Bernstein FC, Koetzle TG, Williams GJB, Meyer EF Jr, Brice MD, Rogers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Databank: A computer based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Borenstein N, Freed N. 1993. MIME (Multipurpose Internet Mail Extensions). Part 1. Mechanism for specifying and describing the format of internet message bodies. RFC 1521.

Chemscape Chime software, MDL information systems. 14600 Catalina Street, San Leandro, California 94577. http://www.mdli.com/chemscape/chime.html.

Fielding R, Irvine UC, Gettys J, Mogul J, Frystyk H, Berners-Lee T. 1997. HyperText Transfer Protocol HTTP/1.1. RFC 2068. http://www.w3.org/Protocols/rfc2068/rfc2068.

Netscape Communication Corporation. 1997. JavaScript Guide. http://developer.netscape.com/library/documentation/communicator/jsguide4/contents.htm.

Nicholls A, Sharp KA, Honig B. 1991. Protein folding and association: Isights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct Funct Genet 11*:281–296.

Radzicka A, Wolfenden R. 1988. Comparing the polarities of the amino acids: Side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry 27*:1664–1670.

Raggett D. 1997. HTML 3.2 Reference Specification. W3C recommendation. http://www.w3.org/TR/REC-html32.

Richards F. 1977. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng 6*:151–176.

Sander C, Schneider R. 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins 9*:56–68.

Sitkoff D, Sharp KA, Honig B. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem 98*:1978–1988.

Vollhardt H, Henn C, Moeckel G, Teschner M, Brickmann J. 1995. Virtual reality modeling language in chemistry. *J Mol Graphics 13*:368–372.

Vriend G. 1990. WHAT IF: A molecular modeling and drug design program. *J Mol Graph 8*:52–56.