A graphics program for the analysis and display of molecular dynamics trajectories

Leif Laaksonen

Technical Research Center of Finland (VTT), Biotechnical Laboratory, Espoo, Finland.

The program SCARECROW has been developed to help the molecular modeler to analyze and display the very big and complex data files produced by molecular dynamics programs. The molecular graphics program SCARECROW is written to support the display, animation, and extensive analysis of molecular dynamics trajectories. Using the macro language it is easy to make scripts for video animation and for the automated display and analysis of time series. Extensive coloring and atom selection commands are included to help the user to focus on relevant regions of the molecule. Time series can be produced and viewed on the screen or transferred to other programs.

Keywords: computer graphics, molecular dynamics, conformational space analysis, animation

INTRODUCTION

Interactive computer graphics is a powerful tool for designing and visualizing molecules. Today, however, there is still a problem in using the wide range of computational programs available. There exist no agreed standard data structures for file transfer between programs. The problem is even more severe with the huge amount of data generated during molecular dynamics simulations on biomolecules. There exist today a large number of molecular graphics programs for the display and modification of molecular structures, such as the QUANTA[1] and INSIGHT[2] program packages. The problem with these programs is that dynamics trajectory file transfer between these programs is difficult or impossible. It is impossible to analyze trajectories produced by DISCOVER[3] with QUANTA, and INSIGHT refuses to analyze CHARMm[4,5] trajectories. These commercial packages come without the source code which makes it difficult usually to make one's own interfaces. These problems can be solved partially with the open interface approach. The purpose of the program SCARECROW is to provide the user with an open code, where improvements and modifications can be done easily. The program can analyze and display the trajectories from several dynamics programs. It can display two-dimensional (2D) history plots of variables, such as distances, bond angles, torsion angles, and energies variations, during a dynamics simulation. The time series also can be written to a disk and transferred to more elaborate statistical analyzing packages for the analysis and display. The user can make modifications to the time series and perform graphical correlation analysis between different time series. The molecule can be displayed as sticks, licorice or as CPK spheres. The user interface is both menu and keyboard driven. Pull-down menus are used to choose the desired option. Some of the graphical commands, however, can be accessed only from the menu. The program also can be run entirely from scripts.

Color Plates for this article are on page 24.

Address reprint requests to Dr. Laaksonen at the Technical Research Centre of Finland (VTT), Biotechnical Laboratory, P.O. Box 202, SF-02151 Espoo, Finland.

Received 8 May 1991; accepted 21 May 1991

SYSTEM OVERVIEW

SCARECROW is designed to run on all Silicon Graphics 4D or Personal Iris series machines. The graphics is implemented using GL function calls. Full 24-bit graphics is needed. SCARECROW also supports a nongraphics mode. In this mode it is possible to generate time series to be transferred to other display programs. The nongraphics part can be run on any UNIX machine.

IMPLEMENTATION

SCARECROW is written in C. Dynamic memory allocation is used in most cases. At startup time the user can define the maximum number of atoms the system can handle during that session. Systems up to 50 000 atoms have been successfully displayed and handled. Several structures can be superimposed on the screen. The user also can define a memory limit when the system starts to do the trajectory analysis out of core. In practice this means that there are no size (software) limitations for the trajectory file size.

It is possible to look at the frames as a movie or to step through the frames one at a time, by pressing the up- and down-arrow keys. The user can define by picking distances, angles, and tor-

sion angles, and calculate their time series and calculate the average values and standard deviations. The time series can be displayed as 2D plots on the screen and the history of the time series and correlation plots can be displayed during the display of the frames.

It is also possible to trace molecule movement from frame to frame and to display the translation as a continuous trajectory of the chosen atoms. This shows the diffusion of small molecules.

Atoms can be colored according to their charge, force in the force field, conformation energy, conformation temperature, standard atom coloring, or user selection.

COORDINATE DATA FILE FORMATS SUPPORTED

Currently SCARECROW can read the atom coordinate information from the following programs:

(1) *QUANTA/CHARMm* in the CHARMm[5] text coordinate format or the CHARMm free format
(2) *INSIGHT/DISCOVER*,[3,6] both the Cartesian coordinate files and the output coordinate files can be handled
(3) *Brookhaven Protein Data Bank* files written in the PDB[7] file format
(4) *MOPAC5*,[8] this requires the use of the GRAPH key word to produce the graphics output file
(5) *YASP*[9] (yet another simulation package) coordinate input format
(6) *GAMESS*,[10] this requires the use of the MOLPLT = .TRUE. switch in the GAMESS input deck to produce the output for a molecular structure drawing program.

Molecular dynamics trajectories can be displayed and analyzed from the following programs:

(1) CHARMm dynamics trajectory format
(2) DISCOVER history file format
(3) YASP dynamics trajectory file format.

The display–analyze functions for DISCOVER and YASP trajectories are always done out of core.

THE MOLECULAR DATA STRUCTURE

The molecule–atom information is organized as in the CHARMm program. All molecules have:

(1) A segment name of four characters
(2) A residue name of four characters
(3) A residue number; in fact CHARMm is using two but SCARECROW is only using the first one
(4) An atom name of four characters.

The keyboard commands applying to atoms are supplied as a command and three fields separated by a colon. For example, **set atom colo s1:*:ca yellow** sets all CA atoms in all residues in segment S1 to the color yellow. Instead of the name yellow the RGB code for yellow can be used. The general identifier looks like **segment:residue:atom**, where the segment name is always a character string. The wild character * and the character ? can be used. The character * matches all character strings and ? matches any character at that position. In the residue field and atom field residue numbers and atom numbers can be used, or a combination of characters and integers can be used. For example, **set atom colo s1: tyr,1,10,30-45:ca,n,c blue** sets all CA,N and C atoms in all tyrosine residues, and residues 1,10 and 30–45 in segment S1 to the color blue.

DIFFERENT DISPLAY STYLES

SCARECROW provides six different representations of molecular structures:

(1) Wire frame. The number of pixels defining a line can be changed.
(2) CPK space-filling images. The sphere radius can be defined to the van der Waals or covalent value, respectively.
(3) Licorice images. The cylinder and sphere radius can be changed independently to display ball-and-stick images.
(4) C_α linear distance plots of proteins for the display of protein secondary structure.
(5) Ramachandran plots of proteins for protein secondary structure analysis.
(6) Potential iso-energy surfaces.

It is also possible to combine the first three and the last display styles for molecule. All display styles can be used for the animation.

Hardcopies can be made from the screen dump to PostScript printers.

PROGRAM AVAILABILITY

The program SCARECROW is copyrighted and distributed under the following conditions: The SCARECROW program, including source code and documentation is available from the author to nonprofit organizations. Please use the mail address *laaksone@geeni.bio.vtt.fi* for requests.

ACKNOWLEDGEMENTS

The author gratefully acknowledges financial support from the Academy of Finland.

REFERENCES

1 Polygen Corporation, 200 Fifth Avenue, Waltham, MA 02154, USA
2 BIOSYM Technologies, Inc., 10065 Barnes Canyon Road, San Diego, CA 92121, USA
3 Discover User's Guide, Version 2.6.0, 1990, Biosym Technologies, 10065 Barnes Canyon Road, San Diego, CA 92121, USA
4 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. *J. Comp. Chem.* 1983, **4**, 187–217
5 CHARMm User's Guide, 1988, Polygen Corporation, 200 Fifth Avenue, MA 02254, USA
6 Insight II, Reference Guide, Version 1.1.0 and Reference Guide 2.0.0, 1991 Biosym Technologies, 10065 Barnes Canyon Road, San Diego, CA 92121, USA
7 Protein Data Bank Atomic Coordinate and Bibliographic Entry Format Description, 1985, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973, USA
8 Stewart, J.J.P. *Mopac 5 Manual* QCPE #455
9 Müller-Plathe, F. private communication
10 Schmidt, M.W., Boatz, J.A., Baldridge, K.K., Koseki, S., Gordon, M.S., Elbert, S.T., and Lam, B. *QCPE Bulletin* 1987, **7**, 115