##
#
MSKCC Document Delivery Services
#
Monday, November 28, 2005
#
##

Request ID: DDS36520
User: Gangi-Dino, Rita
Location: Brit Lib
Requested on: 11/21/2005
Needed by: 11/25/2005
Journal Title: Chemical Design Automation News
ISSN:
Article Author(s): Scheeren, M, Hooft, R, Vriend G
Article Title: Molecular modelling using WHATIF on personal computers
Year: 1994
Volume:
Issue:
Pages: 11-13
PMID:
User's Comments: In color, if available
Staff Notes: Bounced by OCLC 11/25, requested from British Library. Needs by 12/1.--KJ

Molecular Modeling Using WHAT IF on Personal Computers

By Milo Scheeren, Ph.D., Emma Scheeren-Groot, Ph.D., Rob W. W. Hooft, Ph.D., Chris Sander, Ph.D., and Gert Vriend, Ph.D.

The program WHAT IF is widely used in the fields of protein engineering, molecular modeling, protein structure determination and analysis. We describe a new version of this versatile software package that can be used on personal computers running under the MS-DOS operating system. This implementation brings professional molecular modeling to the personal computer.

Introduction

Molecular graphics and molecular modeling are rapidly becoming everyday tools in molecular biology laboratories. Molecular graphics packages — such as FRODO,[1] Insight,[2] Quanta,[3],Hydra,[3] BRAGI,[4] Midas,[5] O,[6] WHAT IF,[7] and Sybyl[8] — provide the scientist with many tools, such as:

- visualization of molecules, their surfaces, properties and interactions;
- automatic model building of an unknown structure based on the known structure of a homologous protein;
- prediction and analysis of the effects of point mutations;
- protein sequence analysis and sequence database searches.

These programs are available for workstations from different vendors, but if only a personal computer is available, one needs to resort to less powerful, often unsupported, graphics programs.

In an earlier publication,[9] we described a simple graphics system with very flexible capabilities although it is based on a limited set of graphics primitives. This system is centered on two, almost independent, tasks:

1. The WHAT IF program does all calculations, manipulations, and database searches.
2. A graphics driver displays dots, lines, rectangular boxes and text strings, and returns the position of the cursor on the screen.

We have used several variants of this two-task system to port the program WHAT IF to a diverse range of machines. There is only one version of WHAT IF, but each platform has its own graphics driver. These include Silicon Graphics workstations (using the GL graphics library), VAX + E&S PS300 series (E&S hierarchical data language), NeXT (GL emulator), and Intel 80386 and 80486 based MS-DOS personal computers (GL emulator). An X-windows based GL emulator has been implemented on the following architectures: PC under Linux, SUN, DEC-Alpha OSF/1, and IBM RS6000. The GL emulators were written in the context of this project.

The advantages of the availability of the same program on a multitude of hardware platforms are clear. One can keep working with the same program on different machines without the need to to get used to several programs, each with its own look and feel. The MS-DOS machines are an important addition as they provide a relatively inexpensive and widely available platform, and give the scientist access to ▶

INSIDE STORY

our software development. We have had our first College meetings for both Colleges, and our in-house developers are poised to deliver software later this year. College membership continues to grow, as companies realize that this is a cost-effective way to access software and keep abreast in this fast-moving field.

The Oxford Materials contract research facility has recently completed its first projects. Companies who want to "see before investing" in modeling, and those who wish to top up their own in-house efforts, or to explore a new area outside their expertise, see great advantages in this approach.

The opportunities for industry to use us to access innovative science are immense. Our destination will be reached when industry has realized fully these benefits; Oxford Materials invites you to join us on what should prove an exciting trip.

For an information pack and further details, contact Sophia Zurek, Oxford Materials Ltd., Chestnut Farm, Tarvin Road, Frodsham, Cheshire WA6 6XN, UK (Tel: 44-928-735679; FAX: 44-928-73335352). ■

WHAT IF

Continued from page 11
sophisticated protein structure software when a larger workstation is not available.

Description

The program WHAT IF provides a wide spectrum of capabilities for manipulating and studying biological macromolecules.[10-13] It is beyond the scope of this article to describe its many options. Work on the program has gone on for almost seven years, and will continue in the future. It comprises around 350,000 lines of code written in FORTRAN 77 and ANSI C. Table 1 gives a short summary of the major WHAT IF options, including recent additions. The full functionality of the latest version of the program has been ported to MS-DOS based personal computers.

Discussion

What are the technical problems in porting a major scientific package to MS-DOS? One problem is the size of the program. This created the necessity to work with a DOS-extender and 32-bit compilers, which still leaves several limitations. For example, debugging and core-dump features do not work for very large programs. Other problems are:

- an upper limit on the size of the modules that can be linked;
- several standard FORTRAN 77 features are not recognized,e.g., constructs where function values are used in implied DO-loops;
- the FORTRAN 77 compiler used puts restrictions on the order of declarations;
- a minimum of 16 MByte of main memory is required for linking the program.

MS-DOS based PCs are not optimized for paging. Therefore the executable should be as small as possible to maintain an acceptable performance, especially during program start-up. We minimized memory requirements by using variable equivalencing in FORTRAN wherever possible and by using external files in all options that do not need fast data access. As a result, 12 MBytes of main memory is sufficient for execution on MS-DOS based personal computers, although 16 MBytes provide better performance. The swap-file should be at least 30 MBytes.

The start-up of WHAT IF takes approximately a minute due to initialization of the swap file. However, once started, the MS-DOS based version provides a remarkable performance, comparable to entry-level ▶

Table 1. Examples of major menus provided by WHAT IF

Graphics
Most known three-dimensional macromolecular visualization techniques have been implemented. Two-dimensional graphics facilities such as ϕ-ψ plots (backbone dihedral angles), B-factor plots (crystallography), are available. Any graphical view can be plotted either in black and white or in color via the Postscript language.

Databases
Access is provided to all popular protein structure and sequence databases. An extended relational database for protein structure-function-sequence queries,[13] coupled to a structure fragment database,[17] is an integral part of WHAT IF.

Structure comparison
A range of superposition options is available, including a fully automatic 3-D structure alignment method.[10]

Structure verification
Complete reports about the quality of and errors in protein models can be produced automatically. This includes several geometric checks, as well as protein packing analysis,[12] verification of symmetry information,[18] and energy analysis.

Structure modification
Residues can be mutated, either interactively or automatically using a database-extraction algorithm,[19] complemented by removal of inter-atomic clashes. A fully automatic interface to GROMOS[20] (including a topology maker) allows for in-depth energetic analysis of mutant proteins. The WHAT IF movie options help with the analysis of gromos molecular dynamics trajectories.

Molecular interaction
Many tools are available to evaluate molecular contacts. Contacts can be highlighted in the molecular display. Two dimensional contact plots with "pickable" contact information can be generated.

Sequences
Sequence files can be read in several formats. Elementary sequence operations have been implemented, e.g., alignment, incremental multiple sequence alignment, and graphical evaluation.

Other
Many other calculations can be performed, e.g., surface accessibilities, hydrogen bonds, and atomic contacts. Input and output of coordinates is possible in PDB,[21] GROMOS,[20] and several other formats.

workstations. Using a 33MHz 80486 DX-based PC, an all-atom representation of a protein of 100 amino acids can easily be rotated in real time. The program minimally requires a 80386 DX CPU.

The program is available from the authors for a minimal fee. Extensive documentation is available upon request. 600 pages of documentation are available via anonymous FTP (swift.embl-heidelberg.de, directory /pub/whatif/writeup). External programs accessible from WHAT IF (e.g., GROMOS,[20] GRID,[14] RIBBONS,[15] DSSP[16]), are not distributed, but the interfaces to these programs are an integral part of the program.

We hope that bringing professional molecular modeling to the desk top will aid numerous biological scientists in the understanding of biological structure.

Acknowledgments

We are grateful to Michael Scharf, Daan van Aalten, Georg Tuparev, Anna Tramontano, Mike Carson, Paul Hooikaas, Rob Schilperoord, Vincent Eijsink, the computer group and the biocomputing program at the EMBL, the crystallography department at the University of Groningen and to the more than 200 WHAT IF users for their continuous help and support.

Milo Scheeren, Emma Scheeren-Groot, Leiden University, The Netherlands. Rob W. W. Hooft, Chris Sander, Gert Vriend, E.M.B.L., Heidelberg, Meyerhofstrasse 1, D-69117 Heidelberg, Germany.

References

1. Jones, T. A. A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* 1978, 11, 268-272
2. Dayringer, H.E., Tramontano, A., Sprang, S. R., and Fletterick, R. J. Interactive program for visualisation and modelling of proteins, nucleic acids and small molecules. *J. Mol. Graph.* 1986, 4, 82-87
3. Hubbard, R. E. In Fletterick, R. J. and Zoller, M., eds. *Computer graphics and molecular modeling,* Cold Spring Harbor, 1986, pp. 9-12
4. Schomburg, D. and Reichelt, J. Braggi: a comprehensive protein modeling program system. *J. Mol. Graph.* 1988, 6, 161-165
5. Ferrin, T.E., Huang, C. C., Jarvis, L. E., and Langridge, R. The MIDAS display system. *J. Mol. Graph.* 1988, 6, 13-27
6. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjelgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* 1991, A47, 110-119
7. Vriend, G. WHAT IF: a molecular modelling and drug design program. *J. Mol. Graph.* 1990, 8, 52-56
8. Sybyl is distributed by Tripos Inc.
9. Tuparev, G., Vriend, G., and Sander, C. GCI: a network server for interactive 3D graphics. *J. Mol. Graph.* 1992, 10, 12-16
10. Vriend, G. and Sander, C. Detection of common three-dimensional substructures in proteins. *Proteins* 1991, 11, 52-58
11. Vriend, G. Parameter relation rows: a query system for protein structure function relationships. *Prot. Eng.* 1990, 4, 221-223
12. Vriend, G. and Sander, C. Quality control of protein models: directional atomic contact analysis. *J. Appl. Cryst.* 1993, 26, 47-60
13. Vriend, G., Sander, C., and Stouten, P. F. W. A novel search method for protein sequence-structure relations using property profiles. *Prot. Eng.* 1994, 7, 23-29
14. Goodford, P. J. *J. Med. Chem.* 1985, 28, 849-857
15. Carson, M. Ribbon molecules of macromolecules. *J. Mol. Graph.* 1987, 5, 103-106
16. Kabsch, W. and Sander, C. How good are predictions of protein secondary structure? *FEBS Lett.* 1983, 155, 179-182
17. Jones, T. A. and Thirup, S. Using known substructures in protein model building and crystallography. *EMBO J.* 1986, 5, 819-822
18. Hooft, R. W. W., Sander, C., and Vriend, G. in preparation
19. de Filippis, E., Sander, C., and Vriend, G. in preparation
20. van Gunsteren, W. and Berendsen, H. J. C. GROMOS, Groningen molecular simulation computer package. University of Groningen, The Netherlands, 1987
21. Bernstein, F. C., Koetzle, T. F., Williams, G.J.B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* 1977, 112, 535-542 ■