Computer Corner

Protein Explorer: easy yet powerful macromolecular visualization

Eric Martz

Protein Explorer (PE, www.proteinexplorer.org) enables students, educators and other nonspecialists to visualize macromolecular structures easily. It also offers several advanced capabilities useful to protein structure specialists. Great attention has been given to making PE easy to use. Explanations, color keys and troubleshooting information are displayed automatically. There are also 'Frequently Asked Questions', a one-hour 'Quick-Tour', an alphabetical 'Help/Index/Glossary', and a detailed 'Tutorial'; all making PE much easier to use than either Chime or RasMol. Moreover, it is much more powerful; in addition to basic macromolecular visualization capabilities common to most similar programs, it offers one-click visualization of interfaces between moieties ('contacts'), cation–π interactions and salt bridges, as well as easy-to-use routines to visualize regions of conservation in three-dimensional protein structures based on multiple sequence alignments.

Protein Explorer (PE) is built upon Chime, a molecular graphics browser plugin that is freeware from MDL Information Systems (www.mdlchime.com). It was possible to implement PE within a few years only because of the power inherent in Chime. Chime, in turn, is in part built upon the molecular graphics rendering and command language in RasMol [1,2]. However, Chime has several additional significant capabilities, such as the ability to render solvent-accessible molecular surfaces and animations. The problem is that to get much out of either Chime or RasMol, the user must learn a complicated and extensive command language. This requirement makes the power of Chime and RasMol inaccessible to most of those who could benefit from macromolecular visualization. PE addresses this problem by enabling both basic and complex visualizations from menus, buttons and forms, without requiring the user to learn a single RasMol-style command. Nevertheless, PE accepts RasMol commands as a convenience for those who have learned them (see command input slot in Fig. 1). A detailed ease-of-use comparison of PE with RasMol is available at http://molvis.sdsc.edu/protexpl/pe_v_ras.htm

PE is free, and operates on ordinary Windows or Macintosh computers (also on linux or SGI/Irix in a Windows subsystem, http://molvis.sdsc.edu/protexpl/platform.htm). It is offered as a structure viewer at the Protein Data Bank (http://www.pdb.org). PE can be used on-line, or downloaded for off-line use. Web links can pre-specify the molecules to be displayed, supporting course or textbook websites. PE runs in the Netscape browser (Windows or Macintosh), or in Microsoft Internet Explorer (Windows only).

This article is not meant to provide instructions for the use of PE because the program has extensive built-in instructions. Rather, it is intended to help readers decide whether PE will be useful in their work.

Overview of Protein Explorer

FirstView

The first image of any molecule shown by PE is designed to be maximally informative. It is accompanied by a generic description, with links to illustrated explanations of backbone traces, disulfide bonds, 'hetero atoms', a standard color scheme for identifying elements, the absence or presence of hydrogen atoms, and water in protein crystals. (Typically only 10–20% of the water is tightly enough bound to be resolved and displayed. This kind of information is available through links on the *FirstView* page.) Clicking on any atom reports its element, and the name and sequence number of the residue to which it belongs (in a more explicit report than the one-line report of RasMol). Cn3D [3], the visualization program offered by the US National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) had previously implemented several features similar to those in PE, including an informative first view, animation of nuclear magnetic resonance (NMR) ensembles, and coloring from a sequence alignment.

QuickViews menu system

The menu system of PE (completely separate from the built-in menu of Chime) follows the paradigm of RasMol: atoms are first selected, then displayed (or hidden), then colored (see upper left in Fig. 1). The 'Select' menu of PE offers >25 pre-defined categories of atoms for 'one click' selection, including individual chains and ligands. Atoms, residues and chains can be selected or deselected by clicking on them in the rotatable 3D image. This makes it easy to select, for example, one molecule of ligand from several that are present. Sequences of protein or nucleotide chains are displayed in the 'Seq3D' window. Clicking on the sequence highlights individual residues or ranges in the 3D image, leaving them selected for optional 'Display' or 'Color' menu operations.

The 'Display' menu offers backbone traces, smoothed backbone traces, secondary structure cartoons, 'vines' (backbone traces with sidechains), stick, ball and stick, spacefilling to van der Waals radii, dot surfaces, solid or transparent rolling probe solvent-accessible surfaces, and 'contact surfaces'. It can also hide the selected atoms, or hide everything except the selected atoms. Boolean menu options allow consecutive 'Select' or 'Display' menu operations to be cumulated or subtracted. In addition to standard renderings, the 'Display' menu offers distance-based routines that highlight salt bridges, cation–π interactions, or atoms noncovalently bonded to a selected moiety (contact surfaces). Each of these is a one-click operation, automatically triggering the display of detailed help.

The 'Contact Surface' display (Fig. 1) gives an overview of noncovalent bonding.

 0968-0004/02/$ – see front matter PII: S0968-0004(01)02008-4

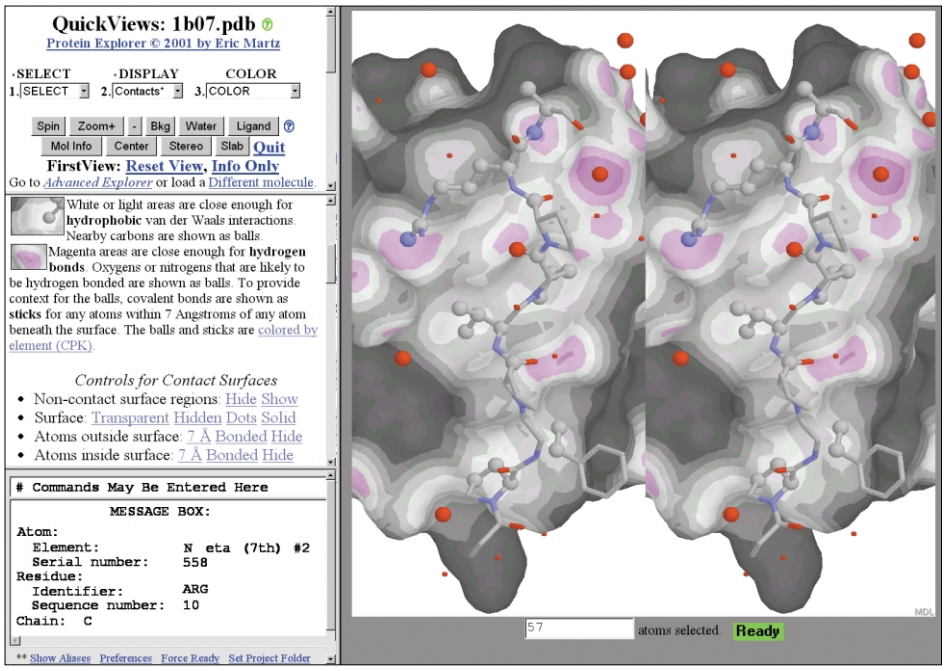

Fig. 1. Overview of noncovalent bonding interactions visualized as a 'Contact Surface' in Protein Explorer (stereo pair from PDB ID 1B07). A peptide ligand (balls and sticks) is bound noncovalently to a protein domain (surface). Gray balls (carbon atoms) over white pockets represent likely hydrophobic interactions, whereas red or blue balls (oxygen or nitrogen, respectively) over magenta patches represent likely hydrogen-bonding interactions (or salt bridges). Red balls not connected to sticks are water oxygens. Clicking on an atom reports its identity, as shown in the lower left frame for an arginine sidechain nitrogen. Using the links under 'Controls for Contact Surfaces' (left middle frame), the surface can be rendered transparent, atoms beneath the surface shown, and distances between atoms are easily displayed (not shown). Thus, bonded pairs of atoms can be identified. This image was obtained in two clicks: select chain A, display contacts. The large object in the background is a protein Src-homology 3 (SH3) domain rendered as a solvent-accessible surface. Balls and sticks represent the portions of a bound peptide ligand inhibitor (chain C) that are within 7 Å of the SH3 domain (more distant portions are hidden). The surface of the SH3 domain is colored to indicate distances from the nearest atoms. The darkest areas are >4.5 Å from the nearest atoms, so unlikely to be bonded. Light gray and white areas are distances suitable for van der Waals interactions (~3.5–4.5 Å), whereas magenta areas are close enough for hydrogen bonds (2.5–3.5 Å). Ligand atoms shown as balls were found at likely noncovalent bonding distances from suitable atoms beneath the surface, whereas sticks are generally too far away. Note the pull-down menus and convenience buttons (upper left frame). (The command entry slot, visible in the lower left frame, is provided for the convenience of those who have already learned RasMol commands, but no commands need be learned to use Protein Explorer effectively.) Hands-on generation of a contact surface can be tried in the QuickTour at http://proteinexplorer.org

It shows the surface of any moiety selected by the user, colored by distance to the closest atoms. Nearby atoms are shown whereas more distant atoms are hidden. This view makes it easy to pick out hydrophobic interactions and hydrogen bonds. It is achieved in two simple steps. First, you select, for example, an atom, ligand, pocket, chain, secondary structure element or sequence range; then one more click displays all contacts to the selected moiety. The 'Noncovalent bond finder', a separate tool released in 1998, is available within PE for a more detailed tour of noncovalent bonding interactions. It is possible to step out in 0.1 Å shells around the selected moiety, displaying the closest atoms or residues.

The 'Color' menu offers standard color schemes to highlight α helices and β strands in secondary structure, N- versus C-termini, temperature (revealing regions of higher disorder in the protein crystal), element (CPK), and schemes that distinguish apolar, polar and charged amino acids. A color key is automatically displayed whenever a color scheme is applied [color schemes follow the proposed DRuMS standards (http://www.umass.edu/molvis/drums)].

A 'Molecule Information Window' offers sequences, the PDB file header, methods for visualizing specific oligomers [4] or fewer or single chains, information on crystal contacts and other information specific to the current molecule.

A cluster of 'convenience buttons' is always available. There are buttons to toggle continuous slow spinning, the display of ligands or water, the color of the background (black or white), and stereo and slab modes. Zoom buttons enlarge or reduce the image 25% per click. A centering button centers the currently selected atoms, or any atom clicked with the mouse; and this center is preserved during rotation and zooming.

Coloring a 3D protein by conservation and mutation

The 'MSA3D' routine of PE accepts a multiple protein sequence alignment, and automatically colors a 3D protein image to reveal regions of conservation or mutation (Fig. 2). Such alignments are readily obtained from websites outside of PE. A guide to using MSA3D is provided, along with ready-made built-in demonstration alignments.

NMR and animations

PE facilitates exploration of multiple-model ensembles that result from NMR studies, and can animate 'morphs' of conformational changes using multiple-model PDB files generated by interpolation. Such 'movies' can be rotated for viewing from any perspective, and can be displayed in a variety of renderings and color schemes. Examples are built into PE.

Comparisons with other software, limitations

PE provides visualization but not modeling. Modeling (changing of protein conformation, mutation of residues, homology modeling) can be done in other software, then saved as PDB files for visualization in PE. An excellent free package for such modeling is DeepView [5] (also known as SwissPDBViewer, http://www.expasy.ch/spdbv; see also related resources indexed at http://molvisindex.org). Chime cannot move two molecules relative to each other ('docking'), but two molecules can be explored side-by-side in 'Protein Comparator' – a two-molecule mode within PE. Two or more molecules can be aligned structurally, and the alignment viewed as a multiple-model file in PE.

Because it is optimized for interactive rotation, PE (Chime) does not produce high resolution 'publication quality' images. In designing the rendering code used in PE, Chime and RasMol, Roger Sayle [1] made an excellent compromise between image quality and speed of rendering, allowing even large proteins to be rotated in real time. For its intended use, his code is unsurpassed today, eight years later.

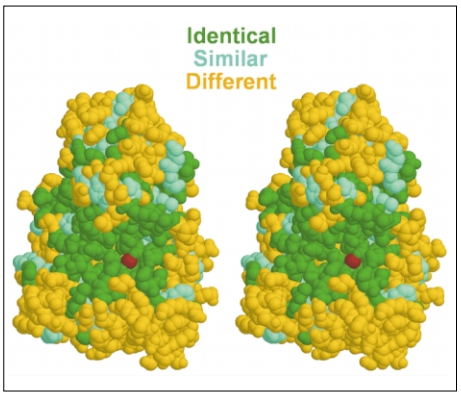

Fig. 2. Conservation of the catalytic pocket of enolase (stereo pair, PDB ID 4ENL). The catalytic site is marked by the dark atom (zinc). The multiple protein sequence alignment included sequences from human, *Drosophila*, yeast, archebacteria and eubacteria (kindly provided by Garry Duncan, Nebraska Wesleyan College). It was pasted into a form in the MSA3D section of Protein Explorer. One click of a button then generated a colored alignment listing (not shown) and the above image. The color scheme, required consensus level, and the definition of 'similar' are customizable. Many checks are built-in to detect and avoid errors. The MSA3D section includes a detailed tutorial with instructions for constructing multiple sequence alignments in Biology Workbench (http://workbench.sdsc.edu). Snapshots of results and built-in examples are viewable at http://proteinexplorer.org

Advanced explorer

So far, I have surveyed features designed for students, educators and occasional users. Here, I shall touch upon some features intended for frequent users and protein structure specialists. 'Preferences' can be customized, and personal settings survive between sessions. Setting the 'expert' preference skips the FirstView screen, hides water in the initial view, shows all models in multiple-model ensembles, and bypasses some of the help intended for novices.

Command aliases

Although one need not learn any commands to use PE effectively, nearly all RasMol commands can be entered in PE, and Chime has additional commands such as those for controlling animations and displaying surfaces. Command aliases are abbreviations for commands or command sequences. For example, 'ss' displays colored and rendered disulfide bonds, and 'bs' displays the selected atoms as balls and sticks. PE comes with >100 aliases, and these are easily customized. The 10 most recently entered commands can be recalled for re-entering or editing. PE displays the commands generated automatically by its menus and buttons; for those interested, observing these, and trying them in the command entry slot, is a good way to learn the command language.

Project folder

A 'project folder' can be specified on the local hard disk, into which files of command scripts and PDB files can be placed. The 'script' and 'load' commands default to the project folder without specifying a path. One can have as many project folders as desired, specifying the appropriate one within PE as needed.

Future plans

Support is being developed for delivering molecular structure 'presentations' in PE. This will enable pre-selected images and molecules to be displayed from labeled buttons in the traditional manner. (Many examples of traditional presentations employing Chime can be found at http://molvisindex.org) However, any image in the presentation can be explored with the full power of PE, after which one can return to the presentation (see demonstrations at http://www.umass.edu/microbio/chime/pipe). Present mechanisms of saving scripts for use in presentations are unsatisfactory. To solve this, a 'script recorder' is under development in collaboration with Tim Driscoll (http://www.molvisions.com). It will record the commands generated by menu and button operations, saving them to a script file for use in a presentation. Research results concerning protein structure can be communicated advantageously by presentation in PE; collaborations of this nature are invited.

Acknowledgements

Thanks to the Division of Undergraduate Education of the US National Science Foundation for grant support, to Roger Sayle for RasMol, to Tim Maffett and others at MDL for Chime, to Paul Stothard for contributing parts of the MSA3D code, and to Joel Sussman and Philip Bourne for encouragement.

References

1 Sayle, R.A. and Milner-White, E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* 20, 374–376
2 Bernstein, H.J. (2000) Recent changes to RasMol, recombining the variants. *Trends Biochem. Sci.* 25, 453–455
3 Wang, Y. *et al.* (2000) Cn3D: sequence and structure views for Entrez. *Trends Biochem. Sci.* 25, 300–302
4 Henrick, K. and Thornton, J.M. (1998) PQS: a protein quaternary structure file server. *Trends Biochem. Sci.* 23, 358–361
5 Guex, N. *et al.* (1999) Protein modelling for all. *Trends Biochem. Sci.* 24, 364–367

Eric Martz
Dept of Microbiology, University of Massachusetts, Amherst MA 01003, USA.
e-mail: emartz@microbio.umass.edu

Protein sequence motifs

Protein Sequence Motif is a regular column for brief reports of new motifs or sequence homologies that have been recognized in published sequences. Contributions to this column should be short (less than 500 words plus one figure) and will be subject to peer review. Preference will be given to reports of motifs or sequence homologies with profound biological significance. All sequences should have been published elsewhere in full and/or be freely available in the appropriate databases (e.g. GenBank, SWISS-PROT, etc.), but the particular motif or sequence alignment noted should not have been described before. Adequate statistical evaluation and, where appropriate, structural correlations should be given.