

# **A Platform-Independent Graphical User Interface for SEQSEE and XALIGN**

David S. Wishart<sup>1</sup>, Scott Fortin<sup>2</sup>, David R. Woloschuk<sup>2</sup>, Warren Wong<sup>2</sup>, Timothy Rosborough<sup>2</sup>, Gary Van Domselaar<sup>1</sup>, Jonathan Schaeffer<sup>2</sup> and Duane Szafron<sup>2</sup>

Protein Engineering Network of Centres of Excellence, Faculty of Pharmacy and

Department of Computing Science

University of Alberta, Edmonton, AB, Canada T6G 2N8

SEQSEE (Wishart et al., 1994a) and XALIGN (Wishart et al., 1994b) are two text-based, menu-driven programs developed specifically for comprehensive protein sequence analysis. Originally compiled to run on SUN and SGI workstations only, SEQSEE and XALIGN have been distributed to more than 300 laboratories around the world. Both programs have been used in a variety of applications ranging from routine sequence analysis to the identification of previously unknown sequence relationships (Upton et al., 1992; 1993; Dulhanty & Riordan, 1994). Since releasing these programs, we have received numerous requests asking if they could be ported to additional computer platforms (Macintosh's and PC's) or if the text-based menus could be replaced with a more friendly graphical user interface (GUI).

In response to these and other requests we have integrated XALIGN and SEQSEE into a single-threaded package with a uniform graphical user interface that is fully supported by SGI (Irix Ver. 5.0 and higher), SUN (Solaris and SunOS 4.1.3 and higher), Macintosh (Power PC OS 7.5 and higher) and PC (Windows 95) platforms. We have chosen the Smalltalk programming language to develop our graphical interface because it allows the creation of sophisticated GUI's that look and operate almost identically across all major platforms and operating systems. In many respects, Smalltalk, which was originally developed by Xerox's PARC in the late 1970's, is a more sophisticated version of the more familiar WWW language called JAVA. In particular, Smalltalk allows the facile creation of object-oriented, platform-independent graphical user interfaces. By designing the Smalltalk GUI to access the computationally intensive back-end routines through Smalltalk function calls, we were able to preserve a substantial portion of the original SEQSEE and XALIGN code (written in C). This separation between the front-end (the GUI) and the back-end has allowed for a more rapid implementation of the front-end while preserving the integrity of well-tested back-end programs.

In building the SEQSEE/XALIGN interface, a total of 11 separate windows or views were constructed, including (1) a sequence editor; (2) an alignment editor; (3) a simple text editor; (4) a graph viewer/editor; (5) a dotplot viewer/editor; (6) a helical wheel viewer/editor; (7) a structure

viewer/editor; (8) a sequence motif viewer; (9) a sequence statistics viewer; (10) a data browser; and (11) a file chooser.

The sequence editor supports auto-spacing (every 10 residues), auto-wrapping and mouse-driven text selection along with the usual cutting, pasting, copying and segment deletion operations. It has a text entry filter (permitting only IUPAC standard one-letter amino acid entry), a sequence "ruler", a sequence length and position monitor and an editable cursor position box. The sequence editor also supports a pop-up sequence reference card and a mouse-driven, color-coded secondary structure "painter". A screen shot of the sequence editor is shown in Fig. 1a.

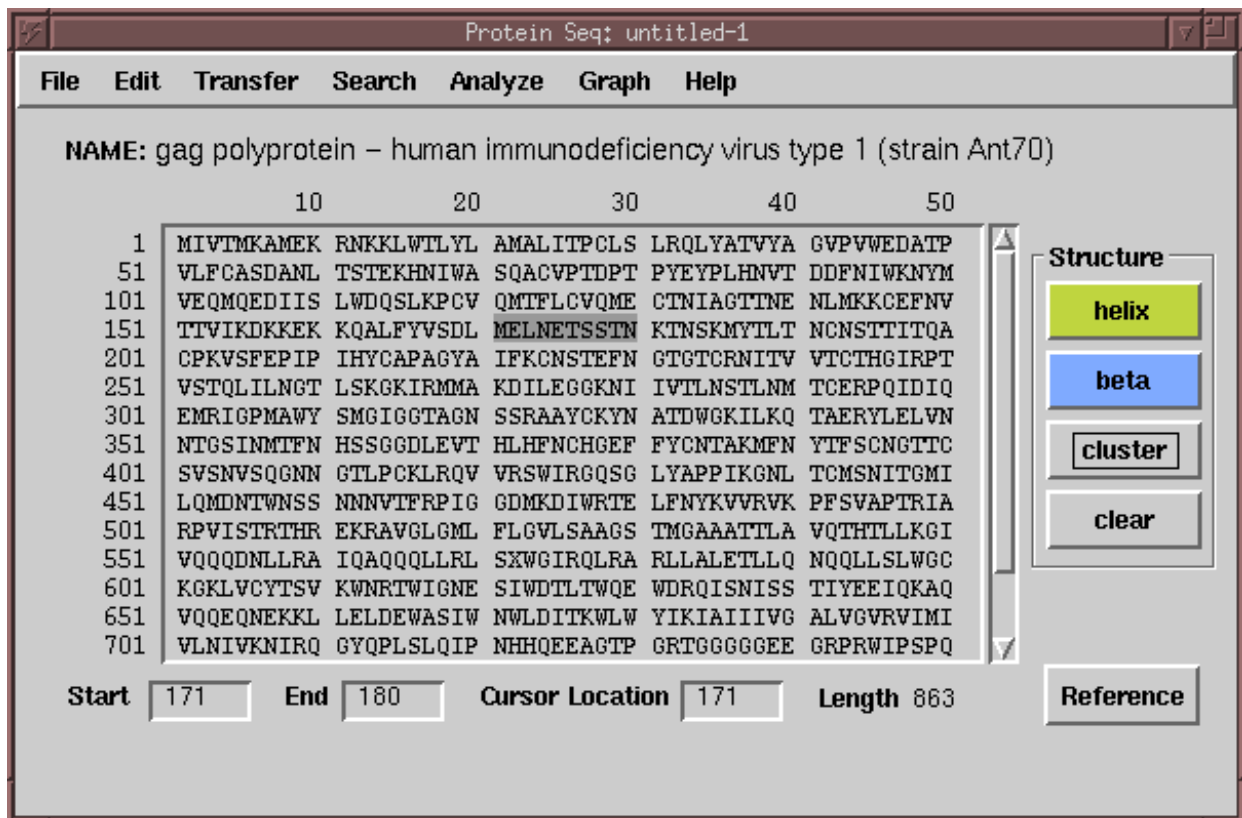


Figure 1a: The Sequence Editor (Unix)

The alignment editor integrates the functionality of XALIGN with a general multiple sequence editor. The alignment editor includes a mouse-activated sequence browser for quick data selection and loading. It also supports sequence fragment editing, auto-alignment, auto-consensus calculation, color-coded secondary structure display and an editable consensus threshold box. Also included is a mouse-driven "painter" for multi-sequence selection. One or more partial or complete sequences can be selected with this highlighter and interactively moved rightward or leftward via mouse-activated

arrow keys, thereby permitting manual multi-sequence alignment. A screen shot of the alignment editor is shown in Fig. 1c.

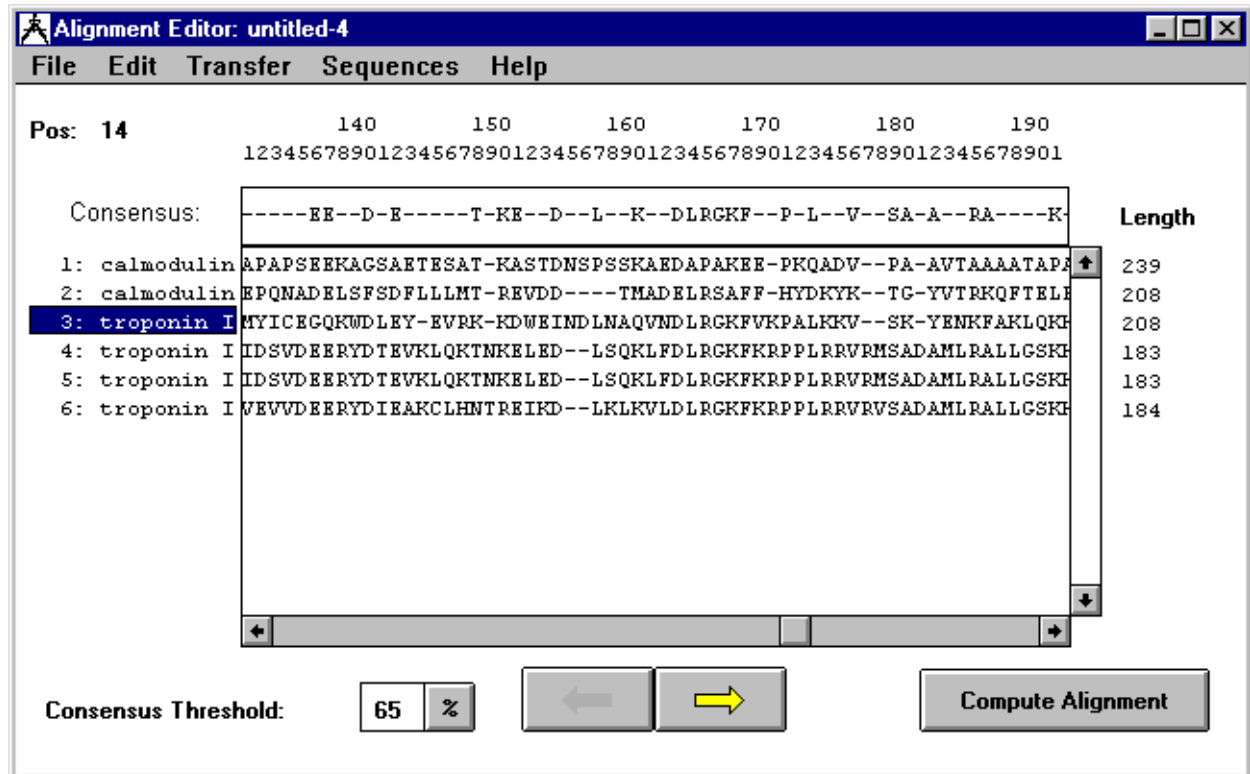


Figure 1c: The Alignment Editor (Windows 95)

The graph, dotplot and helical wheel viewers/editors share many similarities. All three support fully scrollable displays, step-wise and regio-selective (or framed) zooming and auto-scaling. They also offer full annotating and editing options including variable color and adjustable linewidth arrows, boxes, lines and circles. Additionally, the graph viewer supports graph axis, title and line editing, the helical wheel viewer supports user-selectable wheel and text coloring schemes, and the dotplot viewer offers simultaneous diagonal plot and text viewing. A screen shot of the dotplot viewer is shown in Fig. 1b.

The structure viewer displays predicted secondary structure by using standard, colored, 3-D "helix" and "beta-sheet" icons. It also permits selective toggling and re-ordering of up to six different kinds of structure predictions. Membrane spanning helices are displayed separately. The sequence motif viewer supports single or multiple motif selection and viewing including PROSITE patterns, B and T-cell epitopes as well as phosphorylation sites. It also permits toggling between a text view and a colored graphical view of selected motifs. The sequence statistics viewer offers a fully scrollable

multi-view window containing graphs (pI calculations, amino acid content, etc.) along with a scrollable list of dozens of statistical or proteometric calculations.

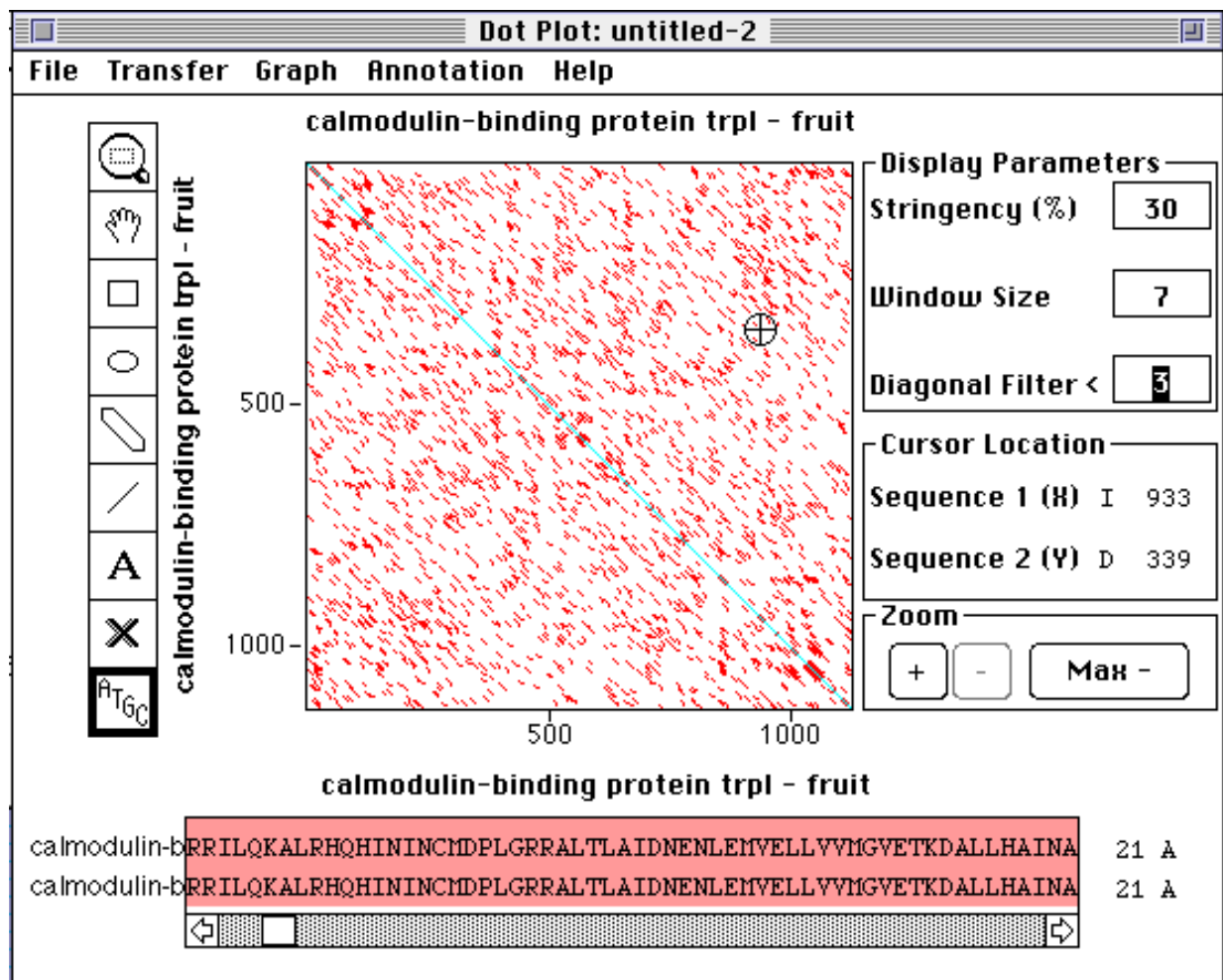


Figure 1b: Dot Plot Viewer (Macintosh)

This integrated package includes both a data-browser and a file chooser as file management tools to select and display files or sequences from database searches. In addition, an "option notebook" is provided, which allows the user to easily select or change various parameters (window sizes, scoring matrices, smoothing functions, etc.) specific to each program or function. An HTML-based help system with a hyper-linked introduction, glossary, index and tutorial is also supplied. Context-sensitivity is supported through image mapping to all major screens and views. The complete program, excluding the sequence databases, requires at least 8 Mbytes of hard disk space and 16 Mbytes of RAM. It is available on request from the authors.

## Acknowledgements

The authors wish to thank Brian Sykes, Chris Upton, Robert Boyko, Leigh Willard, Charles Cruden, Alex Nip and Eowyn Cenek for their hard work in testing, rewriting and upgrading SEQSEE and XALIGN. Financial support by the Alberta Heritage Foundation for Medical Research, The Protein Engineering Network of Centres of Excellence, NSERC and the Industrial Research Assistantship Program (NRC) is gratefully acknowledged.

## References

Upton, C., Stuart, D.T. and McFadden, G. (1993) Identification of a poxvirus gene encoding a uracil DNA glycosylase. *Proc. Natl. Acad. Sci.* **90**, 4518-4522.

Upton, C., Mossman, K., McFadden, G. (1992) Encoding of a homolog of the IFN- $\gamma$  receptor by myxoma virus. *Science*, **258**, 1369-1373.

Dulhanty, A.M. and Riordan, J.R. (1994) A two-domain model for the R domain of the cystic fibrosis transmembrane conductance regulator based on sequence similarities. *FEBS Lett.* **343**, 109-114.

Wishart, D.S., Boyko, R.F. and Sykes, B.D. (1994) Constrained multiple sequence alignment using XALIGN. *Comput. Applic. Biosci.* **10**, 687-688.

Wishart, D.S., Boyko, R.F., Willard, L., Richards, F.M. and Sykes, B.D. (1994) SEQSEE: A Comprehensive Program Suite for Protein Sequence Analysis. *Comp. Applic. Biosci.* **10**, 121-132.