short communications

ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments

Charles Simon Bond[a]* and Alexander Wolfgang Schüttelkopf[b]

[a]School of Biomedical, Biomolecular and Chemical Sciences, University of Western Australia, M310, 35 Stirling Highway, Crawley, WA 6009, Australia, and [b]Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland

Correspondence e-mail:
charles.bond@uwa.edu.au

Marked-up sequence alignments typically provide the central figure in articles describing proteins, whether in the fields of biochemistry, bioinformatics or structural biology. The generation of these figures is often unwieldy: interactive programs are often aesthetically limited and the use of batch programs requires the repetitive iterative editing of scripts. ALINE is a portable interactive graphical sequence-alignment editor implemented in Perl/Tk which produces publication-quality sequence-alignment figures where 'what you see is what you get'. ALINE is freely available for download from http://crystal.bcs.uwa.edu.au/px/charlie/software/aline/.

1. Introduction

A large proportion of papers describing the structure, function or evolution of proteins use sequence-alignment figures to visualize key properties of the relevant protein or protein family. Despite the large number of programs devoted to dealing with these alignments, it is often an awkward task to prepare the appropriate figure in a form suitable for publication. Once an alignment has been created, the existing software for manipulating it largely falls into one of two camps: programs which can edit alignments [e.g. SeaView (Galtier et al., 1996), Jalview (Clamp et al., 2004) and Indonesia (http://xray.bmc.uu.se/~dennis/)] and programs which can be used to mark up these alignments to produce high-quality aesthetically pleasing output [ALSCRIPT (Barton, 1993) and ESPript (Gouet et al., 1999)]. The latter programs typically run in batch mode from command files, which is useful for bulk production of figures in a high-throughput context but can be frustrating when trying to produce a custom figure to illustrate a specific scientific point.

ALINE was designed as a single program that is capable of importing sequence-alignment files, allowing editing and mark up of attributes such as sequence similarity, numbering and secondary structure, with the principal goal of producing a publication-quality figure. A flexible and robust system for adding plugins has allowed us to extend the functionality further to include interfaces to external programs and remote servers to create alignments and add additional mark-up interactively.

2. Implementation and requirements

ALINE is implemented in Perl and uses the Tk toolkit for the graphical user interface. Tk was chosen as it can output an encapsulated PostScript file of the canvas on which objects are drawn, resulting in a truly WYSIWYG figure. This approach proved successful for the Tcl/Tk program TOPDRAW (Bond, 2003), which produces publication-quality protein topology cartoons. ALINE is not designed to operate on huge alignments which may result in slow performance, but rather at the scale typically required for publication.

On most up-to-date Linux and MacOSX systems, *ALINE* should work effectively out of the box. The minimum requirement is a recent Perl/Tk distribution (Perl v.5.8.5 and Tk.pm). Under Microsoft Windows, or if problems arise from Perl or its modules, the user should install the ActivePerl distribution (http://www.activestate.com/ Products/ActivePerl/, v.5.8), which is freely available for most operating systems (including Linux, Solaris, MacOSX and Windows). Extended functionality requires ancillary programs including *Ghostscript* (writing bitmap output; http://www.ghostscript.com), *DSSP* (secondary-structure calculation; http://swift.cmbi.ru.nl/gv/

dssp), *WGET* (retrieving files from the internet; http://www.gnu.org/ software/wget), *ClustalW* and/or *Muscle* and/or *MAFFT* (for aligning sequences within *ALINE*) and *CHAINSAW* (for generating molecular-replacement models; http://www.ccp4.ac.uk).

Up-to-date installation instructions can be found at http:// crystal.bcs.uwa.edu.au/px/charlie/software/aline/.

A flexible plugin system was developed to allow advanced users to add functionality to *ALINE*. A number of examples which enable the functionality described below can be viewed in the plugins subfolder of the distribution.

Figure 1
Main figure: a screenshot of *ALINE* during the preparation of a figure (Ariza *et al.*, 2005). Inset: the same sequence read from a PDB file and coloured by temperature factor. Numerical information is interchangeably displayable as a bar graph using the context-dependent menu.

3. Features

When *ALINE* is started, a window opens containing a menu bar at the top, a control panel at the left and a blank canvas (Fig. 1). An example alignment can be opened by selecting 'Demo' from the Help menu. We recommend that new users make use of this demo in order to become familiar with the look and feel and features of *ALINE*.

3.1. Menu bar

Items under the File menu allow the user to Save and Open marked-up alignments in *ALINE*'s own file format and to export new sequences or sequence alignments. The 'Export Postscript' option produces an encapsulated PostScript file which is suitable for publication-quality graphics. If *Ghostscript* is installed, 'Export PNG' will output a PNG-formatted bitmap. Further settings, such as any nonstandard locations of ancillary programs and default colour schemes, may be altered using 'Configure Aline Defaults'.

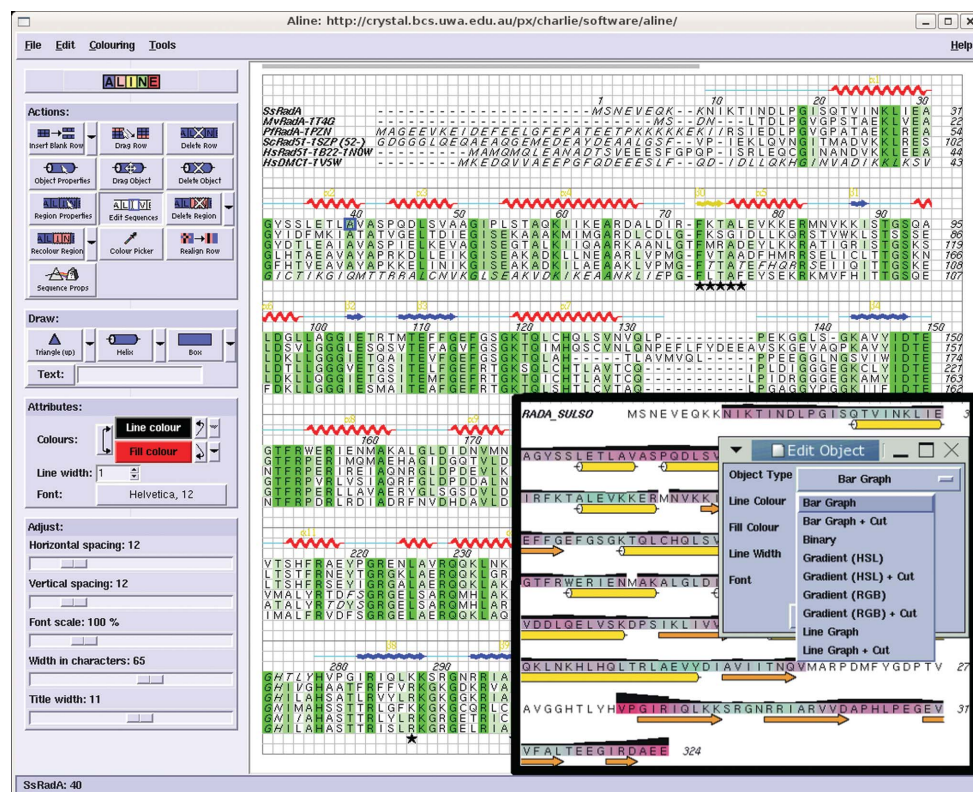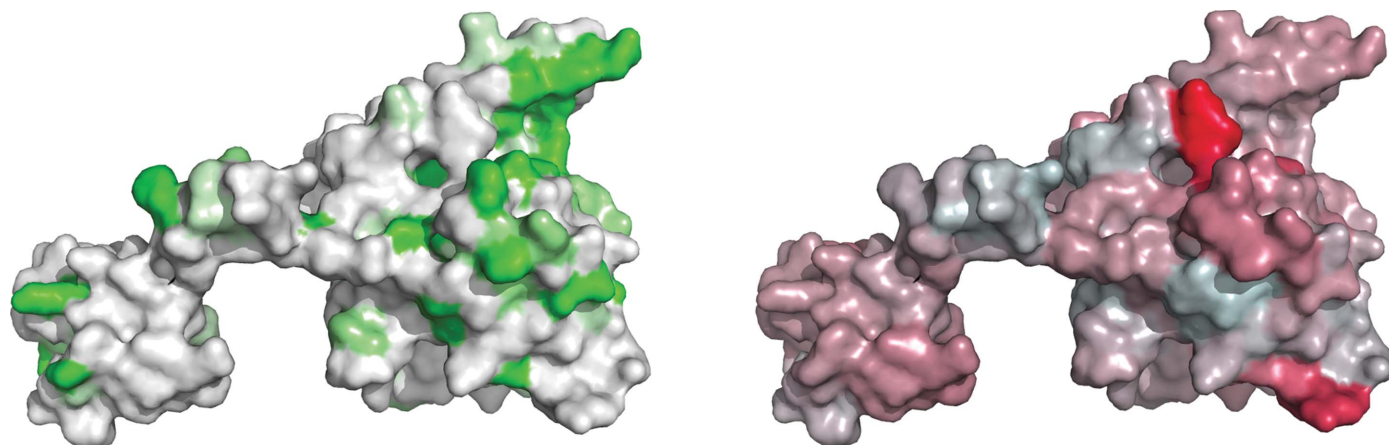The Edit menu includes options related to manually editing the actual

Figure 2
Screenshots from *PyMOL* with colour schemes matched to the alignments in Fig. 1 using scripts generated by *ALINE*.

alignment, adding new sequences from files, external databases or *via* a *BLAST* run and searching for sequence motifs. Options on the Colouring menu allow colouring of the sequence by similarity or bulk-similarity properties (as implemented in the *ALSCRIPT* program CALCONS command; Barton, 1993). Colour schemes for the various shades of similarity can be defined, loaded and saved using items from this menu.

The Tools menu includes tools for altering how the title for each sequence is shown (first line only, or on all lines) and for adding sequence numbers to the alignment. Secondary-structure annotation can be automatically added from a PDB file or from the output of *PSIPRED* (Jones, 1999) or *JPRED* (Cuff *et al.*, 1998) predictions (which must be run manually and saved). A generic graphing option allows one to read in numerical data of format '{residue number} {value}' with one record per line, which can then be plotted on the alignment as a colour gradient, bar or line graph. An additional menu item allows the inclusion of a graph of *B* factor obtained directly from an appropriate PDB file (Fig. 1).

Additional tools allow the user to add mark up directly from webserver submissions (if *WGET* is installed) such as *DisEMBL* (Linding *et al.*, 2003) for disorder prediction. Sequences can be aligned from within *ALINE* if *ClustalW*, *Muscle* or *MAFFT* are installed. A futher command allows one to write a *PyMOL* (http:// www.pymol.org) script enabling the user to easily match the colouring in a *PyMOL*-generated molecular-structure figure to that in the sequence alignment (Fig. 2).

3.2. Control panel

Items on the control panel allow one to alter how the alignment is marked up and its general appearance. The buttons in the Actions panel allow the user to insert new rows, drag them about to change the order and to delete rows. The properties and position of sequence elements and graphical objects can be altered. The Draw Panel allows a series of symbols (stars, triangles, helices *etc.*) or custom text to be drawn onto the alignment to highlight features. The Attributes Panel defines the default colours, line widths and font *etc.* to be used for drawing subsequent objects. The Adjust Panel parameters determine the overall layout of the alignment, such as the spacing between sequence characters, global scaling of font elements and the number of sequence characters displayed per line. These can be manipulated to fit the desired output format, *e.g.* for a figure in a journal article.

In summary, *ALINE* is a portable application which eases the production of typical sequence-alignment figures for publication.

References

Ariza, A., Richard, D. L., White, M. F. & Bond, C. S. (2005). *Nucleic Acids Res.* **33**, 1465–1473.
Barton, G. J. (1993). *Protein Eng.* **6**, 37–40.
Bond, C. S. (2003). *Bioinformatics*, **19**, 311–312.
Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. (2004). *Bioinformatics*, **20**, 426–427.
Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. & Barton, G. J. (1998). *Bioinformatics*, **14**, 892–893.
Galtier, N., Gouy, M. & Gautier, C. (1996). *Comput. Appl. Biosci.* **12**, 543–548.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Jones, D. T. (1999). *J. Mol. Biol.* **292**, 195–202.
Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. & Russell, R. B. (2003). *Structure*, **11**, 1316–1317.