KISS for STRAP: user extensions for a protein alignment editor

Christoph Gille*, Stephan Lorenzen, Elke Michalsky and Cornelius Frömmel

Institut für Biochemie, Charité, Monbijoustr. 2, Berlin 10117, Germany

Received on February 28, 2003; revised on May 17, 2003; accepted on June 24, 2003

ABSTRACT

Summary: The Structural Alignment Program STRAP is a comfortable comprehensive editor and analyzing tool for protein alignments. A wide range of functions related to protein sequences and protein structures are accessible with an intuitive graphical interface. Recent features include mapping of mutations and polymorphisms onto structures and production of high quality figures for publication. Here we address the general problem of multi-purpose program packages to keep up with the rapid development of bioinformatical methods and the demand for specific program functions. STRAP was remade implementing a novel design which aims at Keeping Interfaces in STRAP Simple (KISS). KISS renders STRAP extendable to bio-scientists as well as to bio-informaticians. Scientists with basic computer skills are capable of implementing statistical methods or embedding existing bioinformatical tools in STRAP themselves. For bio-informaticians STRAP may serve as an environment for rapid prototyping and testing of complex algorithms such as automatic alignment algorithms or phylogenetic methods. Further, STRAP can be applied as an interactive web applet to present data related to a particular protein family and as a teaching tool.

Requirements: JAVA-1.4 or higher.

Availability: http://www.charite.de/bioinf/strap/

Contact: christoph.gille@charite.de

The alignment editor STRAP was designed to align large numbers of proteins (Gille and Frömmel, 2001) and contains special tools that aid the handling of very dissimilar amino acid sequences. STRAP aligns a set of protein sequences automatically and allows the user to insert or delete gaps manually. In addition to protein sequences STRAP offers access to atom coordinates which can be transformed by 3D superposition of one protein structure onto another. In contrast to the previous version STRAP is now a genuine desktop application with an intuitive GUI and a comprehensive online documentation (Fig. 1). The wide range of functionality is attained by employing other programs for standard computational tasks. Here we acknowledge CLUSTALW (Aiyar, 2000) which aligns sequences automatically. The final high quality rendering of the alignment including sequence annotations is performed by the LaTeX-extension TeXshade (Beitz, 2000). For displaying 3D structures, RASMOL (Sayle and Milner-White, 1995) and VMD (Humphrey et al., 1996) are used. A particular problem is the limited monitor space when alignments and 3D-structures are viewed simultaneously. To overcome the limits of todays screens, adjacent monitors can be used as so-called remote X-displays. The inter-process communication between STRAP and these programs works unnoticed by the user in the background and does not require additional customization on most computer-platforms.

A large number of analyzing tools for proteins and protein alignments exist, e.g. ModView (Ilyin et al., 2003). Nevertheless, even the most comprehensive program has its limits. Permanently new tools and algorithms are developed but the implementation in program packages is always lacking behind. Further, some users require very special functions which, because of no general interest to the scientific community, will never be included in these packages. For this reason we regard extendibility and scriptability a very important program feature which is already found in many scientific programs. To our knowledge only the sequence alignment editor CINEMA (Lord et al., 2002) is extendable by the user. The modern architecture of CINEMA allows for user written extensions and modules and manages their interdependence.

It is complex and thus is designed for computer experts, but not for writing tiny applications (Lord et al., 2002).

Also life scientists who are no computer experts may want to add a particular functionality related to proteins and alignments. A typical task could be to count and highlight the occurrence of certain amino acid constellations possibly in conjunction with a structure related condition such as the location at the border between helices and turns. To facilitate the design of relatively simple plug-ins like this, we developed the concept KISS which differs markedly from the CINEMA design. The two main features of KISS are that the general purpose of plug-ins is defined by JAVA-interfaces of STRAP and that communication between modules is realized by events. This approach has the advantage that extensions can communicate without accessing the instances of other classes.

*To whom correspondence should be addressed.
Fig. 1. Screen-shot of STRAP showing an alignment of the β1 subunit of the proteasome colored according to charge and secondary structure. The horizontal scrollbar below outlines the entire alignment with the blue glassy area being the visible window. The cursor position is marked as a red square in the alignment and as thin red line in the visible portions of both scrollbars. In the object tree at the left the protein node for yeast (pdb1ryp) is expanded to illustrate affiliated objects: residue selections for the three active site residues and the residues selected in the 3D-wire.

directly, keeping dependence on code level reasonably low. Extensions may listen to events signaling, e.g. changes of the alignment. A typical response of the extension could be to broadcast an event telling that its state is changed. This event might invoke updating of dependent tables, diagrams or alignment panes.

Solvent accessibilities of amino acids is already an integral part of STRAP, nevertheless, let us assume that we create an add-in for plotting it along the sequence of a protein from scratch. We have chosen this complex example to explain the collaboration of three interfaces: ProteinParser, ValueOfResidue and StrapExtension. Furthermore it is a particularly interesting property in the context of sequence alignments because it is tightly correlated to the rate of evolution, the force that drives sequences apart. For simplicity we do not calculate the solvent accessibility. Instead we extract it from DSSP structure files (Kabsch and Sander, 1983) which can be computed from PDB files with the program DSSP. Therefore a suitable protein file parser for DSSP-files needs to be designed as a JAVA-class that implements the interface ProteinParser. For registration of file parsers STRAP maintains a customizable list. Our new parser requires space to store the array of values in. We have prepared a multi-purpose container within the proteins which can be accessed by the methods addObject[s](), getObject[s]() and removeObject[s]() . When a DSSP file is processed the new DSSP-parser stores the solvent accessibility values in the protein using addObject(). Further a menu item for the user to open a dialog is needed. Menu items are provided by classes that implement the interface StrapExtension. When the menu item is clicked the function of STRAP addDialog() should be called passing a third class as a parameter that also needs to be created. This third class should implement ValueOfResidue and basically returns the array
of solvent accessibilities of the residues when its method
\texttt{double[]} \texttt{getValues()} is called. At this point the add-in is
working. A comfortable GUI is created automatically by
STRAP for changing the protein and appearance of the plot.
Bridging numeric calculations and residue selections, an
upper and lower limit can be chosen with two sliders res-
ulting in a set of residues with values ranging within the given
interval.

\textit{StrapExtension} is the general interface for extensions in
STRAP. Its two methods take care for the menu entry and for
the observation of events. In contrast \textit{ValueOfResidue} is one
of the specialized interfaces dedicated to particular purposes.
It assigns a numeric value to each amino acid of a protein.
The user who creates an extension for STRAP should choose
the interface that generalizes the particular task. The easiest
way to do this is to adapt one of the enclosed examples to the
concrete situation.

We tried to keep all interfaces as simple as possible by redu-
cing the number of obligatory methods to a minimum. Further
we reduced the number of STRAP-classes that are relevant for
user developments. There are primarily only three types of
objects the author of plug-ins has to deal with: the alignment
object, the protein objects and the event objects.

\section*{REFERENCES}
multiple sequence alignment. \textit{Methods Mol. Biol.}, \textbf{132},
221–241.
alignments using latex2 epsilon. \textit{Bioinformatics}, \textbf{16},
135–139.
Gille, C. and Frömmel, C. (2001) STRAP: editor for structural align-
Ilyin, V.A., Pieper, U., Stuart, A.C., Marti-Renom, M.A., McMahan,
structure: pattern recognition of hydrogen-bonded and geomet-
ular multiple alignment editor. \textit{Bioinformatics}, \textbf{18},
1402–1403.