BIOINFORMATICS APPLICATIONS NOTE Vol. 17 no. 4 2001
Pages 377–378

STRAP: editor for STRuctural Alignments of Proteins

Christoph Gille and Cornelius Frömmel*

Institute of Biochemistry, Charité, Medical Faculty of the Humboldt University, Monbijoustraße 2a, 10117 Berlin, Germany

Received on May 9, 2000; revised on October 24, 2000; accepted on November 22, 2000

ABSTRACT

Summary: STRAP is a comfortable and extensible tool for the generation and refinement of multiple alignments of protein sequences. Various sequence ordered input file formats are supported. These are the SwissProt-, GenBank-, EMBL-, DSSP- PDB-, MSF-, and plain ASCII text format. The special feature of STRAP is the simple visualization of spatial distances of C_α-atoms within the alignment. Thus structural information can easily be incorporated into the sequence alignment and can guide the alignment process in cases of low sequence similarities. Further STRAP is able to manage huge alignments comprising a lot of sequences. The protein viewers and modeling programs INSIGHT, RASMOL and WEBMOL are embedded into STRAP. STRAP is written in Java. The well-documented source code can be adapted easily to special requirements. STRAP may become the basis for complex alignment tools in the future.

Availability: The tool is available to academic institutions at http://www.charite.de/bioinf. The source code can be requested via e-mail.

Contact: christoph.gille@charite.de

The generation of multiple alignments of protein sequences is a frequent task in biological research. STRAP was developed as a tool for the alignment of more than 400 proteasomal sequences. In only 17 cases is the crystal structure known. This is a typical situation because the sequence databases comprehend much more proteins than the database of protein structures. Why it was necessary to develop the tool STRAP for this task even though lots of software already exists. Various automatic computer programs are available which produce multiple alignments from a number of sequences. Some of them can be used via CGI (e.g. http://www2.ebi.ac.uk/clustalw/, http://www.toulouse.inra.fr/multalin.html, and http://www.ibc.wustl.edu/ibc/msa.html). However, the resulting alignments were not in accordance with common structural features as secondary structure and 3D coordinates of the C_α-atoms. The conclusion was that the amino acid sequences of different proteasomal subunits are too different and the structural information of 17 proteasomal subunits had to be used. Next we searched for freely available alignment software which could utilize the 3D-information. The interactive alignment program CINEMA (Parry-Smith *et al.*, 1998) can show 3D-skeletons of proteins. The programs VISEUR (Campagne *et al.*, 1999) and DINAMO (Bentz *et al.*, 1999) are both alignment programs and homology modeling tools. Their molecular graphics displays are more closely linked to the sequence alignment than in CINEMA. STING is an interconnected 3D and sequence viewer (http://www.asparagin.cenargen.embrapa.br/) with special features facilitating the analysis of interfaces between proteins (Da Silva *et al.*, 2000). InsightII/MODELER is a commercial modeling program (http://www.msi.com/life/products/insight/). In contrast SWISS-MODEL is an automated protein modeling server and freely accessible (Guex and Peitsch, 1997). The SWISS-PDB-VIEWER by Guex *et al.* (1999) is a very user-friendly protein viewer with integrated sequence alignment tools and protein superposition algorithms.

Unfortunately, the programs were not suitable for our task. One reason was that we had 400 sequences which did not fit on the screen. Consequently extensive scrolling makes the manipulation awkward. Modeling programs map the amino acids of a protein sequence onto one protein structure. Here we are dealing with many sequences and a bunch of structures.

Confronted with these difficulties, we got the idea to develop a tool which simply views spatial distance in the sequences alignment and which is able to manipulate alignments containing a high number of sequences.

STRAP is characterized by the following features:

- Structural relationships between aligned sequences can be visualized in the alignment. All residues within a certain distance from the current cursor position are drawn on a blue background (see Figure 1). Supposed, the homologous proteins are transformed into the same coordinate system, equivalent residues are highlighted on a structural basis.

*To whom correspondence should be addressed.

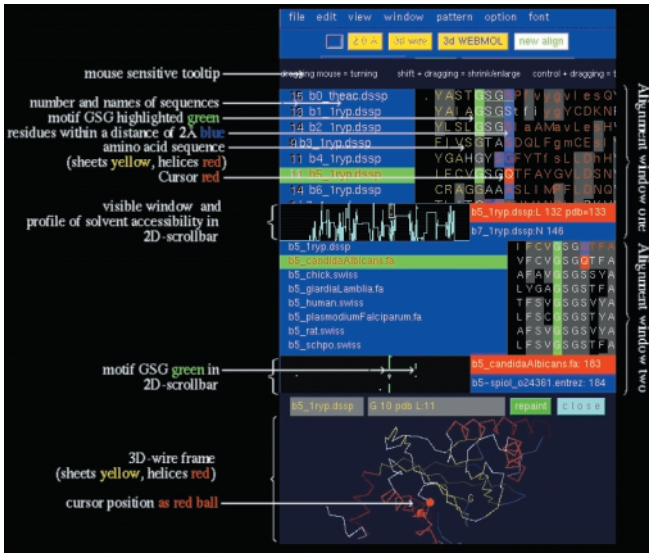

Fig. 1. Screen-shot of the program STRAP. The alignment of 224 proteasomal sequences is loaded and displayed in two windows. In the upper window sequences of the same family are displayed in one single line. In contrast, the representation of the alignment in the lower window requires 224 lines because each protein is printed in a separate line. The alignment does not fit into the windows and the 2D scrollbars are installed for scrolling. Some corresponding amino acids are spatially close to the C_α-atom at the cursor position. They are printed on a blue background.

- STRAP can handle sequence alignments comprising a high number of sequences. This ability is achieved by stacking several proteins in one line, where only the top one is visible. This is demonstrated in the upper half of the figure.
- The alignment can be viewed and edited in several synchronized windows, also known as 'split screen editing'.
- Groups of sequences can be edited simultaneously by forming blocks. This feature was inspired by CINEMA.
- Position based sequence variability (entropy) or PROSITE patterns can be projected onto the 3D C_α-trace.
- The protein viewers INSIGHT, RASMOL (Sayle and Milner-White, 1995) and WEBMOL (Walther, 1997) are integrated into STRAP.

Next we present our strategy in detail. At first, all 17 structures of proteasomal subunits were superimposed using the β-subunit of *Thermoplasma acidophilum* as the reference structure (Goede *et al.*, 2000). The protein-backbones of the 17 subunits are similar though their sequences are not, in accordance with the findings of Chothia and Lesk (1986). We generated 17 sequence alignments automatically. In each alignment we included only sequences of the same proteasomal subtype. Within one subtype the sequence similarities are sufficient for automatic alignment programs. Next we imported the 17 alignments into STRAP. All sequences belonging to one subtype were placed into the same editor line as is shown in the upper window of the figure. The 17 lines were aligned on a structural basis. The resulting alignment contains more than 400 sequences and can be viewed in the demo applet.

REFERENCES

Bentz,J., Baucom,A., Hansen,M. and Gregoret,L. (1999) Dinamo: interactive protein alignment and model building. *Bioinformatics*, **15**, 309–316.

Bentz,J., Baucom,A., Hansen,M. and Gregoret,L. (1999) Dinamo: interactive protein alignment and model building. *Bioinformatics*, **15**, 309–316.

Campagne,F., Jestin,R., Reversat,J.L., Bernassau,J.M. and Maigret,B. (1999) Visualisation and integration of G protein-coupled receptor related information help the modelling: description and applications of the Viseur program. *J. Comput. Aided Mol. Des.*, **13**, 625–643.

Chothia,C. and Lesk,A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.*, **5**, 823–826.

Da Silva,M.C., de Sa,M.F., Chrispeels,M.J., Togawa,R.C. and Neshich,G. (2000) Analysis of structural and physico-chemical parameters involved in the specificity of binding between alpha-amylases and their inhibitors. *Protein Eng.*, **13**, 167–177.

Goede,A., Preissner,R. and Frömmel,C. (2000) A fast algorithm for superposition of arbitrary atomic arrangements, in preparation.

Guex,N. and Peitsch,M.C. (1997) Swiss-model and the Swiss-pdbviewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.

Guex,N., Diemand,A. and Peitsch,M.C. (1999) Protein modelling for all. *Trends Biochem. Sci.*, **24**, 364–367.

Parry-Smith,D.J., Payne,A.W., Michie,A.D. and Attwood,T.K. (1998) Cinema—a novel colour interactive editor for multiple alignments. *Gene*, **221**, C57–C63.

Sayle,R.A. and Milner-White,E.J. (1995) Rasmol: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374.

Walther,D. (1997) Webmol—a Java-based pdb viewer. *Trends Biochem. Sci.*, **22**, 274–275.