

Tips and tricks of the trade

Drag 'n' Drop Proteins



Most Open Source Sequence alignment programmes do not support drag and drop to easily grab selected proteins and shift them to the target desktop application. The free alignment software STRAP is a striking exception to this rule.

Lab Hint

Researchers, who occasionally use alignment programmes, appreciate intuitive graphical user interfaces that follow general conventions, while advanced users go for alignment programmes that offer a wide range of embedded algorithms, data visualisation and work flow to thoroughly analyse protein sequences and structures. The free open source protein alignment programme STRAP, may be worth a try for both novice and advanced alignment programme users, since it is especially designed to streamline repeated programme tasks. STRAP can be started without any download by a simple mouse click on the Java Web-Start button at <http://www.bioinformatics.org/strap/> or <http://3d-alignment.eu/>.

Almost all desktop applications enable loading proteins or alignments via a graphical file selector dialogue (Ctrl-O). Mouse-driven navigation through the file directories is intuitive but cumbersome. Shifting files or other objects with the computer mouse by drag and drop (D&D) to the target location or target desktop application is a more convenient way of importing files that are usually embedded in commercial alignment software packages. The user simply selects one or several sequence files on the desktop or in the file browser and “drags and drops” them into the application window. Unfortunately, most free programmes do not support D&D. Furthermore, many Bioinformatic web services, such as Uniprot’s text search, yield lists of proteins that must be downloaded manually into a directory and, subsequently, imported into the application.

To facilitate the protein sequence import from Web services, STRAP allows dragging of protein IDs directly from the browser into the application. To further reduce the length of the mouse path spanning from the browser to the STRAP window, a tiny drop target may be placed next to the protein IDs on the web page. Though the dropped proteins appear in the alignment, with a short delay of about three seconds, there is no need to wait for the download to

be completed. The user may proceed, grabbing further proteins at any speed. If the downloaded proteins should be subjected to other software tools, they can be copied to any location on the file system. STRAP differentiates between proteins, alignments, residue selections, ligand structures and image icons. Dropping an object in STRAP may evoke different programme actions, depending on the type of object and the target.

Automatically-generated sequence alignments may fail to reflect the true evolutionary relationship of sequence positions. Such “alignment errors” are typical for sequence alignments of remote homologues if the 3D-structures are neglected. These errors become evident when conserved functional positions, such as catalytic sites or ligand binding sites, are not aligned adequately. Many alignment programmes completely ignore 3D-coordinates, only a few command line programmes, e.g. T-coffee, MultiProt and ClustalW and their web-front ends, use 3D-coordinates for multiple sequence alignments.

Structural superposition

STRAP chooses the optimal alignment strategy according to the available data. It utilises a sequence-based approach, if only sequences of proteins exist. If, however, the 3D-structures of at least two of the proteins are known, structural superposition is combined with multiple sequence alignment. STRAP’s button “Associate 3D-structure” in the tool-panel below the alignment window offers three possibilities. The user may directly load the structure if it is already published and deposited in the PDB archive. Since this is rather unlikely, a homologous structure must be chosen in most cases. The classical approach is to conduct a BLAST-search to identify a structure file in the PDB with a similar sequence, which takes about 20 seconds. Proteins having a structure identifier related to one of the public protein sequence collections may be found much faster by looking-up pre-computed search results.

Sequence variations or mutations may affect, e.g. folding, stability and function

of a given protein. Visualisation of the protein sequence often gives a first clue on a genetic-variations possible impact. The variant positions may be related to specific positions called “sequence features” that provide information about ligand binding, post-translational modifications as well as catalysis or DNA binding. STRAP downloads “sequence features” from various sources and stores sequence variations and residue selections as annotated residue selections. Textural information, e.g. the source of information, notes on experimental evidence, web links, database references, rendering hints for protein viewers and for the alignment PDF output, may be attached to the residue selection. Whether a mutation affects the catalytic site or a binding site can be assessed in the 3D-visualisation. STRAP uses the programme Pymol for 3D-visualisation. Pymol is tightly integrated, i.e. mouse clicks on amino acids in Pymol are sensed by STRAP and vice-versa.

Many research teams present illustrations of alignments on their websites to feature their projects. Usually, specific sequence positions are highlighted using marks and underlining. These snap shots often represent “frozen” states of alignments that are not automatically updated if new data is available. STRAP, in contrast, dynamically embeds alignments in web pages to ensure that sequences, structures and sequence features are continuously downloaded from their original sources. Since the integration of STRAP in web pages is very robust and powerful, web services such as CE/CE, Gangsta+, JenaLib, PDBSum, Prodom, Superimpose and ViperDB already use STRAP as a viewer for protein structures, 3D-superpositions and sequence alignments. Currently, integration into Cosmic mutations, Dasty DAS client, Cath, Ensembl, SSAP and SwissVar is in progress.

CHRISTOPH GILLE

Do you have any useful tips?

Contact us at:

editors@lab-times.org