VAMSAS[‡]: Bridging the gaps between the analysis of DNA, Protein Sequences and Protein Structure

Jim Procter[1], Andrew Waterhouse[1], Dominik Lindner[2], Iain Milne[2], Pierre Marguerite[4], David Martin[1], Frank Wright[3], Tom Oldfield[4], David Marshall[2], Kim Henrick[4] and Geoff Barton[1]
[1]University of Dundee; [2]Scottish Crop Research Institute; [3]Biomathematics and Statistics Scotland; [4]European Bioinformatics Institute.
[‡] 'Visualisation and Analysis of Molecular Sequences, Alignments and Structures' is a project funded by the U.K. Biotechnology and Biological Sciences Research Council's Bioinformatics and eScience Initiative.
http://www.vamsas.ac.uk

VAMSAS is a collaboration between three research groups that develop freely available interactive graphical applications for evolutionary analysis, alignment manipulation and analysis, and protein structure visualization. It is increasingly rare that we gain new insight without combining information from at least two of these three fields. Moreover, each has specialised concepts, analysis methods and visualisation requirements that do not fit easily together. Overcoming this problem is the main focus in VAMSAS.

The Applications

Jalview is a widely used protein and nucleic acid multiple sequence alignment editor and viewer. It now has extensive sequence and alignment annotation capabilities, as well as access to web services for sequence retrieval, DAS sequence annotation, multiple sequence alignment and secondary structure prediction.

TOPALi is an application for evolutionary analysis that provides interfaces for handling computationally intensive calculations. It now includes several public web services for phylogenetic (protein and DNA), positive selection and recombination analysis.

AstexViewer@MSD-EBI (AV) is the primary platform independent structural visualization and analysis front end to the Macromolecular Structure Database. It provides access to services for structure comparison and mapping across sequence and structure.

Jalview, TOPAli and AV are independent applications with their own set of services for data retrieval and analysis. However, it is easy to see that results generated by one application are of use in one or both of the others. For example, an accurate phylogenetic analysis of a protein family can aid interpretation of observable physico-chemical variation between structures.

Dynamic Data Exchange and Interoperation

Our initial work involved identifying concepts common to all applications. These include sequences, structures, alignments, trees, and annotation. An XML Schema capable of their representation was then developed. This is a form suitable for both local file based exchange and interoperable web based analysis services. However, these modes of data exchange do not alone support dynamic visualization of calculations carried out in one application but viewed in another. To achieve this, we enabled each application to share a single document. Data and analysis viewed in one application can be written to this document, making it available to all others. An application can also broadcast messages about information contained in the document, so a user may interactively view the same data in two or more independent visualizations simultaneously.

Conclusions

We have simplified studies that involve analysis across three major areas in protein bioinformatics by providing analysis tools which transparently share data and results. The means by which this is accomplished is general and extensible, and provides a way for other analysis tools and services to access and provide data and results to existing VAMSAS application users.