

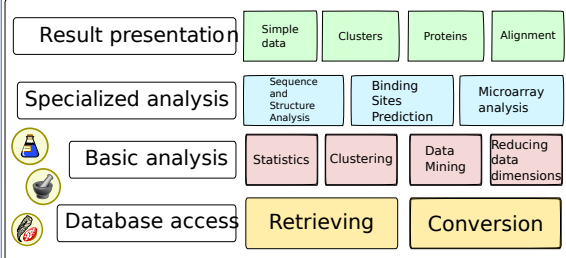
Bioinformatics Applications in the Virtual Laboratory

<http://virolab.cyfronet.pl>

Objectives

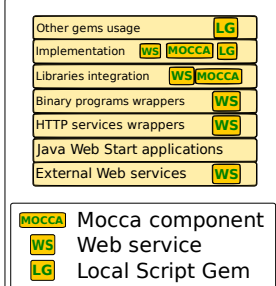
The ViroLab virtual laboratory can be used in many domains of science, among which bioinformatics is perhaps the best example. To show this, we provide a set of applications (*gems*) which can be used in newly created bioinformatics *experiments*. Additionally, we present an analysis of the available software and its classification into a number of categories, taking into consideration application technology, usage scope and applicability to various type of problems.

Bioinformatics Applications



Classification of applications by scope of usage.

Gem technology with relevant Grid Object implementation.



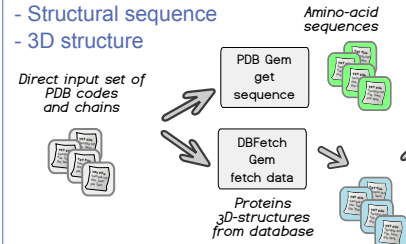
Bioinformatics Gems

Accessing various databases - DbFetch, PDB, Scop, GEO and data format adjustment in FormatConverter. **Statistics** in R, **Data clustering** with Cluto and Cluster 3.0 **Data mining** in Weka, PCA and MDS methods for **Reducing Data Dimensionality**. **Microarray** data clustering, dataset downloading and creation from data samples. Multiple **Sequence Alignment** with services provided by EBI, Multiple **Structure Alignment** (Mammoth, MultiProt, SSM). Services for **predicting ligand binding site** in proteins. **Result visualization**: numerical data in Gnuplot, protein structure (Jmol, ProteinWorkshop), sequence alignment (JalView), microarray (JTreeView)

Comparing Protein Sequence and Structure

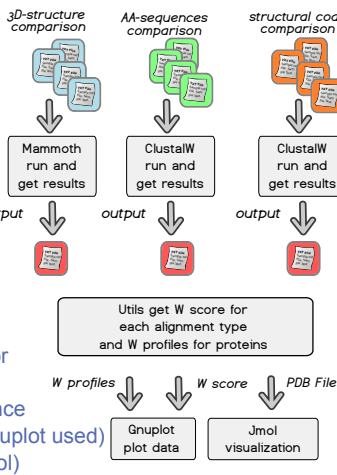
Compare a family of proteins on three levels of protein description
 - Amino acid sequence
 - Structural sequence
 - 3D structure

Possibility of using different gems to solve the same part of problem



Visualization created for conservative regions along amino acid sequence (W profile for protein, Gnuplot used) and for 3D structure (Jmol)

Search for conservative regions on each level.
 „Early Stage” model developed by prof. Irena Roterman and her team



Data acquisition:

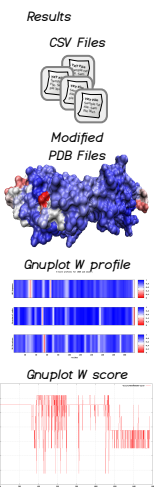
- Pdb codes (ScopDb, direct data)
- AA sequence (Pdb)
- Structural codes (EarlyFolding)
- 3D structures (DbFetch)

Additional data manipulation

- Aligning sequences and structural codes
- ClustalW
- Aligning structures
- Mammoth

Analyzing alignments

- Computing W score
- Creating results
- W score and W profiles plots
- Modified PDB files
- CSV files
- Additional visualization



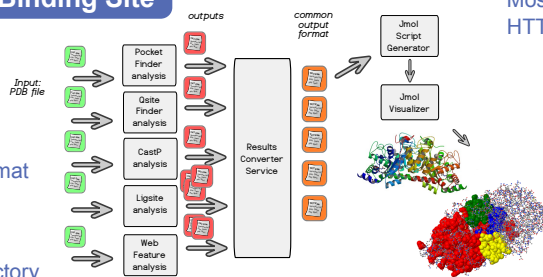
Services for Predicting Ligand Binding Site

Analysis of protein binding sites facilitates determining protein function or searching for substances which will have an effect on a given protein.

Converting results from service-specific format to a common one.

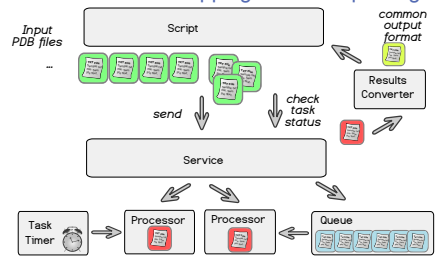
Two experiment modes are available:
 - batch analysis of PDB files from user directory
 - interactive mode, where the user inputs PDB codes

Available services: CastP, ConSurf, FOD, LigsiteCSC, PASS, Pocket-Finder, Q-Site-Finder, SuMo, WebFEATURE



Each service has defined system parameters:
 - number of processors for analyzing tasks
 - peak analysis time

Most services are available only via WWW or email
 HTTP communications wrapping and task queuing system used



With the task queuing system gems developed as Web services provide an asynchronous mode of analysis.

Authors Tomasz Jadczyk (3), Maciej Malawski (1), Marian Bubak (2,3)

(1) Institute of Computer Science AGH, al. Mickiewicza 30, 30-059 Krakow, Poland

(2) Informatics Institute, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

(3) ACC CYFRONET AGH, Krakow, ul. Nawojki 11, 30-950 Krakow, Poland

References - M. Brylinski, L. Konieczny, P. Czerwonko, W. Jurkowski, and I. Roterman. "Early-stage Folding in Proteins (In Silico) Sequence-to-Structure Relation"; *Journal of Biomedicine and Biotechnology*, 2 (2):65-79, 2005

- A. Labarga, F. Valentin, M. Anderson, and R. Lopez. "Web Services at the European Bioinformatics Institute"; *Nucl. Acids Res.*, 35 (Web Server issue)
 - M. Bubak, M. Malawski, T. Gubala, M. Kasztelnik, P. Nowakowski, D. Harezlak, T. Bartynski, J. Kocot, E. Ciepiela, W. Funika, D. Krol, B. Balis, M. Assel, and A. Tirado-Ramos. "Virtual Laboratory for Development and Execution of Biomedical Collaborative Applications"; in: M. Cannataro (Ed.) *Handbook of Research on Computational Grid Technologies for Life Science, Biomedicine and Healthcare*. Information Science Reference, 2009, IGI Global

Acknowledgements Thanks to prof. Irena Roterman-Konieczna, dr. Monika Piwowar and Katarzyna Prymula, Department of Bioinformatics and Telemedicine, Jagiellonian University Medical College.

ViroLab EU-IST 027446

Coordinator: Prof. P.M.A. Stout
 Universiteit van Amsterdam
www.virolab.org

