A Target Selection Informatics Resource for Structural Genomics

Ana Rodrigues[1], Guy G. Dodson[12], Roderick E. Hubbard[13]

Keywords: structural genomics, target selection, informatics resource

A large number of structural genomics programs have been established worldwide with the common aim of large-scale, high-throughput protein structure determination. Due to the considerable challenges posed by the experimental methods of structural determination (primarily X-ray crystallography and nuclear magnetic resonance spectroscopy), it is important to select and prioritize candidate molecules that will maximize the information gained from each new structure.

We are currently developing an informatics resource capable of performing target selection through the implementation of a number of sequence analysis protocols. This framework aims to facilitate the selection and prioritization of candidate proteins for structural determination, by enabling structural biologists to select targets from their genomic sequences of interest, and according to their own research needs.

The work is being developed on three distinct levels. The first focuses on the identification of the kinds of annotations that can be devised for a nucleotide or protein sequence and the assessment of the usefulness of such information for structural genomics. The second involves the establishment of an informatics framework to support these calculations and data through a relational database. The third concerns the refinement of the system and its application to real structural genomics programs.

Preliminary results, driven by the Structural Biology Laboratory's participation in structural and functional genomics projects (such as the EU funded Structural Proteomics IN Europe project, and the Wellcome Trust funded MalariaVac consortium) indicate that the resource will prove useful in performing target selection on a genomics scale [1, 2].

References

[1] Rodrigues, A. and Hubbard, R.E. 2003. Making decisions for structural genomics. *Briefings in Bioinformatics* 4(2):150-167.

[2] URL: http://www.ysbl.york.ac.uk/~rodrigues/targets.html

[1] Structural Biology Laboratory, University of York, York YO10 5YW, United Kingdom. E-mail: rodrigues@ysbl.york.ac.uk

[2] National Institute for Medical Research, Mill Hill, London NW7 1AA, United Kingdom. E-mail: ggd@ysbl.york.ac.uk

[3] Vernalis Ltd, Granta Park, Abington, Cambridge CB1 6GB, United Kingdom. E-mail: rod@ysbl.york.ac.uk