

How to pick the best low-hanging fruit in medically important target genomes.

Helliwell J.R*, Cianci M., Raftery, J and Rosmarin, I.
*Dept of Chemistry, University of Manchester, M13 9PL,
UK.*

*Keywords: Structural genomics; Bioinformatics;
Boolean algebra gene selection*

The original promise of structural genomics, every gene a protein structure, remains unrealised. We have developed a Boolean method for targetting 'high return' genes, which is general, and applied it to the medically important M.tuberculosis target genome along with that of M. leprae so as to keep essential genes, but also deleting eukaryotic homologues [1]. From that analysis, out of thousands of genes, we have a set of 65 Clusters of Orthologous Groups comprising 95 proteins. From within that select group we have now also considered those genes that have a low predicted protein 'foldability' [2] and that are also probable transmembrane proteins, which have a lower crystallisation probability. Thus, overall, we have selected new 'rational-drug-design' gene candidates whilst recognising the problem cases for protein crystallographic study; these problem cases are evocatively referred to as being the 'high hanging fruit' ie the difficult pickings.

This is a contribution from the North West Structural Genomics Consortium (www.nwsgc.ac.uk) *University of Manchester Chemistry Node.*

- [1] **James Raftery and John R. Helliwell** Spoilt for choice: protein target selection in a time of plenty *Acta Cryst* (2002). **D58**, 875-877
- [2] **Uversky VN, Gillespie JR, and Fink, AL** Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 2000; 41:415-427.