STRATEGIES FOR STRUCTURAL GENOMICS TARGET SELECTION

Terry Gaasterland

Laboratory of Computational Genomics, Rockefeller University, 1230 York Ave., New York, NY
10021

gaasterland@rockefeller.edu

We have developed a protein sequence analysis pipeline that ranks proteins as targets for high throughput structure determination. The ranking is designed to maximize both the biological and informational impact of new 3D protein structures solved through the structural genomics initiative. The analysis system accepts proteins from multiple genomes as input, builds sequence families based on remote homology, identifies families with one or more solved structures, and ranks the remaining families according to criteria designed to maximize structure determination efficacy, increase the likelihood of a novel fold, and maximize the number of new protein structure models that can be built from a solved structure.

REFERENCES

1. Gaasterland, T. and Oprea, M. (2001) *Curr. Opin. Struct. Biol.* 11(3), 377–381. Review. PMID: 11406390.
2. Gopal, S., Schroeder, M., Pieper, U., Sczyrba, A., Aytekin-Kurban, G., Bekiranov, S., Fajardo, J.E., Eswar, N., Sanchez, R., Sali, A., and Gaasterland, T. (2001) *Nat. Genet.* 27(3), 337–340.