REGIONAL CENTRE FOR BIOTECHNOLOGY
Seminar series

Protein tertiary structure prediction in the post-genomic era

Shashi Bhushan Pandit, PhD
Genoscope, CEA,
Evry cedex, France

Friday, November 25, 2011
11:00 AM
Seminar Room

Despite significant progress, the prediction of protein tertiary structure from amino acid sequence remains an unsolved problem in computational structural biology. Protein tertiary structure prediction methods can be logically divided into Template-based modeling (TBM) and Template-free (TF) categories. The TBM methods in comparison to TF methods employ structurally related templates to the target sequence as an essential input. Hence, the success of TBM requires similar structures to those adopted by the target sequence to be present in the Protein Data Bank (PDB) and the ability to select such templates and generate high quality alignments. To identify better templates we have developed methods to be used in conjunction with Threading/ASSembly/Refinement (TASSER), which generates full-length models by rearranging the continuous protein structural fragments identified by threading followed by refinement in an optimized knowledge-based force-field.

I have developed two methods *METATASSER* and *TASSER_low-zsc* to select better templates for improving accuracy of structure prediction. *METATASSER* methodology employs 3D-jury approach to select threading templates from three threading methods viz. SP^3, SPARKS^2 and PROSPECTOR. In contrast, *TASSER_low-zsc* uses the structural similarity of the threading aligned region from low z-score ranked templates. The benchmarking of *TASSER_low-zsc* on a representative set of proteins showed TM-scores of best models improved by ~4-9% over original TASSER models. In addition, this procedure also resulted in more foldable targets. *METATASSER* has been successfully used in Critical Assessment of protein Structure Prediction (CASP) experiments and has been ranked within top 10 automated servers. In both procedures, final models show improvement over the initial alignment and are closer to the native structures.

For modeling of homologous sequences (sequence identity between the target and template is ≥35%), I have developed optimized method (TASSERLite) for speed and accuracy of predicted models. In the benchmark study on 901 proteins, TASSERLite showed improvement

in the aligned region of ~10% in the root mean-square deviation from native over the initial templates. Moreover, final models showed good prediction for the unaligned regions (loops). Thus, TASSERLite could significantly refine the initial template structures and provide models that are considerably closer to the native structures. This method has been used successfully in large-scale tertiary structure modeling of proteins encoded in genomes. The retrospective assessment of structure prediction for the sequences with predicted structures and for which experimentally solved structures are available showed that TASSER method has reliable prediction accuracy. I will propose methodology for structure prediction of multi-domain proteins. The modeled protein structure can be used for its function prediction. I will present methodology for protein function prediction.