ISMB2003 Tutorial Proposal

Tutor: Assoc. Prof. Shoba Ranganathan,
Departments of Biochemistry and Biological Sciences & Bioinformatics Centre
National University of Singapore, Singapore
shoba@bic.nus.edu.sg

Relevant qualifications and teaching experience (brief resume at ***surya.bic.nus.edu.sg/shoba***)

- **Research**: I have worked in the area of Biocomputing for over two decades and in protein structure prediction and modeling for over five years: recent publications in the area of molecular modeling are included in the **References** section at the end of this document. As a scientist using molecular modeling to address biological problems rather than a methodology/software generator, I have tried out several available methods and software for model generation and support the view that a single computational approach is usually insufficient in providing solutions to all biological questions.
- **Professional Appointments**:
  - **Chair: S* Life Science Informatics Alliance** - global bioinformatics distance education.
  - **Editorial Board: Applied Bioinformatics**
  - **Previous appointments** (relevant to this submission):
    - **Research Fellow:** Australian Genomic Information Centre & ANGIS, Sydney.
    - **Consultant – Molecular Modeling:** eBioinformatics (now Entigen) Inc.
- **Teaching and training:** With 10+ years as an academic, I have given specialized training courses in Protein Structure, Structure Prediction and Molecular Modeling Australia-wide while at ANGIS, as well as designed the sequence-to-structural model interface for integrated web-based bioinformatics analysis "**BioNavigator" (bionavigator.com)** system. My lecture (#6 - Protein Structure Primer) for the S* Life Science Informatics Alliance (www.s-star.org), provides an overview to a protein bioinformatics and molecular modeling course. At NUS, I oversee the Graduate Program in Bioinformatics and undergraduate bioinformatics modules. This tutorial received the best ranking at ISMB2002.

Title: Molecular Modeling: building a 3D protein structure from its sequence

Expected goals, objectives and motivation of the tutorial

Molecular modeling can be distinguished from all other methods for analyzing relationships among protein sequences because it yields atomic coordinates suitable for direct comparisons with experimentally determined x-ray and solution NMR structures. This technique is very relevant today, as several thousands of new protein sequences are now available, consequent to the worldwide genomic sequencing efforts. The function of these proteins may be ascertained from an insight into their structure at an atomic level. Knowing a protein's three-dimensional (3D) structure suggests clues as to its interactions with other biopolymers and ligands, and has enormous implications in computer-aided drug design, mutagenesis and protein engineering.

Although detailed structural information is available for only ~15% of all known protein sequences, one of the primary goals of the Structural Genomics projects is to generate experimentally determined structures with sufficient diversity to support homology modeling of large numbers of protein sequences. Recent analyses by Sander and coworkers [1] have

suggested that determination of as few as 16,000 selected structures could enable production of ‘‘accurate’’ homology models for 90% of all proteins found in nature. For this subset, a rapid *in silico* approach to 3D structure is via homology modeling [2], since homologous sequences adopt very similar structures. For those proteins that do not have a detectable structural homologue, it is possible to arrive at a putative structural model using secondary and tertiary structure prediction methods, followed by comparative modeling.

The tutorial will provide a route map starting from a protein sequence and arriving at a 3D atomic-scale model structure, based on homology or comparative modeling, using wherever possible, freely available software and web-servers. Specific examples from the literature [3] and from my published work [4-8: simple proteins to a viral capsid assembly] will be provided. The main objectives of the tutorial are to understand the protocols involved in generating a good structural model using analytical principles, rather than using one-step black-box approaches.

Intended audience: This is an introductory level tutorial aimed at those who wish to enter the field as well as to provide the general audience an appreciation of protein structural models. Prerequisite: Introductory Biochemistry such as Biochem 101. A background in introductory computer-based sequence analysis would be advantageous. Pre-tutorial learning material will include S* lectures 2 and 6 (available from www.s-star.org and its mirror sites).

Length: Half day

Detailed outline of the presentation.

Topics that will be addressed include:

- *Understanding protein structure*
  - Levels of protein structure
  - Structural classification
  - Databases
  - Structural families
  - Visualizing protein structure
  - Surface properties
  - Mapping functional regions onto the surface
- *Quick overview of sequence analysis*
  - Siblings and cousins
  - Multiple sequence alignment
  - Conservation of residues important for structure or function
  - Sequence motifs and patterns
  - Databases of aligned protein “families”
  - Protein domains
- *Finding a structural homologue*
  - Criteria for defining a “homologue”
  - Structural template for small proteins stabilized by
    - Disulfide bonds
    - Metal ions
  - What if there is no structural homologue

        - Secondary structure prediction
        - Tertiary structure prediction
- *Template selection*
    - One or many template structures?
    - Differences between sequence alignments derived from
        - Sequence methods
        - Structural alignments
- *Aligning query sequence to template structure(s)*
    - Residue conservation check
        - Functional regions
        - Motifs or patterns
    - What to do with "indels" (insertion/deletion or gap regions in the alignment)
    - To edit the alignment or not
    - Visual inspection of "indel" regions
- *Building the model*
    - Methods available
    - Automated *vs.* manual step-by-step approach
    - Structural quality analysis
    - Improving ill-defined regions
        - Iterative model building
        - Molecular dynamics and/or Monte Carlo simulations.
- *The molecular modeling protocol: a quick recap*
    - Resources required
    - Preparing the required input
    - Decision points
    - Iterative steps
- *Beyond simple model building*
    - Multiple domains in a protein chain
    - Membrane proteins
    - Proteins with multiple chains
    - Proteins with bound ligand and cofactor molecules

References:

1. Vitkup, D., Melamud, E., Moult, J., Sander, C. 2001. *Nat. Struct. Biol.* **8**, 559–566.
2. Marti-Renom, M. A., Stuart, A., Fiser, A., Sanchez, R., Melo, F., Sali, A. 2000. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 291–325.
3. Rodriguez, R., Chinea, G., Lopez, N., Pons, T., Vriend, G. 1998. *Bioinformatics* **14**: 523-528.
4. Ranganathan, S., Singh, S., Poh, C.L., Chow, V.T.K. 2002. *Appl. Bioinformatics* **1**: 43-52.
5. Ranganathan, S. 2001. *Biotechniques* **30**, 50-52.
6. Ranganathan, S., Male, D.A., Ormsby, R.J., Giannakis, E., Gordon, D.L. 2000. *Pac. Symp. Biocomputing* **5**: 155-167.
7. Simpson, K.J., Ranganathan, S., Fischer, J., Janssens, P.A., Shaw, D.C., Nicholas, K.R. 2000. *J. Biol. Chem.* **275**: 23074-23081.
8. S. Ranganathan, D.A. Male, R.J. Ormsby, E. Giannakis and D.L. Gordon (2000): Pinpointing the putative heparin/sialic acid-binding residues in the 'sushi' domain 7 of factor H: a molecular modeling study, Pacific Symposium on Biocomputing, 5, 155-167.