MULTI-SCALE MODELLING OF BIOSYSTEMS: FROM MOLECULAR TO MESOCALE

JULIE BERNAUER
INRIA AMIB – Bioinformatique, LIX,
École Polytechnique, 91128 Palaiseau, France
Email: julie.bernauer@inria.fr

SAMUEL FLORES
Department of Bioengineering,
Stanford University, Stanford, California 94305, USA
Email: scflores@stanford.edu

XUHUI HUANG
Department of Chemistry,
Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
Email: xuhuihuang@ust.hk

SEOKMIN SHIN
School of Chemistry
Seoul National University, Seoul 151-747, Republic of Korea
City, State ZIP/Zone, Country
Email: seokmin.shin@gmail.com

RUHONG ZHOU
IBM Thomas J. Watson Research Center,
Yorktown Heights, NY 10598, USA
Email: ruhongz@us.ibm.com

1. Background

Modeling of Biosystems and their representations is one of the major challenges in current computational biology. Topics such as Structure Prediction, Dynamics and Sampling, Mesoscale Modeling, Molecular Assemblies, Structural Interactions and Systems Biology are important for better understanding biological function. They are still computationally or experimentally expensive and lead to a large amount of data, with attendant analysis challenges. In the structural genomics and systems biology era, models are thus needed at different scales, both in space and time. This session focuses on multi-scale approaches to model biosystems, and in particular those which extend simulation size and time scales.

Recent progress in protein structure prediction and protein folding dynamics has allowed in-silico experiments to reach longer timescales. Knowledge-based techniques have grown in accuracy and efficiency, partly due to the growth in solved structures, but the physics-based simulation so crucial to correct kinetics and thermodynamics is still very challenging. Adequate

sampling and dynamics analysis at different size and time scales is thus of interest in the field and in this session.

Molecular modeling techniques can now model complexes of significant size, while at the macroscopic scale neuromuscular and tissue modeling has gained finer and finer resolution. In the middle stands the mesoscale -- cellular systems which we wish to model accurately and economically. In this session we will exhibit structural and dynamical modeling techniques which represent progress towards that goal.

Along with modeling molecules comes the challenge of understanding how they interact and aggregate, since a complex is often the functional unit in normal cells, or the causative agent in disease. The analysis of such assemblies is key in many therapeutic studies. Several neurodegenerative diseases are known to be correlated to protein aggregates. Understanding the phenomena of complex formation, aggregation, and misfolding requires the combination of various techniques. This session will present novel techniques and results in this field.

A yet higher level of organization exists at the system level. The role of computer modeling at this level has grown in the emerging discipline of systems biology. Different physiological functions occur at different scales, and so this field often requires multiresolution modeling techniques. Due to its crucial high-level perspective of the mesoscale, systems biology is one of the foci of our session.

The Biology, Computer Science, Chemistry and Mathematics communities have a growing interest in multiscale modeling at or near the mesoscale. This session highlights its impact on our understanding of biological systems.

2. Session Summary

This session includes four oral presentations and accepted articles, a tutorial and a discussion session. The tutorial contains two parts: a survey talk and an introduction to the multi-scale macromolecular modeling software RNABuilder.

2.1. *Oral Presentations and accepted articles*

Molecular dynamics simulations of the full triple helical region of collagen type I provide an atomic scale view of the protein's regional heterogeneity

Authors: Dale L. Bodian, Randall J. Radmer, Sean Holbert and Teri E. Klein

This article presents an all-atom explicit solvent molecular dynamics study (10 ns) of the full triple helical domain of collagen to gain insights on the role of variation in the sequence, structure and dynamics of the protein involved in fibril formation. To make this large system tractable, the authors split it into smaller overlapping fragments. A careful analysis of the trajectories obtained showed that key regions of collagen present structural heterogeneity.

Computational generation inhibitor-bound conformers of P38 MAP kinase and comparison with experiments

Authors: Ahmet Bakan and Ivet Bahar

This study focuses on the structural dynamics of an important drug target: the p38 MAP kinase. It compares the dominant changes observed in the 134 structural coordinate sets that are known for

this protein, to molecular dynamics (MD) and elastic network model (ENM) in silico experiments. ENM is shown to sample the observed structural diversity well, compared to an MD simulation improved by inclusion of small organic molecules. The work describes the role of global modes in ligand binding, and suggests improvements to flexible-enzyme docking algorithms.

New conformational search method using genetic algorithm and knot theory for proteins

Authors: Yoshitake Sakae, Tomoyuki Hiroyasu, Mitsunori Miki, Yuko Okamoto

In this article, a previously described conformational search and sampling method for biomolecules is used as a base to propose a new strategy. The authors combine parallel simulated annealing using genetic crossover and knot theory to generate putative protein structures. While the previous parallel simulated annealing and genetic crossover method led to global-minimum energy protein conformations that had bad conformational properties ("knots"), the new method is shown to perform well on protein G.

Structural insights into pre-translocation ribosome motions

Authors: Samuel C. Flores and Russ Altman

Cryo-EM reconstruction and a dynamic model are used in this study to provide structural information on pre-translocation ribosome motions. The entire T.Thermophilus 16S and 23S rRNAs and most of the r-proteins are fitted in the cryo-EM map of the E.coli ribosome in the hybrid state. The fitted model exhibits a contact between P/E site tRNA and the head domain that was predicted by coevolution; it also recovers the intersubunit bridges known to be maintained during the full transition. The rotation of 16S with respect to 23S, and of the head domain with respect to the body, is modeled subject to the constraints that the ribosome pass through three experimentally observed conformations and maintain the head-tRNA contact. The results show that it is geometrically and sterically feasible for the head and tRNA to move in a coordinated fashion, and for a controversial experimentally observed intermediate to be sampled in the course of the motion. The method is applicable to the study of other large complexes.

2.2. *Tutorial*

2.2.1. *Survey Talk*

One of the emerging challenges of modern computational biology is the modeling of large biosystems. The understanding of these systems is key to solving many fundamental biological problems. Our session is focused on multi-scale techniques from molecule studies to cell or organism level. The emphasis is on: structure prediction, dynamics and sampling, mesoscale modeling, molecular assemblies, aggregation and analysis of structural interactions. These approaches may consist of a wide range of tools such as force fields, sampling, structure prediction, and dynamics methods.
In the first part of the tutorial session, we will address the issues and background relevant to the multi-resolution modeling of biological systems. We will start with an example of multi-scale modeling of human heart. After that, the following three topics will be briefly discussed: (1). Molecular Assemblies, Aggregation and Analysis of Structural Interactions. (2) Conformational sampling. (3). Multi-scale modeling of RNA structures.

2.2.2. *Introduction to RNABuilder*

In recent years we have seen an explosion in newly discovered RNA functions in the cell. However as mentioned our understanding of mechanisms of action has been hampered by a lack of structural information -- RNA's large size, flexibility, charge, folding time, and propensity for kinetic trapping challenges both experimental and computational probes of structure. The major thrust of our session is to demonstrate techniques to progress towards the mesoscale in biocomputation. To that end I will describe RNABuilder, a multi-resolution, internal coordinate dynamics code. It gives the user control over the flexibility, sterics, and forces acting on the molecule. It will be shown how this approach can be used to fold moderate-sized RNAs, or model the dynamics of large protein-RNA complexes, in a minutes to hours on a single processor. The presentation includes a short demo to show how to model a simple system.

3. Acknowledgements

Julie Bernauer acknowledges support from INRIA. Xuhui Huang acknowledges support from the Hong-Kong University of Science and technology.
The sessions chairs would like to thank all the authors and anonymous reviewers who contributed to the session.