Pacific Symposium on Biocomputing 15:200-204(2010)

MULTI-RESOLUTION MODELING OF BIOLOGICAL MACROMOLECULES

SAMUEL FLORES, JULIE BERNAUER, XUHUI HUANG, RUHONG ZHOU, AND SEOKMIN SHIN

1. Introduction

The field of molecular modeling has long recognized modeling the folding, assembly, and long-time dynamics of large macromolecules as its biggest challenge, since it is precisely at large size and long time scales that computational methods are most taxed. In response, many methods have been developed to reduce the computational cost by coarsening the granularity of the problem. Inevitably accuracy remains limited for these methods and so in recent years a consensus has emerged that we should work at more than one level of resolution, often simultaneously, in an approach known as multi-resolution modeling. We propose to organize a session on multi-resolution approaches to predict and analyze macromolecular structure, assembly and dynamics. This session will focus on state-of-the-art methodological developments and applications at different levels of molecular organization. Research directions are based on developments arising in the community field of the organizers, like the ones developed in the NIH Center for Biomedical Computation at Stanford University. In this proposal, the emphasis is made on integrative techniques for analysis of molecular structure and organization. Bridging the gap between computer science and structural biology, we represent an emerging community, presenting new efficient representations for the prediction and analysis of molecular structures and dynamics.

2. Session Summary

.
This session includes an invited talk, seven reviewed oral presentations, three additional accepted papers, a discussion session, and a tutorial. We are pleased to have the distinguished Professor Michael Levitt as our invited keynote speaker. Ron Levy will preside at our discussion session.

3.1 *Oral presentations*

Multi-resolution Modeling of Biological Macromolecules toward understanding allosteric signaling mechanisms in the ATPase domain of molecular chaperones
Authors: Ying Liu and Ivet Bahar

This paper represents a biologically important problem: the allosteric signaling pathways of molecular chaperones, HSP70. In order to investigate this problem, the authors adopt a multi-resolution approach by combining methods at different resolutions: sequence alignment, residue contact map, Gaussian network model, and mutual
information. They show that a subset of central residues located at the interface between the two lobes of the Nucleotide Binding Domains near the nucleotide binding site form a putative communication pathway invariant to structural changes.

Multi-resolution and multi-physics approach for interactively locating functionally linked ion binding sites by steering smal molecules into electrostatic potential maps using a haptic device
Authors: Olivier Delalande, Nicolas Ferey, Benoist Laurent, Marc Geroult, Brigitte Hartmann, and Marc Baaden

In this work a haptic device is cleverly used to find ion binding sites on a protein. The user experiences tactile feedback which provides a quick and intuitive assessment of binding affinity.

3D-BLAST: 3D protein structure alignment , comparison, and classification using spherical polar

Fourier correlations
Authors: Lazaros Mavridis and David Ritchie.

The authors represent protein atomic density using spherical harmonics and Laguerre-Gaussian radial functions and find that the lowest-order terms can be used to quickly scan a database of structures and evaluate shape similarity. The results suggest a structural search tool fast enough to be implemented on a web server is within reach.

Structural prediction of protein-RNA interaction by computational docking with propensity-based statistical potentials
Authors: Laura Pérez-Cano, Albert Solernou, Carles Pons, Juan Fernández-Recio

Relatively little work has been done in protein-RNA docking, possibly due to the paucity of solved complexes. In this work a coarse-grained residue-nucleotide potential is devised which is then used to discriminate the native complex from decoys generated by rigid-body rotations and translations of the molecules.

Multi-resolution Modeling of Biological Macromolecules Multiscale dynamics of macromolecules using Normal Mode Langevin
Authors: Jesus A. Izaguirre, Christopher R. Sweet, and Vijay S. Pande

In this paper, the authors introduce a new approach that has addressed an important issue in multi-scale modeling of biological macromolecules: how to project out fast motions of biological macromolecules in an automatic way and explicitly propagate only slow degrees of freedom. In this way, one can achieve a speedup in simulations without sacrificing accuracy. In order to achieve this goal, the authors use the coarse-grained normal mode analysis to decompose the dynamics of the biological macromolecules into slow and fast modes. Only slowest degrees of freedom are then explicitly propagated by a Langevin integrator. They demonstrate the power of
the new method by folding of the Fip35 mutant of WW domain. This method is promising in multi-scale simulations of biological macromolecules.

Constructing multi-resolution markov state models (MSMS) to elucidate RNA hairpin folding mechanisms
Authors: Xuhui Huang, Yuan Yao, Gregory Bowman, Jian Sun, Leonidas Guibas, Gunnar Carlsson, and Vijay Pande

Most multi-scale modeling approaches are based on multi-scale representation of the molecules. These approaches are normally limited by their accuracy by leaving out important degrees of freedoms. In this paper, the authors introduce a completely different multi-scale approach by generating multi-resolution Markov State Models (MSMs) to analyze dynamic atomistic simulation data. MSMs are a powerful tool to predict long timescale dynamics from many short simulations; however, previous constructions of MSMs are all focused on a single resolution. The key insight of their new algorithm Super-level-set Hierarchical Clustering (SHC) is to generate a set of super levels covering different density regions of phase space, and then cluster each super level separately. The authors demonstrate the power of the new algorithm using the folding of a small RNA hairpin. This new algorithm holds its promise to bridge the timescale gap between simulations and experiments, and also opens up a new generation of algorithms to treat the dynamics at multi-resolutions.

3.2 *Papers without oral presentation*

Predicting RNA structure by multiple template homology modeling
Authors: Samuel Flores, Yaqi Wan, Rick Russell, and Russ Altman.

This is the first published application of RNABuilder, a multi-purpose macromolecular modeling code

which uses the Simbody internal coordinate mechanics library. The authors show that a "Frankenstein" molecule comprised of ribozyme fragments from two different species can be used as a template for modeling the structure of a ribozyme from a third. In regions for which no template is available, base pairing forces are used to enforce known contacts. The results show how threading can be done for highly interconnected RNAs with relatively low sequence similarity to molecules of known structure.

Insights into the intra-ring subunit order of TRIC/CCT: a structural and evolutionary analysis
Authors: Nir Kalisman and Michael Levitt

The TRiC complex in eukaryote is composed of 8 different protein products arranged in a ringed structure. The order of the subunits in the ring is still unknown despite its important functional role. In this work we mapped evolutionary information of TRiC onto a structural hypothesis, which led us to formulate several restrictions on the possible ring arrangements. We conclude that 72 ring arrangements (out of the possible 7!=5040) are consistent with those restrictions.

3.3 *Tutorial*

The tutorial aims to address emerging multi-resolution approaches to predict and analyze structure, assembly and dynamics of biological macromolecules. In particular, the following three topics will be discussed: (1) Multi-scale modeling of RNA structures. (2) Macromolecular assembly prediction and analysis, and (3) Conformational sampling.

Multi-scale modeling for RNA structures. Turning the secondary structures of a RNA molecule into 3-dimensional RNA structure can be done with the assistance of various computer programs. For example, Nucleic Acid Builder (NAB) is a modeling language with utilities to create, analyze and manipulate molecular structures. Visual Molecular Dynamics (VMD) is a popular molecular viewer which permits the user to drag atoms or molecules using a mouse. It is also possible to connect to a molecular dynamics package and apply forces with a mouse, a process called Interactive Molecular Dynamics (IMD). Multi-resolution techniques could potentially improve such tools. For example, the user could be empowered to selectively rigidify parts of the molecule, or to apply forces to specific internal coordinates such as torsion angles, or to enforce geometric constraints between molecular subunits such as nucleotide bases. One such tool, RNABuilder, is introduced in this session.

Macromolecular assembly prediction and analysis. The function of biological macromolecules often relies on their interactions with one or many partners. These interactions tend to deform the macromolecules to better adapt their binding partners. This process involves not only shape complementarities but also chemical interactions. For example, protein-protein docking prediction techniques usually include two steps: finding putative complex conformations and scoring them to keep the most biologically relevant. Finding suitable conformers involves large conformational sampling, and it is difficult to be achieved at atomic level for large assemblies. Thus, multi-resolution approaches are necessary to be introduced to deal with large complexes. Such multi-resolution approaches have already shown to be the promising by successfully predicting individual side-chain contacts at the interface in the last CAPRI, the worldwide competition on docking of protein complexes. Furthermore, in the past few years, new algorithms dealing with multiple scales of representations and multiple levels of experimental data have emerged in the docking community for detecting interfaces, sampling and scoring the exploration results. In the tutorial, we will discuss the challenges on developing new multi-resolution techniques for conformation selection of protein-protein docking. For example, multi-scale scoring functions based on geometric criteria will be discussed.

Conformational sampling. Conformational sampling is one of the major changelings for the multi-scale modeling of the biological macromolecules due to the rugged nature of the free energy landscapes. Without adequate sampling, it is impossible to validate the parameters or force fields or to address phenomena that occur on biologically relevant timescales. Many methods have been developed in an attempt to address the sampling problem such as Generalized Ensemble (GE) algorithms like Replica Exchange Method (REM) and Simulated Tempering (ST), Metadynamics, Transition path sampling, and

Adaptive seeding method. Some sampling algorithms such as Resolution Replica Exchange can even directly integrate simulations running at different resolution and greatly enhanced their ability to sample the conformational space.

3. Acknowledgements

We gratefully acknowledge session funding from INRIA, the Stanford Simbios Center, and IBM Research. Samuel Flores and Xuhui Huang acknowledge support from the NIH Center for Physics-Based Simulation of Biological Structures (Simbios) with NIH Roadmap U54 GM072970. Xuhui Huang also acknowledges support from the Hong Kong University of Science and Technology. The session chairs thank the many anonymous referees who read submissions and made recommendations.