Mathematics and Computer Science in Modeling and Understanding of Structure and Dynamics of Biomolecules

Józef Adam Liwo,
Faculty of Chemistry, University of Gdańsk,
Wita Stwosza 63, 80-308 Gdańsk, Poland
Gabriel del Río Guerra,
Instituto de Fisiología Celular, Universidad Nacional Autónoma de México,
Circuito Exterior s/n Ciudad Universitaria, Coyoacán, 04510 México D.F.
François Major,
Département d'Informatique et de Recherche Opérationnelle
Université de Montréal C.P. 6128, succ. centre-ville Montréal, Québec, Canada H3C 3J7

August 9, 2019–August 11, 2019

1 Overview of the Field

Mathematical methods are nowadays extensively used in life sciences. On one hand, they are used to construct models of biological structures and processes at all levels of description, from structure and dynamics of biological macromolecules and their assemblies, through the structures and processes occurring in organelles, cells, organisms to modeling populations and whole ecosystems, as well as evolution. On the other hand, numerical mathematics and statistics provide tools to process sophisticated experimental data, extract correlations from databases, as well as simulate the structure and dynamics of the systems studied at different levels of resolution.

The workshop was focused on modeling biological molecules and their dynamics. The scope of this research comprises prediction of protein and RNA structure from sequence alone or in data-assisted mode, modeling the structure of protein assemblies, protein-protein docking, global-optimization methods in relation to these tasks, modeling biomolecule functions, and evolution of biomolecules. This research is of key importance not only because of the continuous quest of mankind to understand the world but also for purely utilitarian reasons such as, e.g., fighting cancer and hereditary diseases or producing better and more resistant crops. Prediction of the structures of biomolecules, mainly proteins, is of utmost importance in this field, because the experimental methods are not sufficient to provide the structures of the newly-discovered proteins. Therefore, the Community Wide Experiments on the Critical Assessment of Techniques for Protein Structure Prediction (CASP) [1] are organized since 1992 to assess the respective methods which, in turn, helps the biological chemists to select the most reliable ones for their research.

2 Recent Developments and Open Problems

In the last decade, enormous progress was made in the field of molecular modeling. Simulations of protein folding at all-atom level became possible owing to use of distributed computing and construction of dedicated supercomputers [2], and GPU computing [3]. Coarse grained models, in which atoms are grouped into united interaction sites have also been developed, which enable us to extend the simulation time- and size-scales by orders of magnitude [4]. Along with the development of models and simulation techniques, efficient data analysis approaches such as, e.g. principal component analysis, support vector machines, and neural network-based techniques have been developed. Another branch are the modeling methods that use heavy bioinformatics (database) input [5]. These methods integrate database information into modeling by using a variety of approaches; recently the Deep Learning-based methods achieved the top score in CASP13 [6].

Despite these successes, the computational approaches to biomolecular modeling still need improvement, Force fields, both all-atom and coarse-grained, can reproduce the experimental structures of only some proteins and nucleic acids. The availability of low-resolution experimental data (SAXS, CryoEM and ambiguous nuclear magnetic resonance data) pose new challenges to include them in modeling. New methods are also required to analyze the growing amount of experimental data. These topics were addressed during the workshop.

3 Presentation Highlights

The aim of this workshop was to bring together mathematicians, chemists, biologists, computer scientists, physicists so that they could exchange ideas and share experience. There were 22 participants total (out of initially confirmed 25, because of three last-minute cancellations), including‘ the organizers, all presenting a lecture, a talk, a short talk or a poster. The talks were divided into the following four thematic sessions: (1) *Protein structure and function*, (2) *Topology, disorder, and allostery*, (3) *Structure-function prediction*, (4) *Models and algorithms*, and the opening lecture by Robert Jernigan entitled *The importance of correlations in biology*. The poster presentations concerned coarse-grained modeling of protein structure and dynamics.

In his opening lecture, Robert Jernigan described the new protein-sequence alignment algorithm that he developed. The new algorithm utilizes tertiary structure information on the proximity of the amino acids and cab relate directly protein genome-related data to protein functions, based on accurate protein sequence alignment. Evolutionary relationship between proteins was also covered by Banu Ozkan in her talk on evolutionary aspects of allostery. Allosteric properties of proteins are important for protein function and are the next frontier in structural modeling of macromolecular interactions.

Many talks addressed ab initio and data-assisted modeling of the structures of biological macromolecules. Gregory Chirikjian described a new algorithm that he recently developed to retrieve original structure based on combined cryoelectron microscopy (CryoEM) and small X-ray electron scattering (SAXS) data. He showed that, even with large sections of data missing, the combined approach results in reasonable image restoration, not possible when only data of one kind are used. The method has been tested on model examples by now. Practical protein-structure determination from CryEM data was discussed by Daisuke Kihara who has developed a de novo modeling tool named MAINMAST (MAINchain Model trAcing from Spanning Tree), which enables us to obtain models with a very decent resolution. Marek Cieplak presented a talk on coarse-grained studies of the folding of intrinsically-disordered proteins, demonstrating that knots are formed in the process. Ensemble-based modeling of intrinsically-disordered proteins with the aid of ^{13}C chemical-shift data was presented by Yi He. Nina Pastor presented ta talk on modeling the effect of phosphorylation on the conformation of intrinsically disordered proteins, demonstrating that phosphorylation leads to increasing the content of α-helical structure.

A very interesting talk on accurate ab initio modeling of RNA structure with his simRNA approach that uses coarse-grained statistical potentials was presented by Michał Boniecki. Agnieszka Karczyńska presented protein-structure prediction method with the use of coarse-grained UNRES force field and input from bioinformatics approaches; this method achieved some success in the CASP13 experiment. Ilya Vakser presented a talk on his method of comparative modeling of protein complexes, which performed very well in CASP13. Use of machine learning to protein docking was presented by Marcelino Arciniega.

The topic of protein function was covered by Changbong Hyeon in his talk of cost-precision trade-off

and transport efficiency of molecular motors. Using the uncertainty relation, he demonstrated that transport efficiencies of the molecular motors are optimized under experimental conditions. An interesting computational model of kinetochore-microtubule attachments in relation to modeling the process of cell mitosis was presented by Tamara Bidone.

4 Outcome of the Meeting

As pointed out by many participants, many new scientific contacts were made that are likely develop into collaborations. The participants learned about new methods and could talk in person to their developers. On the other hand, method developers could hear from simulation practitioners in what directions to develop their methodologies.

References

[1] http://www.predictioncenter.org/

[2] K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana, D. E. Shaw, Structure and dynamics of an unfolded protein examined by molecular dynamics simulation, J. Am. Chem. Soc. 134 (2012) 3787–3791.

[3] M. S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. Legrand, A. L. Beberg, D. L. Ensign, C. M. Bruns, V. S. Pande, Accelerating molecular dynamic simulation on graphics processing units, J. Comput. Chem. 30 (2009) 864–872.

[4] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. Dawid, A. Koliński, Coarse-grained protein models and their applications, Chem. Rev. 116 (2016) 7898–7936.

[5] A. Tramontano, *The ten most wanted solutions in protein bioinformatics*; Taylor & Francis, Boca Raton, 2005; pp 69–88.

[6] L.A. Abriata, G.E. Tamo, M, Dal Peraro. A further leap of improvement in tertiary structure prediction in CASP13 prompts new routes for future assessments, Proteins: Struct. Func. Bioinf., (2019) (Early View) DOI:10.1002/prot.25787