Mini-symposium BIOMATH
Formal methods in the eyes of biological data

Résumé

The development and application of formal methods to biological problems implies a non- standard assessment of their validity. Indeed, in this context, the verifiers or checkers of the methods are the biological data. Formal models are used to understand the functioning and emergent properties of biological systems and they are revised/improved based on their agreement with the empirical observations obtained on these systems. They also produce data that predict what has not been yet observed and that cannot be obtained by current experimental methods. This mini-symposium highlights recents mathematical developments toward elucidating biological questions or modeling complex biological systems from this perspective.

Organisateur(s)

1. **Elodie Laine**, Laboratoire de Biologie Computationnelle et Quantitative (LCQB), UMR 7238, CNRS-UPMC.
2. **Alessandra Carbone**, Laboratoire de Biologie Computationnelle et Quantitative (LCQB), UMR 7238, CNRS-UPMC. Institut Universitaire de France.

Liste des orateurs

1. **Amaury Lambert**, Stochastics and Biology Group, Laboratoire de Probabilités et Modèles Aléatoires, UPMC Univ Paris 06 - SMILE Group, Center for Interdisciplinary Research in Biology, Collège de France.
   Titre : Chromosome painting.
2. **Laurent Jacob**, "Biométrie et Biologie évolutive", LBBE, UCB Lyon 1.
   Titre : Representing Genetic Determinants in Bacterial GWAS with Compacted De Bruijn Graphs.
3. **Sergei Grudinin**, NANO-D, INRIA Rhone-Alpes Research Center.
   Titre : Using machine learning and fast conformational space exploration techniques for some problems in structural bioinformatics.
4. **Alain Trouvé**, CMLA, Ecole Normale Supérieure
   Titre : Local equilibrium in molecular dynamics : should we believe what we see.
5. **Carole Frindel**, CREATIS, INSA-Lyon, Université de Lyon, INSERM-CNRS
   Titre : Simulation of images for a defined biological purpose.

Elodie Laine, Sorbonne Universités, UPMC University Paris 06, CNRS, IBPS, UMR 7238, Laboratoire de Biologie Computationnelle et Quantitative (LCQB), 75005 Paris, France, `elodie.laine@upmc.fr`
Alessandra Carbone, Sorbonne Universités, UPMC University Paris 06, CNRS, IBPS, UMR 7238, Laboratoire de Biologie Computationnelle et Quantitative (LCQB), 75005 Paris, France, Institut Universitaire de France, `Alessandra.Carbone@lip6.fr`

1 Amaury Lambert : Chromosome painting

We consider the most simple stochastic model of neutral population genetics with recombination : 1) We assume that at time 0, all individuals of a haploid population have their unique chromosome painted in a distinct color ; 2) At rare birth events, due to recombination modeled as a single crossing-over, the chromosome of the newborn is a mosaic of its two parental chromosomes. When t is large, after fixation has occurred, all individuals end up with the same mosaic of colors on their chromosome. How does this mosaic look like ? We will display mathematical results and will show massive simulations as well as data from evolutionary experiments on *C. elegans* in Henrique Teotonio's lab (ENS, Paris). This is joint work with Emmanuel Schertzer and Veronica Miro Pina.

2 Laurent Jacob : Representing Genetic Determinants in Bacterial GWAS with Compacted De Bruijn Graphs

Antimicrobial resistance has become a major worldwide public health concern, calling for a better characterization of existing and novel resistance mechanisms. GWAS methods applied to bacterial genomes have shown encouraging results for new genetic marker discovery. Most existing approaches either look at SNPs obtained by sequence alignment or consider sets of kmers, whose presence in the genome is associated with the phenotype of interest. While the former approach can only be performed when genomes are similar enough for an alignment to make sense, the latter can lead to redundant descriptions and to results which are hard to interpret. We propose an alignment-free GWAS method detecting haplotypes of variable length associated to resistance, using compacted De Bruijn graphs. Our representation is flexible enough to deal with very plastic genomes subject to gene transfers while drastically reducing the number of features to explore compared to kmers, without loss of information. It accomodates polymorphisms in core genes, accessory genes and non coding regions. Using our representation in a GWAS leads to the selection of a small number of entities which are easier to visualize and interpret than fixed length kmers. We illustrate the benefit of our approach by describing known as well as potential novel determinants of antimicrobial resistance in Pseudomonas aeruginosa, a pathogenic bacteria with a highly plastic genome. A pre-print available at http ://biorxiv.org/content/early/2017/03/03/113563.

3 Sergei Grudinin : Using machine learning and fast conformational space exploration techniques for some problems in structural bioinformatics

Computer simulations in structural biology have become an integral part of current research activities. To elucidate some structure or function of biological systems, most often researchers use molecular dynamics simulations with classical forcefields. However, the application of these methods to large biomolecules still faces important practical difficulties due to the combinatorial explosion of possible interactions involved. There have been numerous attempts to reduce the size of the conformational space of a system under study that would increase the time of the simulations. Consequently, the reduced conformational space would require the adapted potential functions.

I my talk will demonstrate how machine learning and optimization in general can be used to design interaction potentials adapted to specific conformational space exploration problems. More precisely, I will present our recent results on the prediction of properties of small molecules, on the prediction of protein-protein and protein-drug interactions, as well as individual protein folds at atomic level. I will also present some methods for efficient space exploration including FFT-accelerated techniques. Some relevant publications can be found on our website at https ://team.inria.fr/nano-d/publications/.

4 Alain Trouvé : Local equilibrium in molecular dynamics : should we believe what we see ?

Molecular dynamics (MD) simulations can produce nowadays huge amount of data using high-throughput CPU/GPU clusters. However, we believe that the systematic use of MD for the study the large molecules of real biological systems is still largely impeded by a lack of adequate modeling and understanding of the produced highly complex signals that emerge at the level of the relevant subsystems for various time scales. We will present here an ongoing work towards the dynamics modeling and detection of local equilibrium for relevant subsystems compatible with the usual practice of MD and aiming at avoiding the detection of spurious artefactual local equilibriums. Such well characterized local equilibriums would be basic descriptive atoms extracted from various MD trajectories. Joint work with Sharad Goulam (PhD Student) and Luba Tchertanov.

5 Carole Frindel : Simulation of images for a defined biological purpose

Our approach is to develop new computer tools for images simulation with a clearly defined medical (or biological) objective [1, 2, 3, 4, 5]. These can include diagnostic and prognostic software [2, 3] that exploits available images and data about the patient and his / her disease, or assessment of therapy through pre- and post-treatment imaging [4, 5]. This latter axis includes the development of imaging biomarkers to quantify the effect of a new drug [5]. Image simulation algorithms are based on mathematical (geometric, statistical), biological, and / or physicochemical models of living organisms on several scales in order to construct a partial numerical model of the patient ?s anatomy and physiology and its pathology.

Références

[1] G. Colson, S. Balmand, C. Boyer, M. Rosa-Calatrava, C. Frindel and D. Rousseau,, *Simulation dimages en microscopie électronique pour la détection de la contamination de cellules par un virus*, XXVI e GRETSI, Juan-Les-Pins, France September, 2017.

[2] M. Giacalone, C. Frindel, M. Robini, F. Cervenansky, E. Grenier and D. Rousseau, *Robustness of spatiotemporal regularization in perfusion MRI deconvolution : An application to acute ischemic stroke*, Magnetic Resonance in Medicine January, 2017.

[3] M. Giacalone, C. Frindel, M. Robini, E. Grenier and D. Rousseau, *Simulations numériques en IRM de perfusion cérébrale : bruit log-ricien ou bruit gaussien ?*, XXV e GRETSI, Lyon, France September, 2015.

[4] H. Rositi, C. Frindel, M. Wiart, C. Olivier, F. Peyrin and D. Rousseau, *Analyse en échelles pour l'optimisation du paramètre de reconstruction en tomographie X de phase ; application à l'imagerie du cerveau du petit animal*, XXV e GRETSI, Lyon, France September, 2015.

[5] H. Rositi , C. Frindel, M. Langer, M. Wiart, C. Olivier, F. Peyrin, D. Rousseau, *Information-based analysis of X-ray in-line phase tomography with application to the detection of iron oxide nanoparticles in the brain*, ptics Express, vol. 21, no. 22, 2013, pp. 27185-27196.