Università della Svizzera italiana

**Faculty of Informatics**

# Workshop on "Computational Biology"

March 4th, 2014 – Room SI-006 (Informatics building)

**08:00** Computational solutions for disentangling genomic sequence complexity
*Maria Anisimova (ETH Zurich, Switzerland)*

**08:45** Computational Design of novel functional proteins - case studies in vaccine design
*Bruno Correia (The Scripps Research Institute, USA)*

***09:30*** *Coffee Break*

**09:45** Computers in biology: from atom to multiscale simulations
*Vittorio Limongelli (University of Naples, Italy)*

**10:30** Ecology and Evolution of Host-Microbe Interactions
*Norman Pavelka (Singapore Immunology Network, A*STAR, Singapore)*

**11:15** New routes towards efficient simulations of the dynamics of biomolecular complexes
*Fabio Pietrucci (EPFL, Switzerland)*

USI - Faculty of informatics
Via Giuseppe Buffi 13, CH-6904 Lugano
tel. +41 58 666 46 90, fax +41 58 666 45 36, e-mail decanato.inf@usi.ch, www.inf.usi.ch

# Computational solutions for disentangling genomic sequence complexity

***Maria Anisimova (ETH Zurich, Switzerland)***

***Abstract:***

Advances in high-throughput technologies are transforming molecular biology and its applications to human health and sustainable agriculture into quantitative disciplines. Computational science plays an increasingly important role in the integration and analysis of large complex genomic data, as there is an urgent need for more accurate and powerful methods to disentangle the observed biological complexity.

In this talk, I will present an overview of several of our most recent projects that contribute to modeling complex sequence features. First, I will discuss new methods for annotating, aligning and phylogenetic/ functional studies of sequences with tandem repeats (TRs). Imperfect TRs with different number of repeated units cause serious problems for sequence analysis; in genome assembly, sequence alignment and all downstream analyses. Moreover, TRs are often found in proteins with fundamental biological functions, in virulence and resistance related genes, and those associated with infectious and neurodegenerative diseases. We developed statistical methods for: (a) Detection and validation of TRs in genomic sequences, (b) Probabilistic graph-based alignment of sequences with TRs, and (c) Phylogenetic method to study conservation of TR unit number and order. The methods were applied to entire proteomes of all available plants and animals. We observed an unusually high conservation of TRs, sometimes dating back to the common ancestors of human and yeast, or green plants and red algae. Such deep conservation implies importance of these TRs in key protein functions.

I will conclude my talk with future projects focusing on other complex molecular processes: modeling somatic hypermutation during the maturation of antibodies, and detecting silent (amino-acid preserving) mutations that affect genetic fitness, often in genes associated with cancers and diabetes.

***Biography:***

Dr. Maria Anisimova develops computational techniques to study how genomic changes affect phenotype, and more recently, started developing bioinformatics methods with applications in immunology and helping to understand genetic associations with disease.

Since obtaining her PhD in statistical genomics at University College London, Dr. Anisimova has contributed several widely used methods in evolutionary genomics, in particular to detect adaptive signal in protein-coding genes and to enable large-scale phylogenomic analyses. She has worked across borders—in the UK, France, Russia and Switzerland—and has a strong interdisciplinary network of collaborators across the globe.

While at the ETH Zürich, Dr. Anisimova's research activities were supported by grants from SNSF, SATW, ETH Zurich and SIB. She has supervised PhD and master students, several of whom received competitive awards for their research. Since 2014, she has been appointed at the Zurich University of Applied Sciences (ZHAW) to develop a new research programme in computational biology.

As world expert in phylogenomics and evolutionary genomics, Dr. Anisimova is an editor of several international journals in the field and regularly teaches at international courses in bioinformatics. In 2012, she edited the two-volume book "Evolutionary genomics: computational and statistical methods", Springer Verlag, an in-depth survey of the field with contributions by the leading scientists in the field.

# Computational Design of novel functional proteins - case studies in vaccine design

***Bruno Correia (The Scripps Research Institute, USA)***

***Abstract:***

The full capability to manipulate protein structure and function holds the promise to make transformative contributions in diverse scientific domains. Novel nanomaterials, enzymes and proteins with biomedical applications are all within the reach of protein engineers. Computational protein design methodologies have made critical contributions to enable structure-based design. Vaccine research could greatly benefit from new strategies to produce cheap and efficacious vaccines against pathogens that remain elusive to traditional approaches, becoming serious global health problems. Advances in the discovery of potent and/or broadly neutralizing antibodies (bnabs) are quickly uncovering sites of vulnerability in pathogens. Structural biology has revealed the atomic details of these neutralization epitopes. Computational methodologies have been developed to perform structure-based design of immunogens to re-elicit neutralizing antibodies against conserved epitopes, providing the foundation for epitope-focused vaccines. The designed immunogens mimic relevant epitope conformations recognized by bnabs and were named epitope-scaffolds (ES). I will present two methodologies for the design of epitope-scaffolds: Rosetta Multigraft and Rosetta Fold From Loops. Given that a large fraction of the neutralization epitopes encompass multiple protein segments, Rosetta Multigraft was conceived to perform the transplantation of multi-segment epitopes to protein scaffolds. I applied this method to a Human Immunodeficiency Virus (HIV) epitope and designed one ES holding two discontinuous structural segments. Here, we used computational design to guide in vitro evolution experiments and showed that this joint computational-experimental approach can accomplish challenging protein design problems.

The second methodology I will present, Rosetta Fold From Loops, was developed to fold and design novel proteins around functional sites of interest. Using this algorithm we designed ESs to mimic an epitope from Respiratory Syncytial Virus (RSV), a relevant vaccine target. Biophysical and structural characterization of the designs revealed high binding affinity to the target antibody and excellent structural mimicry of the viral epitope. To test whether the ESs could elicit RSV neutralizing antibodies we proceeded with macaque immunizations. Remarkably, neutralization activity was detected in 12 out of 16 macaques. In some animals the neutralization titers were comparable to those of natural infection in humans, which typically confer protection. These results provide a proof of principle for the use of ESs for vaccine development and support its potential to target other pathogens that remain elusive to vaccine development, like HIV and Influenza. More generally, the results support that computational protein design can play an important role in generating novel proteins relevant for biomedical purposes but also for basic research applications.

***Biography:***

I'm originally from Portugal and in 1998 I started my studies in chemistry at the Universidade de Coimbra. My passion for molecular modeling started during my undergraduate research while investigating the structural features of amyloid fibrils generated by the aggregation of Transthyretin. I was able to publish my first research paper as an undergraduate and won my first young investigator award from the Protein Society Symposium. In 2004 I was selected amongst 172 applicants to join the 1st Portuguese program in Computational Biology. In 2005 I joined the University of Washington – Seattle under the supervision of Bill Schief and David Baker to do my dissertation work. My research topics were focused on the development of methodologies for protein design and their application to immunogen design. In 2010 I completed my PhD, which was awarded by Universidade Nova de Lisboa. In 2011, I started my post-doctoral studies at The Scripps Research Institute under the supervision of Benjamin Cravatt. Ever since the focus of my research has been developing quantitative mass-spectrometry methods to characterize protein-small molecule interactions at the proteome level. In addition, I've been also developing web-based platforms to enable the access of the scientific community to large datasets of protein-small molecule interactions, which can be critical to provide new leads for small-molecule development.

# Computers in biology: from atom to multiscale simulations

***Vittorio Limongelli (University of Naples, Italy)***

***Abstract:***

Computer simulations have increased over the years their role in studying biologically relevant phenomena. However, in-silico models still suffer of some limitations such as the size of the simulated systems and the time scale of the investigated processes. Here, I illustrate how such limitations can be overcome using the most advanced computational techniques in a number of cases.

We first demonstrate how one can use atomistic simulations to describe the binding of ligands to protein and design new drugs. We focus on the G-protein coupled receptor of bile acids GP-BAR1, which is a member of a large receptor family, GPCRs, involved in many diseases and targeted by approximately 40% of marketed drugs. Despite their biological relevance, the ligand binding mechanism to these receptors is poorly understood owing to the limited structural data available. Combining homology modeling, molecular docking and molecular dynamics simulations we have elucidated the binding mechanism of a potent ligand (1). This understanding has allowed designing the most potent inhibitor so far discovered.

The methods used in this study are however based on approximated estimation of the binding free energy, while this information is crucial to improve computer-aided drug discovery (2,3). To this end, we have developed a new method called funnel-metadynamics (4), that allows describing the binding path of a ligand to its target and leads to an accurate estimate of the binding free energy. This new approach has been successfully used in many ligand/protein binding studies (e.g. cyclooxygenase enzymes) and is suitable to study more complex cases such as peptide/membrane and protein/protein interaction.

I will also illustrate how long time-scale processes, such as protein and DNA folding, can be described using enhanced sampling techniques. In the present case, we have identified a new DNA structural motif, named "G-triplex", through metadynamics unfolding simulations on the Gquadruplex DNA aptamer TBA (5). After this important discovery I have coordinated the efforts of a group of experimentalists that has been able to confirm my hypothesis. The abundance of guanine-rich regions in human genome, able to form G-triplex, makes this discovery of great interest for future studies.

In an ideal situation, the target structures should be simulated in their biological context. However, the size and complexity of the physiological environment represent a major hurdle for computational approaches. Here, I present an important advance in this field by a very recent protocol that combines multiscale and enhanced sampling methods (coarse-grained/metadynamics) to simulate long time-scale events in very large systems (6). Using such approach the dimerization process in membrane of the transmembrane helices of epidermal growth factor receptors has been elucidated, computing also its free energy profile. This protocol allows reaching millisecond time scale, opening new opportunities to study larger protein clusters in membrane (e.g. GPCRs, ion channels) and more complex systems like antigen/antibody interaction.

*References*

1. Limongelli V and coworkers J. Med. Chem. 2014, 57, 937-954.
2. Limongelli V et al. Proc. Natl. Acad. Sci. USA 2010, 107, 5411-5416.
3. Limongelli V et al. Proc. Natl. Acad. Sci. USA 2012, 109, 1467-1472.
4. Limongelli V, Bonomi M, Parrinello M. Proc. Natl. Acad. Sci. USA 2013, 110, 6358-6363.
5. Limongelli V et al. Angew. Chem. Int. Ed. Eng. 2013, 125, 2325-2329.
6. Lelimousin M, Limongelli V, Sansom MSP. Biophys. J. 2013, 104, 32.

***Biography:***

Vittorio Limongelli obtained his master's degree in Medicinal Chemistry with special commendation in 2004 and his PhD in Pharmaceutical Science in 2007 at University of Naples, Italy. He did his postdoc at ETH Zurich in the group of Prof. Parrinello. From 2010 he is Assistant Professor at University of Naples where he teaches "Analytical Chemistry" and "Advanced Methods in Medicinal Chemistry". His research focuses on the study of systems of biological interest through advanced computational approaches and the development of new methods like funnel-metadynamics, published on PNAS in 2013, and a yet unpublished protocol which combines coarse-grained molecular dynamics and metadynamics. His research led to more than 26 publications, 3 PNAS, 2 Angewandte Chemie, 2 JACS, and 14 scientific projects granted. His hindex is 13. In 2013 he was awarded the "Scrocco" prize from the Italian Chemical Society for "his original contribution to the study of biological phenomena using advanced computational methods". In 2014 he got the qualification to function as Associate Professor in Italian Universities.

# Ecology and Evolution of Host-Microbe Interactions

***Norman Pavelka (Singapore Immunology Network, A*STAR, Singapore)***

***Abstract:***

The immune system has evolved to recognize and distinguish different types of microbes, such as bacteria or parasites, and mount appropriate host responses to fight off infections. Using functional genomics approaches we interrogated host-microbe interactions from the host's perspective and developed novel algorithms and software for microarray and proteomics data analysis. This body of work led to the discovery of IL-2 production by dendritic cells (DCs) upon encounter with Escherichia coli bacteria and of the involvement of a type I interferon pathway in the interaction between DCs and the helminth Schistosoma mansoni. On the other end, microbes share the remarkable ability to adapt to a wide variety of challenging environments. We therefore begun to study adaptive evolution in the model eukaryote Saccharomyces cerevisiae and investigated its ability to evolve creative workarounds to the deletion of an essential gene required for cell division. By coupling laboratory evolution experiments and cell biology with integrated genomic-transcriptomic data analysis, we discovered an unexpected role for aneuploidy, i.e. the state of bearing an unbalanced number of chromosomes. We then integrated whole-genome re-sequencing, transcriptional profiling, mass-spectrometry based proteomics and high-throughput phenotypic profiling data on a large collection of isogenic wild-type aneuploid yeast strains and found this chromosomal aberration to be both required and sufficient to induce specific changes at the transcriptome, proteome and phenome level, some of which were shown to be beneficial especially under the most selective conditions. By applying phylogenetic reconstruction and network analysis to high-throughput karyotyping data from a large number of progeny derived from above-mentioned aneuploid yeast strains, we further showed that some aneuploid karyotypes, by creating unbalances between specific components of the mitotic machinery, could themselves generate chromosomal instability, i.e. the propensity of gaining and losing chromosomes at high frequency. Taken together, these integrative 'omics' studies uncovered a role for aneuploidy as both a driver and a catalyzer of adaptive evolution in yeast. We now hypothesize this to be a conserved evolutionary strategy shared across fungi and that it could play an important role in the ability of fungal pathogens to adapt to mammalian hosts through evolutionary processes.

Indeed, mammalian hosts represent complex ecological environments in which microbes continuously adapt, compete and evolve. Microbial virulence is one of the complex traits under evolutionary pressure that emerges as a result of this dynamic host-microbe interaction. In my lab, we aim to dissect the nature of the host-derived selective pressures that control fungal virulence in the mammalian gastrointestinal (GI) tract, as well as the evolutionary strategies that fungi can come up with to counteract these host-derived control mechanisms. Currently, we have set up mouse models of GI colonization and oral-faecal transmission of the human commensal and opportunistic pathogen Candida albicans, to study its virulence evolution when transmitted from host to host. Preliminary results support the hypothesis that prolonged exposure to the mammalian GI tract drives phenotypic and karyotypic diversity in initially clonal fungal populations, which might act as a substrate for further selection and evolutionary adaptation. By applying deep-sequencing-based microbiome analysis and whole-genome re-sequencing strategies coupled with genetic or chemical perturbations of the host, we are now beginning to dissect the role played by the host immune system and the GI microbiota as selective forces driving this evolutionary process.

***Biography:***

Dr. Pavelka received his Degree in Biotechnology from the University of Milano-Bicocca in 2001 and his Ph.D. in Immunology from the University of Rome "Tor Vergata" in 2006. He formerly worked as a graduate student in Prof. Paola Castagnoli's lab at the University of Milano-Bicocca, where he developed his microarray data analysis skills and contributed to numerous Functional Genomics projects aimed at understanding the role of dendritic cells in host-pathogen interactions. Dr. Pavelka then joined Prof. Rong Li's group at the Stowers Institute for Medical Research as a Postdoctoral Research Associate in May 2006, where he published seminal papers on the role of aneuploidy in evolution of yeast cells. While at the Stowers Institute, he collaborated with Dr. Michael Washburn, publishing important contributions to the proteomics data analysis field. In October 2010, Dr Pavelka was awarded the A*STAR Investigatorship (Biomedical Sciences) and set up his lab at the Singapore Immunology Network in May 2011, to study evolution of host-pathogen interactions and the genome changes that underlie such evolutionary processes.

# New routes towards efficient simulations of the dynamics of biomolecular complexes

***Fabio Pietrucci (Ecole Polytéchnique Fédérale de Lausanne, Switzerland)***

***Abstract:***

We witness a time of opportunity: experimental methods in biology are developing fast, disclosing a wealth of phenomena and interactions at the cellular and molecular levels. On the other side, atomic---resolution simulations of biological systems, traditionally limited in system size/time scale and in accuracy, are also sizably improving. This is due only in part to the continuous rise of computer performances: the development of powerful new algorithms is the key to bridge the gap between computational and experimental biology. I will discuss current limitations of computational biology at the molecular scale and present innovative approaches able to boost the field. In particular, I will consider enhanced sampling methods combining together molecular dynamics, biasing potentials and tools borrowed from graph theory. This class of methods lends itself to implementation on massively parallel computing facilities, and it is able to provide the detailed microscopic mechanism of a process as well as the associated thermochemical and kinetic observables that can be directly compared with experiments. The new approaches will allow tackling a number of challenging problems, such as the effect of drugs on protein---protein interactions, the discovery of cryptic binding sites in biomolecular complexes, and the aggregation of amyloids. The interplay between simulations and experiments is crucial and will be highlighted throughout.

***Biography:***

Fabio Pietrucci was born in Milan (Italy) in 1979. In 2003 he obtained a master's degree in materials science from the University of Milano---Bicocca. In 2006 he obtained a PhD degree from the European Doctoral School of Nanostructures and Nanotechnologies of the same university, under the supervision of Professor Marco Bernasconi. During his doctoral studies, in 2005 he joined for six months the Computational Science Group of Professor Michele Parrinello at the Università della Svizzera Italiana of Lugano. In 2006 he obtained a post---doctoral research grant at the International School for Advanced Studies (SISSA) of Trieste, in the group of Professor Alessandro Laio within the Statistical and Biological Physics Sector. In 2009 he moved as a post---doctoral researcher to the Ecole Polytéchnique Fédérale de Lausanne, first within the Centre Européen de Calcul Atomique et Moléculaire (CECAM) led by Professor Wanda Andreoni and since 2013 at the Institute of Theoretical Physics, where he was granted the title of teacher in charge.