DY 12: Focus Session: Statistical Physics-Based Methods in Molecular Evolution - organized by Alexander Schug and Martin Weigt (joint session BP/DY)

Time: Monday 15:00–17:00 Location: H 2013

Invited Talk DY 12.1 Mon 15:00 H 2013

Evolution of quantitative traits and non-equilibrium matrix ensembles — Simone Pompei, Torsten Held, and •Michael Lässig — Institute for Theoretical Physics, University of Cologne

Evolution affects molecular quantitative phenotypes, such as stability, binding affinities, and metabolic activities of cellular proteins. Linking sequence data to phenotypic and functional changes remains a critical gap in our understanding of evolutionary processes. In this talk, we present new methods to infer *a priori* unknown quantitative phenotypes from their correlation signature in time-resolved sequence data, using non-equilibrium statistical mechanics and random matrix theory. We use these methods to map the phenotypic evolution of the human influenza virus.

DY 12.2 Mon 15:30 H 2013

Big Data in Structural Biology: Predicting Protein and RNA Structures by inferring residue co-evolution — •Alexander Schug — John von Neumann Institute for Computing, Jülich Supercomputer Centre, Forschungszentrum Jülich

To gain any detailed understanding of biomolecular function, one needs to know their structure. The structural characterization of many important biomolecules and their complexes remains experimentally challenging. Novel statistical tools based on statistical physics such as Direct Coupling Analysis (DCA) take advantage of the explosive growth of sequential databases and trace residue co-evolution to infer secondary and tertiary contacts for proteins [1] and RNAs [2]. These contacts can be exploited as spatial constraints in structure prediction methods leading to excellent quality predictions [1,2,3]. Going beyond anecdotal cases of a few protein families, we have applied our methods to a systematic large-scale study of nearly 2000 PFAM protein families of homo-oligomeric proteins [4]. Also, we can apply DCA to infer mutational landscapes by capturing epistatic couplings between residues and can assess the dependence of mutational effects on the sequence context where they appear [5].

[1] Weigt M et al., PNAS (2009); F. Morcos et al., PNAS (2011)
[2] E. De Leonardis et al., NAR (2015)
[3] Schug A et al., PNAS (2009); Dago A et al., PNAS (2012)
[4] G. Uguzzoni et al., PNAS (2017)
[5] M. Figliuzzi et al., MBE (2016)

DY 12.3 Mon 15:45 H 2013

Coevolution based inference of allosteric architectures — •Barbara Bravi[1], Carolina Brito[2], Riccardo Ravasio[1], and Matthieu Wyart[1] — [1]Institute of Theoretical Physics, Ecole Polytechnique Fédérale de Lausanne, Switzerland — [2]Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

We analyze maximum entropy approaches to infer the functional design of elastic materials exhibiting allostery, i.e. the property of highly specific responses to ligand binding at a distant active site. To guide and inform protocols of de novo drug design, it is fundamental to understand what architectures underlie such a transmission of information and whether their features can be predicted from sequence data alone. We consider the functional designs of in silico evolved allosteric architectures which propagate efficiently energy (including shear, hinge, twist) or strain (resulting in a less-constrained trumpet-shaped region between the allosteric and the active site). We show that maximum entropy approaches, built to capture statistical properties such as conservation and correlations, can provide predictive information on the cost of single and double mutations while their performance at reproducing the original allosteric fitness is strongly design-dependent. We benchmark existing maximum entropy inference methods on these computationally evolved functional architectures and we propose an improved framework accounting for a multiplicity of co-evolutionary factors which is aimed at disentagling allostery-based correlations from extrinsic ones.

DY 12.4 Mon 16:00 H 2013

Architecture of allosteric materials — Carolina Brito[1], Solange Flatt[2], •Riccardo Ravasio[2], Matthieu Wyart[2], and Le Yan[3] — [1]Universidad Federal do Rio Grande do Sul, CP 15051, 91501-970 Porto Alegre RS, Brazil — [2]Ecole Polytechnique Federale de Lausanne, CH-1015, Lausanne, Switzerland — [3]Kavli Institute for Theoretical Physics, Santa Barbara, CA 93106, USA

Allosteric proteins transmit a mechanical signal induced by binding a ligand. However, understanding the nature of the information transmitted and the architectures optimising such transmission remains a challenge. We show using an in-silico evolution scheme and theoretical arguments that architectures optimised to be cooperative, which propagate efficiently energy, qualitatively differ from previously investigated materials optimised to propagate strain. Although we observe a large diversity of functioning cooperative architectures — including shear, hinge and twist designs, they all obey the same principle of nearly displaying a mechanism, i.e. an extended zero mode with a predicted optimal frequency. Overall, our approach leads to a natural explanation for several observations in allosteric proteins, and suggests a path to discover new ones. On this line, we study the extended soft modes of hessians defined from 46 couples of proteins for which the active and inactive structures are available and compare them with the aforementioned principle. Moreover, the set of architectures that are evolved through the in-silico scheme defines a well controlled ground where to benchmark the results of co-evolutionary methods, usually applied to protein sequences.

DY 12.5 Mon 16:15 H 2013

Direct Coupling Analysis on the genome scale — •Erik Aurel — Royal Institute of Technology in Stockholm, Sweden

Direct Coupling Analysis (DCA) is a powerful tool to find pair-wise dependencies in large biological data sets. It amounts to inferring coefficients in a probabilistic model in an exponential family, and then using the largest such inferred coefficients as predictors for the dependencies of interest. A main success story has been predicting spatially proximate residues in protein structures from sequence data.

From a population genetics point of view DCA should be viewed as inferring epistasis, synergistic effects on fitness, from samples. I will discuss applications of DCA to the genome scale in bacteria and how that allows to find unexpected (and expected) dependencies between genes in trans i.e. that are not close on the genome.

This is joint work with many people, most recently with Chen-Yi Gao and Hai-Jun Zhou, available as arXiv:1710.04819.

DY 12.6 Mon 16:30 H 2013

Interprotein coevolution: bridging scales from residues to genomes — Giancarlo Croce[1], Thomas Gueudre[2], Hendrik Szurmant[3], Matteo Figliuzzi[1], and •Martin Weigt[1] — [1]Université Pierre & Marie Curie, Sorbonne Université, Paris, France — [2]Human Genetics Foundation, Turin, Italy — [3]Western University of Health Sciences, Los Angeles, USA

Interacting proteins coevolve at multiple but interconnected scales, from the residue-residue over the protein-protein up to the family-family level. The recent accumulation of enormous amounts of sequence data allows for the development of novel, data-driven computational approaches. Notably, these approaches can bridge scales within a single statistical framework [1,2], which is built upon idea from the inverse statistical physics [3,4]. While being currently applied mostly to isolated problems on single scales, their immense potential for an evolutionary informed, structural systems biology is steadily emerging.

[1] H. Szurmant and M. Weigt, Current Opinion in Structural Biology 50, 26-32 (2017).
[2] G. Croce, T. Gueudre, MV Ruiz Cuevas, H. Szurmant, M. Figliuzzi, M. Weigt, submitted (2017).
[3] S. Cocco, C. Feinauer, M. Figliuzzi, R. Monasson, M. Weigt, Rep. Prog. Phys. (2017), https://doi.org/10.1088/1361-6633/aa9965.
[4] H. Chau Nguyen, Riccardo Zecchina, Johannes Berg, Advances in Physics, 66 (3), 197-261 (2017)

DY 12.7 Mon 16:45 H 2013

The evolutionary consequences of population spread on curved surfaces — Daniel A. Beller[1], Kim M. J. Alards[2], Ricardo A. Mosna[3], Federico Toschi[2], and •Wolfram Möbius[4] — [1]Brown University, Providence, RI, USA — [2]TU Eindhoven, Eindhoven, The Netherlands — [3]Universidade Estadual de Campinas, Campinas, SP, Brazil — [4]University of Exeter, Exeter, United King-

dom

We investigate the evolutionary dynamics of populations growing and expanding on curved surfaces. Using a combination of individual-based simulations and theory we characterize the effect of individual features (cones and spherical caps) on the shape of the population front and the genetic composition of an expanding population. We find that, on sufficiently large scales, geodesics allow us to describe both population and evolutionary dynamics quantitatively. Using these findings, we characterize the consequences of large-scale surface roughness on genetic diversity and compare to the case of heterogeneous but flat environments.