

Editorial

Information Theory in Molecular Evolution: From Models to Structures and Dynamics

Faruck Morcos^{1,2,3} 

¹ Department of Biological Sciences, University of Texas at Dallas, Richardson, TX 75080, USA; faruckm@utdallas.edu

² Department of Bioengineering, University of Texas at Dallas, Richardson, TX 75080, USA

³ Center for Systems Biology, University of Texas at Dallas, Richardson, TX 75080, USA

Historically, information theory has been closely interconnected with evolutionary theory. The work of Ronald Fisher in population genetics [1] and the formulation of the principle of minimum Fisher information [2] are just two early examples of such connections. In recent years, with the advent of high-throughput sequencing technologies, the field of molecular evolution has been able to take advantage of large amounts of samples from evolution to improve models and applications to understand structural, dynamical, and functional aspects of biomolecules. Information metrics have been prevalent, in recent years, to estimate the likelihood that two amino acid sites in a protein are coevolving. A relevant example of such metrics is *Direct Information* (DI) [3,4] used in the context of Direct Coupling Analysis to estimate if two positions in a multiple sequence alignment are likely to be proximal in the 3D structure of a protein or RNA molecule. Other standard information metrics like *Mutual Information* have been applied and are particularly useful for the case of molecular complexes and interactions [5,6].

This special issue focuses on important aspects of the study of molecular evolution through the statistical features of sequence data, molecular simulation, and evolutionary convergence towards specificity in signaling networks. Three articles [7–9] investigate how phylogenetic relationships in sequence data have an effect in the inference procedure of a joint probability distribution $P(a_1, a_2, a_3, \dots, a_L)$ of a given sequence of length L . Particularly, these studies are centered under the premise that a preprocessing step for multiple sequence alignment analysis might reduce phylogenetic bias and could improve the inference procedure. These methods, ultimately, improve prediction of amino acid contacts and functional connections among amino acid sites.

In [7], Hockenberry et al. conducted a systematic study of previously relevant reweighting schemes that have been useful in other applications. These methods contrast versus the current practices of identity-based sequence reweighting used in Potts model inference. They find that previous applications do not add considerable value for the inference task and leave open the question for novel schemes that might improve the inference of coevolving residue pairs. Interestingly, in [8,9], the authors propose novel schemes to account for phylogenetic bias. First, Horta et al. [8] introduce a new inference method which uses a priori information about phylogeny to enhance contact prediction and fitness effects in simulated data. Second, Maliverni et al. [9] propose another scheme called continuous sequence reweighting (SR) that reveals structural features that are unique to subfamilies as opposed to determining global properties common to all family members. These articles as a whole provide an in-depth and useful picture on how to deal with phylogenetic correlations in the task of contact inference and the estimation of the effects of mutation.

A second set of articles in this issue [10–12] deals with the complex problem of evolutionary dynamics in protein structures and sequences. Cadet et al. [12] study formal statistical properties of sequence change and show how fluctuations follow a $-5/3$ Kolmogorov power and behave like an incremental Brownian process. In another study,



Citation: Morcos, F. Information Theory in Molecular Evolution: From Models to Structures and Dynamics. *Entropy* **2021**, *23*, 482. <https://doi.org/10.3390/e23040482>

Received: 15 April 2021

Accepted: 15 April 2021

Published: 19 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Wang et al. [10] investigate members of the family of β -Lactamases, enzymes involved in antibiotic resistance. In this study, they uncovered, via molecular simulations, important amino acid positions that share functional and dynamical features with another class of evolutionarily related proteins called Penicillin-binding proteins (PBP), enhancing our understanding of the dynamics of catalytic residues in the context of antibiotic resistance. In a third article, also concerned with the dynamics of protein evolution, Campitelli et al. [11] devise accurate metrics to quantify epistasis upon amino acid perturbations (EpiScore) and the asymmetric Dynamic Coupling Index (DCI_{asym}) to measure how connected residues are affected depending on which residue has been perturbed. These metrics are relevant contributions to the study of allostery and the evolutionary forces that shape this important functional phenomenon.

In a final study, Sinner et al. [13] construct another information metric to predict the degree of specificity between molecules in two-component signaling networks. Molecular interactions between histidine kinases (HK) and response regulators (RR) have evolved towards amino acid specificity at the physical interface in the HK-RR complex where phosphotransfer occurs. A degree of coevolutionary strength at this interface can be quantified for a large number of organisms. The authors created a public web server called ELIHKSIR.org (Evolutionary Links Inferred for Histidine Kinase Sensors Interacting with Response regulators) to facilitate the prediction and analysis of these links and to assess the effect of mutations in interacting specificity.

All together, the methodological contributions presented in this issue of *Entropy* will help advance the study of molecular evolutionary dynamics through the lens of information theoretical metrics and a combination of structural modeling and molecular dynamics simulations.

Funding: The author's research is funded by the University of Texas at Dallas, NIH grant number R35GM133631, and NSF grant number MCB-1943442.

Acknowledgments: We acknowledge all author contributions to the Special Issue in Information Theory and Molecular Evolution: From Models to Structures and Dynamics.

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Fisher, R.A. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. R. Soc. Edinb.* **1919**, *52*, 399–433. [[CrossRef](#)]
2. Grandy, W., Jr.; Milonni, P. *Physics and Probability: Essays in Honor of Edwin T. Jaynes*; Cambridge University Press: Cambridge, UK, 1993. [[CrossRef](#)]
3. Weigt, M.; White, R.A.; Szurmant, H.; Hoch, J.A.; Hwa, T. Identification of direct residue contacts in protein–protein interaction by message passing. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 67–72. [[CrossRef](#)] [[PubMed](#)]
4. Morcos, F.; Pagnani, A.; Lunt, B.; Bertolino, A.; Marks, D.S.; Sander, C.; Zecchina, R.; Onuchic, J.N.; Hwa, T.; Weigt, M. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, E1293–E1301. [[CrossRef](#)]
5. Bitbol, A.F. Inferring interaction partners from protein sequences using mutual information. *PLoS Comput. Biol.* **2018**, *14*, e1006401. [[CrossRef](#)] [[PubMed](#)]
6. Marmier, G.; Weigt, M.; Bitbol, A.F. Phylogenetic correlations can suffice to infer protein partners from sequences. *PLoS Comput. Biol.* **2019**, *15*, e1007179. [[CrossRef](#)] [[PubMed](#)]
7. Hockenberry, A.J.; Wilke, C.O. Phylogenetic Weighting Does Little to Improve the Accuracy of Evolutionary Coupling Analyses. *Entropy* **2019**, *21*, 1000. [[CrossRef](#)] [[PubMed](#)]
8. Rodriguez Horta, E.; Barrat-Charlaix, P.; Weigt, M. Toward Inferring Potts Models for Phylogenetically Correlated Sequence Data. *Entropy* **2019**, *21*, 1090. [[CrossRef](#)]
9. Malinverni, D.; Barducci, A. Coevolutionary Analysis of Protein Subfamilies by Sequence Reweighting. *Entropy* **2019**, *21*, 1127. [[CrossRef](#)] [[PubMed](#)]
10. Wang, F.; Zhou, H.; Wang, X.; Tao, P. Dynamical Behavior of β -Lactamases and Penicillin-Binding Proteins in Different Functional States and Its Potential Role in Evolution. *Entropy* **2019**, *21*, 1130. [[CrossRef](#)]

-
11. Campitelli, P.; Ozkan, S.B. Allosteric and Epistasis: Emergent Properties of Anisotropic Networks. *Entropy* **2020**, *22*, 667. [[CrossRef](#)] [[PubMed](#)]
 12. Cadet, X.F.; Dehak, R.; Chin, S.P.; Bessafi, M. Non-Linear Dynamics Analysis of Protein Sequences. Application to CYP450. *Entropy* **2019**, *21*, 852. [[CrossRef](#)]
 13. Sinner, C.; Ziegler, C.; Jung, Y.H.; Jiang, X.; Morcos, F. ELIHK SIR Web Server: Evolutionary Links Inferred for Histidine Kinase Sensors Interacting with Response Regulators. *Entropy* **2021**, *23*, 170. [[CrossRef](#)] [[PubMed](#)]