Programming Fitness Landscapes by Sparse Epistasis

Kabir B Husain[1] and Arvind Murugan[1]

***Short Abstract* —The effect of a mutation in a genetic sequence depends on the context in which it is made. This phenomenon, known as epistasis, leads to rugged fitness landscapes that proteins explore over evolutionary timescales. The form of evolutionary pathways in these landscapes are not well characterised. Inspired by recent experimental findings, here we construct a class of fitness landscapes characterised solely by their epistatic statistics. We analyse the structure of mutational pathways in this space, and investigate both its global and local topology. Our results serve as theoretical baselines against which experimental data may be assessed.**

***Keywords* — Epistasis, Protein evolution, Statistical landscapes, Mutational pathways**

I. Background

Proteins evolve towards fitter variants by a process of mutation and selection. In principle, these local transformations occur in the background of a fitness landscape defined over an underlying sequence space. Global features of this landscape are therefore key in determining how well and how quickly proteins evolve. However, we do not quite understand what the defining structure of a `generic' or representative fitness landscape is -- this is related to the problem of predicting the form of the genotype-phenotype map.

Nonetheless, experiments have, by exhaustive mutagenesis, begun to map this out. In particular, a recent study [1] found that the landscape could be encoded by a small number of appropriately-defined epistatic contributions. This sparsity suggests a simple but non-trivial structure, and leads to the question of whether particular mutational pathways in genotypic space can be understood as arising just from this underlying epistatic structure.

II. Methods and Results

To address this question, we resort to a minimal, statistical approach. We represent a genotype as a bit string, so that the genotypic space is a Boolean hypercube. On this space, we generate an ensemble of fitness landscapes by randomly generating a (small) number of epistatic terms; as described in [2], this is analogous to specifying a landscape by its Fourier modes. Our main control parameter is the extent of sparsity: the fraction of epistatic terms that we allow to be non-zero.

We first consider local properties of the resulting ensemble, measuring correlations lengths as a function of the sparsity and magnitude of epistasis. We then go on to characterise more global properties: in particular, the 'ruggedness' of the landscape (intuitively, the number of local fitness maxima), and the connectivity of mutational pathways in this space [3]. We contrast our results with prior statistical landscapes, such as the NK model, and discuss the effects of higher-order epistatic contributions. Finally, we consider potential applications to existing and future experiments [4].

III. Outlook

The form of the sequence-function map, and its connection with the mutational pathways undertaken by an evolving protein, has been a long standing puzzle. Our theory provides a simple prescription to generate ensembles of rugged landscapes, and therefore a systematic investigation of the link between epistasis, ruggedness, and evolutionary paths. Our results may be important for identifying a minimal, conserved set of features shared by diverse fitness landscapes.

IV. References

[1] Poelwijk, F. J., Socolich, M., & Ranganathan, R. (2017). Learning the pattern of epistasis linking genotype and phenotype in a protein. *bioRxiv*.

[2] Poelwijk, F. J., Krishna, V., & Ranganathan, R. (2016). The Context-Dependence of Mutations: A Linkage of Formalisms. *PLoS Computational Biology*, *12*(6), 1–19.

[3] Nowak, S., & Krug, J. (2015). Analysis of adaptive walks on NK fitness landscapes with different interaction schemes. *Journal of Statistical Mechanics: Theory and Experiment*, *2015*(6). https://doi.org/10.1088/1742-5468/2015/06/P06014

[4] Starr, T. N., Flynn, J. M., Mishra, P., Bolon, D. N. A., & Thornton, J. W. (2018). Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. *Proceedings of the National Academy of Sciences*.

[1]Acknowledgements: KH is supported by a postdoctoral fellowship from the JSMF.

[1]James Franck Institute, University of Chicago. E-mail: kabirh@uchicago.edu

[2]Same address. E-mail: amurugan@uchicago.edu