A Machine Learning approach to Biophysics of proteins

Katebsaber M.A*

*Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

Abstract

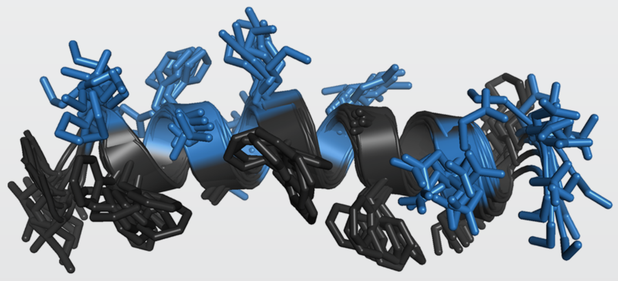
Introduction: Computational prediction of protein structure is one of the most promising tools for de novo protein design. Despite significant improvements in this area, most of the tools do not share the same success. The main challenge in the computational design of proteins is the Inverse Protein Folding problem which connects the hierarchy of structures in one protein. Hence, there is an as-yet unmet need for the development of tools and methodology of protein design.

Methods: In Machine Learning approaches, protein descriptors are either obtained experimentally or computationally. Then suitable models are selected based on the nature of the problem (i.e. classification, clustering, and regression). Finally, the results will be investigated using Biophysical golden standard methods.

Results and discussion: The success of a computational tool is typically based upon: (1) Biophysical model for describing proteins and (2) prediction algorithm which works on top of the Biophysical model. Machine Learning fits the scope of the second phase and helps in the determination of Biophysical parameters, sequence, and structure. Accuracy of Machine Learning approaches heavily depends on Biophysical models and there is an urgent need for Biophysical protein descriptor indices.

Conclusion: Despite the challenges facing in the utilization of Machine Learning approaches in Biophysics, it successfully created reliable solutions for designing customized proteins. In this seminar it would be discussed that how designing better Biophysical descriptors of proteins, results in accuracy and efficiency improvements in Machine Learning approaches.

Keywords: Computational protein design, Biophysical protein models, Machine Learning

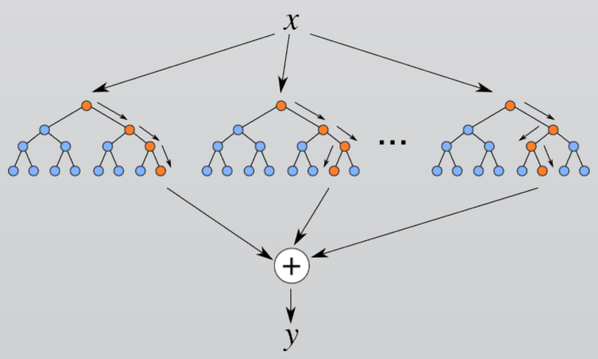


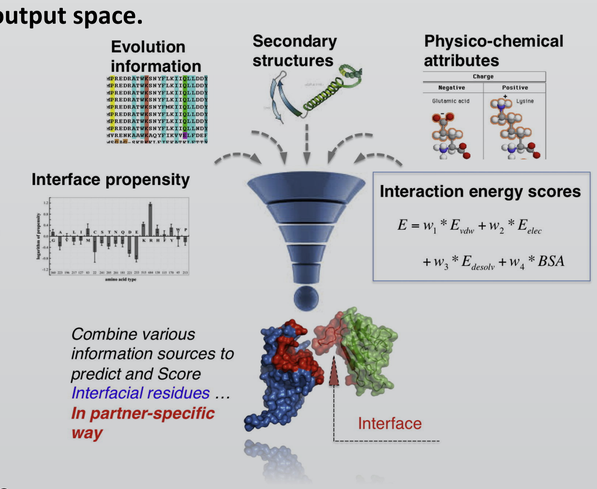
SCAN ME

Objective — Designing a protein sequence which is functional and has a high specificity and affinity to the target.

Shortcomings — For modeling interaction specificity we have to evaluate multiple competing complexes, which presents computational and experimental challenges. even with customized hardware, simulations of dynamics take months and in most cases require non-physiological conditions.

Machine Learning Solution — Incorporation of **Machine Learning** approaches can help us overcome computational and experimental expenses. Machine learning is a computer science discipline that evolved as a sub-field of Artificial Intelligence. The main purpose of Machine learning is to learn and predict the rules and patterns that govern the data. Machine Learning has different flavors and supervised learning is the most used one in Biophysics. In this approach an algorithm (Learner) tries to find a mapping function between Biophysical input features and and output space.





Protein Biophysical profiles, help Machine Learning resolve structure to function problem

Protein — **Machine Learning** — **Biophysics**

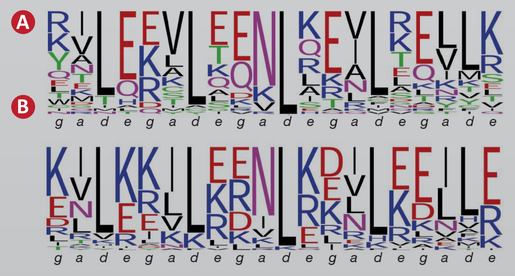
An elegant solution — **Marrying Biophysics and Machine Learning**

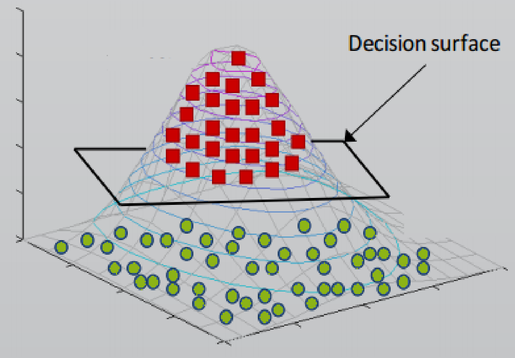
1. We first create biophysical profile of proteins.
2. We train our Machine Learning model to learn patterns among our experimental data.
3. Finally we have to **evaluate** our model with Biophysical golden standard methods.

Evaluation — Circular Dichroism and Micro-Array assay are two methods used for evaluation of stability. In utilizing these two methods, an interaction profile was defined using –log(F/Fmax) scores derived from micro-arrays, and profiles were clustered using Euclidean distance as the dissimilarity metric.

What about results?

Amino acid distribution is similar in experimental and computational methods.

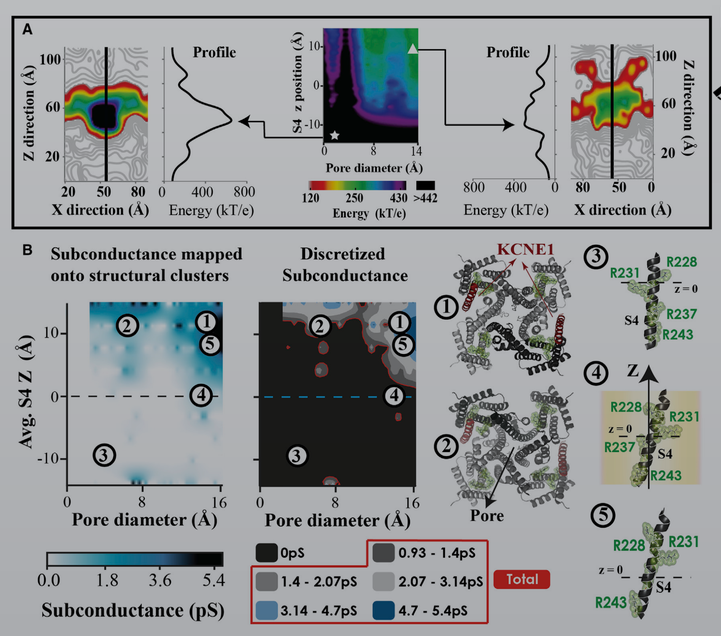




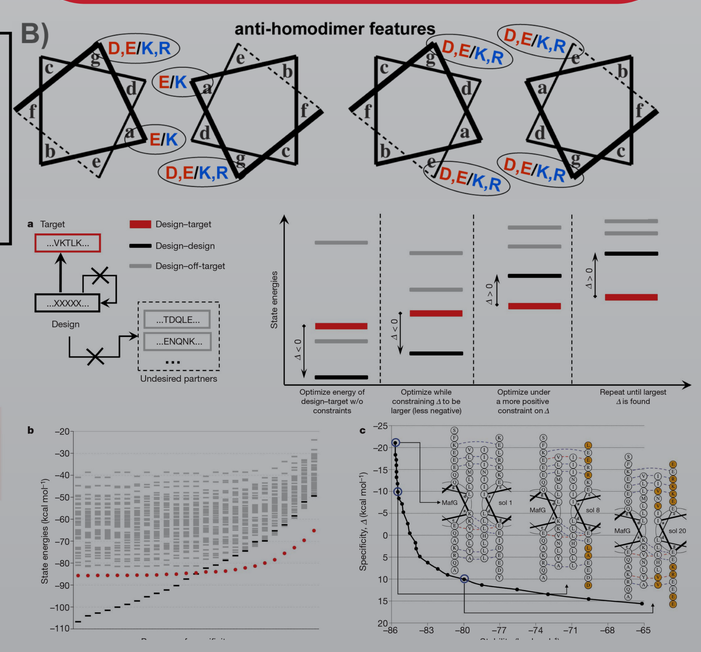
Discussion

Machine Learning provides a way to analyze and optimize stability/specificity trade-offs in protein design. This framework is applicable to most proteins and is a powerful tool for understanding protein behavior. This methodology can also be applied to study the effects of mutations, ligands, and drugs. Determinants of protein-interaction specificity are not yet as well understood for other complexes, but significant progress in this area is evident. **Amino acid distributions demonstrate an underlying pattern** for different structures of proteins is available. Finding these patterns innately depends on Biophysical models for description of protein molecules.

Simulations show a GOOD correspondence with experimental results.



Learning paradigm and its outcome interpretation in Biophysics



Reference

1. Ramasubramanian, Smiruthi, and Yoram Rudy. 2018. “Article The Structural Basis of IKs Ion-Channel Activation: Mechanistic Insights from Molecular Simulations.” Biophysj :(11) 114 94–2584. https://doi.org/10.1016/j.bpj.2018.04.023.
2. Grigoryan, Gevorg, Aaron W. Reinke, and Amy E. Keating. 2009. “Design of Protein-Interaction Specificity Gives Selective BZIP-Binding Peptides.” Nature 64–859 :(7240) 458. https://doi.org/10.1038/nature07885.
3. Fischer, Axel W., David M. Anderson, Maxx H. Tessmer, Dara W. Frank, Jimmy B. Feix, and Jens Meiler. 2017. “Structure and Dynamics of Type III Secretion Effector Protein ExoU As Determined by SDSL-EPR Spectroscopy in Conjunction with de Novo Protein Folding.” ACS Omega 84–2977 :(6) 2. https://doi.org/10.1021/acsomega.7b00349.
4. Karamzadeh, Razieh, Mohammad Hossein Karimi-Jafari, Ali Sharifi-Zarchi, Hamidreza Chitsaz, Ghasem Hosseini Salekdeh, and Ali Akbar Moosavi-Movahedi. 2017. “Machine Learning and Network Analysis of Molecular Dynamics Trajectories Reveal Two Chains of Red/Ox-Specific Residue Interactions in Human Protein Disulfide Isomerase.” Scientific Reports 11–1 :(1) 7. https://doi.org/10.1038/s5-03966-017-41598.
5. Lee, Yoonji, Jimin Pei, Jordan M Baumhardt, Yuh Min Chook, and Nick V Grishin. 2019. “Structural Prerequisites for CRM-1 Dependent Nuclear Export Signaling Peptides: Accessibility , Adapting Conformation , and the Stability at the Binding Site.” Scientific Reports, no. April: 13–1. https://doi.org/10.1038/s0-43004-019-41598.