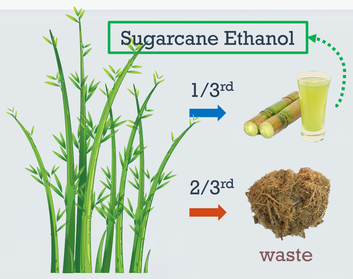
IMPROVING THE THERMOSTABILITY OF ENZYMES USING BIOINFORMATICS AND ELECTROSTATICS ANALYSIS

KHOA NGO [AB] FERNANDO BRUNO [C] VITOR LEITE [C] JOSÉ ONUCHIC [A] VINÍCIUS CONTESSOTO [A]

[A] Center for Theoretical Biological Physics, Rice University | [B] Department of Physics, University of Houston | [C] Department of Physics, São Paulo State University

This project was supported by the National Science Foundation (Grant: NSF PHY-14-27654).

1 Background



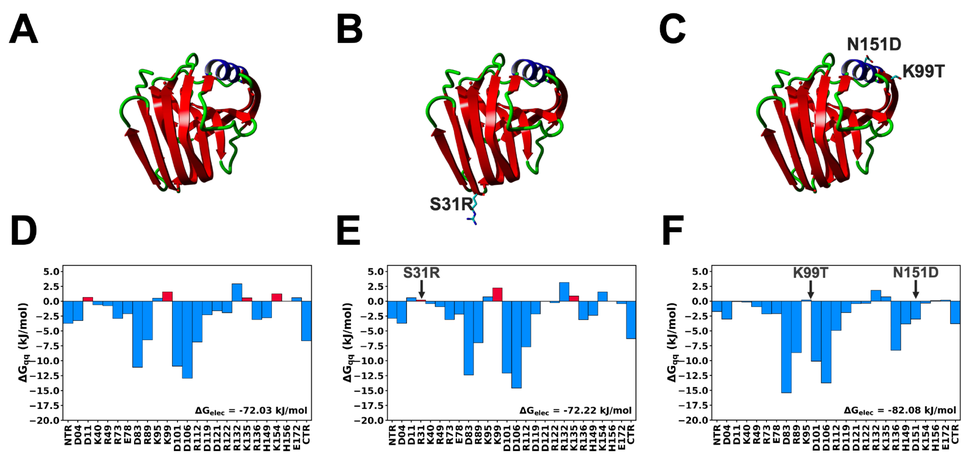
- As humanity faces looming threats from climate change, renewable sources of fuels, such as **biofuels** derived from microbial, plant, or animal materials, are getting popular as a mean to help reduce CO2 pollution.
- Produced from existing food crops, **1st generation biofuels** helps reduce GHG emissions, but indirectly contributes to a rise in food price [1]. By using only lignocellulosic biomass (corn stover, sugarcane bagasse etc.), **2nd generation biofuels** eliminate such problem but technical barriers exist that make their production uneconomical.
- In nature, certain enzymes help convert cellulose/hemicellulose into sugars that can be fermented to turn in bioethanol. In this work, we propose a novel approach to predict mutations that may **improve the thermostability** of such enzymes. Doing so can help reduce the cost of production and make 2nd generation biofuel a viable alternative.

2 New Mutation Approach

1	Acquire the Wild Type (WT) amino acid sequence of the enzyme of interest (Xylanase A from Bacillus subtilis)	• PDB ID: 1XXN. • Xylanases are enzymes that catalyze the hydrolysis of beta-1,4 glycosidic linkages of xylans, present in many plant cell walls, to release oligo- and disaccharides containing the sugar xylose.
2	Reconstruct the ancestral & consensus models of the WT enzyme from homologous sequences	• Ancestral approach: using statistical inference to reconstruct ancestral enzymes from extant enzymes. • Consensus approach: generating a model consisting of the most frequent residues found at each position when aligning extant sequences together.
3	Optimize electrostatics interactions in the WT enzyme model to reveal destabilizing residues	• The TKSA-MC web server [2] reveals the electrostatics free energy contribution toward the native (folded) state of each ionizable residue in the protein. • Residues that are destabilizing with ≥50% of surface area exposed to the solvent are candidates to be replaced.
4	Replace destabilizing residues in the WT w/ those found in the ancestral & consensus models at the same locations	• Research has shown that ancestral proteins, derived from a warmer past, were often more thermostable than their modern counterparts. • Evolution conserves amino acid residues that contribute to protein functions and stability, hence the consensus sequence was also used to make mutations.
5	Predict the 3D structures of the new mutants & test their thermostability using simulations	• The 3D structures of the mutants were predicted using the MODWEB web server [3]. • The mutants along with the WT undergo molecular simulations in GROMACS [4] using structure-based Cα models to evaluate their thermodynamic characteristics.

3 Mutants Exhibited Higher Thermostability than the Wild Type

Figure 1: Sample usage of **electrostatics analysis** to identify candidate amino acid residues to be mutated using the TKSA-MC web server.



- **A**, **B**, **C** are the predicted 3D structures of the Xylanase A WT variant & the S31R, K99T/N151D mutants, respectively.
- **D**, **E**, **F** show the charge-charge energy contribution of each ionizable residue to the native (folded) state stability when compared to the unfolded state of the structures above them.
- Bars colored red represent destabilizing residues that should be mutated to increase protein thermostability.
- Bars colored blue are either stabilizing residues or "destabilizing" residues that are not exposed to the solvent much ($SASA < 50\%$).

Table 1: Thermodynamic Parameters Obtained from Simulations for All Mutants and Rationale for Mutant Selection

Rank	Model	T_F (reduced units)	ΔG from WT (kT)	ΔG_{elec} (kJ/mol)
1	K99T [b]	1.2400	-7.22	-63.13
2	Q7H/G13R/I107L	1.2392	-6.74	-82.68
3	S31Y	1.2383	-6.52	-64.44
4	K99T/N151D [a]	1.2350	-5.42	-82.08
	S31R [c]	1.2350	-5.19	-72.22
5	N181R	1.2325	-4.12	-65.31
	N32D	1.2325	-4.28	-83.03
6	S22E/N32D	1.2317	-4.09	-75.47
7	Q7H/S22P	1.2300	-3.47	-84.84
8	N151D [c]	1.2283	-2.71	-91.07
9	Q7H	1.2267	-2.04	-77.01
10	Q7H/V150A	1.2258	-1.67	-64.33
11	N54E	1.2233	-0.57	-68.33
	S27E	1.2233	-0.51	-86.87
12	K154A [b]	1.2225	-0.42	-79.91
	S22E	1.2225	-0.63	-72.89
13	WT	1.2217	0.00	-72.03
14	D106K	1.2208	0.31	-35.53
	D106R	1.2208	0.37	-47.91

All unlabeled mutants are **control mutants** or the WT variant.
[a] Mutant obtained by combining multiple mutations determined to be stabilizing.
[b] Mutant obtained by replacing a **destabilizing** charged residue with a residue located in the Ancestral or Consensus sequence at the same location.
[c] Mutant obtained by replacing an uncharged residue with a charged residue located in the Ancestral or Consensus sequence at the same location and exhibit lower ΔG_{elec} than the WT.

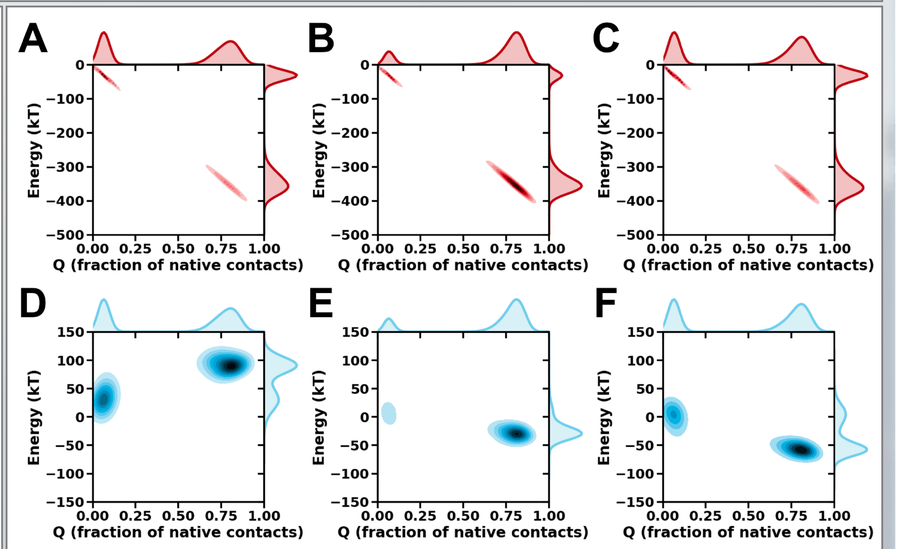


Figure 2: Contribution of different potentials to the folding process.

- **A**, **B**, **C** are the distributions of the **Lennard Jones 10-12 potentials** over the reaction coordinate Q (i.e., fraction of native contacts) at the folding temperatures of D106R, WT, and K99T protein models, respectively.
- **D**, **E**, **F** are those for the **electrostatics potentials**.

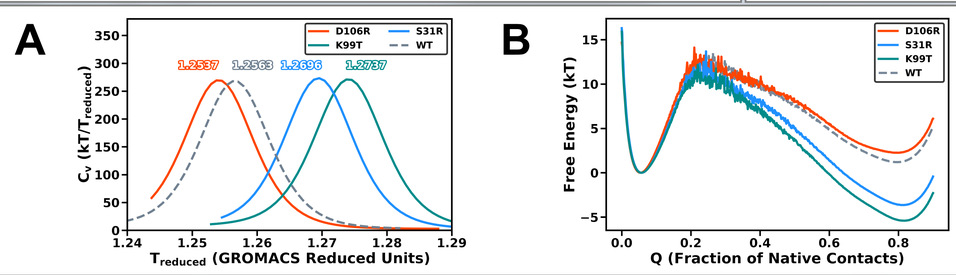


Figure 3: Contribution of different potentials to the folding process.

- **A** is the **heat capacity at constant volume (C_V)** at different simulation temperatures.
- **B** is the changes in the Gibbs free energy (ΔG) at different Q values, recorded at the folding temperature of the WT enzyme.

4 Conclusions

- Unlike conventional approaches, our approach requires substantially less time and cost. In exchange, **the structure of the enzyme** needs to be known and there must **an abundance set of homologs** available to construct reliable ancestral and consensus models. This approach is not only useful for improving the efficiency of enzyme used in biofuel production, but can also help make many enzymes viable for large scale biotechnological and medical applications.

5 References

- [1] Binod, P., E. Gnansounou, R. Sindhu, and A. Pandey, 2019. Enzymes for second generation biofuels: Recent developments and future perspectives. *Bioresource Technology Reports* 5:317–325.
- [2] Contessoto, V. G., V. M. D. Oliveira, B. R. Fernandes, G. G. Slade, and V. B. P. Leite, 2018. TKSA-MC: A web server for rational mutation through the optimization of protein charge interactions. *Proteins: Structure, Function, and Bioinformatics* 86:1184–1188.
- [3] Pieper, U., N. Eswar, F. P. Davis, H. Braberg, M. S. Madhusudhan, A. Rossi, M. Marti-Renom, R. Karchin, B. M. Webb,D. Eramian, and et al., 2006. MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Research* 34:D291–D295.
- [4] Bekker, H., H. Berendsen, E. Dijkstra, S. Achterop, R. Vondrumen, D. Vanderspoel, A. Sijbers, H. Keegstra, and M. Renar-dus, 1993. Gromacs - A parallel computer for molecular-dynamics simulations. World Scientific Publishing, 252–256.