Designing a Thermostable Cellobiohydrolase: A Novel Approach to Sustainable Ethanol Production

By Sagarika Samavedi

Abstract

The annual global consumption of petroleum by the transport sector is 36 billion barrels. This leads to increasing concentration of atmospheric CO2 contributing to climate change. These hydrocarbons can be replaced by renewable biofuels. The most common biofuel is ethanol made from cellulose. In nature, cellulose is converted to glucose with different enzymes, one of which is cellobiohydrolase. This enzyme breaks cellulose polymer into 2-carbon sugars called cellobiose. The cellobiohydrolase found in fungi, however does not catalyze at a commercially viable rate. The goal of this project is to create a modified fungal cellobiohydrolase with increased thermal stability which will degrade cellulose at a higher rate and a higher temperature. The enzyme I chose for this study is Cellobiohydrolase Cel7a from *Hypocrea jecorina.* The baseline thermal stability of cellobiohydrolase will be found through an experiment testing the catalytic activity of the enzyme at different temperatures. This data will be plotted on a graph to determine the optimal temperature for cellobiohydrolase and quantify the rate of cellobiose production per unit enzyme. Using protein modelling software I-TASSER 5.0 and CHIMERA 1.2, a 3-D model of the enzyme will be created and 3D printed for reference. Using UGENE 1.25, multiple cellobiohydrolase FASTA sequences will be compared to identify conserved domains and active sites. Using the software ROSETTA, the amino acid sequence of cellobiohydrolase will be modified using site directed mutagenesis to increase the thermal stability without mutating the active site. The newly created protein will be retested with STRUM, and a ΔΔG score will be calculated. A more positive ΔΔG score will correspond to increased thermal stability.

Background

The projected daily world oil demand by 2020 is 97 million barrels of crude oil for various energy-producing uses. Over 64% of crude oil is used in the transportation sector and causes unsustainable levels of greenhouse gas emissions of 138 million metric tons annually, a major factor of climate change. Commercially viable methods of ethanol production can revolutionize the fuel industry by reducing the demand for crude oil and gas. Each year, 150 billion tons of cellulose is formed by plants. This organic polysaccharide is found in the cell walls of plants, the indigestible fiber in foods, in cotton and wood, and even in algae. Cellulose from plants stores solar energy in nontoxic ways within chemical bonds, as it is a highly concentrated polymer of glucose. Cellobiohydrolase is a cellulase enzyme that effectively degrades and fragments cellulose through the breakdown of glyosidic linkages.

Cellulose is ubiquitous and cellulosic ethanol has the potential to replace gasoline and stop the dependency on fossil fuels. The process to make ethanol from corn is complex, expensive, and has a lengthy processing time. Food-based ethanol feedstock competes for food resources, and is thus not sustainable. Technology to develop this process on a commercially viable scale is still being research and refined. Cellobiohydrolase is not naturally effective in cellulosic catalysis at the elevated temperatures required for substantial energy production

Research Goals

Question: What are the best possible single point modifications that can be made to the enzyme Cellobiohydrolase Cel7a from *Hypocrea jecorina* that can improve the enzyme's thermal stability and enzymatic catalytic activity?

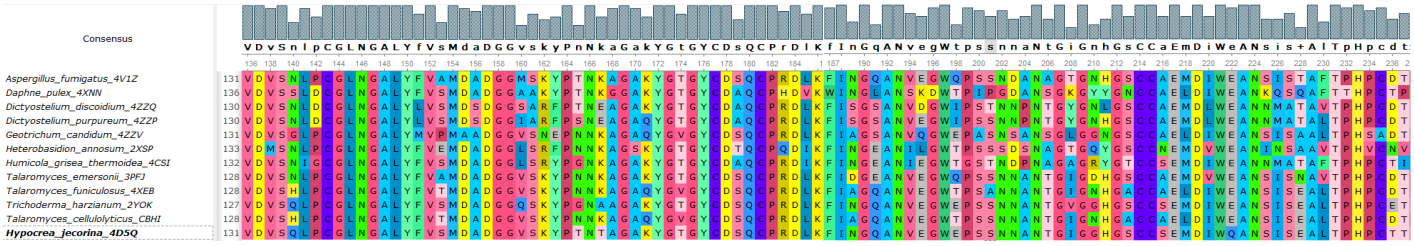
The goal of the project is to find the modifications that are beneficial to the thermal stability of Cellobiohydrolase. The impact of these modifications will be measured with a ΔΔG value. If the modifications increase thermal stability, the ΔΔG value will be a positive number. An efficient method to produce cellulosic ethanol could revolutionize the energy sector and decrease dependency on fossil fuels

Methodology

1. **Compare** FASTA sequences of exo-cellobiohydrolase using the software UGENE
2. **Build** a 3D model using the software I-TASSER of both the wild and modified Cellobiohydrolase Cel7a and visualize it with CHIMERA
3. **Convert** the 3D model to a stereolithographic file and 3D print for visual confirmation
4. **Analyze** this protein with STRUM to determine the ΔΔG value for each modification that is introduced in the protein structure
5. **Compare** the modified model to the original model to determine whether conserved domains were affected. If the active site changed, try a different point modification and re analyze it.
6. **Analyze** both molecules in detail to identify the modified bond structures that increase thermal stability

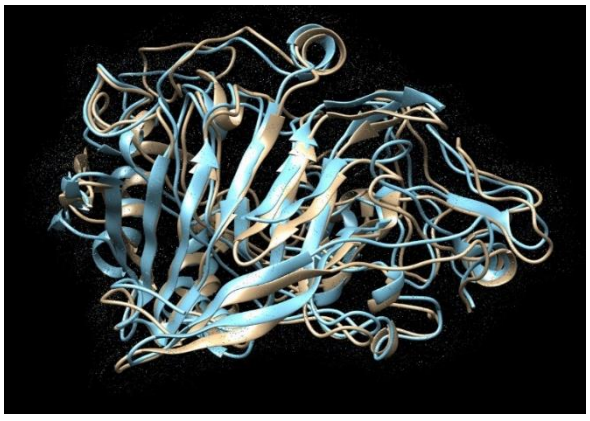
Results

Amino Acid Multi-Segment Analysis and Mapping of Cellobiohydrolase Cel7a

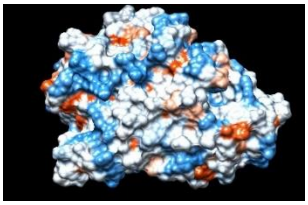


Detailed analysis of the FASTA sequence of *Hypocrea jecorina Cel7a* Cellobiohydrolase was compared with 11 other fungal cellobiohydrolase sequences. Over 4500 amino acids were aligned to identify conserved sections, and sites highly conserved were likely vital to the enzyme active site and were not modified. Sites where the *HJCel7a* did not match the consensus were target areas for single point mutations to the structure. Amino acids were changed to the most common amino acid found in the other fungal sequences that were compared. 16 modifications were selected that caused a positive impact to the thermal stability

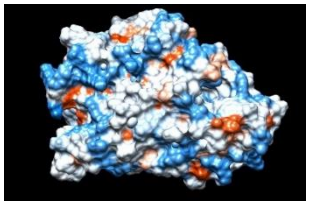
Superimposed comparison of Wild type and Modified Proteins

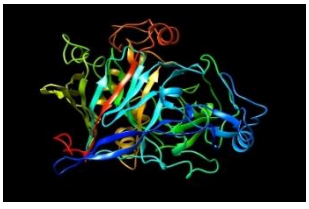


Wild Type Cellobiohydrolase



Modified Cellobiohydrolase





Discussion and Conclusions

Position	Wild HJCel7a	Modified HJCel7a	ΔΔG
2	Serine	Glutamine	0.51
57	Serine	Aspartic Acid	0.71
77	Alanine	Aspartic Acid	1.45
93	Isoleucine	Leucine	1.58
162	Threonine	Lysine	1.03
217	Glutamine	Glutamic Acid	1.16
236	Glutamic Acid	Threonine	0.81
309	Valine	Lysine	1.31
310	Threonine	Arginine	1.30
318	Leucine	Valine	1.52
321	Tyrosine	Asparagine	0.78
328	Aspartic Acid	Glycine	0.30
335	Glutamic Acid	Lysine	0.26
352	Phenylalanine	Methionine	1.97
385	Glutamic Acid	Alanine	0.87
393	Valine	Alanine	1.81
TOTAL			17.73

After analyzing the amino acid sequences of Cellobiohydrolase Cel7a from *Hypocrea* jecorina, and comparing it to other sequences, 16 modifications were carefully chosen. Each modification was assessed using VADAR to determine the exact changes in bonding structure of the protein. For every salt bridge created the thermal stability increased by 1 Kcal per bridge. The structure was made as tightly packed as possible to increase hydrogen bonding and London dispersion forces. Hydrogen bonding makes up 40% of the protein's thermal stability. The impact on the enzyme's thermal stability was assessed using STRUM analysis. STRUM is a program that calculates the protein's ΔΔG value using gradient boosting regression analysis. The Gibbs Free Energy, ΔG, is calculated between the unfolded version of the protein and the folded version of the protein. The ΔG value is first calculated for both the wild and mutant proteins. The ΔΔG value is the change in the Gibbs Free Energy between the two models of proteins. A high ΔΔG value attests that the modification increased thermal stability. Having a positive ΔΔG means it takes more energy to denature the protein.

Variable (Bonds, Chains)	Wild HJCel7a	Modified HJCel7a
# Residues with H Bonds	331	335
Total ASA	15703.4 Å	15678.0 Å
Exposed Non-Polar ASA	9690.4 Å	9594.2 Å
% Side ASA Hydrophobic	19.30 %	18.07 %
Total Volume	57821 Å	56009.9 Å
Free Energy of Folding	-358.61	-355.4
Buried Charges	15	9
Number of Disufide Bonds	7	10
Total H Bonds	265 + 260	268 + 265
% Residues 95% buried	141	146
# Ion Pairs	26	33
Cation-Pi Interactions	47	54
Conf. Strained Residues	10	8

$$\Delta\Delta G = \Delta G_{mutant} - \Delta G_{original}$$

The total number of ionic pairs increased by 28%, cation pi interactions increased by 33.33%, the disulfide bridges increased by 500%, and hydrogen bonds increased by 12.63%. these added bonds strengthen the overall secondary and tertiary protein structure, resulting in a 17.73 Kcal/mole increase in $\Delta\Delta G$. This leads to a stronger protein that can withstand denaturation at temperatures about 17 degrees above normal

Future Impact

Cellulosic ethanol is the leading next generation biofuel with strong economic importance. My research design will potentially increase the Cellobiose yield, and hence the Ethanol production by 300 % of the current rate. Many environmental benefits will occur from emission reduction as well. Cellulosic Ethanol can lower GHG emissions by over 98% relative to petroleum based gasoline. This is a carbon-cycle balanced and renewable way to produce cellulosic ethanol which means we recycle carbon rather than exposing underground carbon reservoirs to the atmosphere. Cellulosic ethanol will not cut into usable food resources and will have an abundant feedstock supply. New E10, E15 and even E100 vehicles are being explored. These Flex-Fuel vehicles are bringing personal transportation one step forward to a greener future.

Current Cost Structure	
Cost per gallon gasoline	$1.16 per gallon
Cost per gallon ethanol from thermostable cellobiohydrolase	$0.75 per gallon
Cost savings per gallon fuel	$0.41 per gallon
Total fuel use per day	359 million gallons

Economic Benefits if all Gasoline is Replaced by Cellulosic Ethanol (based on carbon cost)	
Savings per day	$155.39 million
Annual Savings	$57 billion
Greenhouse Gas Emissions	98.3% reduction
Production Cost	35% in savings

REFERENCES

Wang, Nam S. "Cellulose Degradation." Cellulose Degradation. University of Maryland Department of Chemical & Biomolecular Engineering, n.d. Web.

Center for Climate and Energy Solutions. (2009). Cellulosic Ethanol Factsheet. Retrieved from Cellulosic Ethanol: http://www.c2es.org/technology/factsheet/CellulosicEthanol

Kubicek, C. P., Mikus, M., Schuster, A., Schmoll, M., & Seiboth, B. (2009). Metabolic Engineering Strategies for the improvement of Cellulase Production by Hypocrea jecorina. BioMed Central.

Moumita, S., Hong, G., & Smith, J. C. (2010). Catalytic Mechanism of Cellulose Degradation by a Cellobiohydrolase.

Quan, L., Ly, Q., & Zhang, Y. (2016). STRUM: structure-based prediction of protein stability changes upon single-point mutation. Zhang Lab.

Ivan Getov †, Marharyta Petukh † and Emil Alexov. (2016). SAAFEC:Predicting the Effect of Single Point Mutations on Protein Folding Free Energy Using a Knowledge-Modified MM/PBSA Approach. International Journal of Molecular Sciences.

Bruce Watt, University of Maine, Bugwood.org (copyright owner photograph Hypocrea jecorina)

http://www.brenda-enzymes.org/enzyme.php?ecno=3.2.1.91

http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0016162

http://zhanglab.ccmb.med.umich.edu/STRUM/output/SM100615/

https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

http://vadar.wishartlab.com/index.html

https://www.eia.gov/tools/faqs/faq.cfm?id=23&t=10

https://www.eia.gov/todayinenergy/detail.php?id=11331

http://irdp.ncl.res.in/CompStabAna.html

Trudeau, D. L., Lee, T. M. and Arnold, F. H. (2014), Engineered thermostable fungal cellulases exhibit efficient synergistic cellulose hydrolysis at elevated temperatures. Biotechnol. Bioeng., 111: 2390–2397. doi:10.1002/bit.25308

Mingardon F, Bagert JD, Maisonnier C, Trudeau DL, Arnold FH. Comparison of Family 9 Cellulases from Mesophilic and Thermophilic Bacteria . Applied and Environmental Microbiology. 2011;77(4):1436-1442. doi:10.1128/AEM.01802-10.

https://www.c2es.org/docUploads/Cellulosic-ethanol-11-09.pdf