Homology Modeling and Protein Purification of the Enzyme BaiK for Structural And Mechanistic Studies

Carolina Mata, Roman Aguirre, Stephen Khuu, Kathryn McCulloch
Department of Chemistry and Biochemistry, Cal Poly Pomona
Kellogg Honors College Capstone Project

Abstract

Primary bile acids play a major role in digestion by assisting in the absorption of fats and oils entering the body. Once in the gastrointestinal tract, naturally occurring gut bacteria such as *Clostridium scindens* convert primary bile acids into secondary bile acids via a pathway encoded by the bile acid inducible operon, which includes a putative coenzyme A transferase enzyme, BaiK.

Pre-COVID, a recombinant approach using *E. coli* was chosen to characterize BaiK. First, the *E. coli* Tuner expression line was transformed using a plasmid containing the *baiK* gene for overexpression. Large-scale overexpression was used to obtain large quantities of the protein, and the resulting cell pellet was collected for purification via column chromatography. Protein purification strategies have been optimized. Challenges in purifying larger amounts of protein to yield a single peak in chromatograms were circumvented by performing both affinity and gel filtration chromatography on the same day to avoid aggregation. The concentrated protein will next be used in crystallization screenings.

Homology modelling studies were used to obtain a hypothetical BaiK structure, which can help determine function and characteristics. This was done using available web-based tools such as BLAST, SWISS model, EXPASY, and Xtalpred. These servers were used to convert the genomic sequence to a protein sequence, remove the histidine tag using enterokinase, and align the sequence with published structures of PDB proteins. After investigations based on sequence similarity, conserved areas, and proposed function, multiple proteins were selected as templates for BaiK's homology model. The model was generated using SWISS model, which generated a homo dimer model consisting of a 37% sequence identity. The model had a favorable QMEAN score, with highly conserved areas in the active site. Verification of the model was achieved using the built-in functions from SWISS model.

Background

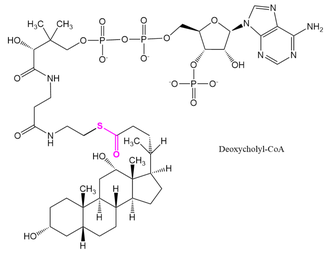
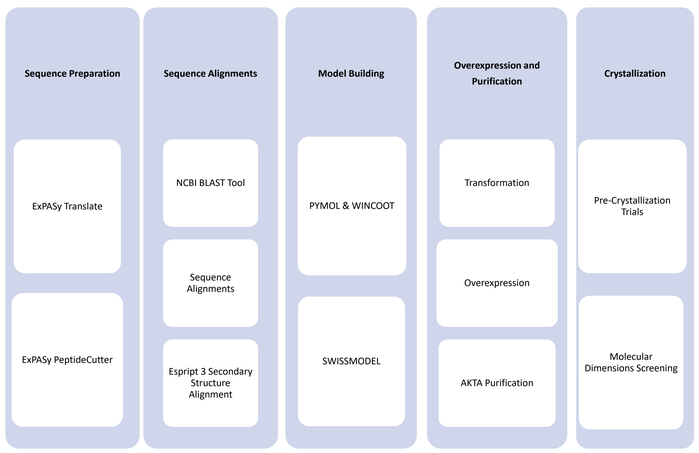


Figure 1- Demonstrates a Deoxycholoyl bile acid linked to Coenzyme A, with the thioester linkage highlighted in pink

Bile acids are converted to secondary bile acids via a multistep 7α-dehydroxylation pathway which is encoded by a pathway of at least eight genes, including *baiK*. The BaiK protein serves as a putative Coenzyme A transferase in the operon. This transferase operates by catalyzing the transfer of Coenzyme A, an important carrier molecule needed for various metabolic processes, to and from other compounds. *baiK* is being investigated because the accumulation of secondary bile acids in the gut is thought to be linked to cancers and other gastrointestinal tract diseases.

Methods



Bioinformatics and Homology Modeling

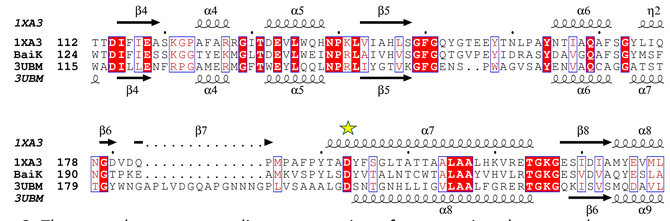


Figure 2- The secondary structure alignment consists of a comparison between the sequences of proteins similar to BaiK including 1XA3, 3UBM, and BaiK itself using 1XA3 as the secondary structure prediction template. This is used to predict the folding pattern at certain sequences, which is demonstrated by the helices and beta sheet arrows above the sequences. The blue boxes with red residues inside are an indicator of conserved residues, while a completely red box with white letters indicate absolutely conserved residues. The catalytic Asp residue is starred in yellow.

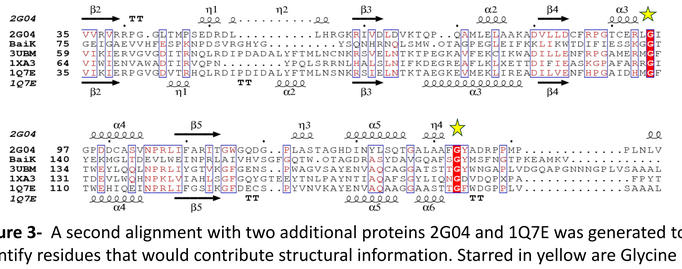


Figure 3- A second alignment with two additional proteins 2G04 and 1Q7E was generated to identify residues that would contribute structural information. Starred in yellow are Glycine residues that are absolutely conserved. When the homology model in **Figure 4** was generated, a glycine loop was also conserved. While these residues do not contribute to catalysis they contribute to the protein's structure and flexibility.

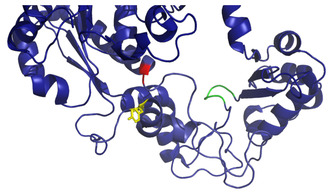


Figure 4- 3UBM was used to generate the model to the right. This model indicates potentially catalytic residues and structures in 3UBM. In red is the key catalytic Aspartic acid residue, in yellow is a potentially catalytic tyrosine residue, and in green is a glycine loop that is proposed to shield enzyme bound intermediates. By comparing Global Model Quality Estimation (GQME), Quality Model Estimation (QME), Root Mean Square Deviation (RMSD), percent similarity, and Gap values the accuracy of a model can be determined.

Protein PDB Code	Protein Name	GMQE	QMEAN	RMSD score	Percent Identity /similarity/Gap
1XA3	CaiB	0.67	-1.60	0.070	36.8/5/8
3UBM	Formyl CoA: oxalate CoA transferase	0.56	-2.37	0.394	28.0/44/10

In Vitro Purification and Analysis

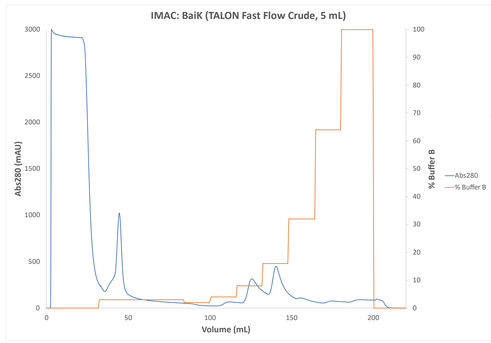


Figure 5- In this trial, IMAC purification was successful in separating BaiK from other elements in the clarified lysate using a 5 mL TALON crude fast flow column. The two peaks farthest to the right were collected and concentrated for purification via size exclusion chromatography. The two buffers used in this experiment included Buffer A- a lysis buffer (no imidazole), and Buffer B- an elution buffer (750 mM imidazole)

Analysis Continued

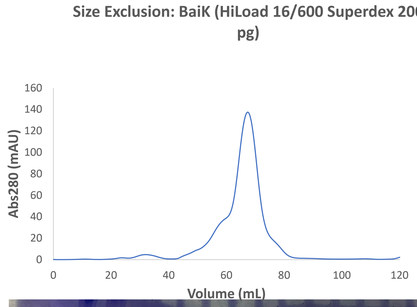


Figure 6- The larger peak from IMAC was injected for further purification onto a HiLoad16/600 Superdex200 pg column. SEC yielded a sizeable peak that contained 8 mg/mL of protein that was then aliquoted into 50 microliter samples for crystallization experiments.

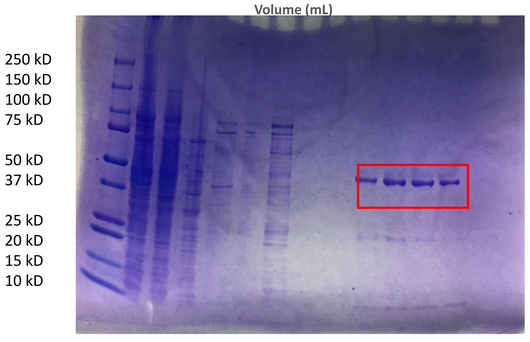


Figure 7- This SDS-PAGE contains crude lysate, clarified lysate, and 10 fractions from IMAC purification. The bands shown in the red box indicate the presence of BaiK protein at approximately 53 kD.

Crystallization Trials and Future Work

Preliminary screenings with Molecular Dimensions screening sets have not yielded a viable crystal for analysis, when an appropriate solvent is found concentrations will be optimized. Future work is focused on optimizing buffers for size exclusion chromatography and protein concentration for crystallization.

Acknowledgements

I'd like to thank Dr. McCulloch for guiding me and answering ALL my questions. Dr. McCulloch has received past support from CSUPERB via a New Investigator award and has currently received support from the Cal Poly RSCA program. Funding has also been provided by the Louis Stokes Alliance for Minority Participation fellowship awarded to Carolina Mata.

References

Rangarajan, E.s., et al. "Crystal Structure- P1 Form- of Escherichia Coli Crotonobetainyl-CoA: Carnitine CoA Transferase (CaiB)." *American Chemical Society*, 2005, doi:10.2210/pdb1xk6/pdb.

"CoA-Transferases." *CoA-Transferases - an Overview | ScienceDirect Topics*, www.sciencedirect.com/topics/agricultural-and-biological-sciences/coa-transferases (accessed March 4, 2020)

Lee, Sangbae, et al. "Sampling Long Time Scale Protein Motions: OSRW Simulation of Active Site Loop Conformational Free Energies in Formyl-CoA:Oxalate CoA Transferase." *Journal of the American Chemical Society*, vol. 132, no. 21, Feb. 2010, pp. 7252–7253., doi:10.1021/ja101446u.

Vital, Marius, et al. "Diversity of Bacteria Exhibiting Bile Acid-Inducible 7α-Dehydroxylation Genes in the Human Gut." *Computational and Structural Biotechnology Journal*, vol. 17, 2019, pp. 1016– 1019., doi:10.1016/j.csbj.2019.07.012.

"Digestive Diseases." *Mayo Clinic*, Mayo Foundation for Medical Education and Research, www.mayoclinic.org/medical-professionals/digestive-diseases/news/identifyingdiarrhea-caused-by-bile-acid-absorption/mac20430098 "ESPript." *ESPript 3.x / ENDscript 2.x*, espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi.
Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST. BMC Bioinformatics. 2009 Dec 15;10:421.

Stefan Bienert, Andrew Waterhouse, Tjaart A. P. de Beer, Gerardo Tauriello, Gabriel Studer, Lorenza Bordoli, Torsten Schwede, SWISS-MODEL, *Nucleic Acids Research*, Volume 45, Issue D1, January 2017, Pages D313–D319, https://doi.org/10.1093/nar/gkw1132

Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, Philip E. Bourne, The Protein Data Bank, *Nucleic Acids Research*, Volume 28, Issue 1, 1 January 2000, Pages 235–242, https://doi.org/10.1093/nar/28.1.235

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. ExPASy. Nucleic Acids Res. 2012 Jul;40(Web Server issue):W597-603.
PyMOL The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2010, *66* (4), 486–501.