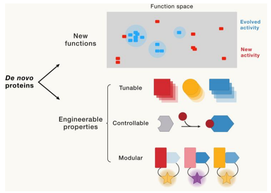
Computationally Driven De Novo Design and Engineering of a β-D-Glucose Binding Protein

Arjun Ganesan[1], Qinyi Zhong[1], Amber Middleton[1], Lee Solomon[1]

[1]Department of Chemistry and Biochemistry, College of Science, George Mason University

Aspiring Scientists Summer Internship Program 2025

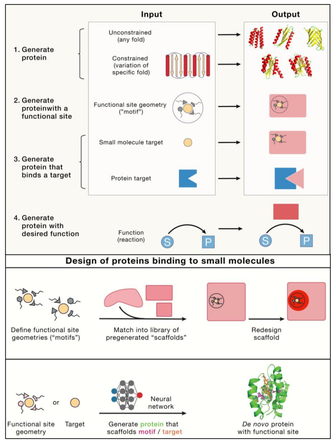
Introduction



(Kortemme, 2024)

Figure 1: De novo protein design is critical in computational biochemistry, referring to the development of novel proteins from the ground up. This bypasses the evolutionary complexity of natural proteins, enables studies on undetermined biochemical pathways, and proposes biomedical and technological solutions, such as biosensors and similar devices.

Figure 2: De novo protein design occurs in three to four primary steps: generation of a stable backbone through secondary structures, peptide sequence design for the backbone to drive proper tertiary folding, and validation through computational analyses, followed by in-vitro testing. In the context of small molecule binding, methods incorporate the iteration of motifs driven by previous models and AI to develop a crystal protein structure.



(Kortemme, 2024)

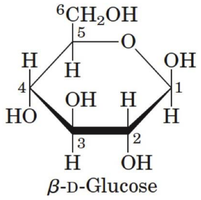
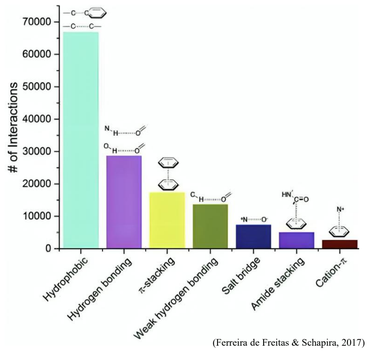


Figure 3: β-D-Glucose is a critical cyclic glucose parallel in present in various eukaryotic metabolic systems. It is a product of photosynthesis and is utilized in various metabolic process, such as glycolysis and the citric acid or Krebs cycle. However, various mechanisms associated with β-D-Glucose are unknown, and developing a binding protein allows for more effective studies on its behavior and utility.

Design Parameters

Figure 4: Structurally stable beta-barrel scaffold protein architecture. Drawing upon this scaffold, the objective was to optimize the protein binding pocket for β-D-Glucose. This scaffold is thermally stable and does not aggregate. Given that glucose is composed of many hydroxyl groups and is therefore polar, polar sidechain substitutions to the binding site is imperative, but it must not compromise overall protein integrity.

Figure 5: Holistic protein design encompasses optimizing protein-ligand interactions and minimizing unfavorable interactions. As β-D-Glucose consists of various hydroxyl groups, maximizing hydrophobic contacts and hydrogen bonding between these groups and participating polar or charged sidechains is paramount. Previous literature denotes the use of stacking and sugar-aromatic interactions to develop a viable binding surface for β-D-Glucose. Similarly, minimizing unfavorable interactions, such as clashes, is critical in proper folding and structure.



(Ferreira de Freitas & Schapira, 2017)

Methodology Architecture

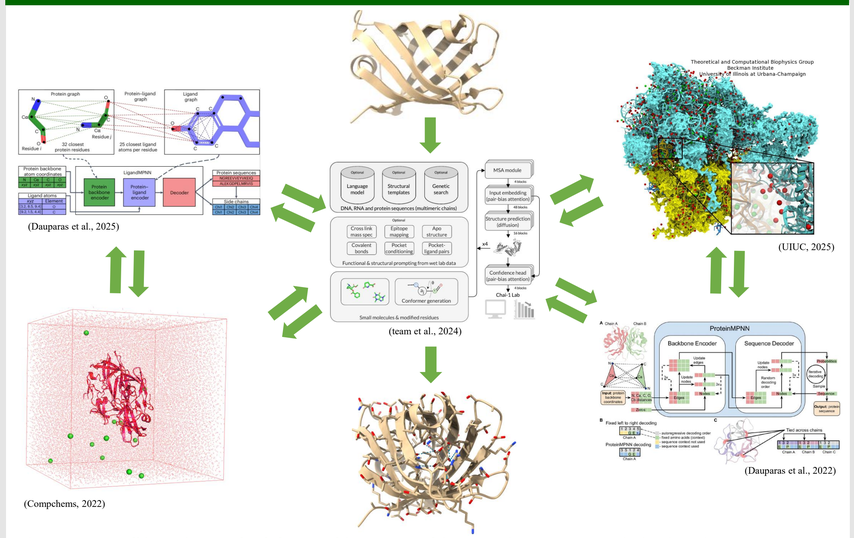


Figure 6: Overall iterative methodology architecture to drive protein structure optimization. Optimization to the given scaffold occurs through sidechain mutations (ChimeraX), protein-ligand folding (Chai Discovery), deep learning sequence prediction (LigandMPNN), and molecular dynamic simulations in addition to docking (VMD & AutoDock Vina). Utilizing these insights and iterating upon it contributes to the development of a hyper-optimized, universally-compliant protein.

Results

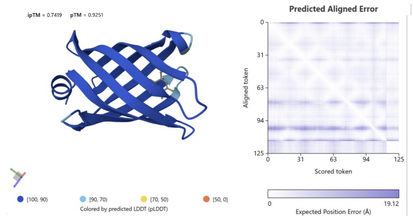


Figure 7: Chai Discovery Metrics. The protein achieved an ipTM = 0.7419 & pTM = 0.9251, indicating modest binding complex structure and high protein viability, respectively.

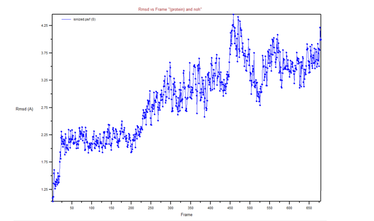


Figure 8: MD Simulation RMSD Data. RMSD, root mean square deviation, measures the variation of two superimposed proteins (angstroms).

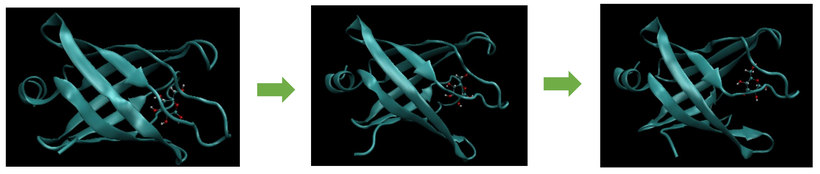


Figure 9: Visual inspection of molecular dynamics simulation. Throughout the simulation the protein begins to deform, indicating lack of backbone stability. Regardless, the ligand remained tightly in the pocket, indicating high pocket affinity at the expense of overall protein integrity.

Donor	Acceptor	Occupancy
SegBGLC-BGLC1-Side	SegAP1-ASP73-Side	63.38%
SegBGLC-BGLC1-Side	SegAP1-GLU97-Side	59.41%
SegAP1-ARG76-Side	SegBGLC-BGLC1-Side	47.06%
SegAP1-ALA48-Main	SegBGLC-BGLC1-Side	16.47%
SegBGLC-BGLC1-Side	SegAP1-GLN43-Side	5.00%
SegBGLC-BGLC1-Side	SegAP1-ASP19-Side	3.82%
SegAP1-GLN43-Side	SegBGLC-BGLC1-Side	3.38%
SegAP1-THR17-Side	SegBGLC-BGLC1-Side	2.06%

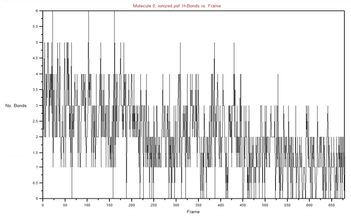


Figure 10: Molecular Dynamic Simulation Hydrogen Bonding Data. Hydrogen bonds with high occupancy indicate greater stability. The right graph depicts number of hydrogen bonds over simulation time (frames).

Conclusions & Future Work

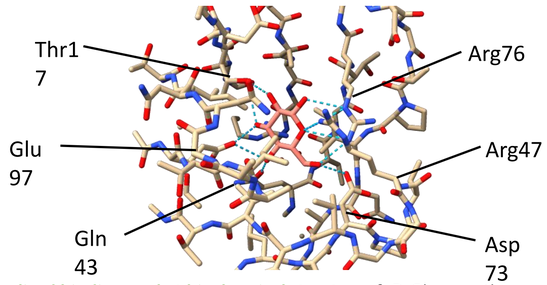
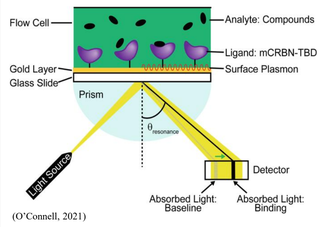


Figure 11: Finalized binding pocket biochemical structure. β-D-Glucose, the orange central molecule, is bound to the protein binding pocket with six amino acids: Thr17, Gln43, Arg47, Asp73, Arg76, and Glu97. These amino acids amount to approximately thirteen hydrogen bonds between the bound ligand and the pocket on ChimeraX.

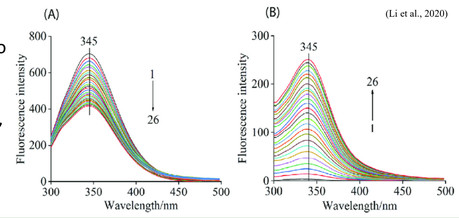
In-Vitro Validation



(O'Connell, 2021)

Figure 12: Surface Plasmon Resonance (SPR). SPR is an optical method that quantifies the alteration in refractive index near a sensor surface, which is a sheet of gold on glass support that composes the floor of flow cell through which an aqueous solution flows. Protein receptor molecules are immobilized on the sensor surface, and the ligand is injected into the aqueous solution to bind to the receptor. This induces a change in refractive index and quantifies protein-ligand binding kinetics.

Figure 13: Fluorescence Titration & Quenching. This adds a fluorescent tag to the protein, in which a ligand would interact with and alter the fluorescent properties. In protein-ligand interactions, quenching would occur in which, upon ligand binding, the fluorescent emission from the protein would decrease.



Key References

Dauparas, J., Lee, G. R., Pecoraro, R., An, L., Anishchenko, I., Glasscock, C., & Baker, D. (2025). Atomic context-conditioned protein sequence design using LigandMPNN. *Nature Methods*, *22*(4), 717–723. https://doi.org/10.1038/s41592-025-02626-1

Ferreira de Freitas, R., & Schapira, M. (2017). A systematic analysis of atomic protein–ligand interactions in the PDB. *MedChemComm*, *8*(10), 1970–1981. https://doi.org/10.1039/c7md00381a

Kandavelu Palani, Kumaran, D., Burley, S. K., & Swaminathan, S. (2012). Structure of a periplasmic glucose-binding protein from*Thermotoga maritima*. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, *68*(12), 1460–1464. https://doi.org/10.1107/s1744309112045241

Kortemme, T. (2024). De novo protein design—From new structures to programmable functions. *Cell*, *187*(3), 526–544. https://doi.org/10.1016/j.cell.2023.12.028

Mocz G, Ross JA. Fluorescence techniques in analysis of protein-ligand interactions. Methods Mol Biol. 2013;1008:169-210. doi: 10.1007/978-1-62703-398-5_7. PMID: 23729253.

O'Connell, N. (2021). Protein Ligand Interactions Using Surface Plasmon Resonance. In: Cacace, A.M., Hickey, C.M., Békés, M. (eds) Targeted Protein Degradation. Methods in Molecular Biology, vol 2365. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-1665-9_1

Pan X, Kortemme T. Recent advances in de novo protein design: Principles, methods, and applications. J Biol Chem. 2021 Jan-Jun;296:100558. doi: 10.1016/j.jbc.2021.100558. Epub 2021 Mar 18. PMID: 33744284; PMCID: PMC8065224.

team, C. D., Boitreaud, J., Dent, J., McPartlon, M., Meier, J., Reis, V., Rogozhonikov, A., & Wu, K. (2024). Chai-1: Decoding the molecular interactions of life. *BioRxiv*. https://doi.org/10.1101/2024.10.10.615955

Acknowledgements

This research was made possible through the support of George Mason University's College of Science, which supports the ASSIP Program and the GMU Hopper Cluster for their contributions to the computational environment utilize for testing and data modeling.