Linking GPCRs with type 2 diabetes: novel ADORA2B model utilizing AR2A multi-sequence templates

Sashrika Saini

Personal Statement

From Purple Oranges to Protein Ribbons

I remember the ineffably unique Crayola baby portraits plastered onto the bulletin board of my 5th grade classroom. Today's topic was genetics, and the day concluded with a drawing of a dark-skinned baby with freckles, red hair, and blue eyes. We observed nature's masterpiece rendered as a vigorously colored sketch. That day, I left school thinking of the impossible: designer babies and possibly designer food. In the 5th grade, I wanted to be a Fashion Designer in the likes of Giorgio Armani and Karl Lagerfeld. I had only dreamt of creating wearable art, but today's lesson on genetics had opened up so many doors. By 6th grade, I had changed my mind. I wanted to be a geneticist. I imagined myself experimenting and creating purple orange peels with a pink banana-like texture on the inside.

As I grew older, I went on many internet excursions. By the end of the day, I had gone from reading about DNA to cell lineage and embryology. I learned that most of our body was constituted of proteins. It seemed as if the different proteins that made our appearances unique were equally capable of making us fatally sick as well. That is when it began: my research endeavor and countless science fairs.

Ever since the start of high school, I had been placed in science classes that devoted part of the curriculum to original, independent research methods, and the process of writing scientific papers. When I switched schools in junior year, I was introduced to very rigorous STEM curriculum. It was time to pick a project yet again. Over the last 2 years I had focused my projects in the field of medicine accompanied with wet lab techniques. This year, I wanted to try something new. I had always been fascinated with proteins, so I decided to venture into the land of computational biology and experiment with protein structure and modeling.

What is Computational Biology?

Computational biology, also referred to as bioinformatics is a discipline that pulls concepts from math, computer science, engineering to develop a deeper understanding of molecular biological systems. Computational biology allows for the usage of large amount of data that can be analyzed or modeled using algorithms. This data can be collected experimentally, or it can also be accessed through online libraries.

Location of Research

I performed my research on my laptop, independent of any lab. I met with my advisors, Maria Borowski and Dr. Sakthikumar Ambady from WPI to discuss my progress and to tweak my methods.

The Math Behind the Madness

Throughout the research, there were many aspects of the project that incorporated math directly, and indirectly. In order to determine which combination of algorithms I would use for structure and template based alignment, I had to understand concepts underlying geometry based algorithms (spherical surface area, Voronoi diagrams, and energy based equations for

minimization). I also had to learn basic statistics: ANOVA, standard deviation, t-tests, degrees of freedom, and Ramachandran Plots. Besides the math, understanding a little bit of biochemistry was also necessary. I read countless literature on rotamers, bond angles and lengths, torsion, covalent and non-covalent bonding, energy minimization, and properties of the 20 amino acids.

Advice

Conducting research is quite an intensive and tasking process, and it is unique for every individual. Having gone through this process, the most important piece of advice I can offer is to be intuitive. Researching is worlds apart from any other kind of learning. It is a process of self-discovery and trial and error. I learned that failure is inevitable and giving up is not an option. I learned very quickly that everything is not black in white. There is too much grey in the entirety of it. You may not always know if you are going in the right direction. Therefore, the best thing to do is embrace the grey, think out of the box, and take a chance. Science is indebted to imagination.

Research

Introduction

A significant amount of scientific research has been devoted to discover a cure for diabetes; however, no treatments have been completely successful. One of the most debilitating aspects of diabetes comes from a condition known as insulin resistance. With this condition, the body does not utilize the hormone properly. Ongoing research has shown positive results in locating defects in the signaling pathways in cells. Specifically, transmembrane proteins that interact with extracellular macromolecules are the roots of the issue. In adipocytes, connective fat tissue, tyrosine kinases are the receptors that play the largest part in transporting insulin into the cell.

Thorough research in the signaling pathways between receptor proteins, extracellular molecules, and the insulin hormone, has concluded that non-specific reception primarily via the Adenosine Receptor A_{2B} (ADORA2B) causes insulin resistance. Due to insulin resistance, glucose accumulates in the blood stream. This eventually leads to type II diabetes mellitus.

Methodology

Method Goal

This homology model will predict a structure from its sequence with an accuracy that is comparable to the best sequence identity tested experimentally in the year of 2014. This will allow users to safely use *in silico* generated protein models in scientific fields mostly only experimental structures provide a solid basis for: structure-based drug design, analysis of protein function, ligand-ligand interactions, docking studies, and MD Simulations. Protein modeling is the alternate way to obtain structural information when experimental techniques fail (Rost, 1999).

Method Summary

There are two aspects to the most efficient approach to homology modeling, and 3D structure

prediction.

1. The structure of a protein is uniquely determined by its amino acid sequence (Epstein et al. 1963), and therefore the sequence should, in theory, contain suffice information to obtain the structure.

2. Similar sequences have been found to adopt practically identical structures while distantly related sequences can still fold into similar structures. Two protein sequences are highly likely to adopt a similar structure provided that the percentage identity between these proteins for a given length is above the threshold.

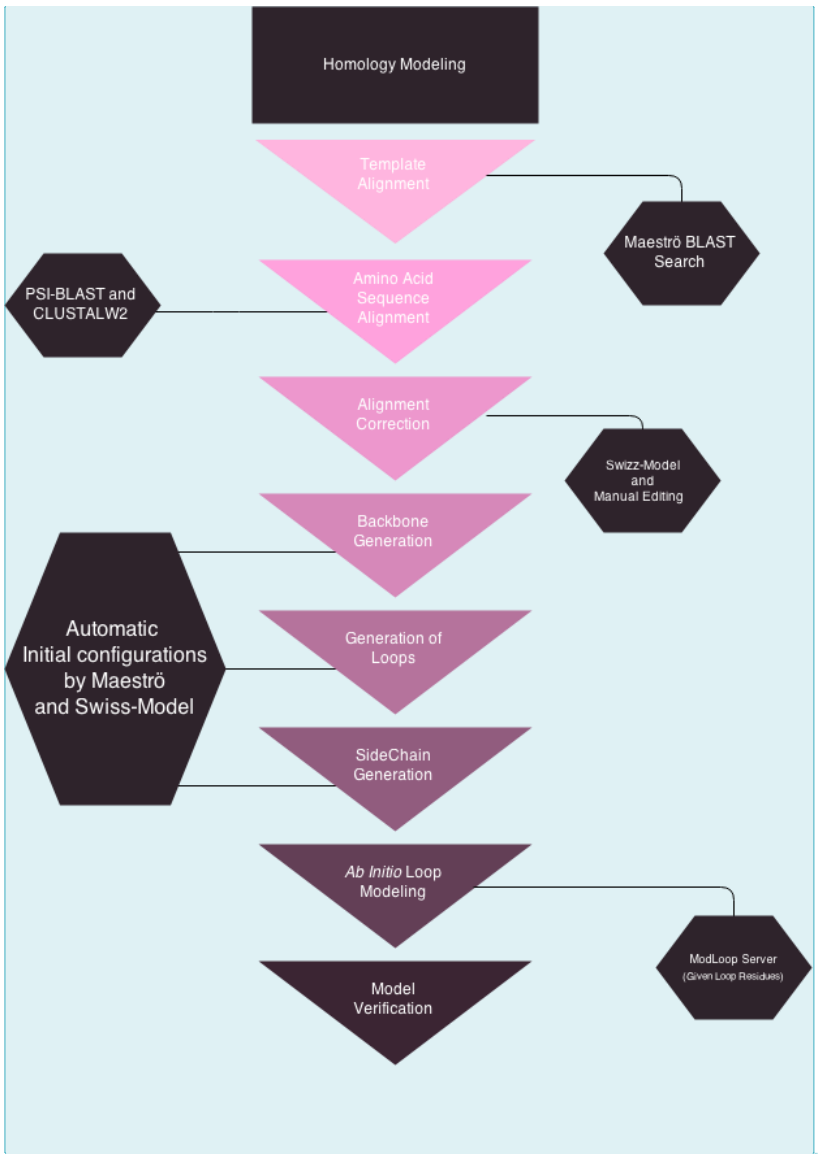

Figure 1. Flowchart of the steps in Homology Model (Template-Based Modeling)

Commented [SS1]: Side chains were added either using YASARA or PELE

This homology model was made using Maeströ (Schrodinger LLC), MacPyMOLEdu, CLUSTALW/CLUSTAL OMEGA, Swiss Model, ModLoop and PELE. Since there is no reserve for the crystal structure of ADORA, other sequences from this superfamily (7TM_1) were used as template to construct a 3D molecular structure of ADORA2B.

PSI-BLAST was then used to screen false positives and obtain other alignments that were previously not visible, with the aid of three PSSM iterations. SWISS MODEL was used to align the sequences of the homologs and secondary structures were generated from the highest sequence identity matches. Multiple sequence alignments were useful for placement of deletions or insertions only in areas where the sequences are strongly divergent (loops).

The most important part of the protein is the backbone, which consists of the main chains (or helices) of the protein. The backbone was generated using the secondary structure of the multiple ADORA2A sequences, in alignment with the template sequence. The combination of SWISS MODEL and Maeströ were used to combine several different algorithms to predict the backbone torsion angles and dihedrals in the sequence. The regions of the model-sequence that contain a homologous domain in the PDB were modeled while those parts without were predicted *de novo* by manually mutating residues after looking at sequence alignments from CLUSTALW2. This method compared small fragments of the sequence with the PDB and inserted them with the same local conformation into the model. Inserting appropriate residues into a continuous backbone closed the gaps in the sequences. The insertions were shifted to the left, out of the helices, and strands and placed into elements that accommodate structural changes such as loops and turns. Gaps were identified and treated using a PDB residue search for loops containing congruent endpoints to those where the residue was inserted. As insulin was the only concerning ligand, which did diffuse through the protein itself, other binding pockets were not considered (A. Sakthikumar, personal communication, January 19, 2015). The identified coordinates of the loop were

then transferred. ModLoop by MODELLER was used to create a new PBD text file with minimized loops. For short loops (five to eight residues), the small fragments were compared to the PDB (as implemented by the Rosetta Method), which resulting in an *ab initio* modeled loop. The loops being the most variable part of the protein were not conserved. By implementing algorithms used by ModLoop, the loops' conformations were changed. Therefore, torsion angles, covalent bonds, non-covalent bonds, and rotamers were optimized.

Next, the side chains of the protein, which determine the conformation of the entire polypeptide, due to the intramolecular interactions were modeled. The side chain conformations (rotamers) of globally conserved residues were extracted and placed in the Cγ location, in between the C_α-C_β bond, which has the same orientation. The model was then optimized using an iterative model to predict minimized energies with a self-parameterizing force. To reach experimental accuracy, the expected minimization proceeded all the way down to ~0.5 Å, which is the uncertainty in experimentally, determined coordinates. The side chains were configured and minimized using unconstrained exploration and binding site search (necessary in order to avoid alterations in the active site) with YASARA.

The validation of the model was the last step, where the protein sequence identity between the template and the model-sequence was determined as a percent using Maeströ. The inside/outside distributions of the residues were checked, to detect completely improperly folded regions. Maeströ and MacPyMOLEdu were used to measure radial distanced for given atoms, bond lengths, and bond-and torsion angles. A basic protein report was created. Ultimately, PROSESS was used to validate the overall quality, covalent bond quality, non-covalent (packaging) quality, and torsion angle quality. Graphs were generated with means and standard deviations.

Results/Data Analysis

All data was collected through PROSESS, a structure validator. The overall quality of the structure of ADORA2B was a value of 5.5 on a scale of one to ten. Even thought the torsion angles, which represented the two dihedrals between the bonds of residues in the backbone, loops, and side chains had a value of 7.5; it did not impact the overall quality very much. The non-covalent bonds, which make up the majority of all 3D protein structures, were at a value of 4.5, therefore lowering the quality of the entire model. This makes up all the side chains of the proteins, which are the most variable parts of the entire structure, thus altering the entire protein slightly. The percentages for abnormal bond lengths were .17, along with multiple outliers in CA-C-O bonds and CA-C-N calculated to a high covalent bond quality. All the torsion angle energies were a high .09 percent, and 93 percent of residues that fell in the favored region of the Ramachandran plot as shown in Figure 4. This together achieved a score of 7.5. The threading scores of the residues as shown in Figure 3 are in the acceptable region, between one standard deviation and the mean. This signifies that the local environment of the protein was assessed well during structural loop and side chain alignments.

The b-factor measured from the proteins is quite consistent, with a pattern of drops after approximately every twenty-five residues. The Ramachandran plots indicate that generally, the majority of the residues fell in the favored regions, and only a few landed in the unfavorable regions.

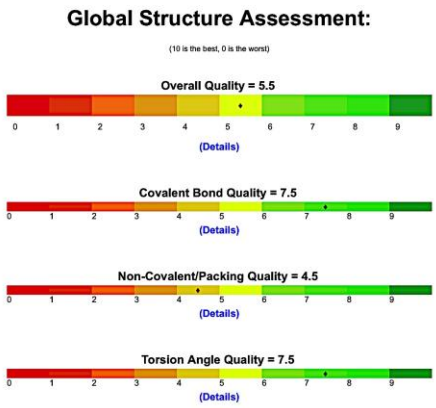

Figure 2. Above the Global Structure Assessment of the homology model of ADORA2B is given. Overall Quality is 5.5.

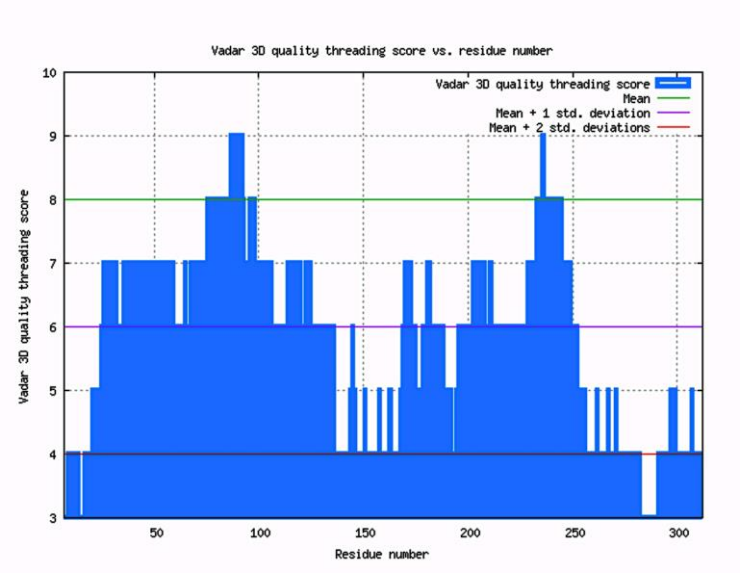

Figure 3. Above the Vadar Threading Score is given to all the residues in the homology model of ADORA2B. The local environment, packing and hydrophobic energy of the given structure are part of this calculation.

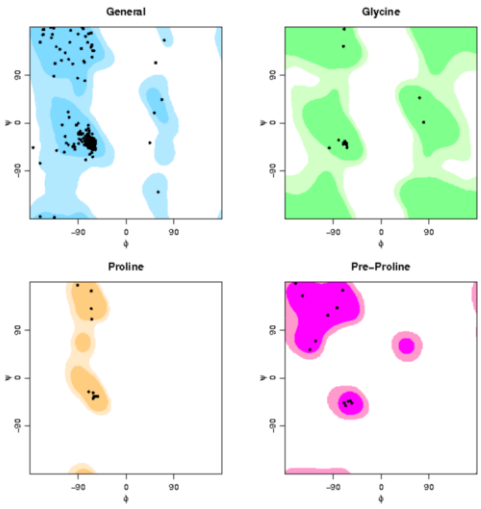

Figure 4. Ramachandran plot are shown above for 3 amino acids as well as the general plot for the entirety of the construction.

Applications and Future Extensions

As for future experimentation, the first task is to delve into wet lab experiments using the appropriate cell line: 3T3-L1, along with insulin dosages in various concentrations to allow for the blockade of the ADORA2B and IR receptors. Then, and ELISA immunoblot assay will be conducted to see which protein is responsible for the hindrance of the cascade of glucose molecules. Also, to decrease the chances of an inaccurate homology model, a verified lab will be contacted to aid in the process of creating a scaffolded tertiary structure of the GPCR. The model will be further refined to conduct a MD simulation as well as perform ligand-docking studies with Insulin, and the various antagonists of ADORA2B. As a continuation, the researcher would also like to focus a more drug-discovery related aspect of this project, where algorithms that search for the accurate antagonist, can be processed real-time in a lab. This way, the theoretical aspect of the

project can be visualized easily. Plus, a discovery like this would make way for synthesized pharmaceutical drug manufacturers to conduct clinical trials using the conformations of the antagonists found from the study; they would test for a decrease in insulin resistance from ADORA2B blockade.

References

1. Diabetes (type 1, 2, gestational). (2014), Retrieved September 18, 2014 from http://www.uchospitals.edu/online-library/content=P01513
2. Duckworth WC, Bennett RG, Hamel FG (1998). "Insulin degradation: progress and potential". Endocr. Rev. 19 (5): 608–24.
3. Epstein CJ, Goldberger RF, Anfinsen CB (1963). *Cold Spring Harbor Symp. Quant. Biol.* 28:439
4. Kim D E, Chivian D, Baker D (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*. 32:W526-W431
5. Figler, R., Wang, G., Srinivasan, S., Jung, D., Zhang, Z., Pankow, J., ... Linden, J. (2011). Links Between Insulin Resistance, Adenosine A2B Receptors, and Inflammatory Markers in Mice and Humans. *Diabetes, 60*, 669-679.
6. Hansen, K (1991). Determinants of microvascular complications. In J.Pickup and G. Williams (Eds.), *Textbook of Diabetes* (pp. 519-525). Oxford: Blackwell Scientific Publications.
7. Karnik, S. K., Chen, H., McLean, G. W., Heit, J. J., Gu, X., Zhang, A. Y., Fontaine, M., Yen, M. H., and Kim, S. K. (Nov. 2, 2007). Menin controls growth of pancreatic β-cells in

pregnant mice and promotes gestational diabetes mellitus. *Science*, 318 (5851), 806-809. Retrieved from: http://www.jstor.com

8. Kjeldsen, T., Anderson, A., & Wiber, F. (1991). The Ligand Specificities of the Insulin Receptor and the Insulin-Like Growth Factor I Receptor Reside in Different Regions of a Common Binding Site. *Proceedings of the National Academy of Sciences, 88*, 4404-4408.
9. Longo N, Wang Y, Smith SA, Langley SD, DiMeglio LA, Giannella-Neto D (2002). "Genotype-phenotype correlation in inherited severe insulin resistance". Hum. Mol. Genet. 11 (12): 1465–75.
10. Maegawa, H. (03/1994). "Insulin Receptor Kinase Phosphorylates Protein Tyrosine Phosphatase Containing Src Homology 2 Regions and Modulates Its PTPase Activity in Vitro". Biochemical and biophysical research communications (0006-291X), 199 (2), p. 780.
11. Monod, J. (1977, January 1). Introduction to Protein Homology / Comparative Modeling, Step in Homology Modeling. Retrieved February 18, 2015.
12. Nichols, E. K. (1988). *Human gene therapy.* Cambridge, MA: Harvard University Press.
13. Rost, B (1999). Twilight zone of protein sequence alignments. *Protein Eng*. 12:85- 94.
14. Tooke, J. E., and Shore, A. C. (1991). Regulation of microvascular function. In J.Pickup and G. Williams (Eds.), *Textbook of Diabetes* (pp. 546-553). Oxford: Blackwell Scientific Publications.
15. Selim, F. (2011). The Human Adenosine A2B Receptor: Homology Modeling, Virtual Screening, and Computer-aided Drug Design. 34-124Whaley, L. F. (1974). *Understanding inherited disorders.* Saint Louis, MI: The C. V. Mosby company.

Commented [SS2]: Put random ones n the QQ sections.