Modelling of G-Protein Coupled Receptors Using GPCR-I-TASSER

Ali Farhat, Yang Zhang PhD

Department of Computational Medicine and Bioinformatics - University of Michigan Medical Center

Introduction

G-Protein Coupled Receptors (GPCRs), the largest and most diverse proteins located on the surface of eukaryotic cells, are essential cell-surface receptors that serve as a communicator for messages in the form of light energy, peptides, lipids, sugars, and proteins. When activated, GPCRs undergo a conformational change that stimulates other receptors within the phospholipid bilayer. Accurate GPCR structure prediction poses a challenge, as the acquisition of high resolution experimental data remains nearly stagnant. There have been various computational approaches that may produce high-resolution models; however, they are limited to access of a homologous template. In order to address these problems, members of the Zhang Lab overseen by Dr. Yang Zhang aside other primary investigators have programmed a hybrid protocol to construct accurate G-protein coupled receptors models: GPCR-I-TASSER (GPCR Iterative Threading Assembly Refinement).

This project involves the inquiry and acquired knowledge base of protein structure and function. The use of GPCR-I-TASSER and developing a thorough understanding of how to run the program, develop accurate protein structure, and learning to analyze them is the primary objective of the project.

Methods

Template Retrieval:

- GPCR-I-TASSER uses two templates: the FASTA file list of amino acids and the residue coordinates. The Protein Database (PDB) provides the FASTA file or it can be transcribed from the Zhang Lab database that has already been pre-aligned without homologous templates.
- The FASTA file is then used to develop the residue coordinate .gpcr file. The cluster's terminal is directed to the Zhang Labs bin folder, there the mk_gpcr.pl program is run, inputting the .txt file. This program, using Perl coding software, analyzes the sequence and develops coordinates for each amino acid.
- The job is named based upon the proteins coded name, whether it has a four digit PDB ID, a six digit UniProt ID, etc.

Running GPCR-I-TASSER:

- The job name is listed under the running list in the GPCR-I-TASSER log directory. Via the command terminal, GPCR-I-TASSER is activated using the ./run_git.pl Perl program, and all proteins that are on the running list are set into the query.
- The program will start running each protein structure shortly. GPCR-I-TASSER threads the sequence through the Local Meta Threading Server (LOMETS) and outputs threaded templates.
- If the homologous templates are located and identified in the server, the template-based assembly procedure is applied; consequently, a full-length model is constructed. If a homologous template is not identified, the procedure is extended from I-TASSER with a GPCR-specific, knowledge-based force-field in order to produce a template for structural modeling. An ab initio transmembrane helix folding procedure is used to assemble the seven transmembrane helices bundle from scratch.
- Final simulations undergo fragment-guided molecular dynamic atomic refinement in order to develop the final structure. Each protein needs approximately 72 hours to complete structure development, and a total of between three to five structures are developed.

Conclusions

- GPCR-I-TASSER structure predictions with the use of homologous templates generate a higher TM-score and lower RMSD providing evidence of highly accurate models.
- Non-homologous template-structure predictions have lower TM-scores when run with GPCR-I-TASSER due to the ab initio transmembrane helix folding procedure.
- The extra steps in the production of model prediction are not completely accurate, as it requires the variable of uncertainty compared to a homologous template; therefore, the TM-score will naturally be lower.
- Some models produced by GPCR-I-TASSER have a very low TM-score (e.g. 4RWA having a 0.1619 TM-score and a RMSD of 27.900). Values in this nature indicate zero correlation between the model and the native structure.
- The alignment program named TM-align was used to match the position of the template structures, allowing for a more accurate TM analysis to be carried out.
- The new 4RWA TM-score was then confirmed to be 0.76624, having an RMSD of 3.19. These values indicate similarity with some error between the model and native structures.
- Initial unnaturally low TM-score and high RMSD can be traced back to a difference in positioning between the models.

Results

Native Structure Found in the Protein Data Bank | Primary Model (Model 1) Developed by GPCR-I-TASSER | Non-Homologous Model Developed by a Python Program

TM Score Comparison (Post-Alignment): 0.76905
Root-Mean Squared Deviation (RMSD): 3.50

TM Score Comparison (Post-Alignment): 0.95536
Root-Mean Squared Deviation (RMSD): 1.52

Figure 1. 4U14 Model Protein Analysis

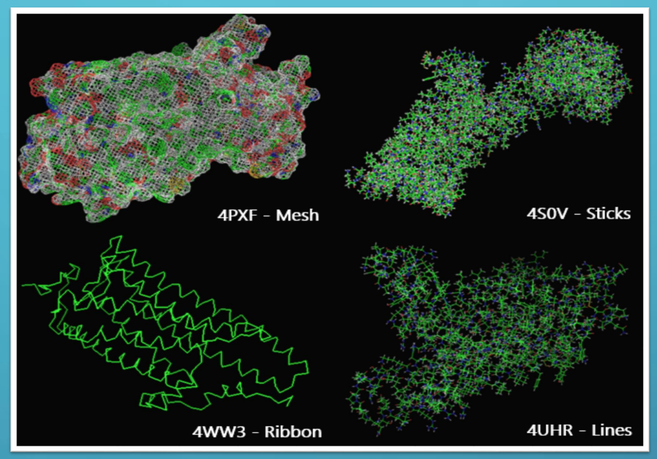

Figure 4. Different Forms of Model Analysis

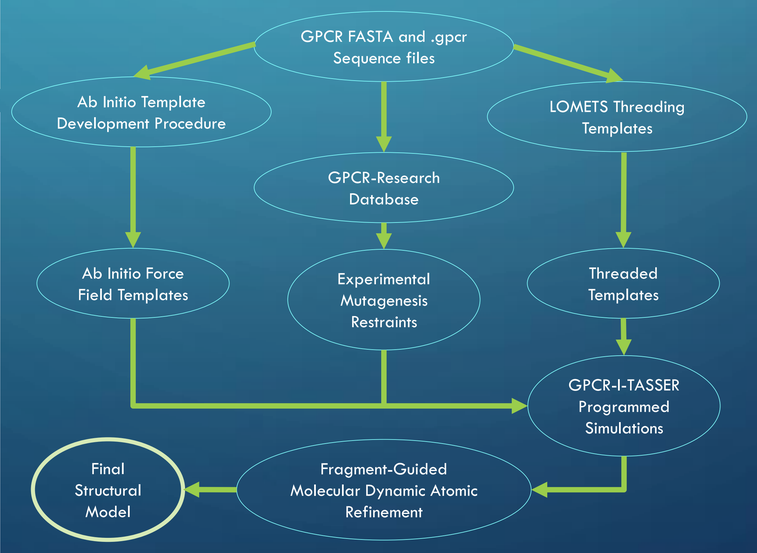

Figure 3. Standard Pipeline Used to Develop Models

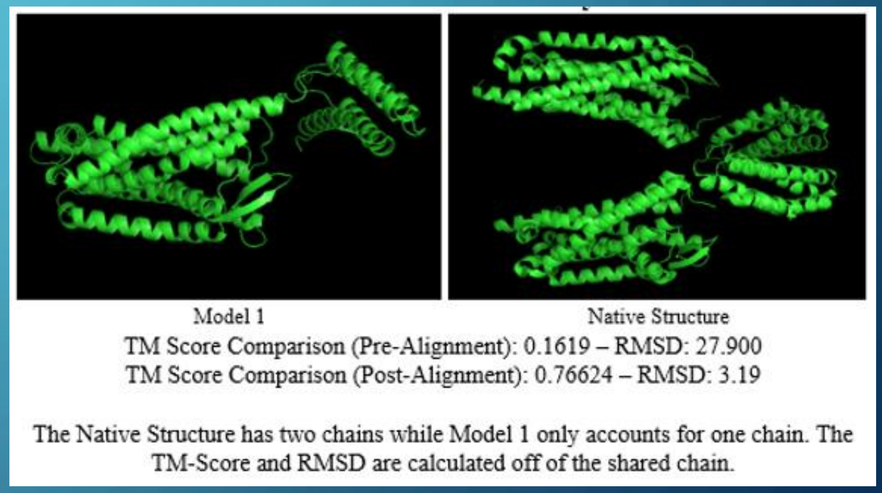

Figure 2. 4RWA Model Protein Analysis

Looking Ahead

- A potential upcoming project would involve updating the GPCR Human Genome Database by rethreading newly developed templates from proteins whose structures have been experimentally solved since the last update of the database.
- With the newly generated threads, the development of specific protein structure using GPCR-I-TASSER will be much more accurate.
- Updating the GPCR-Human Genome Database by developing predictions of higher accuracy allows for more efficient identification of proteins and better application in the medical field.
- Many diseases that involve proliferation, differentiation, angiogenesis, malfunctions in development, and numerous types of cancers involve GPCR errors; consequently, GPCRs are among the most popular targets of pharmaceutical drugs.
- In-depth analysis of GPCR structure prediction is vital for providing the accurate models for utilization in medical applications and pharmaceuticals.

References

Zhang, Jian, Jianyi Yang, Richard Jang, and Yang Zhang. "GPCR-I-TASSER: A Hybrid Approach to G Protein-Coupled Receptor Structure Modeling and the Application to the Human Genome." *Structure* 23.8 (2015): 1538-549. Print.