Opinion

Protein structure prediction by computational homology modeling: a brief explanation

Rafael Trindade Maia
Associate Professor, Federal University of Campina Grande, Brazil

Correspondence: Rafael Trindade Maia, Associate Professor, Federal University of Campina Grande, Brazil, Email rafael.rafatrin@gmail.com

Received: March 07, 2024 | **Published:** September 24, 2024

Introduction

Proteins reign supreme in the realm of life, orchestrating a symphony of complex functions and structures within every living being. These molecular machines, composed of amino acid polymers, are the architects of biological wonders. Unraveling their three-dimensional mysteries is key to unlocking their true potential. Yet, traditional methods like X-ray crystallography and nuclear magnetic resonance (NMR) come with a hefty price tag of complexity and expense. Here will be showed a little on the realm of computational biology, where homology modeling emerges as the beacon of hope, offering a cost-effective and powerful alternative for constructing protein structures with precision and prowess.[1]

Proteins are the architects of complexity, sculpting intricate structures at various levels: 1) the elemental primary structure; 2) the secondary structure; 3) the tertiary structure; and 4) the quaternary structure (Figure 1). The primary structure is the backbone, a linear sequence of amino acids, where the polypeptide begins with the carboxyl group at one end and concludes with the amino group at the other. It's like reading a molecular story encoded in letters, each representing a peptide link. From this raw material, the secondary structure emerges, influenced by the primary sequence, weaving together helices, turns, and β-sheets in a ballet of molecular elegance. But it's the tertiary structure that truly mesmerizes, as these secondary structures fold and twist into a unique three-dimensional masterpiece, defining the protein's biological destiny. And in the realm of multimeric protein assemblies, such as dimers and trimers, we witness the epic formation of the quaternary structure, where these molecular giants unite, forging oligomeric complexes of unparalleled significance.[2]
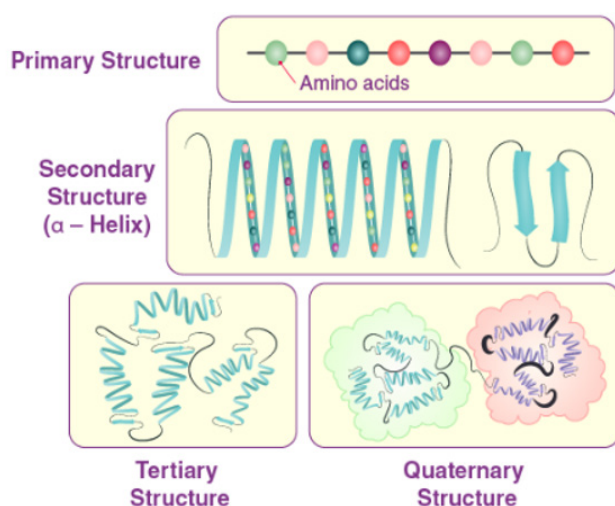
Figure 1 Protein structure organization levels.

In the universe of protein prediction with three thrilling ways: the *Ab initio/De novo* approach, the Threading technique and the Homology Modeling, also called as Comparative Modeling.

Homology modeling, a stalwart among them, operates on the principle that protein structure remains remarkably conserved despite sequence variations. Even as genetic letters shuffle, the architectural blueprint stays "intact", preserving function like a molecular guardian. It's a game of genetic kinship, where proteins from the same lineage share structural secrets, enabling comparative modeling. When two proteins boast homology, they're not just genetic cousins—they're structural siblings, flaunting similar motifs in their molecular makeup. And when faced with an enigmatic protein lacking a structural identity, but boasting homology to a known counterpart, we fashion a three-dimensional blueprint using the known structure as a scaffold. With a mere 25% amino acid identity threshold, we embark on crafting models, but surpassing 40% sparks excitement, and anything beyond 50% yields theoretical marvels fit for scientific epics.[3]

Beyond mere amino acid matches, the quest for the best protein template demands a discerning eye for additional parameters. Picture this: resolution in angstroms and alignment coverage percentages emerge as crucial players in our protein modeling saga. When it comes to crystallographic structures, lower resolutions signal higher quality-yes, you heard that right. While the Protein Data Bank (PDB) flaunts an average resolution of 3.5 Å, structures with resolutions under 2.0 Å emerge as rare, comprising less than a mere 10% of the database. But that's not all; alignment coverage emerges as the unsung hero, with percentages soaring beyond 90% heralding excellence in protein matches, igniting scientific fireworks and paving the way for groundbreaking discoveries.[4]

Gaps deserves a spatial attention. A gap signifies a void in the amino acid lineup, a deletion in the molecular manuscript. But here's the kicker: the number and size of these gaps hold the key to model quality. Brace yourself for a revelation: more gaps, bigger gaps, equals less reliable models, and a higher risk of stumbling upon structural illusions. So, when hunting for the perfect template, researchers must heed the call of the gaps, lest they fall prey to molecular mischief. Once the template is locked in, it's time to unleash the three-dimensional modeling frenzy. Cue the specialized programs and servers, where we submit the blueprint for molecular magic. Picture this: structural carbons of the target protein seamlessly align with those of the template, guided by the intricate dance of amino acid alignments.

Refinement and validation

Step into the realm of protein prediction with homology models–where theory meets reality in a whirlwind of validation, refinement, and optimization. Brace yourself for the ultimate test: the Ramachandran plot, a revered tool that peers deep into the stereo chemical soul of protein structure. Ramachandran plot–analyzer of *phi* and *psi* angles that sorts them into regions. The rule is clear: a model worth its salt should boast a staggering 90% of its residues nestled snugly in favorable and permitted zones.[5] But there's more: energy assessments, both local and global, offer highlights into the molecular balance, that can be performed by PROSA-web quality. The Z-score scrutinizes a structure's energy against a database of its peers, setting the stage for a showdown of molecular prowess.[6]

The accuracy of computational modeling tools is a critical factor in the success of homology modeling. Various studies have evaluated the performance of these tools, revealing that their accuracy can vary based on several parameters, including the quality of the template, the degree of sequence identity, and the specific modeling algorithm used. Homology modeling tools like MODELLER, SWISS-MODEL, and I-TASSER have been benchmarked against known structures, demonstrating varying levels of accuracy. For instance, studies have shown that models produced by these tools often achieve root-mean-square deviation (RMSD) values of less than 2.0 A when the template and target share more than 30% sequence identity. As sequence identity increases—particularly above 50%—the RMSD values typically decrease, indicating greater accuracy and reliability of the model.

Additionally, the quality assessment of models can be corroborated by metrics such as the Ramachandran plot, where models with more than 90% of residues in favored regions are considered high-quality. Furthermore, tools like ProSA-web provide Z-scores that help benchmark the energy profile of a model against a database of known structures, offering insight into its potential accuracy. The introduction of advanced algorithms and machine learning techniques has further enhanced the predictive capabilities of modeling tools. For example, AlphaFold, developed by DeepMind, has set new standards in accuracy for protein structure prediction, achieving high-resolution predictions that often rival experimental methods. Its performance has been validated in several community-wide assessments, solidifying its reputation as a transformative tool in structural biology. In the context of structure refinement two techniques stand out as beacons of enlightenment: energy minimization and classical (atomistic) molecular dynamics. Energy minimization, revered as the optimization of geometry, embarks on a quest to uncover a precise arrangement of atomic coordinates that steer clear of detrimental collisions while concurrently lowering the system's potential energy. Behold, there exist sanctuaries of computational prowess offering free access to energy minimization tools for theoretical models, such as the YASARA and the CHIRON web servers.[7,8]

Molecular Dynamics simulations is another path. This technique, rooted in the principles of Classical Mechanics, orchestrates the atomic ballet of a system through the harmonious integration of Newtonian equations of motion. Thus, a molecular dynamics simulation spanning 5–10 nanoseconds emerges as a veritable cornerstone in the edifice of model optimization and validation by homology. To perform it, tools such as GROMACS and NAMD are excellent options. Once honed and validated, the theoretical model transcends its humble origins, poised to serve myriad scientific endeavors. Its legacy may be enshrined within public repositories of knowledge, such as the esteemed PMDB-Protein Model Data Base

and the venerable SWISS-MODEL repository, forever immortalized for the benefit of scientific inquiry.[9–12] A practical example of the efficacy of computational homology modeling is the study of the human serotonin transporter (SERT), a critical protein involved in the reuptake of serotonin from the synaptic cleft, influencing mood and behavior. Understanding the structure of SERT is vital for drug development, especially for antidepressants and other psychiatric medications. In a study,[13] researchers aimed to model the structure of SERT using homology modeling techniques. They utilized the crystal structure of the bacterial homolog, LeuT, as a template due to its high sequence similarity (approximately 40% identity) with SERT.

The researchers employed MODELLER to generate multiple homology models of SERT based on the LeuT template. They carefully assessed the generated models using Ramachandran plots and energy minimization techniques to ensure that the models adhered to stereochemical constraints and exhibited favorable conformations. Validation of the models revealed that over 90% of the residues were located in the favored regions of the Ramachandran plot, indicating a high-quality structure. Additionally, the energy profiles obtained from ProSA-web showed favorable Z-scores, confirming the models' stability. The refined SERT models were then used to conduct molecular docking studies with various serotonin reuptake inhibitors, providing insights into the binding interactions and affinities of these compounds. The results were consistent with experimental data, demonstrating the models' predictive accuracy and practical utility in guiding drug discovery efforts. Despite the advantages of homology modeling, several limitations and challenges persist. One major challenge is the accuracy of the template: if the template structure is not representative of the target protein, the resulting model may be inaccurate. Additionally, homology modeling typically relies on sequence alignment, which can introduce errors if conserved regions are misaligned. The quality of the generated model can also be affected by the degree of sequence similarity; lower identity can lead to less reliable predictions.

Furthermore, while energy minimization can help refine the models, it may not fully capture the dynamic nature of protein conformations or the effects of ligand binding. Lastly, experimental validation of the models remains essential, as computational predictions cannot replace empirical data. Overall, while computational homology modeling is a powerful tool in structural biology and drug design, its limitations highlight the need for careful selection of templates, rigorous validation of models, and integration with experimental techniques.

Conclusion

Behold the marvels of theoretical-computational modeling: swift, cost-effective, and astonishingly adaptable. Within their digital confines lie boundless realms of exploration and application through the lens of homology. These virtual constructs serve as invaluable tools for an array of endeavors, including drug discovery, docking studies, drug and vaccine development, unraveling the mysteries of catalytic and allosteric binding sites, conducting molecular dynamics simulations, probing the quantum realm, and engineering biomolecules to name but a few. The vista of molecular modeling beckons with a promise both captivating and auspicious. As computational prowess burgeons, the veracity and reliability of theoretical models ascend to ever-greater heights. Their burgeoning accuracy fuels a renaissance in biological and biotechnological research, weaving a tapestry of insights that transcends disciplinary boundaries, seamlessly integrating with the realms of bioinformatics and computational biology. Truly, the future of molecular modeling is ablaze with possibilities, illuminating the path toward unprecedented scientific discovery and innovation.

Acknowledgments

None.

Conflicts of interest

The author declare that there are no conflicts of interest.

References

1. Riley R, Ochoa MB. The state of structural bioinformatics: a survey of progress and challenges. *Current Opinion in Structural Biology*. 2020;64:56–63.

2. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signaling and regulation. *Nat Rev Mol Cell Biol*. 2015;16(1):18–29.

3. Cameau DL, Berrisford JM. Advances in protein modeling and prediction. *Nat Rev Mol Cell Biol*. 2018;19(4):199–211.

4. Zhu X, Liao Y. Alignment coverage in homology modeling and its impact on structural predictions. *Journal of Structural Biology*. 2019;206(3):120–126.

5. Murray LJ. Evaluation of protein structure quality and validation methods. *Journal of Structural Biology*. 2018;204(2):195–206.

6. Wiederstein M, Sippl MJ. ProSA-web: ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*. 2007;35:W407–W410.

7. Krieger E, Vriend G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics*. 2014;30(20):2981–2982.

8. Hedberg C, B Johnson. CHIRON: a web server for energy minimization and conformational analysis. *Bioinformatics*. 2018;34(15):2563–2565.

9. David VD, Lindahl E, Hess B, et al. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26(16):1701–1718.

10. Phillips JC, Braun R, Wang W, et al. *Scalable molecular dynamics with NAMD*. *J Comput Chem*. 2005;26(16):1781–1802.

11. Tiziana C, Paolo D, Domenico C, et al. The PMDB protein model database. *Nucleic Acids Res*. 2006;34:D306–D309.

12. Biasini M, Bienert S, Waterhouse A, et al. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 2014;42(W1):W252–W258.

13. Zhou Z. Homology modeling of the human serotonin transporter and identification of binding sites for antidepressants. *Journal of Medicinal Chemistry*. 2009;52(12):4048–4056