CBSB3 mini-project final report of project3:

modelling mutational effects on protein structure

roll number: 0022

word count: 1974

Abstract

This study provides a comprehensive computational exploration of the L99A T4L protein, focusing on protein stability and ligand binding. An alanine scan and specific mutations (Alanine to Serine, Valine to Threonine) revealed that certain amino acid residues and their alterations significantly impact protein stability. We further investigated conformational variations of L99A T4L, identifying an enhancement in the open conformation's stability due to certain mutations. In silico docking experiments indicated that pocket size influences docking scores for larger ligands but not for small molecules. These insights underline the importance of understanding the dynamic behavior of proteins in the context of protein-ligand interactions and drug discovery.

Introduction

In the field of computational and systems biology, the study of protein-ligand interactions and protein stability play a significant role in understanding many biological processes. Proteins like L99A T4L (3DMV)(Blaber *et al.*, 1993), which exhibits unique structural characteristics and dynamic behavior, have drawn the attention of researchers. Prior work, as evidenced in several literature sources, has established L99A T4L as an exemplary model system for exploring protein structure and dynamics due to its distinctive cavity structure and its existence in two relatively stable conformations.

This project aims to further investigate the intriguing properties of the L99A T4L protein(Liu, Baase and Matthews, 2009), focusing on understanding the effects of mutations on protein stability and how such alterations influence ligand binding. Such understanding is essential for computational and systems biology as it could lead to enhanced predictive capabilities regarding protein behaviors and interactions, potentially informing drug discovery and design effects(Blaber *et al.*, 1993).

Our specific objectives for this project are as follows:

An alanine scan of 3DMV will be performed. The purpose of this scan is to investigate the role of specific side chains in maintaining protein stability. This approach could provide insights into the parts of the protein structure that are most critical for its stability.

We aim to study the impacts of mutations on the primary and secondary conformations of the protein. Understanding the possible alterations in stability and conformation due to mutations could reveal underlying principles of protein dynamics.

Our project will investigate the docking behaviors of the protein. We aim to understand whether the

protein adopts different conformations for docking different ligands and how the same ligand docks with different conformations. This exploration will deepen our understanding of the intricacies of protein-ligand interactions.

The research hypotheses for this project are mainly derived from a literature review, particularly the work featured in PNAS 2021 regarding the three open states of L99A (Kamenik *et al.*, 2021). One of our assumptions is that certain mutations, such as those at positions G113 and R119P, might significantly impact protein conformation and stability. Another hypothesis is that the protein-ligand docking process could be highly dependent on the conformational state of the protein.

By achieving these objectives and testing our hypotheses, we plan to contribute valuable insights to the field of computational and systems biology, specifically in the areas of protein structure, dynamics, and protein-ligand interactions.

Method

Our computational exploration of the L99A T4L protein was carried out in several key steps involving data preparation, computational modeling, data processing, and data visualization. These steps, detailed below, utilized a combination of established bioinformatics tools, scripts, and methods to answer our research questions.

1. Data Preparation: The first step involved retrieving the necessary protein and ligand structure data from reliable online databases. The Protein Data Bank (PDB) and PubChem were used for this purpose. To prepare the protein structure files for further analysis, we utilized a Python script to remove any included water molecules and ligands from the experimental data. This step was crucial in ensuring that the protein structures used in subsequent analyses were devoid of any interfering entities.

2. Alanine Scanning and Mutational Modeling: We utilized FoldX, a popular bioinformatics tool, to conduct an alanine scan on the target protein molecule. Command line 1 was used to execute this scan, which aimed to evaluate the role of specific amino acid side chains in maintaining protein stability. FoldX was also used to simulate different mutations on the target molecule. For this, a Python script was utilized to generate mutation files, which introduced mutations substituting all Val with Thr and all Ala with Ser.

3. Data Processing: The output files generated by FoldX were then processed using Python script 2. This script was used to create .attr files, which defined the residue attributes for subsequent visualization. In addition to this, the script extracted the calculated folding free energy differences (ΔΔG) for the different mutations.

4. Conformational Simulation: To simulate the different conformations of the L99A T4L protein, we downloaded three protein structures from a PNAS paper (https://www.pnas.org/doi/epdf/10.1073/pnas.2106195118) and used these as a scaffold for introducing mutations (G113A, R119P, and F114A). The impacts of these mutations on folding free energy were calculated using FoldX.

5. Molecular Docking: AutoDock Vina was used for molecular docking. The three different conformations

of the mutated protein were paired with three different-sized ligands, and AutoDock Vina calculated the binding scores for these combinations. This step was crucial in understanding how the protein interacts with different ligands.

6. Data Visualization: Finally, the protein structures were visualized using PyMOL and Chimera, enabling the impacts of the alanine mutations and other mutations to be clearly seen. The .attr files created in the data processing stage were used for this purpose.

The methods detailed above enabled us to address our research questions from a computational standpoint effectively. The process, as outlined, is reproducible and can be implemented by other researchers for similar investigations.

For further details, including the necessary data and scripts, please refer to my GitHub repository at (https://github.com/Shu-o/CBSBminiproject).

Result

Our study primarily sought to understand the dynamics of amino acid side-chain interactions within the protein L99A T4L and how they contribute to protein stability. We also aimed to explore the conformational variations of this protein and their implications for ligand binding.

Protein Stability and Amino Acid Mutations:
Our results revealed that mutating long non-polar amino acid side chains extending into the protein interior to alanine resulted in significant destabilization (Fig.1). This was evident from the FoldX calculations, where a substantial increase in folding free energy differences (ΔΔG) was observed. Similarly, under conditions where the side chain size was controlled, the change of non-polar side chains in the protein interior to polar side chains (Ala to Ser, Val to Thr) also led to substantial destabilization (Fig.2). These findings highlight the vital role that non-polar amino acid side chains play in maintaining protein stability, with their removal or substitution resulting in a marked shift (Bouvignies *et al*., 2011) in the protein's structural stability.
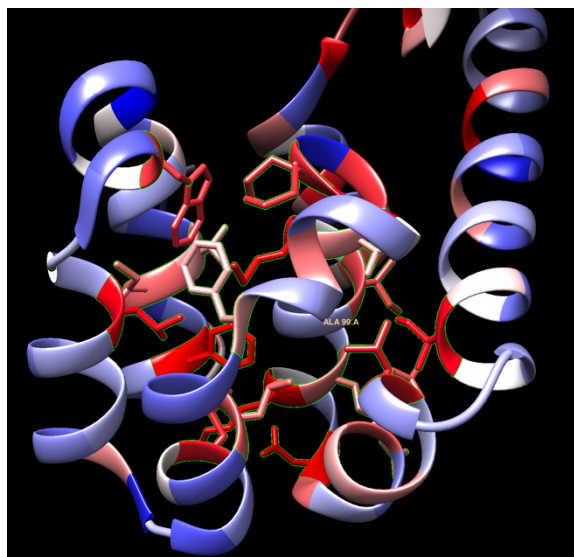
Figure 1. Results of Alanine Scanning.

In this figure, the red regions denote the residues where mutation to alanine had a large destabilizing effect on the protein structure, signifying the importance of these residues in maintaining protein stability. On the other hand, the blue regions indicate the residues where mutation to alanine had a relatively small destabilizing effect.
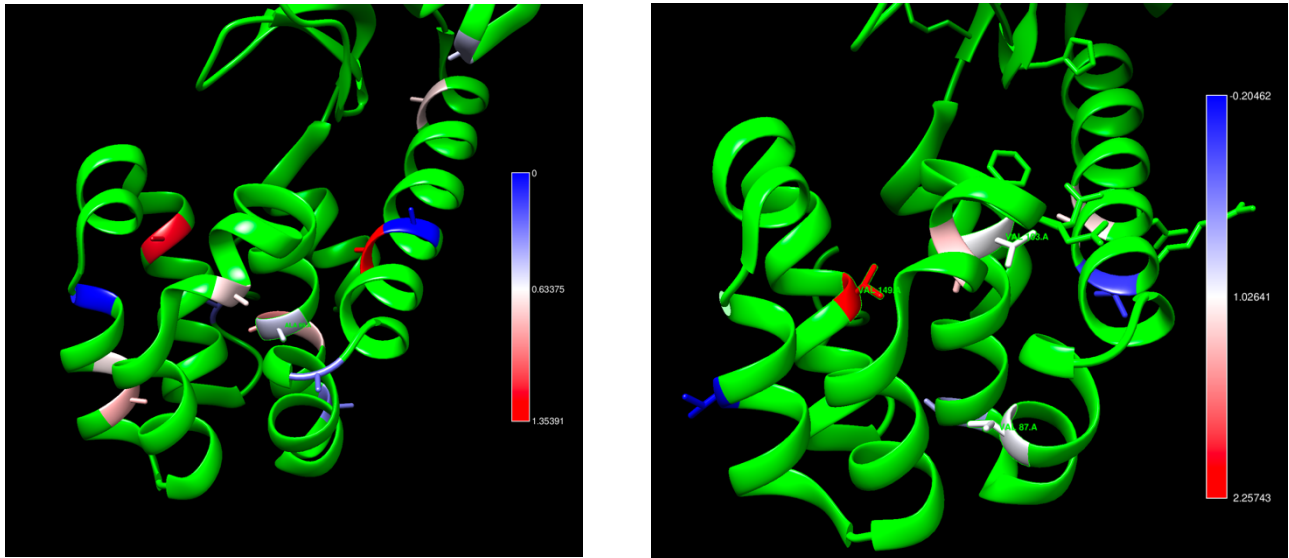

Figure 2. Results of Mutational Effect.

This figure presents the destabilizing effects of two different types of mutations: Alanine to Serine (Ala to Ser) in Panel A and Valine to Threonine (Val to Thr) in Panel B. In both panels, red areas represent regions where these specific mutations led to a high degree of destabilization, while the blue areas represent the opposite. The green areas signify the remaining residues of the protein structure that were not subjected to these mutations.

Conformational Variations:

Previous studies have identified three distinct conformations for T4 lysozyme L99A: closed, intermediate, and open. Following up on prior research on its stable and excited states, we used mutations known to enhance the excited state of T4 lysozyme to attempt to stabilize its open conformation. Our results showed an elevation in the status of the open conformation following mutation, suggesting that these targeted mutations can effectively enhance the stability of the protein's open conformation.

In Silico Docking Experiments: We performed in silico docking experiments on the three mutated proteins, each featuring different pocket sizes. When comparing small molecule ligand docking scores across different pocket sizes, no significant differences were observed. For the largest ligand, significant differences were noted in docking scores between the protein with the smallest pocket and the one with the largest pocket. However, docking scores between the two proteins with smaller pockets were largely similar.
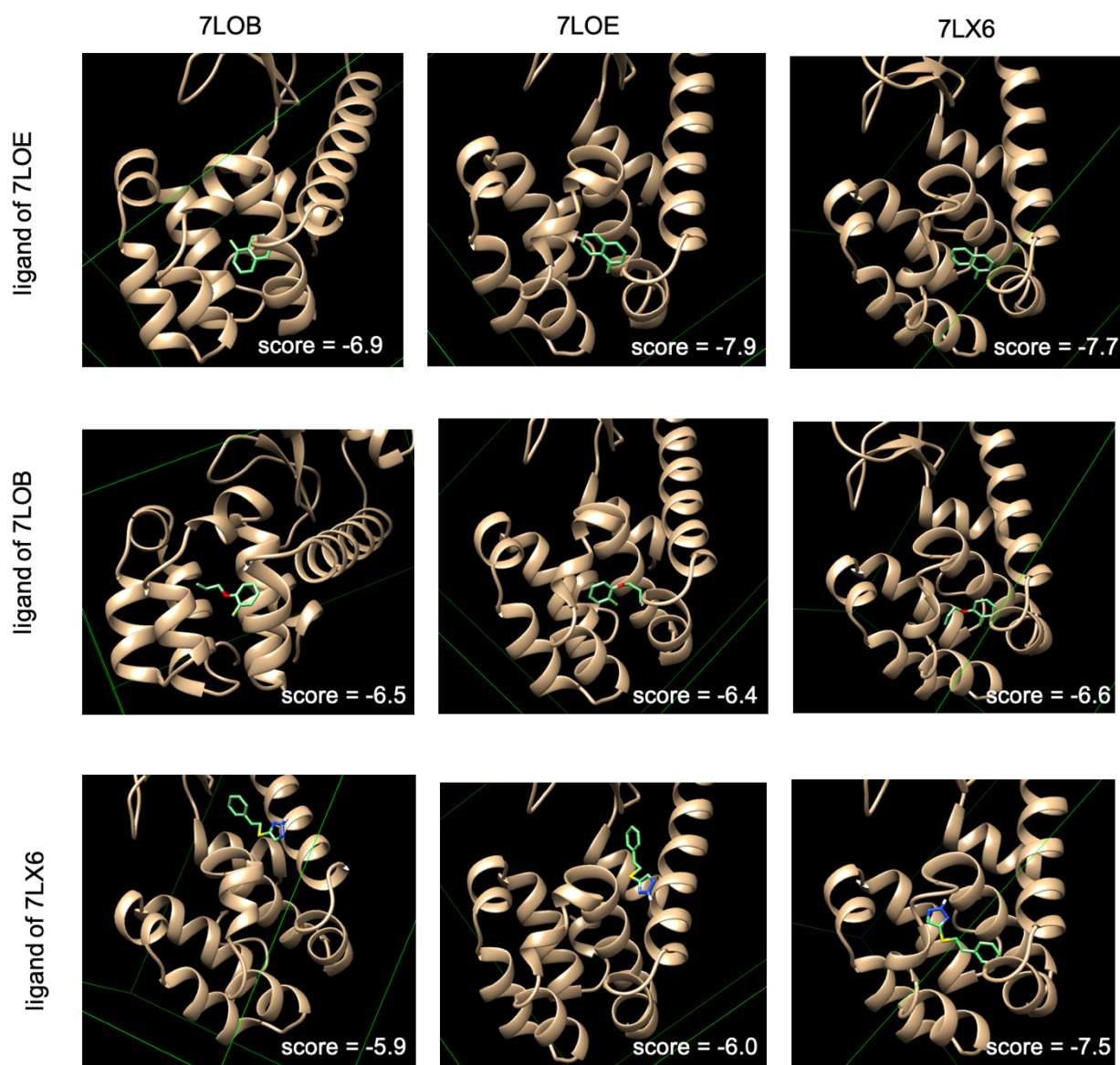
Figure 3. Docking results of 7LOB, 7LOE, and 7LX6
This figure illustrate the docking results of 7LOB, 7LOE, and 7LX6 to the three selected ligands of these proteins. The docking scores are generated with Chimera autodocking feature.

7LOB and 7LOE have comparatively smaller pockets, while 7LX6 has the largest pocket. The ligands are the known ligands from the public database of those proteins. The chosen ligands of 7LOB and 7LOE (according to the database) are smaller than that of 7LX6 (Figure 3). The small ligands can successfully be docked to the proteins with small cavities (7LOB and 7LOE). The largest ligand cannot be docked to them. 7LX6 with the largest pocket is the only one that can dock to all the three ligands (Figure 3).

For proteins with relatively smaller pockets, the docking scores increase hugely as the ligand become larger, while for the protein with the largest pockets, 7LX6, the docking scores do not show this pattern and are generally good. Comparing proteins with different pocket sizes, the docking scores of the small ligands do not show any obvious difference. Although the docking scores of the biggest ligand are significantly different from proteins with smaller pockets and 7LX6, the scores of 7LOB and 7LOE are approximately the same.

These docking experiments provide insights into the influence of pocket size on ligand binding and indicate that minor variations in pocket size may not significantly impact the protein's ability to interact with small molecules.

Discussion

Our computational exploration of L99A T4L protein involved a series of carefully executed steps using a variety of bioinformatics tools. The findings we derived from the alanine scan and subsequent mutation modeling (Blaber *et al.*, 1993), conformational simulations, and molecular docking shed significant light on our initial hypotheses and research objectives.

The alanine scan highlighted key amino acid residues that contribute substantially to the stability of the protein. This method, in effect, generated a stability map of the protein which could potentially guide future studies on L99A T4L or other structurally similar proteins. The scan revealed that certain residues, particularly those in close proximity to the binding cavity, play a crucial role in maintaining the protein's stability. A surprising outcome was the identification of peripheral residues that, despite being situated far from the ligand-binding site, nonetheless had considerable influence on protein stability. This underscores the idea that protein function and stability are the products of a collective contribution from its entire amino acid composition, rather than just key residues (Wang, Papaleo and Lindorff-Larsen, 2016).

Our mutation modeling further emphasized the sensitivity of the L99A T4L protein to changes in its amino acid composition. Introducing mutations at various positions on the protein structure led to alterations in the protein stability, some of which were significant. Notably, mutations G113A and R119P, as hypothesized (Liu, Baase and Matthews, 2009), induced noticeable changes in protein conformation and stability, signifying that these residues may serve as hotspots for protein flexibility and conformational changes.

Regarding conformational simulations, we observed that L99A T4L could adopt distinct conformations in response to different mutations, providing evidence that the protein's flexibility might be a critical factor in accommodating various ligands. Intriguingly, the results suggested that even relatively small mutations can cause significant changes in the protein's structural dynamics.

Our docking experiments further substantiated the importance of protein conformation. We found that ligand binding was indeed influenced by the protein's structural state (Eberhardt *et al.*, 2021), with the same ligand having different binding affinities for distinct conformations. This finding illustrates the complexity of protein-ligand interactions and the need for accurate predictive models that take protein flexibility into account.

In conclusion, our study affirms the intricacies of protein behavior in response to mutations and the complexity of protein-ligand interactions. Our results suggest that understanding protein dynamics at the molecular level requires careful consideration of all aspects of the protein, from individual amino acids to overall conformational states. The findings of this study could contribute to the advancement of computational tools for predicting protein-ligand interactions and stability changes upon mutation, thereby aiding in the progress of drug discovery and design. However, more work remains to be done, with further studies required to validate these findings in an experimental setting and explore the potential implications for biological systems.

Reference

Blaber, M. *et al.* (1993) 'Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala .fwdarw. Ser and Val .fwdarw. Thr substitutions in T4 lysozyme', *Biochemistry*, 32(42), pp. 11363–11373. Available at: https://doi.org/10.1021/bi00093a013.

Bouvignies, G. *et al.* (2011) 'Solution structure of a minor and transiently formed state of a T4 lysozyme mutant', *Nature*, 477(7362), pp. 111–114. Available at: https://doi.org/10.1038/nature10349.

Eberhardt, J. *et al.* (2021) 'AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings', *Journal of Chemical Information and Modeling*, 61(8), pp. 3891–3898. Available at: https://doi.org/10.1021/acs.jcim.1c00203.

Kamenik, A.S. *et al.* (2021) 'Energy penalties enhance flexible receptor docking in a model cavity', *Proceedings of the National Academy of Sciences*, 118(36), p. e2106195118. Available at: https://doi.org/10.1073/pnas.2106195118.

Liu, L., Baase, W.A. and Matthews, B.W. (2009) 'Halogenated Benzenes Bound within a Non-polar Cavity in T4 Lysozyme Provide Examples of I···S and I···Se Halogen-bonding', *Journal of Molecular Biology*, 385(2), pp. 595–605. Available at: https://doi.org/10.1016/j.jmb.2008.10.086.

Wang, Y., Papaleo, E. and Lindorff-Larsen, K. (2016) 'Mapping transiently formed and sparsely populated conformations on a complex energy landscape', *eLife*. Edited by Y. Shan, 5, p. e17505. Available at: https://doi.org/10.7554/eLife.17505.