Investigation of Computational Modeling Accuracy in Malate Dehydrogenase Structure Predictions and Interactions

Andy Kapoor & Dr. Mary von dem Bussche

Abstract

Integrating advanced computational techniques with wet-lab methods presents the potential for studying natural structures and overcoming challenges with traditional techniques, such as accessibility, cost, time, and need for expertise (Darnell). Through a structured computational workflow verified by empirical data, in-silico models offer potential resolutions to these challenges. In this study, the structural effects of the R153C mutation on Malate Dehydrogenase (MDH) were investigated through optimized implementations of AlphaFold, Rosetta Ab Initio, and AutoDock Vina. AlphaFold's success in accurately predicting MDH structures contrasted with Rosetta's inability to generate viable models, illustrating the challenges of current computational approaches. Subsequent docking simulations with AutoDock Vina provided critical insights into the mutation's impact on substrate coordination, revealing an increased affinity for oxaloacetate in the mutant MDH. However, there was decreased structural conformity, as evidenced by the significant rise in Root Mean Square Deviation (RMSD) values between binding modes, which reflects greater divergence from the most energetically favorable conformation. The results emphasize the importance of verifying computational predictions with experimental validation to ensure the precision of the models and gain confidence in their application to laboratory-based research, so computational modeling can supplement laboratory practices with predictive foresight before intensive procedures are implemented.

Introduction

AlphaFold and Rosetta represent significant advancements in computational biology, each utilizing distinct methods for protein structure prediction. AlphaFold, developed by Google’s DeepMind, utilizes deep learning algorithms to predict protein structures, marking a significant

stride in applying artificial intelligence in biological research. Its ability to handle extensive datasets for protein modeling sets it apart as a groundbreaking tool. Rosetta, in contrast, has a longer history in the field and employs biophysical principles (residue statistical potentials, hydrogen bonding, energy minimization, etc.) for protein structure prediction. It is known for its comprehensive suite of programs that handle various tasks in structural biology, making it a robust and versatile tool for researchers (Moretti). In this study, Rosetta is deployed on a Windows PC with OracleVirtualBox hosting an Ubuntu Linux Virtual Machine (VM). To increase the efficiency of the computational pipeline, concurrent modeling in Rosetta will be achieved by designating separate GNU screens for each modeling query to leverage all available computing power of the VM to reduce computation time without compromising accuracy. AlphaFold, accessed through Google Colab's user-friendly and accessible adaptation ColabFold, is deployed within Microsoft Azure's cloud computing platform with HyperDrive, to optimize and apply the full capabilities of the Azure VM for streamlined concurrent execution of multiple simulations.

Malate Dehydrogenase (MDH) is a key enzyme in the citric acid cycle in human cells and the glyoxylate cycle in plants, catalyzing the conversion of malate to oxaloacetate. The enzyme's role is fundamental in cellular metabolism, linking it to various biological processes (Íñiguez, et al). MDH serves as an ideal subject for exploring the capabilities of computational tools, as its well-researched nature in human and plant forms offers a wealth of empirical laboratory data against which model predictions can be validated. Structural and functional aspects of MDH concerning substrate coordination, and simple mutations, have been extensively studied, providing valuable insights into the enzyme's catalytic mechanism. Arg-81, Arg-87, and Arg-153 coordinate oxaloacetate into the active site of MDH, ensuring its proper position for the enzymatic reaction. The residues interact with carboxylate groups on the substrate and facilitate proper orientation and stabilization for catalytic activity. The R153C missense mutation in MDH significantly changes the enzyme's ability to coordinate the substrate, impacting catalytic turnover, substrate alignment, and

substrate specificity. Previous in-lab experimentation suggests that this specific alteration leads to misalignment of the substrate's hydroxyl or ketone group with the catalytic residues. This finding is crucial as it suggests a potential mechanism by which the mutation affects MDH's enzymatic activity (Bell).

In this research, modeling and simulations with AlphaFold, Rosetta, and AutoDock Vina aim first to predict and analyze the structure of MDH, and secondarily to focus on its interactions with oxaloacetate. The study first validates the efficacy of the structural prediction aspect of the computational pipeline by comparing the *in-silico-produced* structures of wild-type MDH against the crystallographic structure from the Protein Data Bank. Successful alignment in the wild-type model establishes a benchmark for the precision of AlphaFold and Rosetta. Subsequently, the approach is applied to the R153C mutation to explore its structural and functional implications through docking simulations by AutoDock Vina, revealing the interaction dynamics and binding affinities between MDH and its substrate.

The study will explore cost-effective implementations of computational tools in their ability to accurately and efficiently model structural changes in enzymes and predict their functional consequences. Such insights are invaluable, as they not only enhance understanding of enzyme mechanics but also emphasize the potential of computational methods in supplementing traditional laboratory techniques. The R153C mutation is a simple and ideal application for comparing the efficacy of computational models against laboratory findings. If successful, this approach and pipeline of algorithms could offer a more efficient and widely accessible pathway for hypothesis testing and experimental design in biological research, accelerating the pace of discovery and innovation in the field.

Methods

Protein Modeling Environment Configurations:

Microsoft Azure’s Cloud Computing Standard_DC48ds_v3 VM with 48 physical cores, 384

GB Memory, and 2400 GiB SSD was used, balancing the system's computational power and cost. The Azure VM had abundant storage for AlphaFold's input dataset and all files inputted or produced from the programs. Ubuntu Desktop version 22.04.3 or similar is the compatible distribution of choice to be paired with Rosetta on a local Linux virtual machine set up through Oracle VirtualBox with 10 CPU cores and 24 GB Memory. Debian is a suitable alternative, and selecting one of the compatible distributions depends on user preferences.

Azure HyperDrive for Concurrent Modeling (AlphaFold):

Microsoft Azure HyperDrive's parallel computing feature was implemented to run multiple queries concurrently. HyperDrive was then linked to the primary VM. Microsoft Azure's Machine Learning Studio tracked HyperDrive runs. The Learning Studio monitored real-time data for run progress and output structures. Predicted structures were saved to the virtual machine home directory, ensuring the output's security in an organized location before another run was initiated. The HyperDrive script to enable concurrent modeling was adapted from one developed by Dr. Colby T. Ford for RBD prediction in SARS-CoV-2, available in his GitHub Repository. The script outlined and fed the parameters to AlphaFold, taking multiple FASTA sequences from the RCSB Protein Data Bank as inputs, for both MDH wt (PDB ID: 1EMD) and MDH R153C (PDB ID: 1IE3). Using AlphaFold through ColabFold, the FASTA files for wt and mutant MDH were linked in the Jupyter Notebook for structure prediction on the Azure VM. The program was executed with default parameters, ensuring Multiple Sequence Alignment (MSA) data was used as an additional input.

Rosetta Configuration:

Rosetta Abinitio required extensive parameters, as it operates on basic biophysical principles to piece together fragments of proteins (Rämisch). Rosetta received six files in this task.

The first file was the Protein Data Bank predicted structure. The algorithm does not use the crystal structure as a reference to develop the prediction, but rather as a comparison to provide RMSD values that describe the deviation between the two structures. The second file was the amino acid sequence in FASTA format, which was downloaded from UniProt. Next, there are two fragment files that can direct Rosetta toward the most logical structure based on pre-existing data for torsion angles of amino acids, and generic structural conformations. These two fragment files can refine the search, and reduce the algorithm runtime to find the lowest energy and most stable structures. These fragment files were downloaded from the Robetta fragment prediction server. Unlike DeepMind AlphaFold, Rosetta utilizes the Monte Carlo method to randomly apply amino acids with variable phi and psi angles to find the optimal addition of amino acids. The optimal addition is determined by Rosetta's "linear combination of weighted score terms that balances physics-based and statistically derived potentials describing van der Waals energies, hydrogen bonds, electrostatics, disulfide bonds, residue solvation, backbone torsion angles, sidechain rotamer energies, and an average unfolded state reference energy" (Leman, et al.) **(Figure 1)**.

$$E = E_{vdW} + E_{hbond} + E_{elec} + E_{disulf} + E_{solv} + E_{BBtorsion} + E_{rotamer} + E_{ref}$$

Figure 1. Rosetta's linear combination equation for reference energy quantities. The total energy E is the sum of van der Waals interactions E_{vdW}, hydrogen bonding E_{hbond}, electrostatic interactions E_{elec}, disulfide bond contributions E_{disulf}, solvation energy E_{solv}, backbone torsion energy $E_{BBtorsion}$, rotamer energies $E_{rotamer}$, and a reference energy E_{ref} that accounts for the intrinsic preference of amino acids for specific conformations.

The next file is secondary structure prediction, which worked along with the fragment files. Finally, the flag file was written to inform Rosetta of the input files, request output data files, and organize the output format. After installing and compiling Rosetta in the Oracle VirtualBox Ubuntu VM, the input files for wt and mutant MDH were separated into respective folders in the home

directory. GNU screens were installed, named, and used to query 50 wt and 50 mutant MDH predictions separately and concurrently, maximizing CPU utilization. The CLI command "Rosetta/main/source/bin/AbinitioRelax.default.linuxgccrelease @<wt/mutant>.txt" used the txt flag files to instruct structure generation **(Table 1)**.

WT Flag File	Mutant Flag File
-database /home/modelingvm/Rosetta/main/database	-database /home/modelingvm/Rosetta/main/database
-in:file:native ./wt_1emd.pdb	-in:file:native ./mut_1emd.pdb
-in:file:fasta ./wt_rcsb_pdb_1EMD.fasta	-in:file:fasta ./mut_rcsb_pdb_1EMD.fasta
-in:file:frag3 ./wt_aat000_03_05.200_v1_3	-in:file:frag3 ./mut_aat000_03_05.200_v1_3
-in:file:frag9 ./wt_aat000_09_05.200_v1_3	-in:file:frag9 ./mut_aat000_09_05.200_v1_3
-psipred_ss2 ./wt_t000_.psipred_ss2	-psipred_ss2 ./mut_t000_.psipred_ss2
-nstruct 50	-nstruct 50
-abinitio:relax	-abinitio:relax
-use_filters true	-use_filters true
-abinitio::increase_cycles 10	-abinitio::increase_cycles 10
-abinitio::rg_reweight 0.5	-abinitio::rg_reweight 0.5
-abinitio::rsd_wt_helix 0.5	-abinitio::rsd_wt_helix 0.5
-abinitio::rsd_wt_loop 0.5	-abinitio::rsd_wt_loop 0.5
-relax::fast	-relax::fast
-out:file:silent ./50x_wt_fold_silent.out	-out:file:silent ./50x_mut_fold_silent.out

Table 1. Formatting of flag files to produce 50 structures each of wt and mutant MDH. All data reported in silent file.

From the designated silent output file, the Root Mean Square Deviation scores were noted from all runs, for future similarity comparisons with the respective crystal structures using the following commands:

```
grep SCORE 50x_mut_fold_silent.out | awk '{print $2 "\t" $27}' | sort -nk 2 > 50x_mut_VALUELIST.txt
grep SCORE 50x_wt_fold_silent.out | awk '{print $2 "\t" $27}' | sort -nk 2 > 50x_wt_VALUELIST.txt
```

Comparative Analysis of AlphaFold and Rosetta Outputs:

To evaluate the accuracy of the predicted structures from AlphaFold and Rosetta, the models were overlaid with the known crystal structure of MDH in PyMOL. Root mean square deviation (RMSD) values were calculated for both models to determine the similarity to the crystal structure before ligand-protein docking simulations. The best models from both programs, as determined by RMSD calculations, were utilized to conduct the docking simulations.

Docking Simulation of MDH and Oxaloacetate:

Using wild-type and mutant models of MDH after comparative analysis, docking simulations were conducted using AutoDock Tools/ MGL Tools, and AutoDock Vina on a Windows machine to utilize the AutoDock Tools GUI. Oxaloacetate was prepared as the ligand in pdbqt format for AutoDock to receive the necessary information for docking. After downloading the oxaloacetate sdf file from Pubchem, the file was imported into pyMOL, and exported in pdb format. The oxaloacetate pdb was imported into AutoDock Tools, selected with the “Choose” function, and output as a pdbqt file. After preparing oxaloacetate as the ligand for protein-ligand docking, the wild-type and mutant MDH structures from AlphaFold or Rosetta were prepared. First, the protein pdb was imported to AutoDock Tools, and all hydrogens (selected noBondOrder since AutoDock uses atom type, not bond order) were added without renumbering atoms. Kollman charges were added to account for the protein chemical environment, then the protein was saved in pdbqt file format. Grid boxes were set around the three arginine residues involved in coordination to inform AutoDock where oxaloacetate should be assessed for binding. Selecting the three residues in the sequence highlighted them yellow, and allowed the grid box size and offset to be oriented around them, ensuring extra space remained around the residues. The grid box size, spacing in angstroms, and offset centers were noted to be inputted in a configuration file. Docking

was executed in the windows command prompt with the executable default location: ““C:\Program Files (x86)\The Scripps Research Institute\Vina\vina.exe" --receptor mutant_MDH.pdbqt --ligand oxalo.pdbqt --config config_mut.txt --log Vina_Mut_Log.txt --out Vina_Mut_Output.pdbqt,” specifying the names of the input protein, ligand, configuration file **(Table 2)**, log, and structure outputs (Trott, et al.).

AutoDock Configuration File
receptor= mutant_MDH.pdbqt ligand= oxalo.pdbqt center_x = -2.584 center_y = 8.236 center_z = 7.818 size_x = 48 size_y = 70 size_z = 48 energy_range= 3 exhaustiveness= 20

Table 2. Formatting of AutoDock configuration file reporting input protein, ligand, grid size and offset values, affinity energy range, and number of poses/ modes to be calculated (exhaustiveness).

AutoDock determined the best poses and the most favorable configurations of the ligand-protein complex based on energetic favorability and binding affinity. These poses were evaluated for energy minimization and stability of the ligand within the active site, focusing on configurations that exhibit the lowest energy states and optimal interactions between MDH and oxaloacetate.

Analysis of Docking Results:

The binding affinities for wild-type and mutant MDH simulations were gathered from the command prompt output. During this analysis, key parameters were quantified by the software to assess the impact of the mutation. The analysis observed binding affinities and deviation from the most energetically favorable binding mode for adequate docking.

Results

Concurrent runs of wt and mutant MDH through the ColabFold Jupyter Notebook took 33 minutes to complete. The predicted structures were compared against the crystal reference structure (PDB ID: 1EMD) from the Protein Data Bank through PyMOL's 'orient' and 'align *object1, *object2' commands **(Figure 2)**.

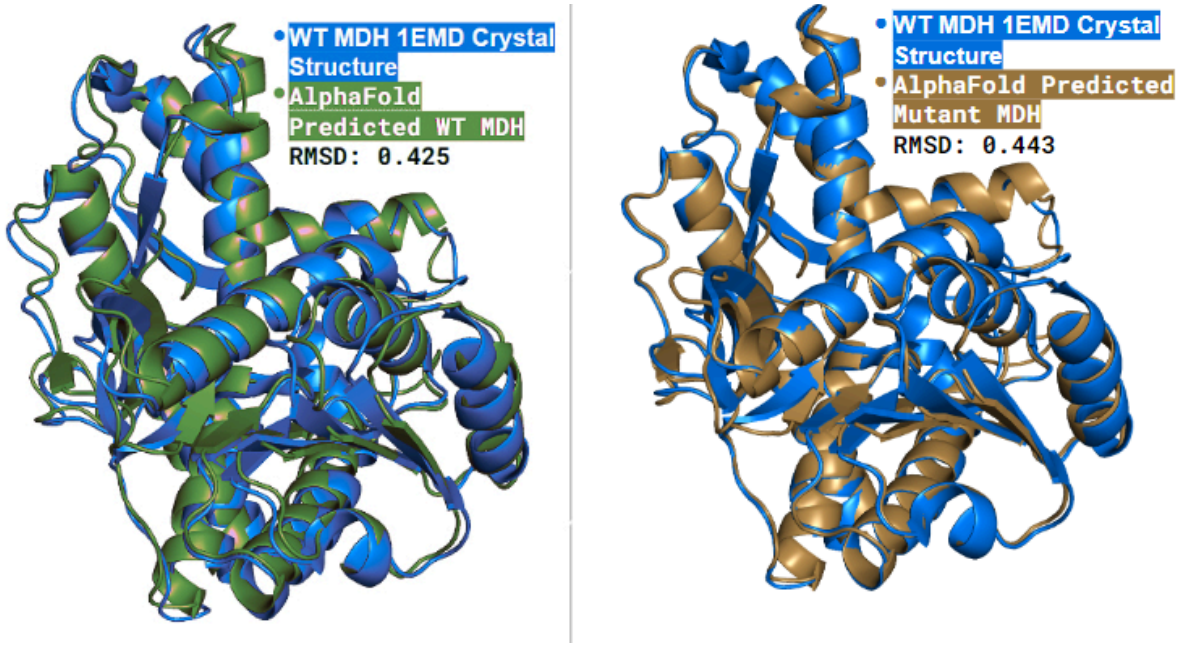

Figure 2. Aligned wt (left) and R153C MDH (right) structures against crystal reference structure, showing high correlation in structure and atom composition.

The resulting RMSD values for the predicted wild-type and mutant MDH were reported as 0.425 and 0.443, respectively. Well under the general threshold of RMSD < 2 Å (angstroms), these values indicate a high degree of accuracy compared to the reference structure. These values corresponded to comparisons over a complete matched atom count, with both simulations

covering a comprehensive set of 1794 to 1794 and 1810 to 1810 atoms for each protein model.

Rosetta was unsuccessful in producing models with RMSD values within the desired threshold for accurate structural predictions. This outcome is represented in the RMSD distribution graph for the 50 models of both wild-type and mutant MDH **(Figure 3)**.

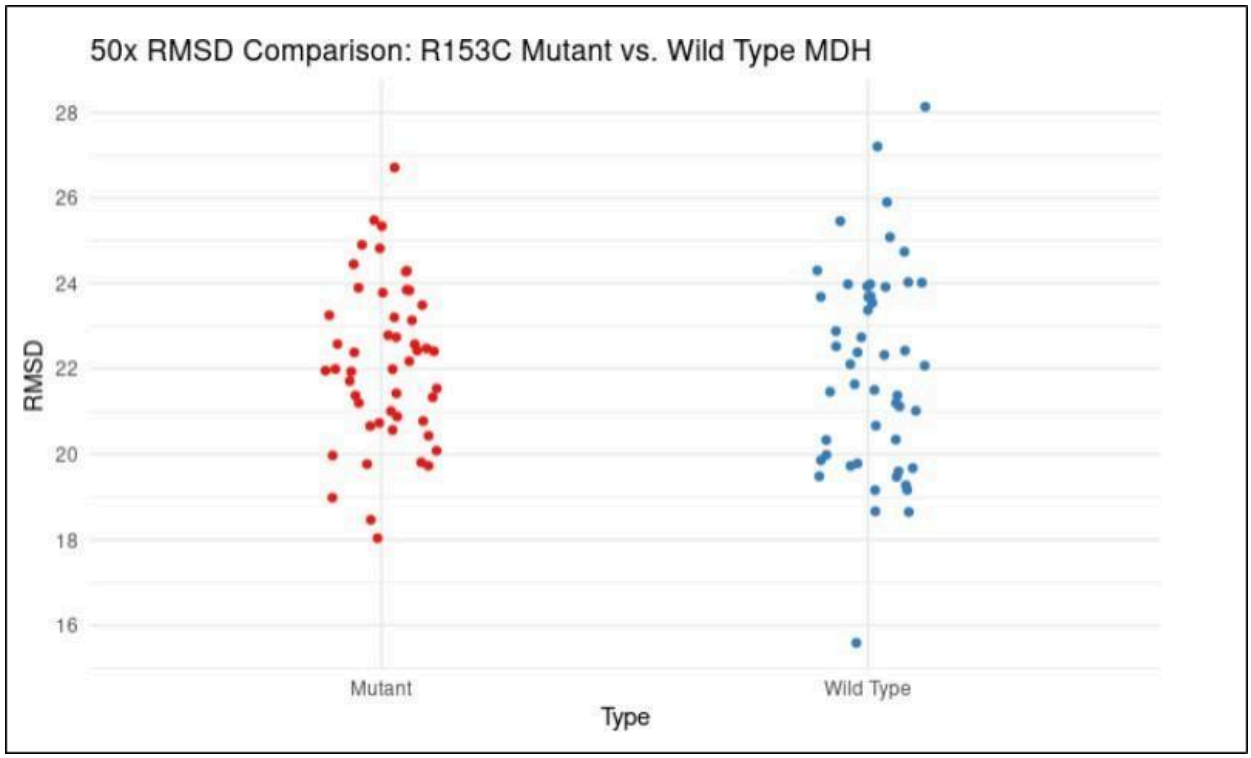

Figure 3. Strip Chart illustrating distribution of root-mean-square deviation (RMSD) values for 50 models each of wt, and R153C mutant MDH.

In reference to the respective MDH structures from the Protein Data Bank (1EMD and 1IE3), all 100 models of MDH produced by Rosetta indicated high variance between the simulated protein structure and the reference crystal structure when compared in PyMOL. The lowest RMSD of the mutant runs was 17.975, and 15.681 for wild type. The rendered structures are unusable for docking simulations.

Affinity (kcal/mol) and RMSD deviations from upper and lower bounds for 20 tested/ 9 reported modes (binding poses) were noted from AutoDock Vina **(Table 3)**. The best of the 9 reported modes was the most energetically favorable model and was set by the software as the

reference mode for comparison with all other poses (presented in RMSD from best mode). The docking analysis for wild-type MDH produced a range of binding affinities, with the top 9 modes presenting affinity values from -4.4 to -3.8 kcal/mol, and 4 out of 8 non-reference modes within the accepted RMSD range **(Table 3)**. The docking analysis for mutant R153C MDH produced a range of binding affinities from -5.8 to -4.8 kcal/mol, with significantly higher RMSD's for upper and lower bounds, pushing all non-reference modes well out of the acceptable RMSD range **(Table 3)**.

Wild Type Docking Results				**Mutant Docking Results**			
Mode	Affinity	RMSD Lower Bound	RMSD Upper Bound	Mode	Affinity	RMSD Lower Bound	RMSD Upper Bound
1	-4-4	0.000	0.000	**1**	-5.8	0.000	0.000
2	-4.1	0.538	2.733	**2**	-5.1	24.884	25.615
3	-4.1	1.524	2.398	**3**	-5.0	35.368	36.027
4	-4.1	1.243	2.986	**4**	-5.0	24.334	24.923
5	-4.0	26.827	27.144	**5**	-4.9	20.821	21.230
6	-3.8	30.842	31.557	**6**	-4.9	30.581	31.115
7	-3.8	12.359	13.196	**7**	-4.8	30.528	31.200
8	-3.8	1.264	2.760	**8**	-4.8	24.854	25.587
9	-3.8	12.774	13.865	**9**	-4.8	20.865	21.455

Table 3. Docking results of wt and mutant MDH, presenting binding affinities (kcal/mol), and deviation from the most energetically favorable identified mode (mode 1).

Discussion

This experiment aims to evaluate the precision of computational modeling by comparing its predictions with empirical laboratory findings. The study applies a structured computational workflow to analyze the efficacy of the software in producing both the wild-type MDH and the R153C mutant, aiming to align with laboratory data to validate the model's accuracy in reflecting

biological phenomena. By examining the docking results of the wild-type, the pipeline's reliability can be established and extended to understand the mutant's behavior.

The exploration of Malate Dehydrogenase and the effects of mutagenesis through computational models have provided important insights into the structural and functional dynamics of wild-type and R153C mutant Malate Dehydrogenase. AlphaFold's predictions referenced against the PDB crystal structure provided RMSD values indicative of high-precision modeling, with both variants of MDH having high structural similarity to the reference, presenting values well below the RMSD threshold (Predicted Model RMSD < 2). The Rosetta software suite, however, encountered unforeseen obstacles in deploying and compiling the software within the Azure VM environment. The inability to compile Rosetta in the VM points towards potential version incompatibilities, or Azure's strict security parameters preventing proper function of the Rosetta installer.

To preserve a core objective of the study -modeling proteins **concurrently** to enhance efficiency and reduce computation duration- the methods were adapted to set up and use a local Linux Ubuntu VM to run Rosetta for later comparative analysis against the AlphaFold results. Completion of the queries in Rosetta took close to two hours to produce 50 models for wt and mutant MDH. The output structures showed highly elevated RMSD values for both protein variations, eliminating their reliability for docking analysis. A potential cause of the discrepancy was using the old Robetta fragment server, as opposed to the current version. The old Robetta fragment server delivered files quicker (30-40 minutes) than the new server with greater traffic (up to 1 week). This highlights the need to integrate more current or potentially more accurate input fragment files, in future studies.

The AutoDock Vina simulations evaluated the binding interactions of MDH. Low RMSD ranges across multiple modes for the wt MDH reinforced oxaloacetate's coordination predictability in multiple positions with repeated stability. The negative affinity of -4.4 kcal/mol

suggests a good binding affinity between ligand and protein, as a strong affinity is represented by -5 to -10 kcal/mol. The low RMSD values from each alternate binding mode propose the binding modes are relatively similar to the most energetically favorable mode (mode 1). This suggests the wild-type protein maintains consistent interactions with oxaloacetate across different binding modes. In contrast, the reference mode of Mutant MDH demonstrated more negative binding affinity (-5.8 kcal/mol), with additional modes in the 'strong binding affinity' threshold, suggesting that alteration in the enzyme binding site geometry increased affinity for oxaloacetate. However, the high RMSD values of all modes in reference to the most energetically favorable binding mode suggest a significant divergence from native or expected binding conformations. This implies that the mutation may have caused structural alterations that led to reduced/ altered binding specificity and orientation. The mutation may have caused disorder in the protein, allowing it to adopt a sporadic and wide range of unfavorable conformations. This idea is supported by laboratory findings that report more flexible substrate specificity due to mutations of 1 of the 3 arginine residues (Bell). However, more thorough data on the docking study, potentially with ligands other than oxaloacetate, is required to gain more information towards a conclusion on the effect of the mutagenesis. The nuanced finding of high binding affinity but severely decreased correlation from the optimal binding formation for mutant MDH points towards a complex interplay between enzyme structure and function under the influence of mutations.

While the computational approach faced challenges, it also exhibited the potential to model protein structures and interactions accurately. Future directions should focus on refining the methodologies for updated computational tools and expanding the data output of docking studies for a deeper understanding of ligand-substrate interactions. Leveraging computational methods, researchers can develop a basis for understanding the effects of mutagenesis on proteins to hypothesize the impacts of mutations before empirical validation. Computational

modeling as a supplement to traditional laboratory methods can accelerate and refine current understandings to guide experimental design. There is a compelling justification for the increased adoption of computational methods, not as replacements, but as essential supplements to laboratory experiments. Contributing to the broader scientific dialogue by integrating computational biology into standard wet-lab practices of biological research, the dual approach can enhance the efficiency and scope of research for a more comprehensive exploration of the mechanisms underlying the fundamentals of life.

References

Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network | Science. https://www.science.org/doi/10.1126/science.abj8754. Accessed 27 Oct. 2023.

Baek, Minkyung, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose. pp. 871–76. *PubMed Central*, https://doi.org/10.1126/science.abj8754.

Bell, Jessica K., et al. "Structural Analyses of a Malate Dehydrogenase with a Variable Active Site *." *Journal of Biological Chemistry*, vol. 276, no. 33, Aug. 2001, pp. 31156–62. *www.jbc.org*, https://doi.org/10.1074/jbc.M100902200.

Bitencourt-Ferreira, Gabriela, et al. "Docking with AutoDock4." *Methods in Molecular Biology (Clifton, N.J.)*, vol. 2053, 2019, pp. 125–48. *PubMed*, https://doi.org/10.1007/978-1-4939-9752-7_9.

Bonneau, R., et al. "Rosetta in CASP4: Progress in Ab Initio Protein Structure Prediction."

Ford, C. (2022). azureml-alphafold2 [Software]. GitHub, from https://github.com/colbyford/azureml-alphafold2

H. Pereira, Andria V. Rodrigues, Alberdina A. van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea Pavkov-Keller, Manoj K Rathinaswamy, et al. "Accurate Prediction of Protein Structures and Interactions Using a 3-Track Neural Network." Science (New York, N.Y.), vol. 373, no. 6557, Aug. 2021,

Hung, Chih-Hung, et al. "Crystal Structures and Molecular Dynamics Simulations of Thermophilic Malate Dehydrogenase Reveal Critical Loop Motion for Co-Substrate Binding." *PLoS ONE*, vol. 8, no. 12, Dec. 2013, p. e83091. *PubMed Central*, https://doi.org/10.1371/journal.pone.0083091.

Jumper, John, et al. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature*, vol. 596, no. 7873, 7873, Aug. 2021, pp. 583–89. *www.nature.com*, https://doi.org/10.1038/s41586-021-03819-2.

Koehler Leman, Julia, et al. "Macromolecular Modeling and Design in Rosetta: Recent Methods and Frameworks." *Nature Methods*, vol. 17, no. 7, July 2020, pp. 665–80. *PubMed Central*, https://doi.org/10.1038/s41592-020-0848-2.

Lyskov, Sergey, and Jeffrey J. Gray. "The RosettaDock Server for Local Protein–Protein Docking." Nucleic Acids Research, vol. 36, no. Web Server issue, July 2008, pp. W233–38. PubMed Central, https://doi.org/10.1093/nar/gkn216.

Moretti, Rocco, et al. "Web‐accessible Molecular Modeling with Rosetta: The Rosetta Online Server That Includes Everyone (ROSIE)." *Protein Science : A Publication of the Protein Society*, vol. 27, no. 1, Jan. 2018, pp. 259–68. *PubMed Central*, https://doi.org/10.1002/pro.3313.

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022 Jun;19(6):679-682. doi: 10.1038/s41592-022-01488-1. Epub 2022 May 30. PMID: 35637307; PMCID: PMC9184281.

Murthy, Mukundh. “Ab Initio Protein Folding.” *Medium*, 11 May 2020, https://mukundh-murthy.medium.com/ab-initio-protein-folding-be27509d134a. ---. “Ab Initio Protein Folding.” *Medium*, 11 May 2020, https://mukundh-murthy.medium.com/ab-initio-protein-folding-be27509d134a.

Ó Conchúir S, Barlow KA, Pache RA, Ollikainen N, Kundert K, O'Meara MJ, Smith CA, Kortemme T. A Web Resource for Standardized Benchmark Datasets, Metrics, and Rosetta Protocols for Macromolecular Modeling and Design. PLoS One. 2015 Sep 3;10(9):e0130433. doi: 10.1371/journal.pone.0130433. PMID: 26335248; PMCID: PMC4559433.

O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry 31 (2010) 455-461

Proteins, vol. Suppl 5, 2001, pp. 119–26. *PubMed*, https://doi.org/10.1002/prot.1170.

dinesh-supreme. “Why Structure Prediction Matters.” *DNASTAR*, 21 July 2020, https://www.dnastar.com/blog/protein-analysis-modeling/why-structure-prediction-matters/.

Pricing - Linux Virtual Machines | Microsoft Azure. https://azure.microsoft.com/en-us/pricing/details/virtual-machines/linux/.

Rämisch, Sebastian. "De Novo Structure Prediction." RosettaCommons, n.d., https://www.rosettacommons.org/demos/latest/tutorials/denovo_structure_prediction/Denovo_structure_prediction.

Structural Analyses of a Malate Dehydrogenase with a Variable Active Site - PubMed. https://pubmed.ncbi.nlm.nih.gov/11389141/. Accessed 10 Nov. 2023.

Takahashi-Íñiguez T, Aburto-Rodríguez N, Vilchis-González AL, Flores ME. Function, kinetic properties, crystallization, and regulation of microbial malate dehydrogenase. J Zhejiang Univ Sci B. 2016 Apr;17(4):247–61. doi: 10.1631/jzus.B1500219. PMCID: PMC4829630.