MODELING AND SIMULATION

SYSTEM BIOLOGY

Final project report

Group members

Oneza Hassan Alvi (04281913032) – taking both courses
Tooba shahid Abbasi (04281913020) – taking modeling and simulation
Ayesha Saddiqua (09282113002) – taking system biology

Submitted to

Miss. Afifa Naveed
Mrs. Faisal Ahmed

Table of Contents

Introduction

Rickettsia parkeri

Rickettsia parkeri is an intracellular bacterium that has a gram negative cell wall. It infects humans through the bite of hard ticks belonging to the Amblyomma genus. The infection results in rickettsiosis, a mild febrile illness. The symptoms are rashes, headaches, eschar, and muscle aches. (Kristen L. Herrick, 2016). R. parkeri infections are relatively more common and occur from the southern US to Argentina. The affected individuals are treated with doxycycline, and they return to normal within a month. Many preventive measures can be taken to avoid this disease, including the use of permethrin and insect repellent with 20% to 30% N, N-diethyl-m-toluamide (DEET) on skin and clothes. (Marina E. Eremeeva).

Lysine Biosynthesis

Lysine is an indispensable amino acid (IAA) required for growth and positive nitrogen balance. (Matthews, 2020). The Diaminopimelate (DAP) and alpha aminoadipic acid pathway (AAA) are the two main pathways that have evolved for lysine biosynthesis. DAP belongs to the aspartate family, and has an important role in eubacterial cell wall biosynthesis (A.M. Velasco, 2002) Aspartate produces the amino acid lysine via the diaminopimelate (DAP) pathway in many bacteria and higher plants. Aspartokinase and aspartate semialdehyde dehydrogenase are involved in the catalyzation of the first two stages of the DAP pathway that biosynthesizes the amino acid lysine. (Dmitry A. Rodionov, 2003) The importance of bacterial biosynthesis of lysine is that it provides lysine for protein synthesis and lysine and maso-diaminopimelate to form the bacterial peptidoglycan cell wall.

PROTEIN TARGET: Aspartate Kinase

AIMS AND OBJECTIVES:

The aim of this project was to use homology modeling to construct a three-dimensional model of aspartate kinase, an enzyme involved in the lysine biosynthesis pathway in bacteria. The objectives were to identify a suitable template structure for modeling, to align the query sequence of aspartate kinase with the template sequence, to generate and evaluate multiple models using Modeler, and to compare the best model with the template using superimposition and analysis of key residues.

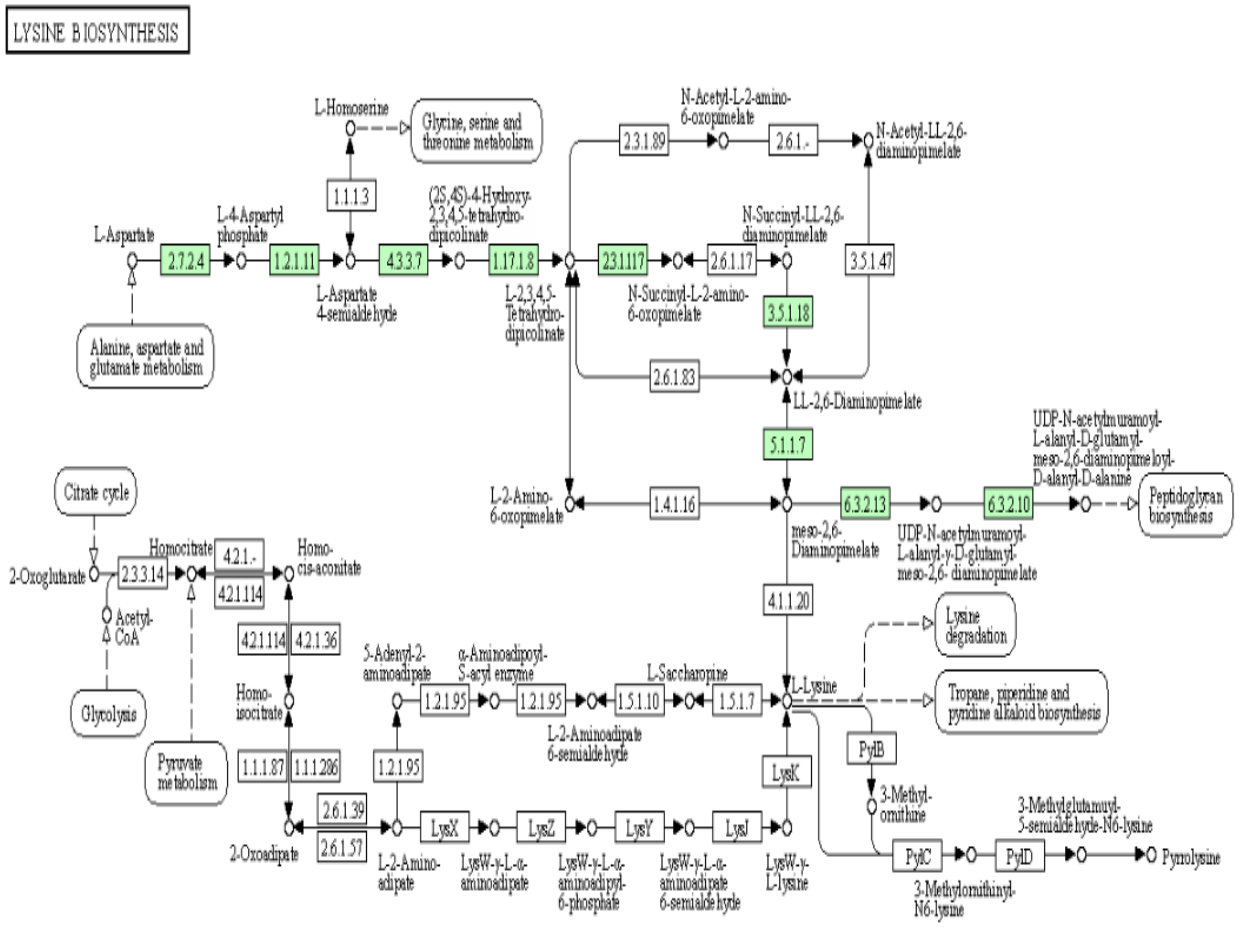


Figure 1 Lysine Biosynthesis Pathway (Kanehesia Laboratories, 2021)

Aspartate Kinase

Aspartate kinase [EC: 2.7.2.4] participates in the lysine biosynthesis. It speeds up the first step of the pathway in all bacteria, i.e., the conversion of aspartate into L-aspartyl-phosphate. (Nobuyuki Kobashi, 1999). It is involved in the production of both leucine and isoleucine. Aspartate kinase is regulated by "end-metabolites through feedback inhibition" (Wang, 2015). It produces aspartyl phosphate by speeding up the aspartate's phosphorylation. The structure of template 3LE1 has two chains, but we have chosen chain A here due to the presence of an active site on it. It has three residues: lysine, threonine, and a sulphate ion.

Methodology

KEGG:

1. KEGG is a comprehensive resource that contains information on genomes, biological pathways, diseases, drugs and chemical substances.
2. For this project, we selected aspartate kinase as our protein of interest from the lysine biosynthesis pathway in KEGG.
3. Aspartate kinase is an enzyme that catalyzes the first step of the pathway, which is the conversion of aspartate to beta-aspartyl phosphate.
4. We obtained the amino acid sequence of aspartate kinase from KEGG and used it as our query sequence for homology modeling.

BLAST:

1. BLAST is a tool that allows us to compare our query sequence to a database of protein sequences and find the most similar ones.
2. We used protein BLAST (BLASTp) to search for homologous sequences of aspartate kinase in the protein database.
3. We analyzed the BLASTp results and selected the best template sequence for modeling based on the criteria of high query coverage, high sequence identity and low E-value.
4. The template sequence that we chose was 3L76, which is the crystal structure of aspartate kinase from Escherichia coli.

Modeler:

1. Modeler is a program that uses comparative modeling to construct three-dimensional models of proteins based on their alignment with known structures.
2. We used Modeler to generate five models of aspartate kinase using the template 3L76 and the query sequence.
3. We evaluated the quality of each model using the DOPE score, which is a statistical potential that measures the compatibility of the model with the template and its stereochemistry.
4. We selected the model with the lowest DOPE score as the best model for further analysis and

validation.

Super Imposition:

1. Super Imposition is a process of overlaying two or more structures to compare their similarities and differences.
2. We used Chimera, a molecular visualization software, to superimpose the best model of aspartate kinase with its template 3L76.
3. We calculated the RMSD value, which is a measure of how well the two structures fit together. The lower the RMSD value, the higher the structural similarity.
4. We also visualized the alignment of the key residues involved in catalysis and substrate binding in both structures.

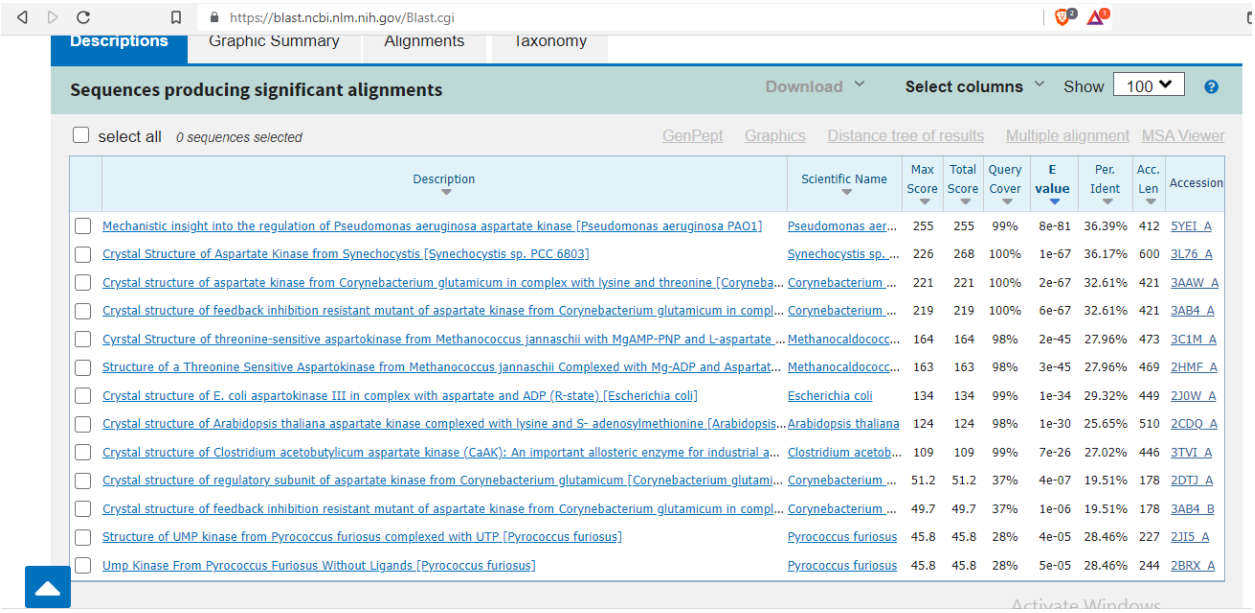
Active Sites:

1. Active sites are regions of proteins where the enzymatic reactions take place. They usually consist of a few residues that interact with the substrate and cofactor molecules.
2. We identified the active site residues of aspartate kinase by reviewing the literature and comparing them with the residues shown in Chimera.
3. We also used online tools such as Castp and Cport to predict and analyze the active sites in cases when literature was not available or sufficient.
4. We discussed the implications of the active site residues for inhibitor design and drug discovery.

Results and discussion

Blastp

The query sequence of aspartate kinase was submitted to blastp to search for similar sequences in the protein database. The blastp results showed the following hits.



Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Mechanistic insight into the regulation of Pseudomonas aeruginosa aspartate kinase [Pseudomonas aeruginosa PAO1]	Pseudomonas aer...	255	255	99%	8e-81	36.39%	412	5YEI_A
Crystal Structure of Aspartate Kinase from Synechocystis [Synechocystis sp. PCC 6803]	Synechocystis sp. ...	226	268	100%	1e-67	36.17%	600	3L76_A
Crystal structure of aspartate kinase from Corynebacterium glutamicum in complex with lysine and threonine [Corynebа...	Corynebacterium ...	221	221	100%	2e-67	32.61%	421	3AAW_A
Crystal structure of feedback inhibition resistant mutant of aspartate kinase from Corynebacterium glutamicum in compl...	Corynebacterium ...	219	219	100%	6e-67	32.61%	421	3AB4_A
Cyrstal Structure of threonine-sensitive aspartokinase from Methanococcus jannaschii with MgAMP-PNP and L-aspartate ...	Methanocaldococc...	164	164	98%	2e-45	27.96%	473	3C1M_A
Structure of a Threonine Sensitive Aspartokinase from Methanococcus jannaschii Complexed with Mg-ADP and Aspartat...	Methanocaldococc...	163	163	98%	3e-45	27.96%	469	2HMF_A
Crystal structure of E. coli aspartokinase III in complex with aspartate and ADP (R-state) [Escherichia coli]	Escherichia coli	134	134	99%	1e-34	29.32%	449	2J0W_A
Crystal structure of Arabidopsis thaliana aspartate kinase complexed with lysine and S- adenosylmethionine [Arabidopsis...	Arabidopsis thaliana	124	124	98%	1e-30	25.65%	510	2CDQ_A
Crystal structure of Clostridium acetobutylicum aspartate kinase (CaAK): An important allosteric enzyme for industrial a...	Clostridium acetob...	109	109	99%	7e-26	27.02%	446	3TVI_A
Crystal structure of regulatory subunit of aspartate kinase from Corynebacterium glutamicum [Corynebacterium glutami...	Corynebacterium ...	51.2	51.2	37%	4e-07	19.51%	178	2DTJ_A
Crystal structure of feedback inhibition resistant mutant of aspartate kinase from Corynebacterium glutamicum in compl...	Corynebacterium ...	49.7	49.7	37%	1e-06	19.51%	178	3AB4_B
Structure of UMP kinase from Pyrococcus furiosus complexed with UTP [Pyrococcus furiosus]	Pyrococcus furiosus	45.8	45.8	28%	4e-05	28.46%	227	2JI5_A
Ump Kinase From Pyrococcus Furiosus Without Ligands [Pyrococcus furiosus]	Pyrococcus furiosus	45.8	45.8	28%	5e-05	28.46%	244	2BRX_A

Figure 1.1 Results of blastp obtained by running the sequence of aspartate kinase.

The template **3L76 A** is chosen among others because it has the highest amount of query coverage (100%), which means that it covers the largest fraction of the query sequence. Query coverage is important because it indicates how much of the query sequence can be modeled by the template. The higher the query coverage, the more complete the model. The template also has a low E-value (1e-67), which means that it has a high statistical significance and a low probability of being a random match. The template has a moderate percentage of identity (36.17%), which means that it has a reasonable degree of similarity to the query sequence. Percentage of identity is not the only criterion for selecting a template, because sometimes a lower identity can still result in a good model if the structure is conserved. The other templates have either lower query coverage, higher E-value, or lower percentage of identity than 3L76, and therefore are less suitable for homology modeling.

Table 1

S.no	**Protein Name**	**Template**
1	Aspartate kinase	3L76

The above table shows protein and its best template, which was retrieved by running the query sequence of protein on blast. The template was selected on the basis of high coverage and sequence

Modeling

Table 2

Protein	Models	DOPE Score
Aspartate kinase	protein.B99990001.pdb	-69686.74219
	protein.B99990002.pdb	**-69726.15625**
	protein.B99990003.pdb	-68973.85938
	protein.B99990004.pdb	-69315.06250
	protein.B99990005.pdb	-69464.62500

This table shows the five models of aspartate kinase that were generated by Modeler, along with their DOPE scores. DOPE stands for Discrete Optimized Protein Energy, and it is a statistical potential that evaluates the quality of a protein model based on its stereochemistry and its compatibility with the template. The lower the DOPE score, the better the model. Among the five models, protein.B99990002.pdb has the lowest DOPE score of -69726.15625, which means that it has the best fit to the template and the least structural errors. Therefore, protein.B99990002.pdb is chosen as the best model for further analysis and validation.

Superimposition:

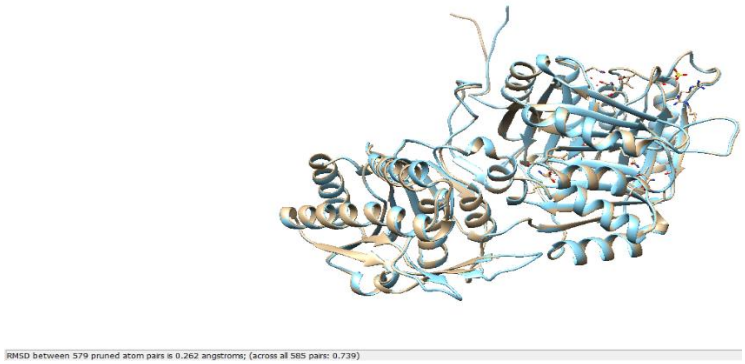


Figure 2.1 superimposition done by using chimera.

Superimposed Model (RMSD = 0.620° A)

In the above figure, the best models of the protein was superimposed with its respective template to see the Root Mean Square Deviation (RMSD). The lower the value, the similar the two structures. The superimposition of the best model of aspartate kinase (protein.B99990002.pdb) with its template 3L76 shows a low RMSD value of 0.620° A, indicating a high degree of structural similarity between the two proteins. This suggests that the model was built with a high level of accuracy and reliability, and that the template was a suitable choice for homology modeling. The superimposition also reveals the alignment of the key residues involved in catalysis and substrate binding, such as Lys-9, Asp-10, Thr-11, Arg-12, Asp-13, Lys-165, Asp-166 and Arg-167. The residue Asp-166 is especially important as it acts as a base that activates the incoming substrate hydroxyl. These residues are conserved across different bacterial aspartate kinases and are essential for their enzymatic activity. These residues form hydrogen bonds, salt bridges and electrostatic interactions with the substrate and the cofactor (Mg2+) to facilitate the catalysis. The superimposition shows that these residues are well aligned between the model and the template, indicating that they have similar roles and conformations in both proteins. The superimposition also shows some minor differences in the loop regions and the C-terminal domain, which could reflect the variations in the amino acid sequences and the environmental factors affecting the protein folding and stability.

Conclusion

The project successfully achieved its aim and objectives by using homology modeling to build a reliable and accurate model of aspartate kinase. The model showed a high degree of structural similarity with the template, as evidenced by the low RMSD value and the alignment of the catalytic and substrate-binding residues. The model also revealed some minor differences in the loop regions and the C-terminal domain, which could reflect the sequence variations and the environmental factors affecting the protein folding and stability. The model could be further refined by using molecular dynamics simulations and docking studies to explore the dynamics and interactions of aspartate kinase with its substrates and inhibitors. The model could also be used to design novel inhibitors or modulators of aspartate kinase, which could have potential applications in antibacterial therapy or metabolic engineering.

References

A.M. Velasco, J. L. (2002). Molecular Evolution of the Lysine Biosynthetic Pathways. *Molecular Evolution of the Lysine Biosynthetic Pathways*.

Dmitry A. Rodionov, A. G. (2003). Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch?*, 6748–6757.

Kristen L. Herrick, S. A. (2016). Rickettsia parkeri Rickettsiosis, Arizona, USA. *Rickettsia parkeri Rickettsiosis, Arizona, USA*, 780-785.

Lilian Hor, R. C. (2013). Dimerization of Bacterial Diaminopimelate Epimerase Is Essential for Catalysis. *Dimerization of Bacterial Diaminopimelate Epimerase Is Essential for Catalysis*, 9238–9248.

Marina E. Eremeeva, G. A. (n.d.). RICKETTSIA PROWAZEKII (CLASSIC LOUSE-BORNE EPIDEMIC TYPHUS; BRILL-ZINSSER DISEASE OR FLYING SQUIRREL-ASSOCIATED TYPHUS). *RICKETTSIA PROWAZEKII (CLASSIC LOUSE-BORNE EPIDEMIC TYPHUS; BRILL-ZINSSER DISEASE OR FLYING SQUIRREL-ASSOCIATED TYPHUS*, 1-14.

Matthews, D. E. (2020). Review of Lysine Metabolism with a Focus on Humans. *Review of Lysine Metabolism with a Focus on Humans*, 2548S–2555S.

Nobuyuki Kobashi, M. N. (1999). Aspartate Kinase-Independent Lysine Synthesis in an Extremely Thermophilic Bacterium, Thermus thermophilus: Lysine Is Synthesized via α-Aminoadipic Acid Not via Diaminopimelic Acid. *Journal of bacteriology*, 181(6), 1713-1718.

Wang, J. G. (2015). Evolution of a chimeric aspartate kinase for L-lysine production using a synthetic RNA device. ,. *Applied microbiology and biotechnology*, 99(20), 8527-8536.