COMPUTATIONAL STUDY ON ALLOSTERY IN BACTERIAL RIBOSOME

by
Hatice Zeynep Yildirim
B.S., Chemical Engineering, Gazi University, 2013

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
The requirements for the degree of
Master of Science

Graduate Program in Computational Science and Engineering
Bogazici University
2019

COMPUTATIONAL STUDY ON ALLOSTERY IN BACTERIAL RIBOSOME

APPROVED BY:

Prof. Levent Kurnaz ……...……
(Thesis Supervisor)

Assoc. Prof. Burak Alakent ……...

Assist. Prof. Ozge Kurkcuoglu Levitas……...

DATE OF APPROVAL: 24.10.2019

ACKNOWLEDGEMENTS

I would like to thank my supervisors Prof. Levent Kurnaz, Prof. Pemra Doruker and Prof. Özge Kürkçüoğlu. I would like to express my gratitude to Pemra Doruker. I have learned many things since I became Pemra Doruker's student. She always supported me and taught me how to search literature, write project and reach information. I am also grateful to Özge Kürkçüoğlu for reading this thesis and providing very useful suggestions, comments and corrections. She is always guiding spirit for me.

I would like to thank Emine Yıldırım and Atabey Kaygun. I really appreciate their valuable comments on this thesis. Special thanks to my friend Büşra Özgüney, she always encouraged me during my master program. I also would like to extend my thanks to Doğa Fındık and Gökçehan Kara for their help in my academic studies.

I owe more than thanks to my family members which include my parents, my sisters, my brother and my love Cankut Kibar for their encouragement throughout my life.

ABSTRACT

COMPUTATIONAL STUDY ON ALLOSTERY IN BACTERIAL RIBOSOME

Molecular machines in a cell have signal processing to perform their function using specific sites such as active or allosteric sites. Ligand binding to active sites or signal transferring from allosteric sites affect their function and dynamics. In this thesis, firstly crystal structure of bacterial ribosome (4kdk-4kdj) and its conformers which are generated by ClustENM are investigated to determine allosteric communication pathways. Targets on ribosome are determined as the Decoding Center (DC) – the Sarcin Ricin Loop (SRL), DC - the Peptidyl Transferase Center (PTC) and the PTC – Tunnel. On the allosteric pathways between DC and SRL, EF-G stands out with critical sites on its domains IV which has a significant function in blocking back translocation of tRNA. Some significant nucleotide and aminoacid like A1493 and Met580 appear on pathways between DC-SRL, which help EF-G hydrolysis. On the DC-PTC pathways drug binding site is observed. On the PTC-ribosomal tunnel pathway has a highly conserved non-Watson-Crick base pair and binding pocket for antibiotic is found.

Secondly, the bacterial ribosome of E.coli, 4v5h, is analyzed to investigate allostery between Secretion Monitor (SecM)–PTC and TF-Ribosomal Tunnel. A76 of tRNA and nascent chain which consists of alanines seems significant between U2585 and A2451 to provide allosteric communication between SecM and PTC. On the shortest pathways on the TF-Ribosomal Tunnel, GLY91 from L22 has a high frequency of occurrence on all pathways. GLY91 is significant for elongation arrest and for the turn of the β-hairpin of L22 which is important since antibiotic resistance appears when a mutation on the β-hairpin occurs.

ÖZET

BAKTERİYEL RİBOZOMDA ALLOSTERİK İLETİŞİM YOLLARI ÜZERİNE HESAPLAMALI CALIŞMALAR

Bir hücrede bulunan moleküler makineler, işlevlerini yerine getirmek için aktif veya allosterik bölgeler gibi belirli bölgeleri kullanarak sinyalleşme sistemine sahiptir. Aktif bölgelere ligand bağlanması veya allosterik bölgelerden sinyal aktarımı moleküler makinelerin fonksiyonlarını ve dinamiklerini etkiler. Bu tez çalışmasında öncelikle bakteriyel ribozomun kristal yapısı (4kdk-4kdj) ve ClustENM tarafından üretilen konformerleri allosterik iletişim yollarını belirlemek için incelenmiştir. Ribozomdaki sinyalleşmenin başladığı ve bittiği bölgeler, Kod Çözme Merkezi (DC) - Sarcin Ricin Döngüsü (SRL), DC - Peptidil Transferaz Merkezi (PTC) ve PTC – Ribozomal Tünel olarak belirlenmiştir. DC ve SRL arasındaki allosterik yollarla da, tRNA'nın geri trans lokasyonunu bloke etmede önemli bir işlevi olan EF-G nin domain IV bölgesi dikkat çekmistir. DC-SRL arasındaki yollarda ortaya çıkan A1493 ve Met580 gibi bazı önemli nükleotit ve aminoasitler, EF-G‘ nin hidrolizine yardim ederler. DC-PTC yollarında ilaç bağlama bölgesi gözlenmistir. PTC-Ribozomal Tünel yolu üzerinde yüksek oranda korunan Watson-Crick bazlı olmayan bir baz çifti ve antibiyotikler için bağlanma bölgesi bulunmuştur.

İkinci çalışma olarak, E. coli'nin bakteriyel ribozomu (4V5H) üzerinde, Sekresyon Monitörü (SecM) - PTC ve TF - Ribozomal Tünel arasındaki allosterik iletisim yolları incelenmiştir. SecM ve PTC arasında allosterik iletişim sağlamak için, U2585 ve A2451 nükleotitleri arasında çıkan alaninlerden oluşan yeni zincir ve tRNA’ nin A76 rezidusu önemli olarak bulunmuştur. TF-Ribozomal Tüneli'ndeki en kısa yollarda çıkan GLY91, tüm yollarda yüksek bir oluşum sıklığına sahiptir. GLY91, L22 proteinin önemli bir bölgesinde bulunuyor ve bu bölgedeki mutasyonlar antibiyotik direnci ortaya çıkarıyor.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

a_{ij}	Affinity between i and j
N_i	The number of heavy atoms of residue i
N_j	The number of heavy atoms of residue j
N_{ij}	Node pair within a cutoff distance
Å	Distance unit, Angstrom

LIST OF ABBREVIATIONS

DC	Decoding Center
SRL	Sarcin Ricin Loop
PTC	Peptidyl Transferase Center
SecM	Secretion Monitor

1. INTRODUCTION

Allostery is a process of signal transferring between the active site of a macromolecule and any other site. These other sites – called the allosteric sites – affect macromolecule dynamics and regulates its function. Since active sites are conserved in many species, allosteric sites enable us to find an effective solution in case of a disease. It is claimed that all proteins might be allosteric [1].

Allostery is defined as a big complex system whose underlying mechanisms are hard to understand. Earlier studies focused on conformational changes between static snapshots of different conformational states of the macromolecules to understand allosteric mechanism. With the advances in computational techniques, allosteric studies have gradually became very efficient. One computational technique to find allosteric sites is to interpret macromolecules as networks.

When networks are formed communication pathways appear and information is transported on these pathways. The network may have hub nodes or high degree links. We hypothesize that finding such hubs and links in macromolecules viewed as networks helps us to determine allosteric communication pathways. In this study, we are going to focus on the shortest pathways to determine allostery.

The shortest paths are like busy bridges. Lots of information and interaction flow on this bridge. For biological systems, we convert proteins to undirected weighted graph in which aminoacids that constitute the proteins are nodes and their interactions with each other are edges. We weight these edges with respect to the probability of their binding, their entropy and energy change, if they are charged or not.

For this task we consider bacterial ribosome of T. Thermophilus (4v9m) and E.coli (4v5h). Ribosome is a complex macromolecule which has numerous signaling processes to synthesize protein. Trigger Factor, on the other hand, is crucial for protein folding. In this thesis, we are interested in studying Decoding Center, Sarcin Ricin Loop, Peptidyl Transferase Center, ribosomal tunnel, TF and SecM.

2. THEORY

2.1. Allostery

Allostery is a mechanism which governs the propagation of information between different sites, i.e. *allosteric signaling* is a type of long distance signaling in a protein within a protein. Allostery can take place due to small-molecule binding, covalent modifications, protein–protein interactions and mutations [2].

Allostery can be thought as a remote controller in biomolecules. An orthosteric site, such as the active site of an enzyme, is a specific region which performs the function in a protein. There are also distant sites that affect the functionality via “the allosteric effect”.

The first studies on allosteric sites appeared in 1904. Christian Bohr discovered that the carbon dioxide molecule affects binding affinity of the oxygen molecule in the hemoglobin [3]. This is called “Bohr effect”. In general terms, it means that one molecule affects the binding affinity of another molecule to a protein. Nowadays it is known that “allosteric effect”. The terminology “allostery” is used by Jacques Monod for the first time in 1961 in a study on the conformational changes of biomolecules [4]. The first allosteric model which does not include a conformational change is introduced by Cooper et.al. in 1984 in a study in which they introduced the notion of “dynamic allostery” by adding the concept of entropy into allosteric studies [5]. In 1999, Nussinov et.al. combined the concept of entropy with the free energy landscape [6] to explain conformation ensembles and allosteric networks. In 2004, Nussinov et.al. claimed that all proteins are allosteric [1]. In the same year, the US Food and Drug Administration (FDA) approved the first allosteric drug. This is the important development in terms of healing of diseases. These developments are summarized in Figure 2.1.

Allostery mechanism has been studied both experimentally and computationally for many years. X-ray crystallography is the most popular experimental method. This is because based on the structural differences between the apo and bound structures, conformational changes can be estimated which in turn help us to determine allosteric sites. Another important experimental method is Nuclear Magnetic Resonance (NMR), which is a more effective since one can obtain dynamical information from NMR data.

Computational techniques, on the other hand, complement experimental techniques, and have some practical advantages over experimental techniques. Simulations provide more information about dynamics of biomolecules than any of the experimental methods. Since one can generate many snapshots and motion in a computational simulation, allosteric sites can be identified.

Network analysis can be used to find allosteric pathways after specifying the allosteric site. In order to use such analyses, biomolecules are first converted to a network. Residues (amino acids or nucleotides) are represented as nodes and their interactions with each other are represented as edges. Using these networks, one can discover important nodes that have high potential to transfer signals. These signals are going to help us to understand how allosteric sites affects orthosteric sites.

The balance between the accuracy of the theoretical models and their computational cost is one of the key factors in developing new methods for protein allostery. With the rapidly decreasing computational cost, we can expect more accurate computational methods in the future.

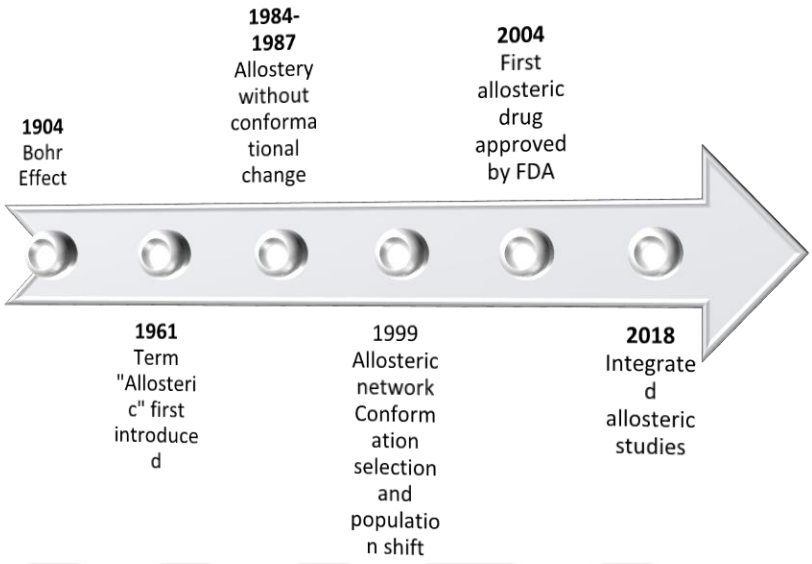


Figure 2.1. History of Allostery.

2.2. Allostery in Ribosome

Allosteric studies on the biomolecules also accelerated by the development of computational methods. There are lots of different computational techniques. Firstly, dynamic simulations help to find allosteric sites and understand the mechanisms. After determining allosteric site, it is possible to investigate how these sites affect the active sites and through which pathways the signal begins to transmit. Network analysis methods are also useful in detecting these pathways. One can model a biological structure as a set of nodes/vertices and a set of the ties/edges connecting these nodes. After biomolecules are converted to graphs, network properties might yield information about signal pathways, i.e. fast routes of spreading of information.

Ribosome is a supramolecular machine that performs protein synthesis in all living cells. Bacterial ribosome (70S) is formed by the association of small (30S) and large (50S) subunits, which in total contain more than 50 different proteins and three RNA chains. Protein synthesis is accompanied by several conformational changes of the ribosome, which are challenging to observe by classical molecular dynamics (MD) simulations. The translocation process still needs clarification in terms of allostery and molecular mechanism [7,8]. Several methods used to investigate ribosome dynamics include elastic network model (ENM) [9-11] coarse-grained MD [12] and MD using a multi-basin structure-based model [1].

2.3. Translocation and Allostery

Translocation process includes several stages as shown in Figure 2.2a. When the initiator tRNA that carries the aminoacid Met (red sphere), collectively named as formylmethionine-tRNA, binds to the P site of the 70S ribosome complex, protein synthesis begins. Then aminoacyl-tRNA-EF-Tu-GTP interacts with the A site, thereby starting the elongation cycle. In this step, the decoding center (DC) performs ably, EF-Tu-GDP and inorganic phosphate quit from the A site of the ribosome complex via hydrolyzing GTP. As a result, the nascent chain is moved from the P site to the A site and the chain is elongated by one amino acid, which is called “peptidyl transfer”. After EF-G hydrolyzes to GTP, the tRNA-mRNA complex is translocated to P site and E site of the 70S ribosome complex and EF-G-GDP and inorganic phosphate quit from the ribosome. In this GTP hydrolysis step, Sarcin Ricin Loop has a significant role in anchoring EF-G to the 70S ribosome complex, while the ribosome conformation changes during the translocation process [13].

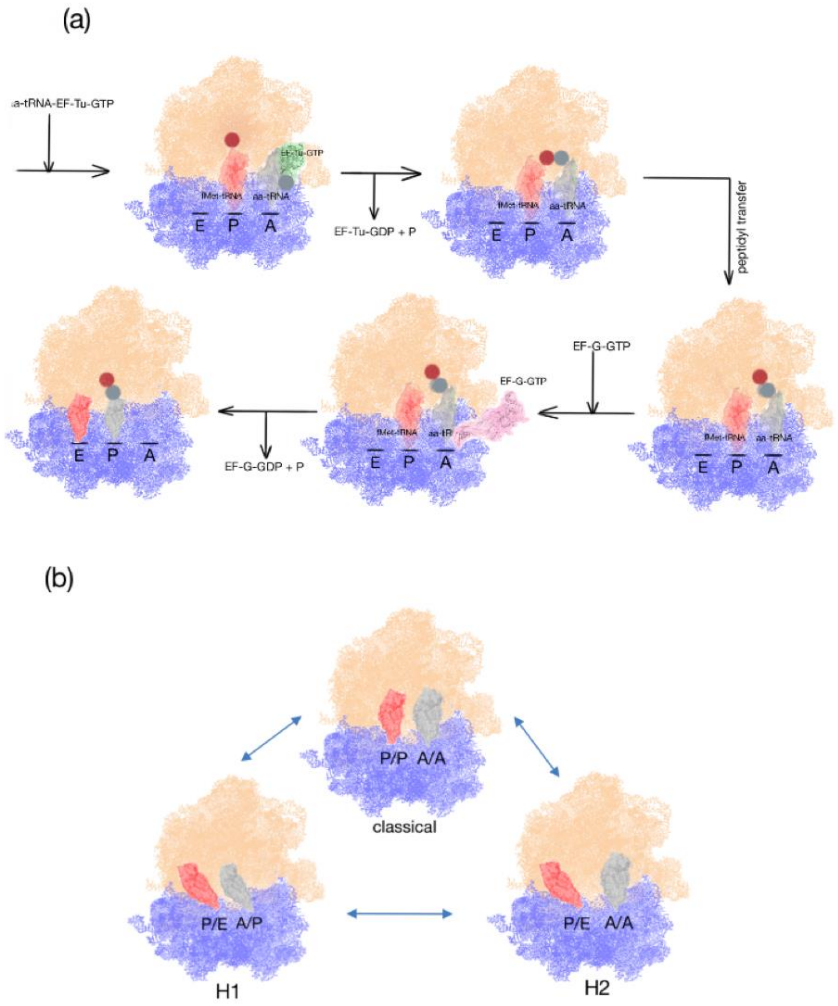


Figure 2.2. Steps of elongation cycle. (a) Elongation steps for the 70S ribosome complex. (b) Classical and hybrid states of tRNA.

After peptidyl transfer, the tRNAs transfer between classical and hybrid states in Figure 2.2b. While the tRNAs settle into A/A - P/P positions in the classical pre-translocational state, they settle into A/P - P/E positions in hybrid 1 state (H1) and A/A – P/E positions in hybrid 2 state (H2). Nearly 12° rotation of the 30S subunit and 3° rotation of head (a negligible swiveling) is related to tRNA movement between binding sites. This rotation indicates a ribosome-EF-G complex before translocational intermediate [14].

The translocation process involves several allosteric communication processes. These processes work about the global dynamics and conformational transitions of the 70S complex. Therefore, the aims were to develop first an efficient method for sampling conformations of the ribosome and then to investigate the allosteric pathways using the generated conformers. As a result, key sites could be revealed on the communication paths, which can be further utilized as alternative docking sites for antibacterial drug design.

3. METHODS

The allosteric communication pathway analysis in this thesis based on the crystal structure of *E. coli* EF-G-ribosome complex (with PDB id: 4V9M, original id: 4kdk-4kdj), which is trapped in an intermediate state of translocation with bound mRNA and tRNA. In order to include the conformational flexibility of the supramolecular structure ribosome, 101 atomistic conformers are used, which have been previously generated by applying the ClustENM conformational sampling method [15] on complex structure, 4V9M. ClustENM is an iterative algorithm, which generates alternative conformers by deformation along the collective modes using elastic network model, further clustering of the generated conformers and then energy minimization of the representative conformer of each cluster. In the elastic network, the center of mass of a residue represents a node and close-neighboring residue pairs/nodes are linked to each other by springs. Magnitude of the spring constant is proportional to number of atom-atom pairs that for within a cutoff distance of 10 A^{o} and the resulting so-called anisotropic network model is used to determine collective modes of motion. Detailed information about the ClustENM method and the ribosome conformers generated can be found in Kurkcuoglu's thesis and their recent study [16].

The communication pathways method, which is developed at the Istanbul Technical University was used for calculation of the shortest pathways. In this method, ribosome is described as a weighted graph which includes C alpha, P atoms as nodes and lengths with linking nodes as edges, which represents the local interaction strength or the affinity. If two residues interact with each other strongly, they are called "close" and they transmit an information using a conformational change [17]. This affinity between node i and j is calculated from equation [16] below.

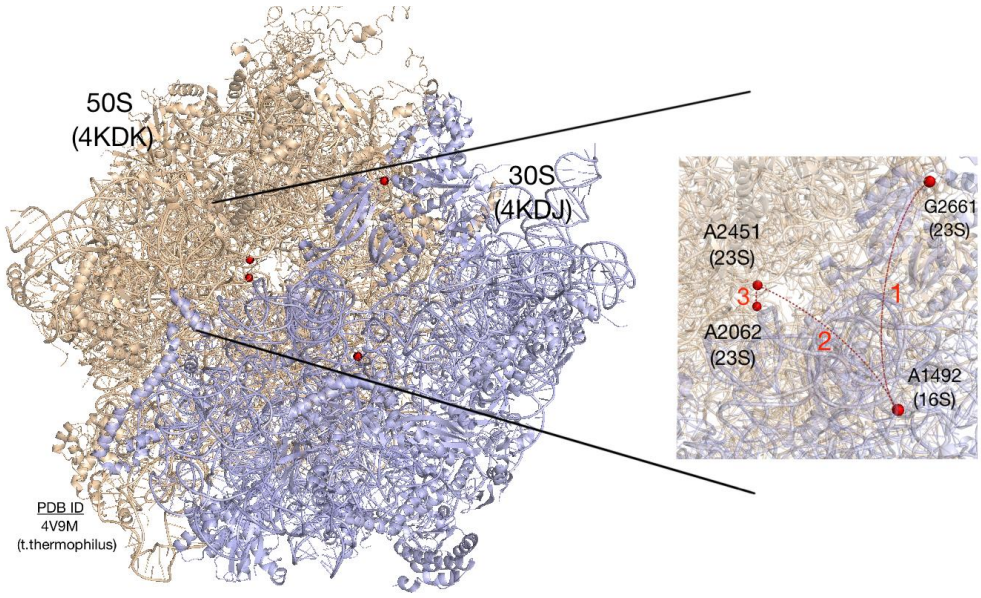
$$a_{ij} = \frac{N_{ij}}{\sqrt{N_i . N_j}} \tag{3.1}$$

Ni and Nj are the number of heavy atoms of residues i and j. Nij is a node pair within a cutoff distance of 4.5 Å Weight of edges between neighboring nodes i and j is set as the inverse of the affinity aij. The cost of a pathway is calculated by summing the weights of the visited edges.

In this thesis, the shortest pathways between DC-SRL, DC-PTC and PTC-Tunnel are calculated by using crystal structure (4V9M) and its 101 ClustENM conformers. On the other hand, the shortest pathway calculation between SecM-PTC and TF-Tunnel is done by using the ribosome complex (4V5H).

For the crystal structure 4V9M, the starting point of the first pathway is DC represented with node A1492 on 16S rRNA and the end point is at the SRL corresponding to G2661 on 23S rRNA. While PTC is located on chain A and residue A2451, the ribosomal tunnel is located on chain A and residue A2062. Both are located on the 50S large subunit of the ribosome structure. These start and sink nodes are shown in Figure 3.1a. For the crystal structure 4V5H, SecM is represented with U2586 on 23S rRNA and TF is represented with the aminoacid Glu18 of the protein L23. These start and sink nodes are shown in Figure 3.1b.

a)



b)

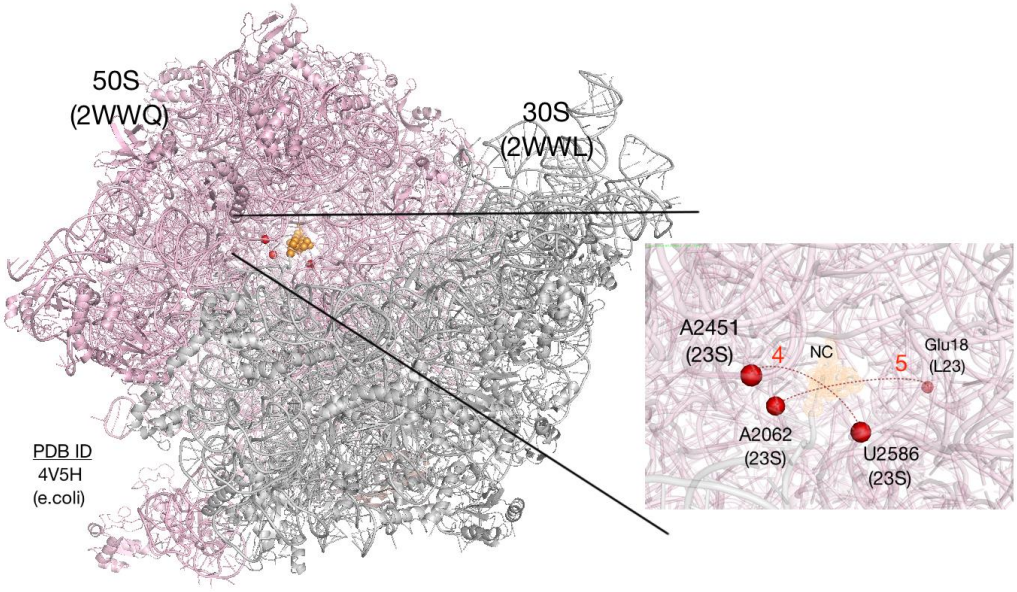


Figure 3.1. The start and sink points of the allosteric communication paths. a) DC, PTC and Tunnel in the crystal structure 4V9M. b) SecM-PTC-TF-Tunnel in the crystal structure 4V5H.

The shortest pathways between the start and sink nodes are calculated by using Yen's algorithm [18] and Dijkstra's algorithm [19] In a previous study, k = 20 shortest pathways were shown to be sufficient for the ribosome complex in order to reveal different residue paths between its functional sites by Kurkcuoglu *et al.* [16]. Here, the authors first calculated k=100 shortest pathways between DC-PTC and ribosomal tunnel-PTC for various bacterial ribosome structures. They clustered all pathways according to node similarity, and obtained 3 main clusters, which were distributed among 20 shortest pathways. Therefore, in this thesis 20 shortest pathways are calculated for the crystal structures (4V9M and 4V5H) and 101 conformers generated with ClustENM.

The frequency of amino acid and nucleotide residues in pathways is calculated to reveal the nodes that are preferred the most. In addition, the shortest pathways between DC-SRL, DC-PTC and PTC-Tunnel based on the crystal structure are compared to those of conformers. After amino acid/nucleotide frequency calculations, costs of these shortest pathways are analyzed. Instead of investigating only the shortest path with the lowest cost, the pathway having a cost with the highest frequency is investigated.

Finally, all paths are converted to one dimensional vectors, *A(1XN).* Here, *N* is the number of nodes. The element *A(1,i)* is equal to *i* if the *i*th node is in the calculated pathway, otherwise it is equal to zero. Then, all vectors are converted to unit vectors by dividing with their magnitudes. The cosine similarity between each pathway is calculated. The value of the similarity score changes between 0 and 1, indicating non-similarity and full-similarity, respectively. The similarity scores are clustered by using the K-Means method. The python code which is used for similarity and clustering calculation is given at Appendix A.

4. RESULT and DISCUSSION

Twenty shortest pathways are calculated for each of the ClustENM conformers, 4V9M and 4V5H crystal structure. Five distinct allosteric sites, which have been reported for ribosome in the literature, are considered in this study as follows: (i) the Decoding Center (DC) and the Sarcin Ricin Loop (SRL) [20], (ii) the DC and the peptidyl transferase center (PTC) [21], (iii) the PTC and the ribosomal tunnel [22], (iv) the secretion monitor (SecM) [23] and the PTC, and (v) the Trigger Factor (TF) [24] and the ribosomal tunnel. These starting points/targets of the pathways are illustrated on the 4V9M and 4V5H crystal structure in Figure 3.1.

4.1. The Shortest Pathways between DC and SRL

The start region is the decoding center (DC) represented with node A1492 on 16S rRNA and the end is at the Sarcin Ricin Loop (SRL) corresponding to G2661 on 23S rRNA. In Figure 4.1, the 50S subunit (4kdk) consists of 23S rRNA (brown), 5S RNA (pink) and the ribosomal proteins (light orange/yellow) color. The 30S subunit (4kdj) consists of 16S rRNA (wheat), tRNA (red, located behind the protein), mRNA (cyan), EF-G (grey) and the ribosomal proteins (blue).

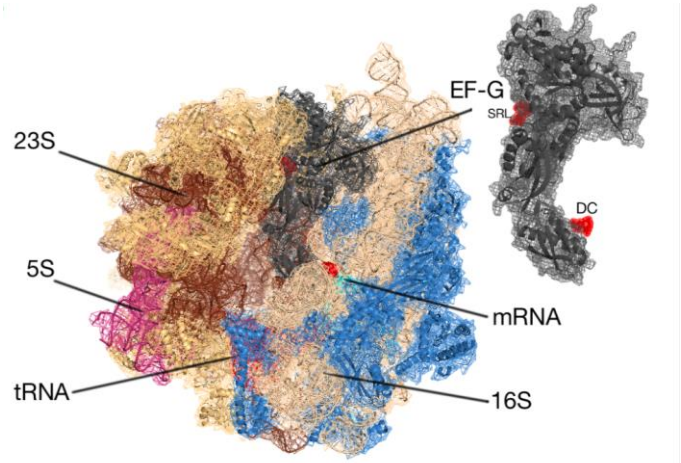


Figure 4.1. 70S crystal structure (4V9M) and its components. EF-G is shown with neighboring SRL and DC on the right from a different angle.

Twenty shortest pathways are calculated for each of the 101 conformers, summing up to 2020 pathways. The occurrence frequency of each residue on these pathways were calculated based on all conformers and plotted in Figure 4.2. In this way, the most visited residues are determined, which are observed in most of the pathways. Almost half of all pathways have the highly frequent residues. One pathway includes about 20 residues between initial and final points.

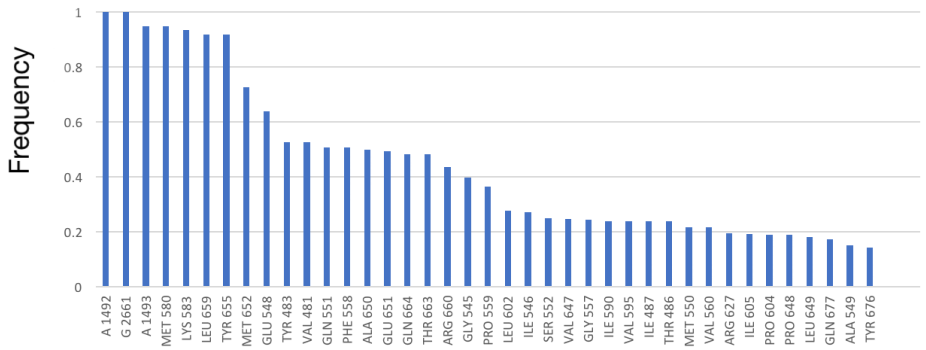


Figure 4.2. Residue frequency based on 2020 shortest pathways between DC-SRL.

In Figure 4.3, the crystal structure is colored according to frequencies of the nodes using Pymol software. Here, red indicates the highest frequency with decreasing order of orange- yellow- green- cyan- blue (lowest frequency). This Figure 4.3 shows that the calculated pathways are passing through the same residues pointing to an optimal pathway between DC and SRL. The spherical nodes describe the shortest pathway for the lowest scored ClustENM conformers. Their score is 6.52. Red nodes that are frequently visited in all 2020 shortest pathways are 16S|A1493, EF-G|MET580, EF-G|LYS583, EF-G|TYR655, EF-G|LEU659.

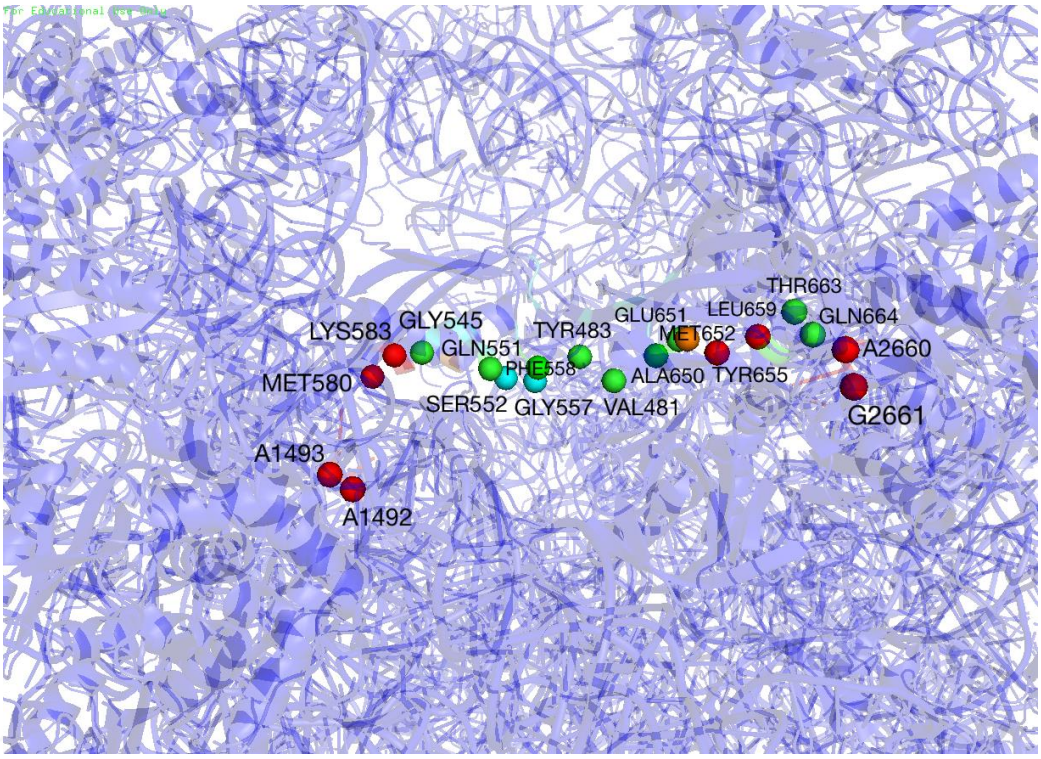


Figure 4.3. The shortest pathway for one ClustENM conformer is shown with spheres and colored according to frequencies based on all paths.

Twenty shortest pathways are also calculated for the crystal structure in order to compare with generated conformers. Figure 4.4 displays the shortest pathways calculated from for the crystal structure, 4V9M, (blue spheres) and the ClustENM conformer (red spheres), which seem to pass through similar regions.

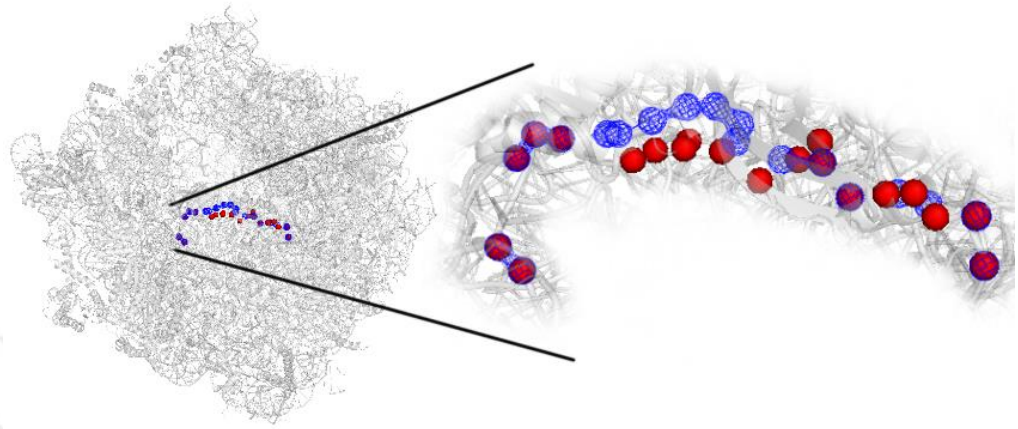


Figure 4.4. The shortest pathways for one ClustENM conformer and the 70S crystal structure (4V9M) are distinguished with red and blue spheres.

The difference between the shortest pathways can be observed in the sequences and their total costs given below. The cost of the 70S crystal structure is clearly higher than that of the conformer, which indicates that the conformer generation procedure may be a key component for analyzing pathways.

ClustENM conformer (total cost: 6.52)

16S|A1492 > 16S|A1493 > EF-G|MET580 > EF-G|LYS583 > EF-G|GLY545 > EF-G|GLN551 > EF-G|SER552 > EF-G|GLY557 > EF-G|PHE558 > EF-G|TYR483 > EF-G|VAL481 > EF-G|ALA650 > EF-G|GLU651 > EF-G|MET652 > EF-G|TYR655 > EF-G|LEU659 > EF-G|THR663 > EF-G|GLN664 > 23S|A2660 > 23S|G2661

Crystal structure (total cost: 10.12)

16S|A1492 > 16S|A1493 > EF-G|MET580 > EF-G|LYS583 > EF-G|GLY545 > EF-G|ALA549 > EF-G|MET550 > EF-G|VAL560 > EF-G|GLU485 > EF-G|LEU602 > EF-G|GLU603 > EF-G|PRO604 > EF-G|LEU649 > EF-G|MET652 > EF-G|TYR655 > EF-G|ASP658 > EF-G|SER661 > 23S|A2660 > 23S|G2661

All the shortest paths are located on the EF-G, which needs to be investigated in detail. EF-G has significant function as a catalyst in translocation during protein synthesis. EF-G consists of five domains, as shown in Figure 4.5. The first domain is known as G domain, which is the nucleotide binding domain. The others are denoted by numbers from II to V. G domain and domain II are similar to EF-Tu [25]. Domains III and V that belong to $\alpha - \beta$ sandwiches group have the same size [26]. Domains II, III, V are related to ribosome binding. On the other hand, Domain IV has a different topology and a crucial function in translocation [25]. After peptidyl transfer, peptidyl tRNA, which is located at the A site, moves towards the P site. In this step, Domain IV of EF-G settles into the A site and functions as a doorstop since it blocks back translocation of tRNA [13].

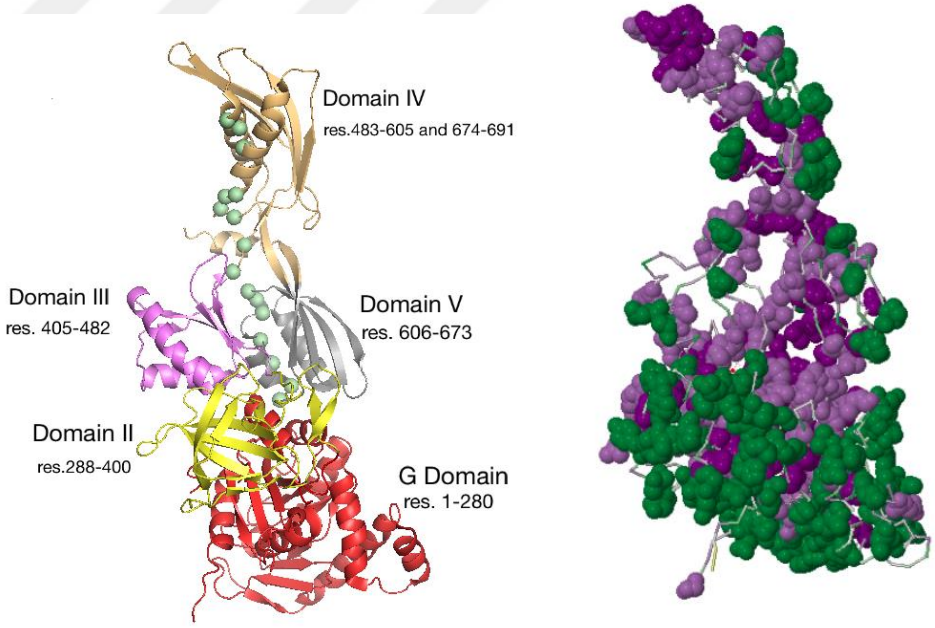


Figure 4.5. a) Structural domains of EF-G with the nodes representing the shortest path for a ClustENM conformer. b) Conservation analysis for EF-G using Consurf server.

The green nodes in Figure 4.5a represent one of the shortest pathways of ClustENM conformers. Residues on the pathways are located in Domain IV and V. Moreover, Domain IV includes highly conserved residues, shown in Figure 4.5b, which points to its functional importance. Consurf server [27] is used for conservation analysis, where scores range from 1 (variable) to 9 (highly conserved). In Figure 4.5b, the inner region of EF-G is highly conserved (light and dark purple corresponding to 8 and 9), whereas its outer region is variable (green,1). Most frequent residues on the pathways are plotted with their conservation scores in Figure 4.6.

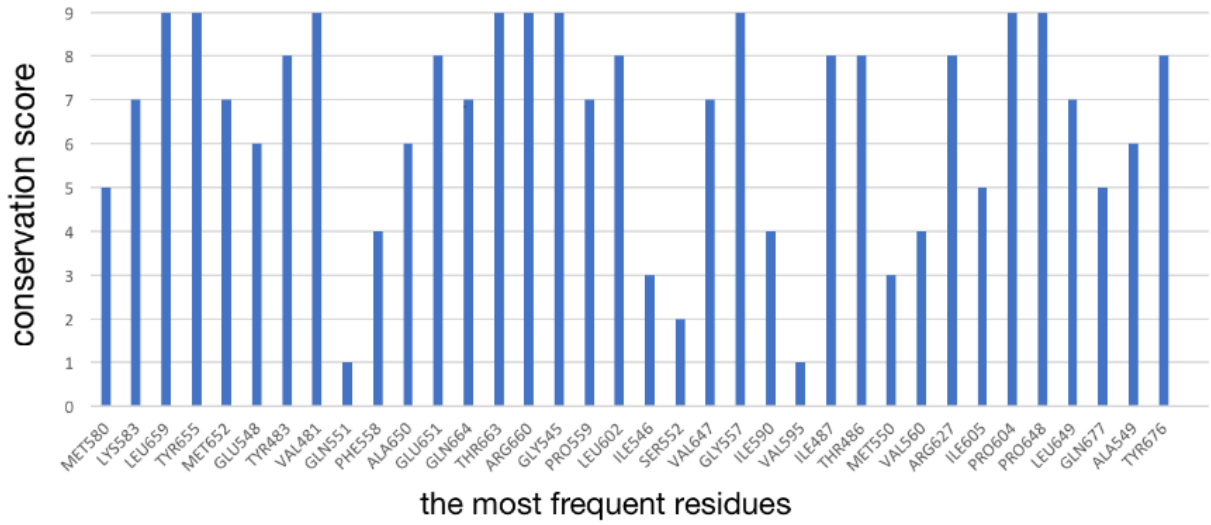


Figure 4.6. Conservation scores of the most frequent amino acids on the pathways.

The most frequent residues A1493 and Met580 have crucial function in GTP hydrolysis on EF-G. During the protein synthesis, EF-G helps the translocation of tRNAs and mRNA by one codon. After GTP hydrolysis and trigger of EF-G, subunits of the ribosome makes a ratchet rotation. A1493 on 16S rRNA of 30S subunit contacts the side chain of Met580 which is located on EF-G and forms a hydrogen bond with Ser578 which is located on EF-G [28]. Domain IV of EF-G interacts with bridge B2a and stabilize this location of B2a. This situation helps ribosome's ratchet rotation. These residues and interactions are shown in Figure 4.7. Here, 50S subunit of ribosome is in pink color, 30S subunit is in gray, Domain IV

of EF-G is in yellow color. Residue Met580 of EF-G is green, Ser578 is blue and A1493 of 16S is orange color.

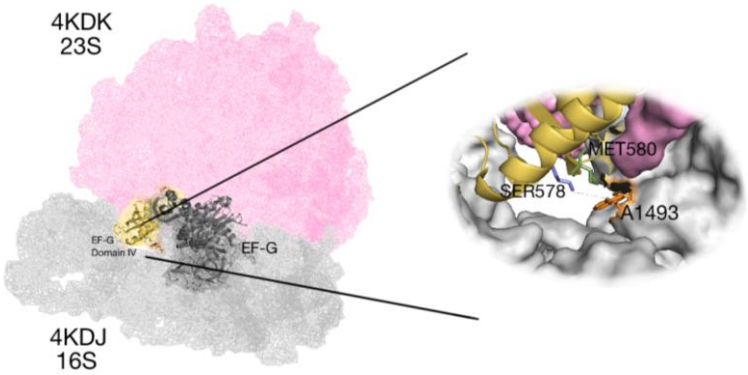


Figure 4.7. Interaction of domain IV of EF-G with B2a bridge on ribosome.

4.2. The Shortest Pathways between DC and PTC

The Peptidyl Transferase Center (PTC) contains the cluster of nucleotides that function together in the process of peptide bond synthesis between the amino acids attached to CCA ends of A- and P-tRNAs. In this section, we investigate the communication pathways between the Decoding Center (DC) and the PTC. Residues A1492 of the Chain A and A2451 of the Chain A are selected for the DC and the PTC, respectively.

We calculated twenty shortest pathways for 100 conformers. As total we obtain 2020 pathways including the shortest pathways of crystal structure. Then overall frequencies of nodes which are located in these 2020 pathways are illustrated in Figure 4.8a-b. The transition from blue nodes to red nodes represents increasing frequency of occurrence. We illustrated one pathway from ClustENM conformers and the crystal structure 4kdk-4kdj, and colored the nodes according to their frequencies in Figure 4.8a-b.

While the cost of the shortest pathway on ClustENM conformer is 6.78, that of the crystal structure is 11.01, as given below. Less frequent nucleotides, namely G1415, G1416, G1417, A1418, are unique to the crystal structure and appear as an alternative for the connection between G1486 and G1959. In contrast, nucleotides U1485, C1484, A1483 form the primary connection between G1486 and G1959 in the ClustENM conformer. As the rest of pathways is mainly consistent for both structures, the difference in the costs arises from this region. Both pathways are shown in the Figure 4.8c, which represents this difference clearly.

The high frequent residue G2553 in the ClustENM conformer is highly conserved and makes Watson-Crick base-pair with the CCA end of A-tRNA [29]. There are some significant residues common to both structures, such as U2506 that has an important role in peptide bond synthesis [30] and is at the same time a drug binding site [31].

ClustENM conformer (total cost: 6.78)

16S|A1492 > 16S|A1491 > 16S|C1490 > 16S|G1489 > 16S|G1488 > 16S|G1487 > 16S|G1486 > 16S|U1485 > 16S|C1484 > 16S|A1483 > 23S|G1959 > 23S|C1958 > 23S|C1957 > 23S|U1956 > 23S|U2552 > 23S|U2554 > 23S|G2553 > 23S|C2507 > 23S|U2506 > 23S|C2452 > 23S|A2451

Crystal structure (total cost: 11.01)

16S|A1492 > 16S|G1491 > 16S|C1490 > 16S|G1489 > 16S|G1488 > 16S|G1487 > 16S|G1486 > 16S|G1415 > 16S|G1416 > 16S|G1417 > 16S|A1418 > 23S|G1959 > 23S|C1958 > 23S|C1957 > 23S|U1956 > 23S|U1955 > 23S|U2552 > 23S|U2554 > 23S|U2555 > 23S|G2508 > 23S|C2507 > 23S|U2506 > 23S|C2452 > 23S|A2451

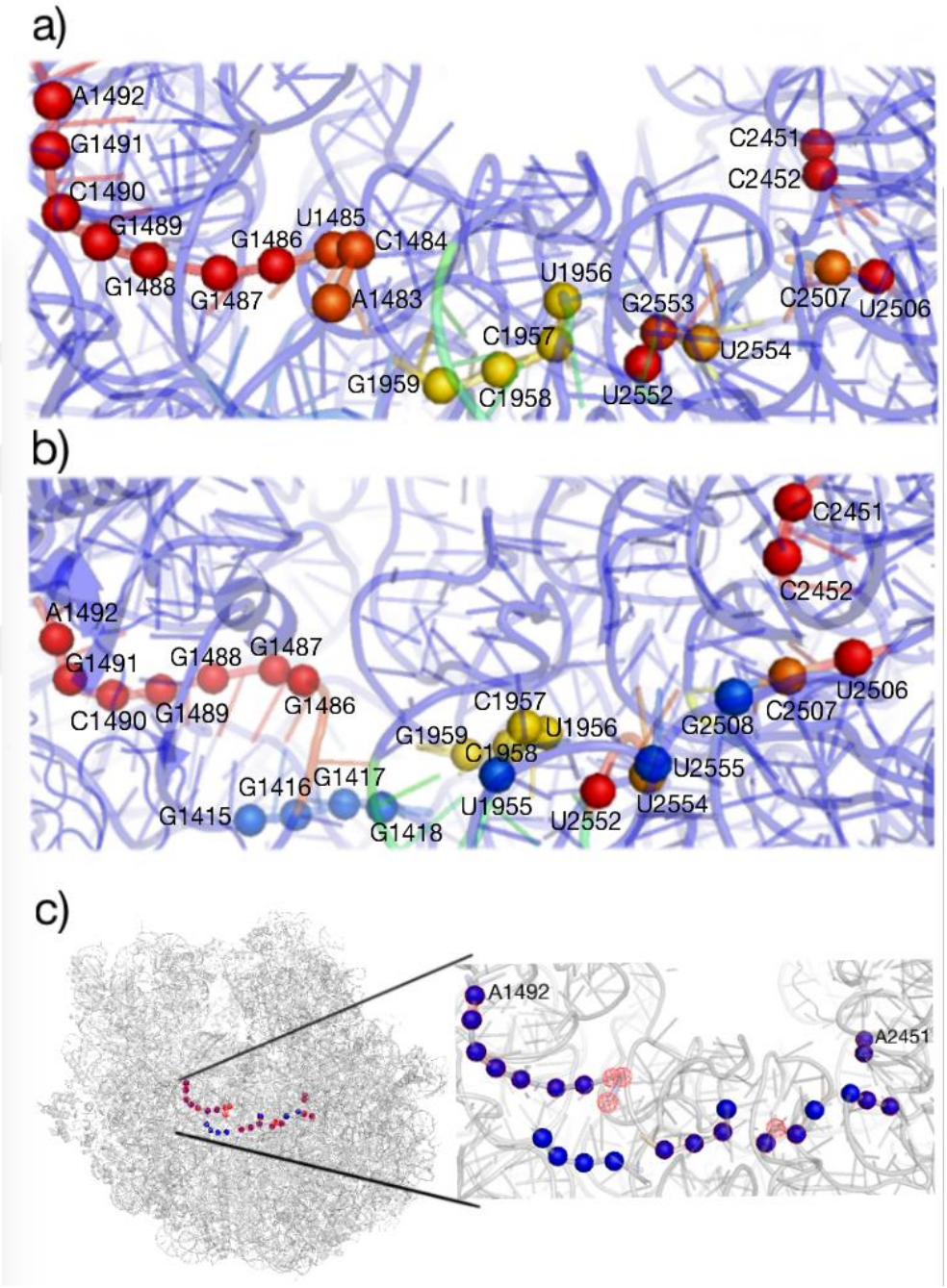


Figure 4.8. Shortest pathways between DC and PTC. (a) The shortest pathway for a ClustENM conformer. (b) the shortest pathway for the crystal structure (4V9M). (c) Comparison of the shortest pathways for crystal structure (blue) and the conformer (red).

In Figure 4.8, one ClustENM conformer with cost equal to 6.78 is represented. The shortest pathway which has the minimum score may not indicate the most probable communication path since biomolecules are flexible. Proteins and their complexes may use short optimal and slightly longer suboptimal pathways for the allosteric communication between distant functional sites. In this regard, there may exist both optimal and suboptimal communication pathway between the DC and the PTC. In order the reveal the lengths (i.e. costs) of the calculated pathways, score distribution for all pathways are analyzed in Figure 4.9. When all conformers and their scores are investigated, frequent scores fall in the region between 7.07-7.14 with maximum frequency observed for score 7.09, shown in Figure 4.9. Again, this distribution of scores is much lower than that observed for the crystal structure.

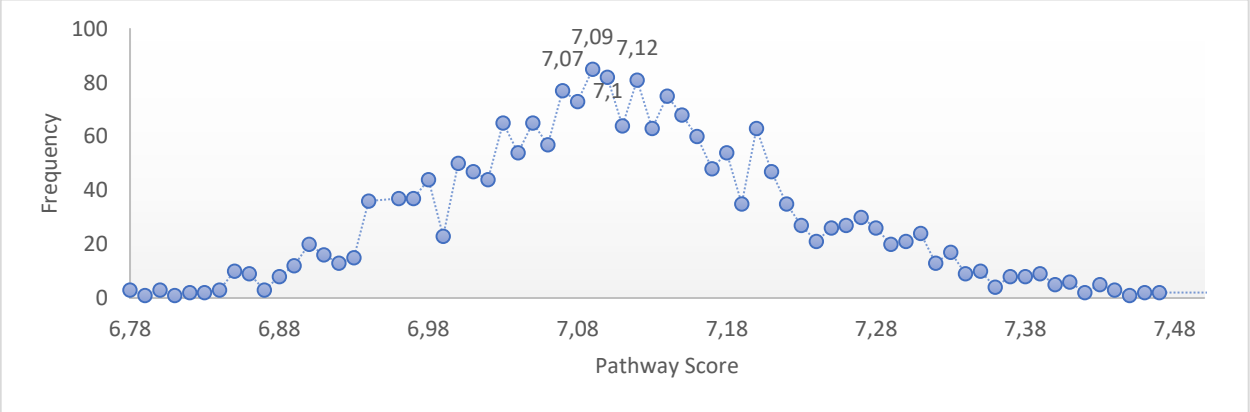


Figure 4.9. Frequency distribution for the scores of pathways between DC and PTC.

85 of all the pathways have the score of 7.09. Among the pathways with score 7.09, the most frequent two paths were as follows :

First path(%36 of 7.09 scored pathways):

16S|A1492 – 16S|A1491 – 16S|C1490 – 16S|G1489 – 16S|G1488 – 16S|G1487 – 16S|G1486 – 16S|U1485 – 16S|C1484 – 16S|A1483 – 23S|G1959 – 23S|C1958 – 23S|C1957 – 23S|U1956 – 23S|U2552 – 23S|C2507 – 23S|U2506 – 23S|C2452 – 23S|A2451

Second path(28% of 7.09 scored pathways):

16S|A1492 – 16S|A1491 – 16S|C1490 – 16S|G1489 – 16S|G1488 – 16S|G1487 – 16S|G1486 – 16S|G1042 – 16S|G1043 – 16S|G1044 – 16S|G1045 – 23S|G1959 – 23S|C1958 – 23S|C1957 – 23S|U1956 – 23S|U2552 – 23S|U2554 – 23S|G2553 – 23S|C2507 – 23S|U2506 – 23S|C2452 – 23S|A2451

When we shifted our attention to the 82 of all pathways that have the score 7.10, the most frequent path which forms the 40% of the whole subset was as follows:

16S|A1492 – 16S|A1491 – 16S|C1490 – 16S|G1489 – 16S|G1488 – 16S|G1487 – 16S|G1486 16S|G1042 – 16S|G1043 – 16S|G1044 – 16S|G1045 – 23S|G1959 – 23S|C1958 – 23S|C1957 – 23S|U1956 – 23S|U2552 – 23S|U2554 – 23S|G2553 – 23S|G2583 – 23S|C2507 – 23S|U2506 – 23S|C2452 – 23S|A2451 (40%)

There were 81 of all pathways with the score of 7.12. Among these the most frequent path formed the 27% of this subset was as follows:

16S|A1492 – 16S|A1491 – 16S|C1490 – 16S|G1489 – 16S|G1488 – 16S|G1487 – 16S|G1486 – 16S|U1485 – 16S|C1484 – 16S|A1483 – 23S|G1959 – 23S|C1958 – 23S|C1957 – 23S|U1956 – 23S|U2552 – 23S|U2554 – 23S|U2555 – 23S|G2508 – 23S|C2507 – 23S|U2506 – 23S|C2452 – 23S|A2451 (27%)

On the opposite side, there were 77 with the score of 7.07 score. The most frequent two paths constitute 32% of the whole sample. These were as follows:

16S|A1492 – 16S|A1491 – 16S|C1490 – 16S|G1489 – 16S|G1488 – 16S|G1487 – 16S|G1486 – 16S|U1485 – 16S|C1484 – 16S|A1483 – 23S|G1959 – 23S|C1958 – 23S|C1957 – 23S|U1956 – 23S|U2552 – 23S|U2554 – 23S|U2555 – 23S|G2508 – 23S|C2507 – 23S|U2506 – 23S|C2452 – 23S|A2451 (32%)

16S|A1492 – 16S|A1491 – 16S|C1490 – 16S|G1489 – 16S|G1488 – 16S|G1487 – 16S|G1486 – 16S|U1485 – 16S|C1484 – 16S|A1483 – 23S|G1959 – 23S|C1958 – 23S|C1957 – 23S|U1956 – 23S|U2552-23S|G2553–23S|G2583 – 23S|C2507 – 23S|U2506 – 23S|C2452 – 23S|A2451 (32%)

As a result, we determined the most frequent scored 6 divergent pathways among the total 2020 pathways, whose scores are close to the mean of the score distribution. When we analyze the nucleotides in these paths, we detected 21 common nucleotides that are in all of these frequent scored pathways.

From our analysis, we conclude that there are no distinct alternative pathways based on the generated conformers. The result is shown in Table 4.1.

Table 4.1. Aminoacids which are located on the frequent scored 6 pathways.

Nucleotides	**%Occurence**	**Nucleotides**	**%Occurence**
A|G1486	100	A|U2552	100
A|G1487	100	A|C2507	100
A|G1488	100	A|C1958	100
A|G1489	100	A|G1959	100
A|C1490	100	A|A2451	100
A|A1491	100	A|C2452	100
A|A1492	100	A|U2506	100
A|G1486	100	A|C2507	100
A|G1487	100	A|G2553	83
A|G1488	100	A|A1483	67
A|U1956	100	A|C1484	67
A|C1957	100	A|U1485	67

Generated conformers from ClustENM are mainly like each other. Therefore, (i) most frequent scored pathway, (ii) the shortest pathway of conformers which have minimum score and the shortest pathway of the crystal structure are shown in Figure 4.10. When these 3 pathways are compared with each other, 16 nucleotides are common in all of them.

Common nuclotides:

16S|A1492, A1491, C1490, G1489, G1488, G1487, G1486

23S|G1959, C1958, C1957, U1956, C2507, U2506, C2452, A2451

Some nodes which are located in the pathway of the crystal structure listed below are not located in the pathway of the ClustENM conformers. The crystal structure presents a partially different path on a region shown in Figure 4.10 (Green nodes). These nodes are:

16S|G1415, G1416, G1417, A1418 23S|U1955, U2554, G2508

(i) The most frequent scored pathway of ClustENM conformers (7.09)

A|A1492 > A|A1491 > A|C1490 > A|G1489 > A|G1488 > A|G1487 > A|G1486 > A|U1485 > A|C1484 > A|A1483 > A|G1959 > A|C1958 > A|C1957 > A|U1956 > A|U2552 > A|G2553 > A|C2507 > A|U2506 > A|C2452 > A|A2451

(ii) The minimum scored pathway of ClustENM conformers (6.78)

A|A1492 > A|A1491 > A|C1490 > A|G1489 > A|G1488 > A|G1487 > A|G1486 > A|U1485 > A|C1484 > A|A1483 > A|G1959 > A|C1958 > A|C1957 > A|U1956 > A|U2552 > A|U2554 > A|G2553 > A|C2507 > A|U2506 > A|C2452 > A|A2451

(iii)The shortest pathway of Crystal structure (11.01)

A|A1492 > A|G1491 > A|C1490 > A|G1489 > A|G1488 > A|G1487 > A|G1486 > A|G1415 > A|G1416 > A|G1417 > A|A1418 > A|G1959 > A|C1958 > A|C1957 > A|U1956 > A|U1955 > A|U2552 > A|U2554 > A|U2555 > A|G2508 > A|C2507 > A|U2506 > A|C2452 > A|A2451

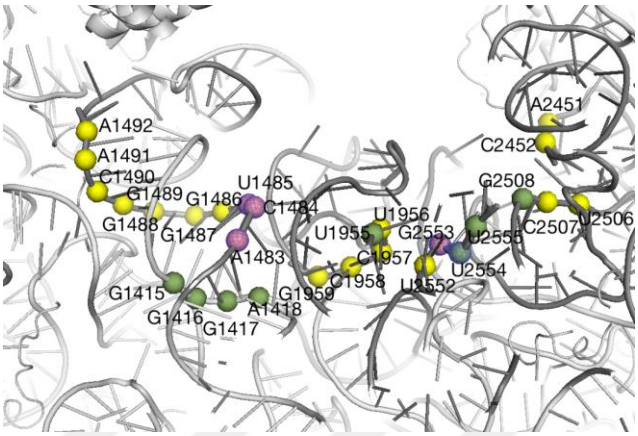


Figure 4.10. The shortest pathways which have different score of ClustENM and the shortest pathway of Crystal Structure, 4V9M (green nodes).

In Figure 4.10, yellow spheres represent the common nodes of 3 pathways. Green nodes show the shortest pathway of the crystal structure. Pink nodes show the most frequent scored pathway of ClustENM structures. Blue mesh nodes show the minimum scored pathway of the CLustENM structures. When average costs of shortest pathways from ClustENM conformers are evaluated, same nucleotides are depicted from the analysis.

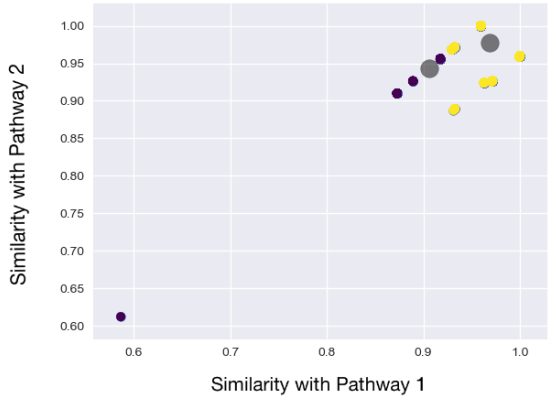


Figure 4.11. K-Means clustering of similarity scores of all shortest pathways between DC-PTC.

We computed 20 paths for each of the 100 clustENM conformers that yields 2000 paths in total. In order to calculate the similarities of the resulting paths we assigned a unique vector to each of these paths.

We used the normalized dot product of the assigned vectors which calculates the cosine of the angle between them to compare these paths. This is also known as the cosine similarity. The similarity measure varies between 0 and +1. If the similarity is 1 the vectors point in the same direction while they are orthogonal if the similarity is 0.

Since we have 2000 paths in total, we constructed a 2000x2000 symmetric matrix with entries between 0 and 1 that contains all possible 2-way comparisons. We use the python code listed in Appendix A to calculate the similarity matrix. The resulting matrix is passed through the k-means clustering algorithm to cluster similar paths.

For a fixed number k, the algorithm starts by selecting k random centroids. In each iteration the algorithm determines which cluster a data point belongs to by using the closest centroid, and then recalculates the centroids for each cluster after going through all data points. Algorithm stops when the centroids stabilze. For our data set, we set the number of clusters to be 2 to split our data set into 2 separate clusters. The gray spheres in Figure 4.11 represent the centroids we calculated while spheres with two distinct colors represent two distinct clusters.

Since each vectors represented by a 2000 dimensional vector, we project them to their first two components in order to sketch them. One of these pathways has lower similarity score, about 0.6. This pathway and one pathway which represents to rest of pathways are shown in Figure 4.12.

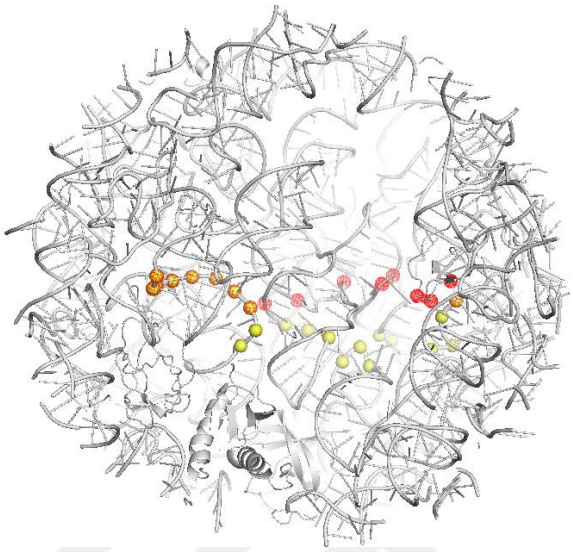


Figure 4.12. The shortest pathway (red) which have 0.6 similarity score while other pathway (yellow) have more then 0.9 similarity.

One of the higher similarity scored pathways:

16S|A1492 – 16S|A1491 – 16S|C1490 – 16S|G1489 – 16S|G1488 – 16S|1487 – 16S|G1486 – 16S|U1485 – 16S|C1484 – 16S|A1483 – 23S|G1959 – 23S|C1958 – 23S|C1957 – 23S|U1956 – 23S|U2552 – 23S|U2554 – 23S|G2553 – 23S|G2582 – 23S|G2583 – 23S|C2507 – 23S|U2506 – 23S|C2452 – 23S|A2451 (cost:7.26)

Lower similarity scored pathway:

16S|A1492 – 16S|A1491 – 16S|C1490 – 16S|G1489 – 16S|G1488 – 16S|1487 – 16S|G1486 – 16S|U1485 – 23S|C1971 – 23S|A1946 – 23S|U1950 – 23S|G2603 – 23S|A2602 – L27|HIE3 – L27|ALA2 – L16|ARG82 – 23S|A2451 (cost:13.30)

The lower similarity scored pathway (cost: 13.30) has less nodes than the other shorter pathways between the DC and the PTC. High cost is due to relatively lower interaction strength between the nodes. This pathway uses the tertiary interaction on two ribosomal proteins, specifically ALA2 and HIS3 on L27, and ARG82 on L16 to reach the PTC. These nodes are not noted in the other 1999 paths.

4.3. The Shortest Pathways between PTC and Ribosomal Tunnel

The shortest pathways between the flexible ribosomal tunnel entrance and the PTC are analyzed. Residues A2062 and A2451 are chosen on the tunnel and the PTC, respectively. Both are located on the 50S large subunit of the ribosome structure. Overall frequencies of 2020 pathways are illustrated on the 4kdk-4kdj crystal structure in Figure 4.13. There is not a significant difference among shortest pathways on the two structures since this region is relatively rigid as it is located at the core of the complex.

ClustENM conformer (total cost: 1.16)
23S|A2062 > 23S|C2063 > 23S|C2064 > 23S|A2450 > 23S|A2451

Crystal structure (total cost: 2.39)
A|A2062 > A|C2063 > A|A2450 > A|A2451

Here, A2450-C2063 is a highly conserved non-Watson-Crick base pair [32] and affects flexibility of the ribosomal tunnel. These nuclotides have effect on peptide bond formation. Also they are essential for effective tRNA translocation [33].

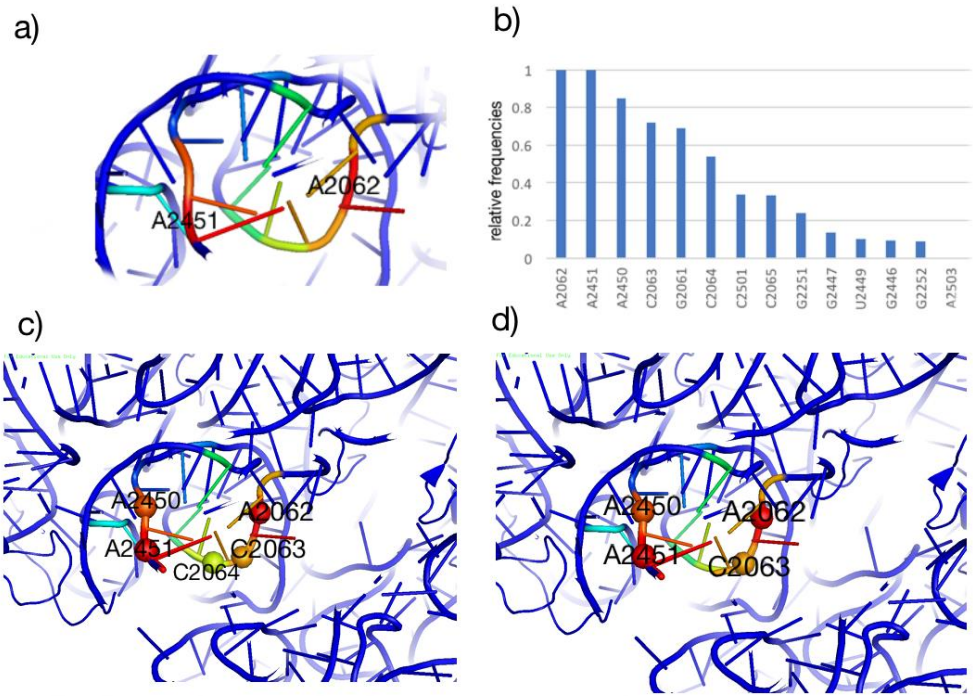


Figure 4.13. Pathways between PTC and tunnel. (a) Overall frequencies of residues (red to blue indicates decreasing frequency), (b) Frequencies of residues on all pathways, (c) The shortest pathway for conformer, (d) The shortest pathway for the crystal structure.

Although the shortest pathway on the crystal structure has a higher cost than the conformer, it crosses four nodes. On the other hand, there are five nodes for the shortest pathway of the conformer. Four of these nodes have the same the crystal structure. Only residue C2064 of A chain is different, and it decreases the communication pathway score. C2064 is one of the conserved nucleotides of PTR (Peptidyl Transferase Ring). It affects a multibranched loop of 23S as rotational or translational shifts of above 2 Å [34].

We have 101 conformers from ClustENM, and we calculated 20 shortest pathways for each conformer. As total, we obtain 2040 pathways and their pathway costs. We considered the shortest pathway which has lower cost as a communication pathway between PTC – ribosomal tunnel. Then we analyzed nucleotides according to their frequencies among these

2040 pathways. Lastly, we calculated costs' frequencies because the lower cost pathways may not be the most efficient ways for communication. We charted these frequencies in Figure 4.14.

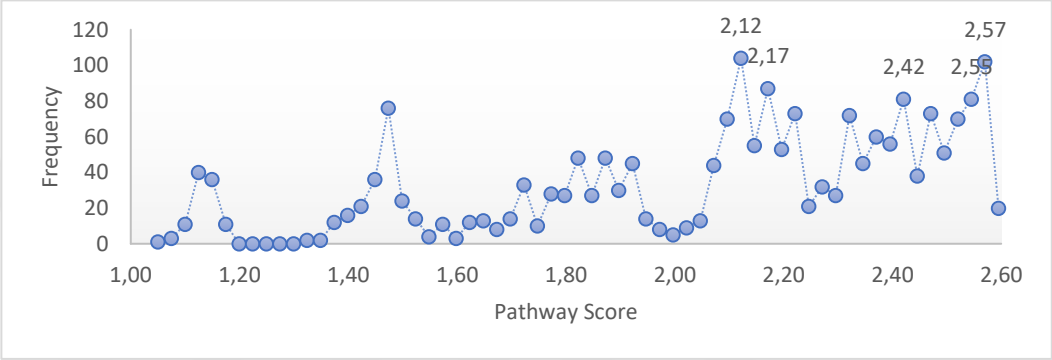


Figure 4.14. Frequency of scores of pathways between PTC and Ribosomal Tunnel

Among all pathways calculated for generated conformers, 104 paths have the cost equal to 2.12. We found 6 subgroups for these pathways, as shown in Table 4.2. While they have the same cost, they include nucleotides at different quantities. Among these 104 conformers, 39% of them is the first pathway shown in Table 4.2.

Below, we observe 5 pathways in each conformer calculation. There are 102 conformers and 102 pathways. Only the fifth pathway shown on Table 4.2 constitutes the 20% of all conformers. Their percentages of occurance are summarized in Table 4.2.

Table 4.2. Pathways which have most frequent score as 2.12 between PTC-Tunnel.

No	Initial Target	Pathways	Final Target	%Occurance	Score
1	A2062	|G2061>|C2501> A2450	A2451	%100	2.11
2	A2062	|C2063>|C2064>|C2065>|U2449> A2450	A2451	%100	2.08
3	A2062	|C2063>|C2064>|C2065>|G2252>|G2251> A2450	A2451	%100	2.18
4	A2062	|G2061>|C2501>|G2447> A2450	A2451	%100	2.15
5	A2062	C2063	A2451	%20	2.26
6	A2062	|G2061>|C2063>|C2064>|C2065>|G2251> A2450	A2451	%100	2.19
CS	A2062	C2063 > A2450	A2451	-	2.39

*CS: Crystal Structure

In Table 4.2, there are two remarkable pathways, and these are pathways 1 and 5. The first pathway has a high percentage and its cost, 2.11, is close to overall cost (2.12). It could be communication pathway between PTC and ribosomal tunnel. Although the fifth pathway is not a hundred percent but we compare to the shortest pathway of the crystal structure, we observe that it is the closest pathway to the crystal structure. The crystal structure has only one extra nucleotide: A6130. Therefore Pathway 5 can be an alternative pathway to Pathway 1. As a result, the pathways in Table 4.2 (except Pathway 5) are available in all conformers, and all have the most preferred pathway cost.

After all, there are three pathways to evaluate communication between PTC and ribosomal tunnel entrance. Red mesh nodes represent to the shortest pathway of crystal structure, blue nodes represent to the Pathway 5 and yellow mesh nodes represent to the Pathway 1 of conformers in Figure 4.15. The Pathway 5 and the shortest pathway of the crystal structure are highly similar. The Pathway 1 has two different nucleotides:C2501 and G2061. On the other hand, A2062 is a common nucleotide of these three pathways.

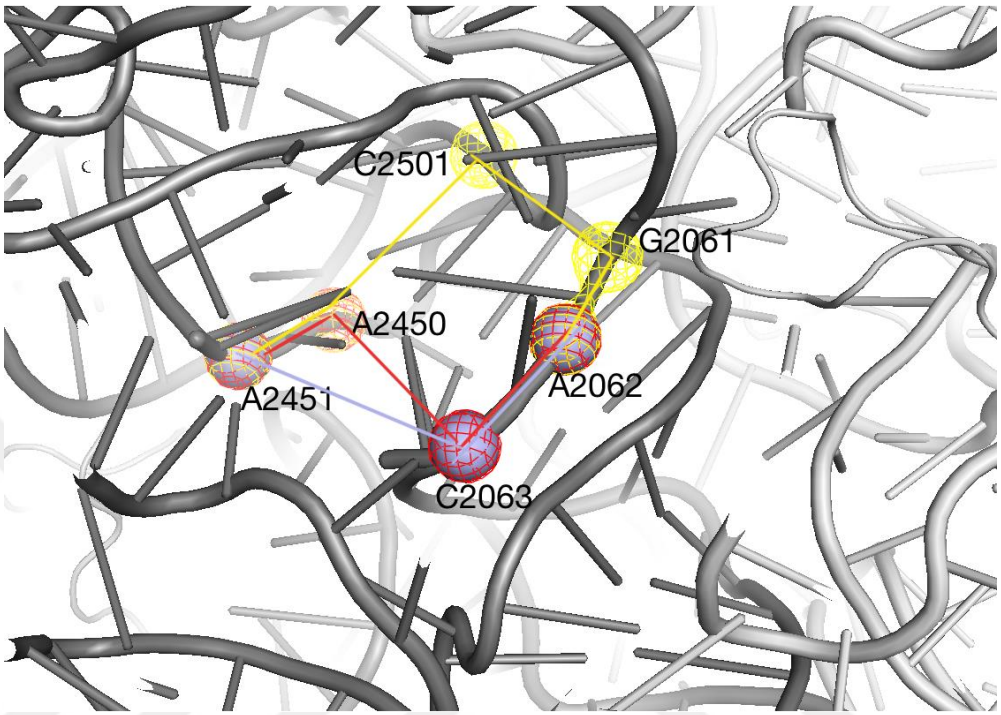


Figure 4.15. The most frequent scored paths between PTC and Ribosomal Tunnel (Red→Crystal Blue→Path5 Yellow→Path1).

To determine optimal pathway and alternative suboptimal pathways, all paths are converted to vectors A(1XN). Dot products between the normalized vectors are calculated. Results are clustered by using the K-Means clustering algorithm. Four clusters are detected. One of these clusters is different from the others with a similarity score of 0.36.

Pathway 1 (2.45)

A2062 – G2061– A2503– U2504– C2452 – A2451

Pathway 2 (1.41)

A2062– C2063– A2450– A2451

Pathway 1 has four nucleotides and a score of 2.45. Pathway 2 consists of only two nucleotides and has lower score as 1.41. These pathways are shown in Figure 4.16. Red nodes represent the Pathway 1 and blue mesh nodes represent the Pathway 2.

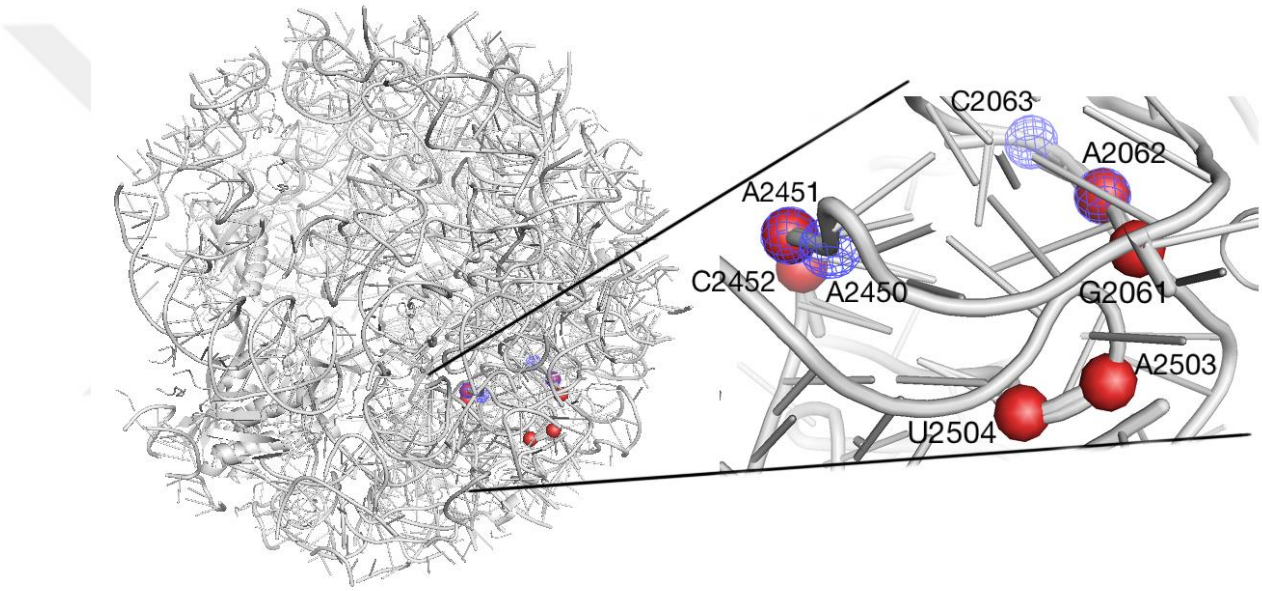


Figure 4.16. Similarity score is 0.36 between above illustrated two pathways between PTC and Ribosomal Tunnel.

According to studies [35], A2602, C2452, A2503, U2504, G2505, U2585, G2061, and U2506 create a binding pocket for drugs. Pathway 1 (cost:2.45) has a higher pathway score but it includes the above-mentioned significant nucleotides. These nodes can form a suboptimal pathway for the communication of PTC and ribosomal tunnel.

4.4. The Shortest Pathways between SecM and PTC

Although newly formed chains are thought to passively pass through the tunnel when exiting the PTC, SecM (Secretion Monitor) interacts with residues on tunnel, affecting the ribosome to stall. It regulates the expression of downstream gene products [23]. SecM inhibits peptide-bond formation in the PTC. It is a secreted protein with 170 aminoacids in length, including 17 amino acids stalling sequence 150FXXXXWIXXXXGIRAGP166 (*E.coli* numbering) near its C terminus, which is enough to stimulate stalling. When SecM homologs in different species are compared, only ILE162, ARG163, and PRO166 are obtained to be invariant [23]. ARG163 makes a hydrophobic interaction with ribosome. Mutation of ARG163 residue can revoke stalling [36]. ARG163 of SecM relocates its side chains between the bases of U2586 and U2609. U2586 is one of the co-responsive elements that receives conformational signals from new chains within the tunnel to modulate the kinetics of translation elongation [37].

There are probable interaction pathways between the critical SecM residue, ARG163, and the PTC [23]. A residual relay via the ribosome that connects SecM to PTC has previously been proposed as a means of communicating the presence of SecM [38]. ARG163 interacts with U2586 and transfer a signal to PTC. The communication pathways between U2586 and A2451 (PTC) are calculated in this thesis. The ribosome 4V5H, which contains a nascent chain in the tunnel is used for the calculation. 20 shortest pathways are calculated. 17 different nucleotides and aminoacids from 23S, tRNA and NC (Nascent Chain) are found in the shortest pathways. Their frequency graph is shown in Figure 4.17.

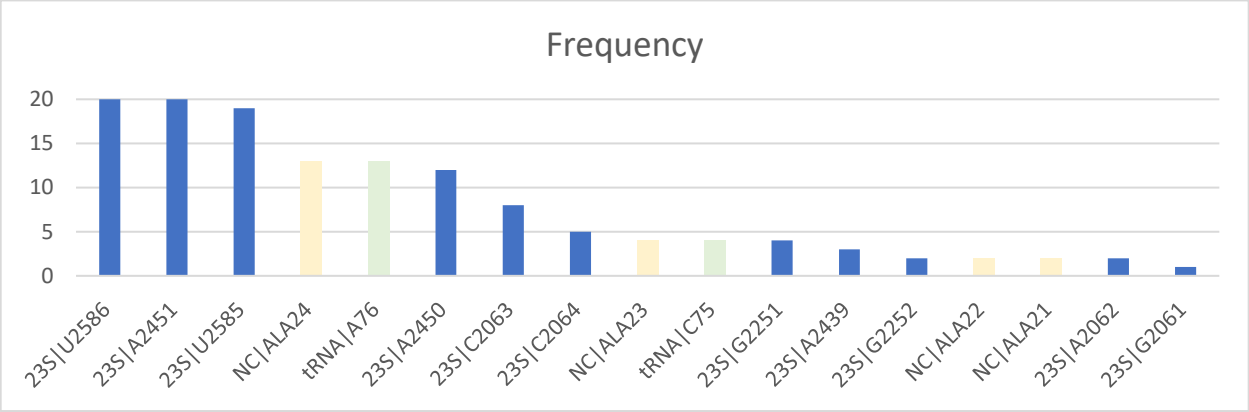


Figure 4.17. Frequency of nucleotides and aminoacids in all pathways. Yellow column represents NC, green column represents tRNA.

Aminoacids ALA21-22-23-24 from the nascent chain are located on shortest pathways. Except ALA24, they have all low frequency. Nucleotides C75 and A76 forming the CCA end of P-site tRNA are surprising results. Each subunit has three binding sites for transfer RNA (tRNA) molecules that are in three different functional states: an A-site which is responsible for the recognition of aminoacyltRNAs (aa-tRNAs), a P-site that binds peptidyl-tRNAs and an E (exit)-site which holds the deacylated tRNA before it dissociates from the ribosome. tRNAs can also effectively transfer perturbation as flexible linkers [16]. Remaining nodes of the pathways belong to 23S.

The shortest pathways accommodates the most frequent nucleotides and amino acids. The shortest pathway score is 2.84, and is given below and in Figure 4.18.

23S|U2586 --> 23S|U2585 --> NC|ALA24 --> P-Site tRNA|A76 --> 23S|A2451 (cost: 2.84)

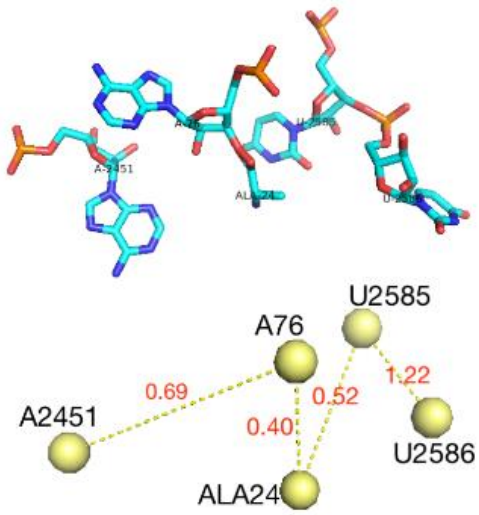


Figure 4.18. The shortest pathway between SecM and PTC.

A76 from P-tRNA and ALA24 from NC are located in the middle of the pathway. Without the polypeptide chain and P-tRNA, the distance between U2585 and A2451 is very high to communicate. Here, it is plausible to think that nascent chain and tRNA could be facilitate the signal transmission.

On the other hand, nascent chains comprise different aminoacids, not only ALA. It may not have any effect on allostery. To analyze this case, we focus on other shortest pathways which don't include the nascent chain. The second shortest pathway, which has 2.96 score, contains U2585 and A76 of P-tRNA. This shortest pathway not only provide an alternative pathway without nascent chain, but also it has a good score comparable with the shortest pathway. The second pathway is, 23S|U2586 --> 23S|U2585 --> P-Site tRNA|A76 --> 23S|A2451 (cost: 2.96). This alternative pathway is shown in Figure 4.19.

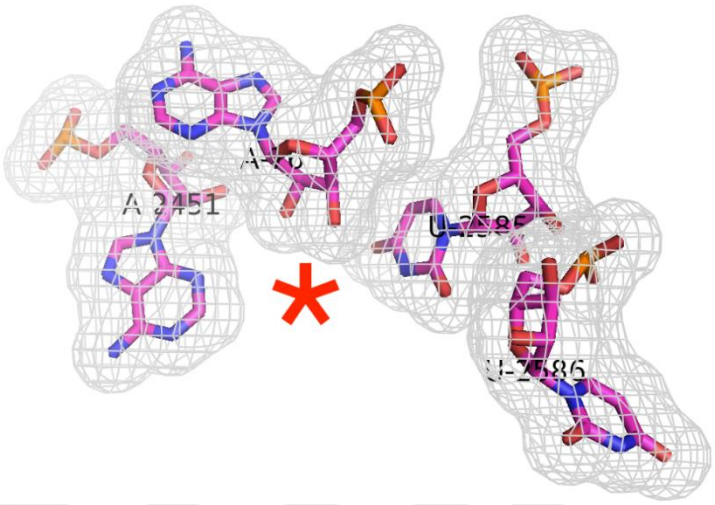


Figure 4.19. The second shortest pathway between SecM and PTC

The base of U2585 contacts the peptidyl bond linkage between GLY165 and A76 of P-tRNA. ARG163 and GLY165 are located at the starred area in Figure 4.19. U2585 and A76 make strong interaction with R163 [37]. They completely disable the PTC in an uninduced state. The position of U2585 and R163 form a steric hindrance to correct accommodation of the incoming aminoacyl-tRNA to the A-site [37].

Moreover, less visited nucleotides G2061 and C2063 seem as significant nucleotides for allosteric communication. G2061 interacts with C2063 through their bases and by shifting of the rRNA backbone. From there, C2063 contacts A2450 and A2451 (PTC), which also interacts with A76 of the P-tRNA. On the other side of SecM, R163 interacts with U2586 and A2587. On the other hand, G2061 is only located in one pathway among twenty shortest pathways. This is pathway 18, shown at Figure 4.20. It has high cost of 4.64, but this pathway includes both nucleotides: G2061 and C2063.

23S|U2586-->23S|U2585-->NC|ALA24-->23S|C2063-->23S|G2061-->23S|A2451 (cost: 4.64)

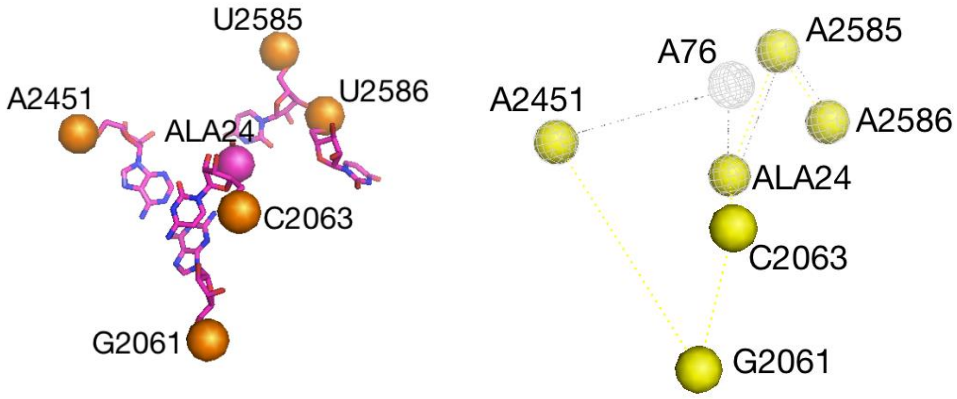


Figure 4.20. Comparison of Pathway 1 (grey mesh) and Pathway 18 (yellow)

Pathway1 and Pathway 18 are shown on the structure in Figure 4.20 for comparison.

4.5. The Shortest Pathways between TF and Ribosomal Tunnel

Trigger Factor (TF) is a molecular chaperone which helps folding into functional three dimensional structures of newly synthesis polypeptide chain. When the nascent chains leave the ribosome exit tunnel, they can form their native structural elements and even fully folded domains while still connected to the PTC [39,40]. Trigger Factor (TF) not only associates with nascent chains, but also with full-length proteins to stabilize native structure until they are fitted into protein [41].

The ribosomal tunnel is mainly composed of ribosomal RNA (~80%) and three ribosomal proteins(L4, L22, L23). The narrowest part of the tunnel is built by a β -hairpin loop of L22 and L4. The nascent polypeptide chain in the the exit tunnel can affect TF recruitment [24]. In addition, the loop of L23 triggers a conformational signal to the TF binding surface of

L23 and regulate the TF recruitment[24]. Studies show that TF recruitment is greatly reduced when signal receive the L23-related region within the tunnel. TF interacts directly with L23; aminoacid Glu18 of L23 protein makes the interaction with TF. Glu18 is a prerequisite for the joining of TF with nascent polypeptide chains [42]. In order to calculate k-shortest pathways, Glu18 is selected as the start node and A2062 on the tunnel is selected as the sink node.

The shortest pathway, where NC is the nascent chain is as follows:

L23|GLU18-->L23|LYS19-->23S|A1392-->23S|U1316-->23S|C1315-->23S|C1314-->23S|G1332-->23S|A1609-->23S|A1616-->23S|C1615-->23S|A1614-->L22|GLY91-->NC|ALA12-->NC|ALA13-->NC|ALA14-->NC|ALA15-->NC|ALA16-->NC|ALA17-->NC|ALA18-->NC|ALA19-->NC|ALA20-->NC|ALA21-->23S|A2062 (cost: 11,71)

Here, the shortest pathway passes through ribosomal proteins L4, L22 and L23. The shortest pathway is indicated on Figure 4.21.

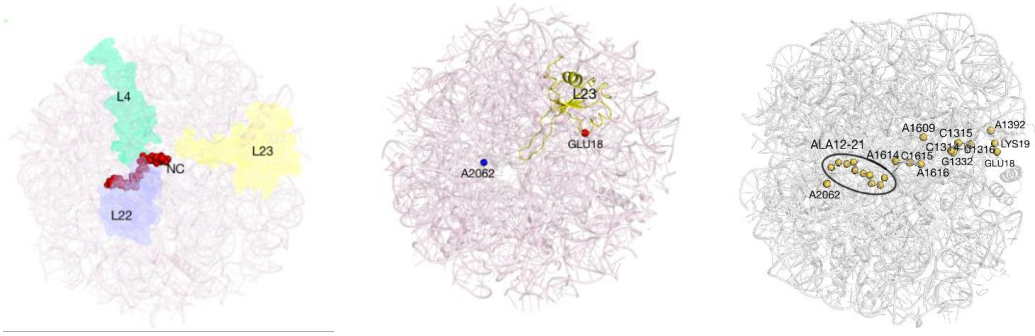


Figure 4.21. a) The crystal structure with NC, L4, L22 and L23, b)The start and sink point for TF-Tunnel, c) The shortest pathway between TF and Tunnel.

The frequency of appearance of all nodes is calculated over all pathways, and shown in Figure 4.22.

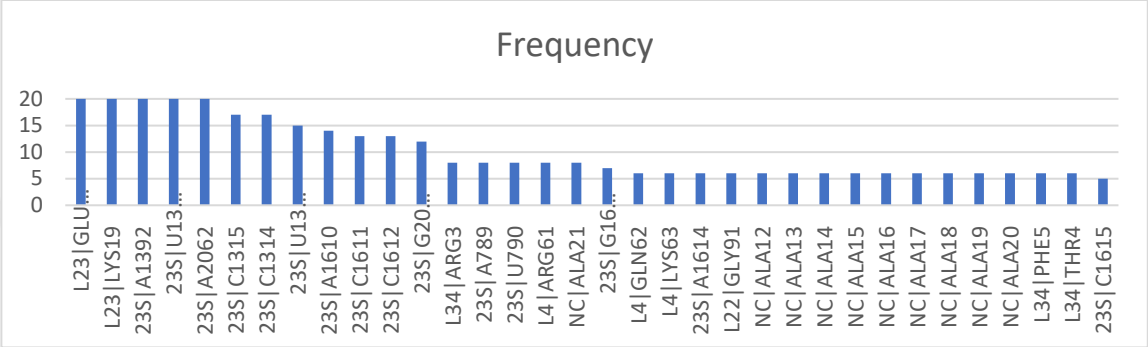


Figure 4.22. The frequency of occurrence of all nodes.

In addition to ribosomal proteins L4, L22, and L23, L34 is also visited with lower frequencies. GLY91, LYS98, ARG99 from L22 (Figure 4.23) are located on the shortest pathways. GLY91 is very frequently visited but LYS98 and ARG99 have low frequency, which are located on two pathways among 20 pathways.

L22 is a core protein. It consists of a one domain containing three α helices packed against a three-stranded antiparallel β sheet forming a well-packed hydrophobic core. Two strands of the β structure create a β hairpin [43]. The mutant beta hairpin is twisted into the ribosomal tunnel that change the shape of narrowest part and influenced the interaction between L22 and 23S rRNA. When the three aminoacids, MET82-LYS83-ARG84 (E.coli numbering), are deleted from the L22 β-hairpin, erythromycin resistance occurs [44]. Erythromycin, a macrolide antibiotics, inhibits the elongation of the nascent chain by assisting dissociation of the peptidyl tRNA from the ribosome. GLY91, which is strictly conserved, is important for the turn of the β hairpin. GLY91 is specifically important for elongation prevention and it eases the SecM stalling effect [36]. The L22 region ILE85-ARG99 is positively charged, Therefore, it can help to stabilize the RNA architecture with coulomb

effects. A deletion of ILE85-ARG99 region in the L22 results in structural rearrangement of the nucleotides, resulting in erythromycin binding affinity decrease [45].

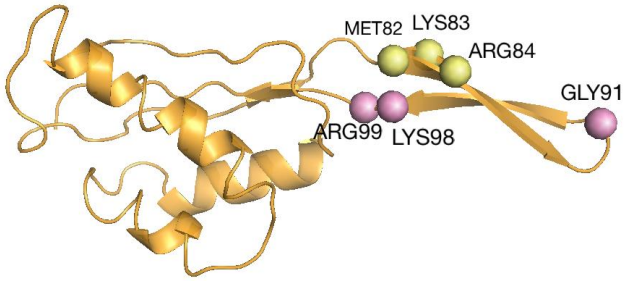


Figure 4.23. Beta hairpin of L22

TRP60, ARG61, GLN62, LYS63, GLY64 and ARG69 from L4 are located on the shortest pathways. LYS63 and GLY64 are important for changing conformation of the tail of L4 [36]. Mutation studies on LYS63 and GLY64 of L4 and MET82, LYS83, ARG84 of L22 proteins not only prevent tunnel constriction but also impair the erythromycin binding pocket [46]. Erythromycin and other macrolides bind to the near the PTC, do not obstruct peptide bond formation by themselves, but prevent entrance of the nascent peptide chain into the exit tunnel. The binding affinity of erythromycin to the L4 protein is reduced by mutation.

ARG3, THR4, PHE5 from L34 are located on the shortest pathways. All of them have high frequency. While ARG3 is located on 8 shortest pathways, others located are visited in 6 other shortest pathways. Besides GLU18, LYS19 from L23 is also visited by the shortest pathways.

Rest of the nodes forming the pathways belong to 23S RNA. Here, C2063 and G2251 are functionally significant. A2450-C2063 is a highly conserved non-Watson-Crick base pair and affects flexibility of the ribosomal tunnel [33]. On the other hand, G2251 and U2585 are highly conserved nucleotides. They interact the CCA acceptor end of tRNA with the 50S subunit P site. Mutations at U2585 decline in the peptidyl transferase activity [47].

5. CONCLUSION and RECOMMENDATION

Allosteric communication pathways between DC-SRL, DC-PTC and PTC-Tunnel, are investigated based on the crystal structure (PDB ID: 4V9M) and ClustENM conformers generated from 4V9M. Also, pathways between SecM-PTC and TF-Tunnel are investigated based on the crystal structure (PDB ID: 4V5H). In general, the costs of shortest pathways on conformers were lower than the cost for the crystal structure. This indicates the utility and relevance of ClustENM algorithm in generating plausible conformers for the study of allosteric pathways in supramolecular.

On the allosteric pathways between DC-SRL, EF-G stands out with critical sites on its domains IV and V, which correspond to quite conserved regions. Domain IV has a significant function in blocking back translocation of tRNA. Domain IV of EF-G interacts with bridge B2a and stabilize this location of B2a. This situation helps ribosome's ratchet rotation. Most frequent residues in pathways, MET580 and A1493, have important function in GTP hydrolysis on EF-G.

On the DC-PTC pathways, G2553 and U2506 stand out as important conserved residues, where the latter appears at a drug binding site. Shortest pathways are quite like each other between DC-PTC. But one unique path is appeared when the similarity analysis is done. All of paths comprised of nucleotide but three aminoacids (L27|HIE3 – L27|ALA2 – L16|ARG82) are located in this unique path. These nodes are not being in other 2000 paths. This can be alternative communication pathway according to rest of them.

A2450-C2063 appears as a highly conserved non-Watson-Crick base pair on the PTC-ribosomal tunnel pathway [32] that affects the flexibility of the tunnel [33]. Nucleotides C2452 – U2504 – A2503 – G2061 are significant since they help to create binding pocket for antibiotics or inhibitors [44]. In future study, docking ligands to this area could be studied.

A76 from the tRNA and the alanines from the nascent chain are appeared on the pathways we calculated for SecM-PTC. We think that nascent chain and tRNA facilitate transferring signals since there is a long distance between U2585 and A2451. Because A76 and U2585 interact strongly with ARG163 of SecM and because U2585 and ARG163 cause a steric hindrance, these nucleotides and their interactions with ARG163 block the PTC activation. This phenomenon prevents the entrance of the incoming aminoacids to the A-site.

The aminoacid GLY91 from L22 is located on the shortest pathways on the TF-Ribosomal Tunnel and is strictly conserved. It also has a high frequency of occurrence on all pathways. GLY91 is significant for elongation arrest and for the turn of the β-hairpin of L22 which is important since antibiotic resistance appears when a mutation on the β-hairpin occurs. We also found that LYS63 and GLY64 from L4 appear on the shortest pathways we calculated. These aminoacids change conformation of the tail of L4, and the tunnel construction is blocked when these aminoacids mutate.

REFERENCES

1. Gunasekaran, K., Ma, B., & Nussinov, R., "Is allostery an intrinsic property of all dynamic proteins?," *Proteins: Structure, Function and Genetics,* vol. 57, pp. 433-443, 2004.

2. Lu, S., Li, S., & Zhang, J., "Harnessing Allostery: A Novel Approach to Drug Discovery," *Medicinal Research Reviews,* pp. 1242-1285, 2014.

3. Bohr KCH, Krogh A., "Concerning a Biologically Important Relationship," 2014, p. 402–412.

4. J. a. J. F. Monod, "General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation," *Molecular Biology,* pp. 289-401, 1978.

5. A. a. D. D. T. F. Cooper, "Allostery without conformational change - A plausible model," *European Biophysics Journal,* pp. 103-109, 1984.

6. Kumar, S., Ma, B., Tsai, C. J., Wolfson, H., & Nussinov, R., "Folding funnels and conformational transitions via hinge-bending motions," *Cell Biochemistry and Biophysics,* pp. 141-164, 1999.

7. Feig, M., Karanicolas, J., & Brooks, C. L., "MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for applications in structural biology," *Journal of Molecular Graphics and Modelling,* vol. 22, no. 5, pp. 377-395, 2004.

8. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Tasumi, M., "The protein data bank: A computer-based archival file for macromolecular structures," *Archives of Biochemistry and Biophysics,* vol. 185, no. 2, pp. 584-591, 1978.

9. Munro, J. B., Sanbonmatsu, K. Y., Spahn, C. M. T., & Blanchard, S. C., "Navigating the ribosome's metastable energy landscape," *Biochemical Sciences,* vol. 34, no. 8, pp. 390-400, 2009.

10. Zhou, J., Lancaster, L., Donohue, J. P., & Noller, H. F., "Crystal structures of EF-G - Ribosome complexes trapped in intermediate states of translocation," *Science,* vol. 340, no. 6140, 2013.

11. Tama, F., Valle, M., Frankt, J., & Brooks, C. L., "Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy," *The National Academy of Sciences,* vol. 100, no. 16, pp. 9319-9323, 2003.

12. Wang, Y., Rader, A. J., Bahar, I., & Jernigan, R. L., "Global ribosome motions revealed with elastic network model," *Journal of Structural Biology,* vol. 147, no. 3, pp. 302-314, 2004.

13. Yamamoto, H., Qin, Y., Achenbach, J., Li, C., Kijek, J., Spahn, C. M. T., & Nierhaus, K. H., "EF-G and EF4: Translocation and back-translocation on the bacterial ribosome," *Nature,* pp. 89-100, 2014.

14. Brilot, A. F., Korostelev, A. A., Ermolenko, D. N., & Grigorieff, N. , "Structure of the ribosome with elongation factor G trapped in the pretranslocation state," *The National Academy of Sciences,* 2013.

15. Kurkcuoglu, Z., Bahar, I., & Doruker, P., "ClustENM: ENM-Based Sampling of Essential Conformational Space at Full Atomic Resolution," *Journal of Chemical Theory and Computation,* vol. 12, no. 9, pp. 4549-4562, 2016.

16. Guzel, P., & Kurkcuoglu, O., "Identification of potential allosteric communication pathways between functional sites of the bacterial ribosome by graph and elastic network models," *Biochimica et Biophysica Acta - General Subjects,* vol. 1861, no. 12, 2017.

17. Chennubhotla, C., & Bahar, I. , "Markov propagation of allosteric effects in biomolecular systems: Application to GroEL-GroES," *Molecular Systems Biology,* 2006.

18. J. Y. Yen, "Finding the K Shortest Loopless Paths in a Network," *Management Science,* vol. 17, no. 11, pp. 712-716, 1971.

19. E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik,* vol. 1, pp. 269-271, 1959.

20. Shi, X., Khade, P. K., Sanbonmatsu, K. Y., & Joseph, S., "Functional role of the sarcin-ricin loop of the 23s rRNA in the elongation cycle of protein synthesis," *Journal of Molecular Biology,* vol. 419, no. 3, pp. 125-138, 2012.

21. Rakauskaite, R., & Dinman, J. D., "rRNA mutants in the yeast peptidyltransferase center reveal allosteric information networks and mechanisms of drug resistance," *Nucleic Acids Research,* vol. 36, no. 5, pp. 1497-1507, 2008.

22. Fulle, S., and Gohlke, H., "Statics of the Ribosomal Exit Tunnel: Implications for Cotranslational Peptide Folding, Elongation Regulation, and Antibiotics Binding," *Journal of Molecular Biology,* vol. 387, pp. 502-517, 2009.

23. Gumbart, J., Schreiner, E., Wilson, D. N., Beckmann, R., & Schulten, K. , "Mechanisms of SecM-mediated stalling in the ribosome," *Biophysical Journal,* 2012.

24. Lin, K. F., Sun, C. S., Huang, Y. C., Chan, S. I., Koubek, J., Wu, T. H., & Huang, J. J. T., "Cotranslational protein folding within the ribosome tunnel influences trigger-factor recruitment," *Biophysical Journal,* vol. 102, 2012.

25. AEvarsson, A., Brazhnikov, E., Garber, M., Zheltonosova, J., Chirgadze, Y., al-Karadaghi, S., … Liljas, A., "Three-dimensional structure of the ribosomal translocase: elongation factor G from Thermus thermophilus," *The EMBO Journal,* vol. 13, pp. 3669-3677, 1994.

26. Orengo, C. A., & Thornton, J. M., "Alpha plus beta folds revisited: some favoured motifs," *Structure,* pp. 105-120, 1993.

27. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. , "ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules," *Nucleic Acids Research,* 2016.

28. Chen, Y., Feng, S., Kumar, V., Ero, R., & Gao, Y. G. , "Structure of EF-G-ribosome complex in a pretranslocation state," *Nature Structural and Molecular Biology,* 2013.

29. Brunelle, J. L., Youngman, E. M., Sharma, D., & Green, R., "The interaction between C75 of tRNA and the A loop of the ribosome stimulates peptidyl transferase activity," *RNA,* vol. 2, no. 1, pp. 33-39, 2006.

30. Martin Schmeing, T., Huang, K. S., Strobel, S. A., & Steitz, T. A., "An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA," *Nature,* vol. 438, pp. 520-524, 2005.

31. Long, K. S., Hansen, L. H., Jakobsen, L., & Vester, B., "Interaction of pleuromutilin derivatives with the ribosomal peptidyl transferase center," *Antimicrobial Agents and Chemotherapy,* vol. 50, no. 4, pp. 1458-1462, 2006.

32. Bayfield, M. A., Thompson, J., & Dahlberg, A. E., "The A2453-C2499 wobble base pair in Escherichia coli 23S ribosomal RNA is responsible for pH sensitivity of the peptidyltransferase active site conformation," *Nucleic Acids Research,* vol. 32, pp. 5512-5518, 2004.

33. Chirkova, A., Erlacher, M. D., Clementi, N., Zywicki, M., Aigner, M., & Polacek, N., "The role of the universally conserved A2450-C2063 base pair in the ribosomal peptidyl transferase center," *Nucleic Acids Research,* vol. 38, no. 14, pp. 4844-4855, 2010.

34. Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., … Yonath, A., "High resolution structure of the large ribosomal subunit from a mesophilic eubacterium," *Cell,* vol. 107, pp. 679-688, 2001.

35. Auerbach, T., Mermershtain, I., Davidovich, C., Bashan, A., Belousoff, M., Wekselman, I., … Yonath, A., "The structure of ribosome-lankacidin complex reveals ribosomal sites for synergistic antibiotics," *Proceedings of the National Academy of Science America,* December 15 2010.

36. Nakatogawa, H., & Ito, K. , "The ribosomal exit tunnel functions as a discriminating gate," *Cell,* 2002.

37. Zhang, J., Pan, X., Yan, K., Sun, S., Gao, N., & Sui, S. F., "Mechanisms of ribosome stalling by SecM at multiple elongation steps," *ELife,* 2015.

38. Bhushan, S., Hoffmann, T., Seidelt, B., Frauenfeld, J., Mielke, T., Berninghausen, O., … Beckmann, R. , "SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center," *PLoS Biology,* 2011.

39. Fedorov, A. N., Friguet, B., Djavadi-Ohaniance, L., Alakhov, Y. B., & Goldberg, M. E., "Folding on the ribosome of Escherichia coli tryptophan synthase β subunit nascent chains probed with a conformation-dependent monoclonal antibody," *Journal of Molecular Biology,* pp. 351-358, 1992.

40. Cabrita, L. D., Hsu, S. T. D., Launay, H., Dobson, C. M., & Christodoulou, J., "Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy," *Proceedings of the National Academy of Sciences of the United States of America,* pp. 22239-22244, 2009.

41. Hoffmann, A., Bukau, B., & Kramer, G., "Structure and function of the molecular chaperone Trigger Factor," *Biochimica et Biophysica Acta - Molecular Cell Research,* vol.

1803, pp. 650-661, 2010.

42. Kramer, G., Rauch, T., Rist, W., Vorderwülbecke, S., Palzelt, H., Schulze-Specking, A., … Bukau, B., "L23 protein functions as a chaperone docking site on the ribosome," *Nature,* vol. 419, 2002.
43. Ban, N., Nissen, P., Hansen, J., Moore, P. B., & Steitz, T. A., "The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution," *Science,* vol. 289, pp. 905-920, 2000.
44. Davydova, N., Streltsov, V., Wilce, M., Liljas, A., & Garber, M., "L22 ribosomal protein and effect of its mutation on ribosome resistance to erythromycin," *Journal of Molecular Biology,* vol. 322, pp. 635-644, 2002.
45. Berisio, R., Corti, N., Pfister, P., Yonath, A., & Böttger, E. C., "23S rRNA 2058A→G alteration mediates ketolide resistance in combination with deletion in L22," *Antimicrobial Agents and Chemotherapy,* pp. 3816-3823, 2006.
46. Lovmar, M., Nilsson, K., Lukk, E., Vimberg, V., Tenson, T., & Ehrenberg, M. , "Erythromycin resistance by L4/L22 mutations and resistance masking by drug efflux pump deficiency," *EMBO Journal,* 2009.
47. Green, R., Samaha, R. R., & Noller, H. F. , "Mutations at nucleotides G2251 and U2585 of 23 S rRNA perturb the peptidyl transferase center of the ribosome," *Journal of Molecular Biology,* 1997.

APPENDIX A: PYTHON CODE FOR SIMILARITY AND CLUSTERING CALCULATION

```
import numpy as np

data = np.genfromtxt("path_DC_PTC.txt")
n=data.shape[0]
S=np.zeros((n,n))

for i in range(n):
        for j in range(i+1):
                dot=np.dot(data[i],data[j])
                normi = np.linalg.norm(data[i])
                normj = np.linalg.norm(data[j])
                S[i,j]=dot / (normi * normj)
                S[j,i]=S[i,j]

import matplotlib.pyplot as plt
import seaborn as sns; sns.set()

plt.scatter(S[:, 0], S[:, 1], s=50);

from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=2)
kmeans.fit(S)
y_kmeans = kmeans.predict(S)

plt.scatter(S[:, 0], S[:, 1], c=y_kmeans, s=50, cmap='viridis')

centers = kmeans.cluster_centers_
plt.scatter(centers[:, 0], centers[:, 1], c='black', s=200, alpha=0.5)

plt.show()
```