Computational Biology and Bioinformatics
2025, Vol. 13, No. 1, pp. 22-41
https://doi.org/10.11648/j.cbb.20251301.13

Review Article

Comparative Study of the Twin Arginine Translocase (Tat) System Across Bacterial Species: Insights into Hydrophobic Interactions, Signal Peptide Binding and Protein Translocation Dynamics

Micael Sousa Correia, Sharon Mendel Williams*

College of Engineering, Environment and Science, School of Science, Coventry University, Coventry, UK

Abstract

This study examines the Twin-Arginine Translocase (Tat) system, especially the TatC subunit's role and variations between Gram-positive and Gram-negative bacteria. It investigates how hydrophobicity affects the Tat pathway, particularly in the interaction of the *Escherichia coli* (*E. coli*) TatC subunit and *Bacillus substilis* (*B. subtilis*) with SufI and TorA signal peptides. Different bioinformatics tools were used in the following research such as NCBI, Clustal Omega, MAFFT for sequence alignment, Phyre2 for structural modelling, and PyMOL, HDOCK, POCASA, KVFinder for protein docking and hydrophobicity analysis. The study provides an in-depth examination of TatC's structure, evolutionary relationships, and interactions with signal peptides. This approach uncovers the crucial balance between hydrophobic and hydrophilic forces in the Tat pathway, challenging the traditional emphasis on the twin-arginine motif in the SufI and TorA signal peptide. The analysis reveals the binding affinities and the pivotal role of the regions of the signal peptide interactions within TatC subunit in particular from Gram-negative *E. coli* and Gram-positive *B. subtilis*, enriching comprehension of the system's flexibility and the fundamental influence of hydrophobicity in protein interactions. The current study also demonstrates that peptides can bind effectively without twin-arginine motifs and suggests a deeper embedding of signal peptides in TatC's hydrophobic zones.

Keywords

Twin Arginine Translocase, Protein Translocation Dynamics, Bioinformatics, Hydrophobic Interactions

1. Introduction

The Twin-Arginine Translocase (Tat) system is a protein export pathway crucial to various microbial processes, including respiration, cell wall maintenance, and pathogenesis. This system is present in all prokaryotes, as well as in chloroplasts in plants. [35, 60]. Unlike the general secretory (Sec) pathway, which exports unfolded proteins, the Tat pathway uniquely exports fully folded proteins, including multi-subunit complexes [44]. While the Sec system can compensate for some metabolic processes, the Tat system is indispensable for bacterial virulence and antibiotic resistance [4, 28, 46].

*Corresponding author: ab6263@coventry.ac.uk (Sharon Mendel Williams)

Received: 31 May 2025; **Accepted:** 17 June 2025; **Published:** 22 July 2025

1.1. Structural and Functional Aspects of the TatC Subunit

The Tat translocase complex typically comprises TatA, TatB, and TatC subunits [44]. However, its composition varies between Gram-positive and Gram-negative bacteria, leading to distinct functional mechanisms. In Gram-negative bacteria like *Escherichia coli*, the Tat system includes all three subunits: TatA, TatB, and TatC, with TatE acting as a supportive paralog to TatA [61]. In Gram-positive species such as *Bacillus subtilis*, the system lacks TatB, instead forming a TatAC complex, and some strains contain two TatC homologs (TatCd and TatCy) specialized for different substrates [40]. These structural variations influence the substrate specificity, complex assembly, and regulation of protein translocation, especially the role and positioning of TatC.

1.2. Tat Signal Peptide Recognition and Binding

Tat-dependent precursor proteins are recognised by N-terminal signal peptides, which initiate translocation through the TatBC complex. These peptides consist of conserved regions, including the positively charged n-region, hydrophobic h-region, and polar c-region [43, 58].

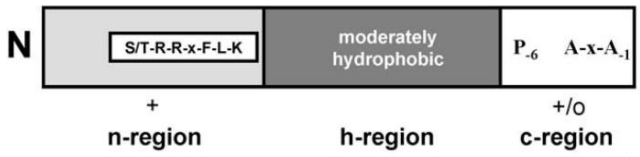

Figure 1. *Signal Peptide SufI regions: Tat Signal Sequence, image provided by Bronstein et al., 2004 [3], presents the n-region with the S/T-RR-x-F-L-K motif, followed by the moderately hydrophobic h-region, and the c-region containing the A-x-A amino acid sequence.*

These signal peptides initiate contact with the TatBC receptor complex. Specifically, the n-region binds between the TM2 and TM3 helices of TatC [1, 2, 14], while the h-region interacts with TatB [66], triggering the recruitment of TatA [57, 58].

1.3. Mechanism of Tat-dependent Protein Translocation

Once the TatBC complex binds a folded substrate's signal peptide, the TatA subunits, initially not associated with TatBC, oligomerize into dynamic, variable-sized pores in the membrane [15]. These pores facilitate substrate passage while preserving membrane integrity by minimizing ion leakage.

Tat transport is powered by the proton motive force (PMF), involving both the membrane potential (ΔΨ) and proton gradient (ΔpH), and does not require ATP—distinguishing it from the Sec pathway [17, 35]. After translocation, a signal peptidase cleaves the signal sequence at the Ala-x-Ala motif, releasing the mature protein into the periplasmic space [43, 16]. Dynamic studies using cross-linking and fluorescence tagging have confirmed that TatA associates transiently with TatBC, forming large assemblies at the translocation site during the transport cycle [39].

1.4. Tat Signal Peptide Binding Sites and Substrate Specificity

The interaction between Tat signal peptides and the TatBC complex has been extensively studied using model substrates such as SufI, a protein involved in septal ring formation during bacterial cell division [10]. Early structural analyses showed that substrates like SufI and CueO bind to the periphery of the TatC subunit, suggesting that the TatBC cavity is too narrow to accommodate fully folded proteins internally [55, 9].

However, subsequent studies using modeling and mutagenesis identified a potential internal signal peptide binding site near the TM1 and TM3 helices of TatC. Substitution of conserved residues (e.g., Glu9 and Glu96) in these regions significantly reduced binding affinity for the signal peptide's twin-arginine motif, while changes to peripheral residues had little effect [47, 5]. These findings suggest a specific internal recognition site that facilitates high-affinity substrate binding.

2. Aim and Hypothesis

The investigation aims to conduct a comparative analysis of bacterial taxa, elucidating their structural relationships and distinctions. Moreover, the research seeks to delve into the significance of hydrophobicity within the twin-arginine translocation (Tat) pathway. It specifically focuses on regions of the TatC in modulating the association with signal peptides, thereby influencing the efficacy of protein translocation. The intent is to enhance our understanding of TatC recognition specificity, highlighting the adaptability of the Tat system in accommodating variations in signal peptide hydrophobicity. Additionally, this study examines hydrophobicity's role in the Tat pathway, with a specific emphasis on docking analyses. These analyses aim to ascertain the impact of hydrophobic and hydrophilic segments of *E. coli str. K-12* TatC and *B. subtilis* on its binding dynamics and interactions with SufI and TorA signal peptides.

3. Methods

3.1. Protein Blast

Subsequent analysis entailed submitting protein sequences from *E. coli str. K-12* and *B. subtilis* into the Protein BLAST tool. The objective was to obtain four additional sequences for each type of bacteria, emphasising varying degrees of identity

similarities. While default settings were primarily maintained, modifications were made to the exclusion criteria to prevent the duplication of the primary bacterial strains. The search was expanded to encompass up to 5000 target sequences to ensure a comprehensive range of results. The analysis of *E. coli* sequences yielded selections from *Pseudomonas aeruginosa, Vibrio cholerae, Salmonella enterica and Klebsiella pneumoniae*. Conversely, examining *B. subtilis* sequences identified relevant sequences from *B. cereus, Paenibacillus, Streptococcus thermophilus, and Streptococcus pneumoniae*, representing the gram-positive bacterial spectrum [63, 25, 45].

3.2. Multiple Sequence Alignment

3.2.1. Clustal Omega

The complete TatC FASTA sequences of each bacterium were processed through the Clustal Omega multiple sequence alignment (MSA) tool to identify amino acid similarities and regions of conservation [53]. The results were colour-coded to highlight the areas of high and low conservation, including substitutions and deletions [37].

3.2.2. Phylogenetic Tree

To construct and examine the phylogenetic trees of the two bacterial groups, this study utilised the MAFFT web-based application for multiple sequence alignment tasks. The advanced algorithms of MAFFT facilitated the alignment of sequences, allowing for an in-depth analysis of evolutionary relationships among the bacterial taxa [29, 30, 33]. MAFFT attributes evolutionary significance and phylogenetic relationships to the sequences by employing sophisticated alignment techniques. The relationships are further clarified through bootstrap values, which aid in discerning the proximity or divergence among the taxa. The alignment process commenced with sequences from Gram-Negative bacteria, followed by those from Gram-Positive bacteria, and ultimately, all ten bacterial sequences were aligned together. While most parameters were kept at their default settings, the bootstrap values were specifically activated to understand phylogenetic relationships better and visualised in Archaeopteryx.js in-built MAFFT function.

3.2.3. Phyre2

Phyre2 plays a crucial role in elucidating protein conformations. This web-based tool processes amino acid sequences to generate a protein's secondary structure model. It achieves this by drawing on databases of similar proteins and employing sophisticated computational algorithms. The resulting model includes a detailed breakdown of the protein's secondary structure components, quantifying the proportion of alpha helices and beta strands. This information is invaluable for understanding the protein's three-dimensional shape and functional implications. Adding to its utility, Phyre2 also aids in predicting protein function and potential interactions based on the modelled structure, offering a comprehensive view of the protein's characteristics [32].

3.3. Protein Docking

3.3.1. PyMOL

Before docking simulations, PyMOL was used to identify, visualise, and manipulate the 3D structures of the TatC protein and signal peptide. Specific chain identifiers were assigned to distinguish the receptor (TatC, chain 'A') from the ligand (signal peptide, chain 'B') using the ALTER command [52]. This ensured correct recognition during docking.

3.3.2. Huang Docking Web Server (HDOCK)

Following the previous modifications, the prepared structure of *E. coli str. K-12* was introduced into the docking web server, HDOCK simulation, for potential binding analyses. The simulation parameters were kept at default settings except for the specified chain identifiers 'A' for the receptor and 'B' for the ligand [64, 24].

3.3.3. Pocket-cavity Search Application (POCASA)

POCASA utilises a computational methodology that employs algorithms for protein-protein interaction prediction and surface probe recognition techniques. It involves a virtual probe that traverses the surface of a given 3D protein structure, scanning for cavities that could signify potential binding sites. These predicted binding regions are characterised by their hydrogen atom composition and are quantified by their volume and depth, with results provided in units of angstroms (Å) [65].

3.3.4. KVFinder

KVFinder was employed to analyse the hydrophobicity of binding pockets in TatC. It facilitated the identification of hydrophobic and hydrophilic regions critical for interpreting binding interactions and docking results [18-20, 49].

4. Results

Sequence identity analyses were conducted with BLAST to compare selected Gram-negative and Gram-positive bacteria against reference proteins from *E. coli str. K-12* and *B. subtilis*.

Table 1. *Gram-Negative Bacteria Sequence identity comparison to E. coli str. K-12 TatC Protein.*

Gram-Negative Bacteria	Classification	Accession	Amino Acids	Identity (%)
Pseudomonas aeruginosa	Opportunistic Pathogen (Diggle & Whiteley, 2019)	PWU38561.1	267	57.60
Vibrio cholerae	Aquatic Pathogen (Halpern & Izhaki, 2017)	EGQ9206430.1	250	66.40
Salmonella enterica	Specialised invasive Pathogen (Hong et al., 2023)	EJJ4019649.1	259	91.12
Klebsiella pneumoniae	Respiratory Pathogen (Mannion et al., 2023)	CDO16241.1	259	83.72

The table presents a selection of Gram-negative bacteria from the NCBI database, detailing their classification, accession numbers, amino acid sequences, and percentage identity to *E. coli str. K-12* [41].

Table 1 presents the sequence identity percentages of TatC proteins in Gram-Negative bacteria relative to the *E. coli str. K-12* TatC, indicating a spectrum of homology. Notably, *Salmonella enterica* exhibited the most significant homology, with a 91.12% identity, whereas *Pseudomonas aeruginosa (P. aeruginosa)* demonstrated the least, with a 57.60% identity [11]. This range reflects both conservation and variation among the Gram-Negative bacterial TatC homologs compared to *E. coli str. K-12*.

Table 2. *Gram-Positive Bacteria sequence identity comparison to B. subtilis TatCd and TatCy Protein.*

Gram-Positive Bacteria	Classification	Accession	Amino Acids	Identity with TatCd (%)	Identity with TatCy (%)
Bacillus cereus	Opportunistic Pathogen (Zheng et al., 2024)	HDR4908830	248	60.34	43.90
Paenibacillus sp. Tmac-D7	Aquatic Pathogen (S áez-Nieto et al., 2017)	WP_141336258.1	241	63.07	46.72
Streptococcus thermophilus	Specialised non-pathogen (Xu et al., 2023)	WP_084825977.1	242	41.95	36.21
Streptococcus pneumoniae	Respiratory Pathogen (Peng et al., 2023)	WP_050251502.1	243	36.71	35.27

The table presents a selection of Gram-positive bacteria from the NCBI database, detailing their classification, accession numbers, amino acid (AA) sequences, and percentage identity to *E. coli str. K-12* (NCBI, 2024).

In contrast to Table 1, Table 2 delineates the sequence identity percentages for Gram-positive bacteria to the TatCd and TatCy proteins of *B. subtilis*. *B. cereus* is highlighted, showing a moderate sequence identity of 60.34% to TatCd and 43.90% to TatCy, illustrating the variances in sequence conservation across different species. Both Tables also present their amino acid lengths, in which Gram-negative bacteria presented a higher amino acid count than Gram-positive bacteria.

Table 3 compares the sequence identities across Gram-positive and Gram-negative groups, categorised according to their pathogenic classifications. Within this framework, *E. coli str. K-12* and *B. subtilis*, categorised as non-pathogenic bacteria [17, 40], exhibited higher sequence identities, with 32.93% to TatCd and 31.35% to TatCy. Conversely, specialised bacteria showed a lower similarity of 24.60%.

Table 3. *Sequence Identity Comparison Between Both Groups.*

Gram-negative Bacteria	Gram-positive Bacteria	Classification	Identity similarity (%)
Pseudomonas aeruginosa	*Bacillus cereus*	Opportunistic Pathogen	28
Vibrio cholerae	*Paenibacillus sp. Tmac-D7*	Aquatic Pathogen	28.57
Salmonella enterica	*Streptococcus thermophilus*	Specialised Pathogen	24.60
Klebsiella pneumoniae	*Streptococcus pneumoniae*	Respiratory Pathogen	27.20

Gram-negative Bacteria	Gram-positive Bacteria	Classification	Identity similarity (%)	
Escherichia coli str. K-12	*Bacillus subtilis*	Non - pathogenic	TatCd 32.93	TatCy 31.35

This table summarizes the bacteria from each group, showing their taxonomic classifications and percentage identity similarities.

4.1. Clustal Omega

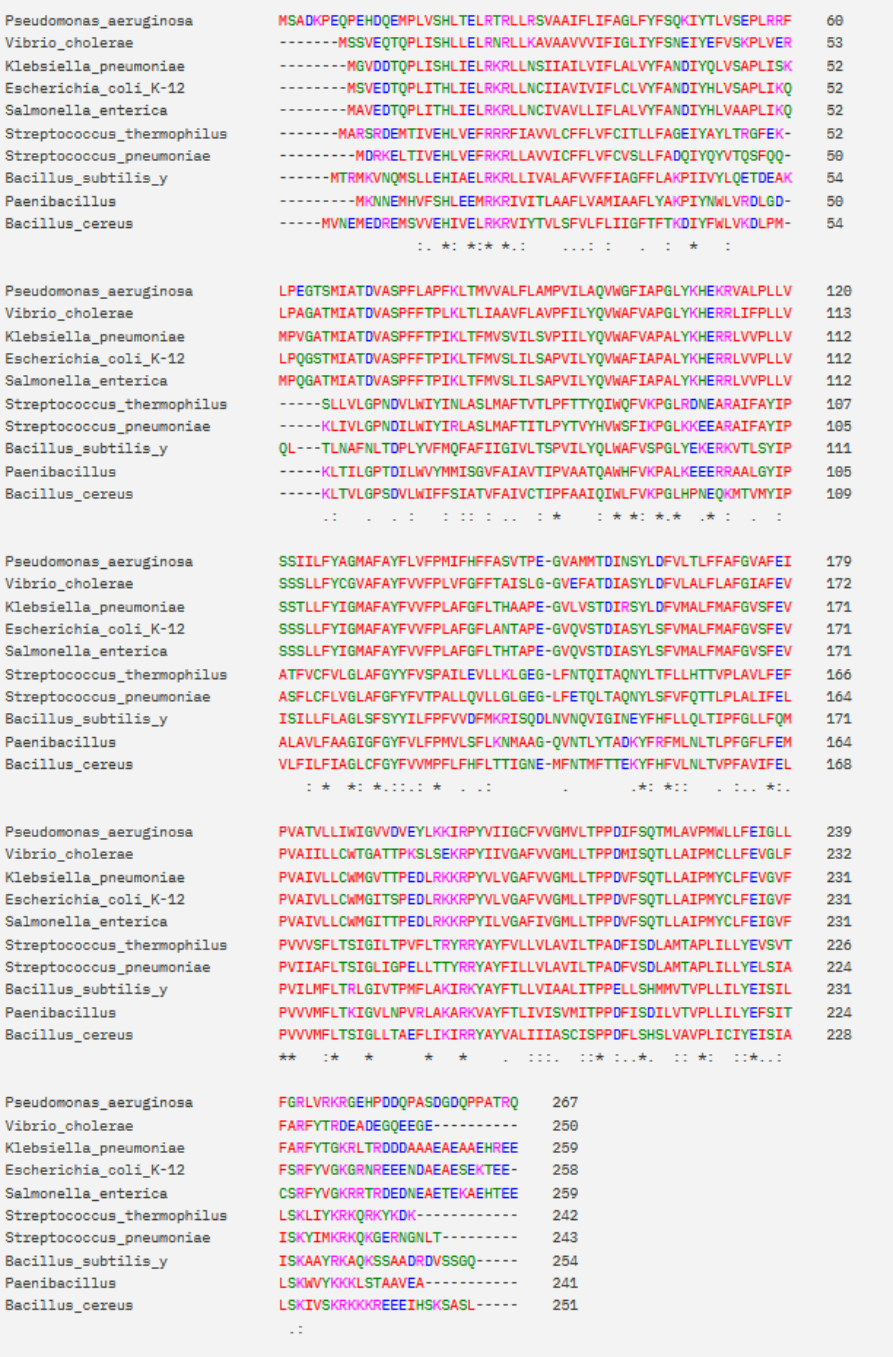

Figure 2. *Gram-negative and Gram-positive Multiple Sequence Alignment of TatC subunit.*

Investigating protein residues is vital for understanding evolutionary connections among taxa and the dynamics of protein-protein interactions [36]. These researchers highlight that these residues are crucial in recognising ligands within a defined pocket of the protein's three-dimensional architecture. This insight enhances our comprehension of protein functions and paves the way for innovative research, potentially leading to the development of novel pharmaceuticals that can promote new biological functions. Figure 2 showcase a comparison of the number of residues across different taxa, symbolised as follows: '*' indicates an exact residue match, for instance, the initial matching amino acid Histidine (H) shown in green in Figure 2; ':' signifies conserved residues, such as Aspartate (D) and Glutamate (E) seen in the same figure shown in Blue; and '.' represents semi-conserved residues like Lysine (K) pink colour, Asparagine (N), and Threonine (T) green colour, also observed in the figure. It should be noted that these tables exclude TatCd.

The figure displays the Gram-negative and Gram-positive multiple sequence alignment of the selected bacteria, colour-coded with their residue's similarities below the sequences. Constructed with Clustal Omega [53].

4.2. Phylogenetic Trees

Figures 3, 4 and 5 introduce phylogenetic trees generated via MAFFT web server from the previous multi-sequence alignment figures to elucidate the evolutionary relationships among the same gram-negative and gram-positive bacteria.

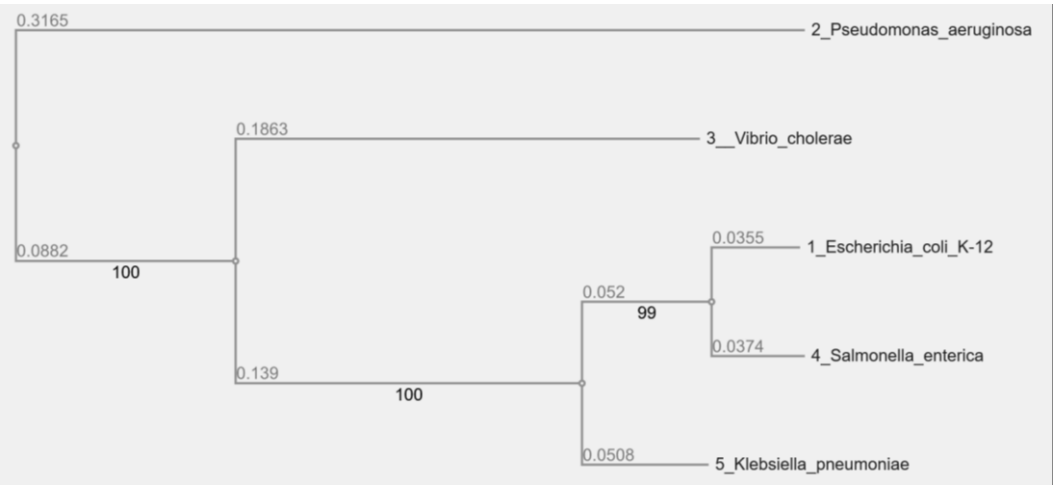

Figure 3. *Gram-negative Bacteria Phylogenetic Tree.*

This phylogenetic tree represents the evolutionary relationships among various gram-negative bacteria. Branch lengths correspond to genetic distance, and bootstrap values are indicated at the nodes to reflect the confidence level of the clade distinctions. This analysis was facilitated by the MAFFT web server for sequence alignment [30].

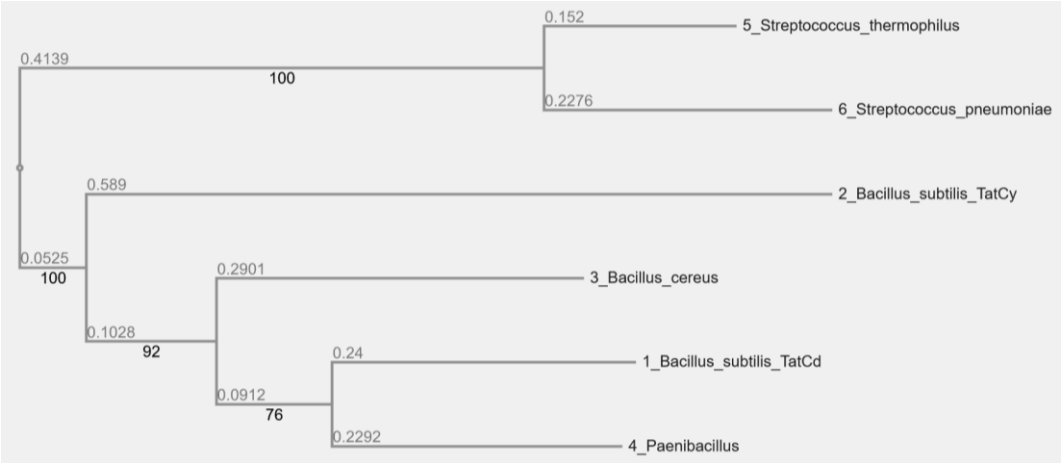

Figure 4. *Gram-positive Bacteria Phylogenetic Tree.*

This figure extends the analysis to gram-positive bacteria, with an additional focus on the TatCd protein. The tree illustrates genetic divergence with bootstrap values provided for clade validation. This analysis was facilitated by the MAFFT web server for sequence alignment [30].

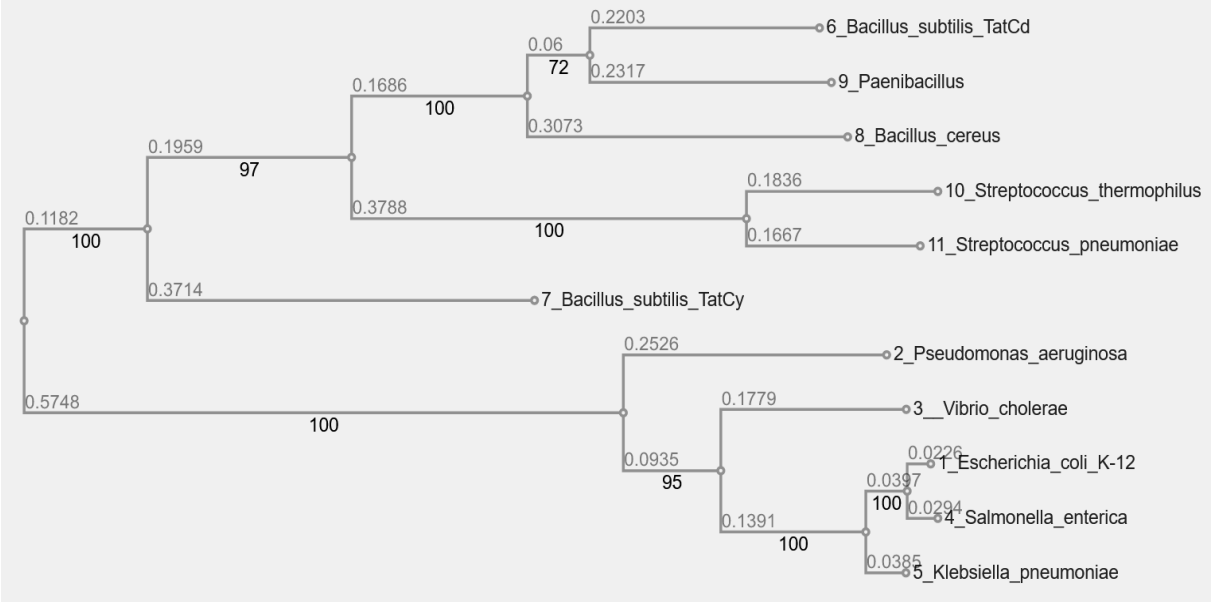

Figure 5. *Gram-negative and Gram-positive Bacteria Phylogenetic Tree.*

The combined phylogenetic tree includes both gram- negative and gram-positive bacteria, with the inclusion of the TatCd protein. This comprehensive view allows for comparison across different bacterial classes, with bootstrap values showing the statistical support for the groupings. This analysis was facilitated by the MAFFT web server for sequence alignment [30].

The derived phylogenetic trees reveal distinct clades corresponding to the two bacterial classifications. Figure 3 illustrates the relationships among gram-negative bacteria, showing clear divergence among species, as evidenced by the bootstrap values. Figure 4 focuses on gram-positive bacteria, introducing the TatCd protein into the analysis. The tree topology indicates a genetic distance between species within this group. Figure 5 provides a detailed approach, incorporating both bacterium types and the TatCd protein into a single phylogenetic framework. The bootstrap values across all trees demonstrate a high confidence level in the branching patterns observed, suggesting a robust evolutionary framework [48].

4.3. Protein-protein Docking Analysis

Protein-protein docking simulations were performed to reveal the interaction between *E. coli str. K-12* TatC and *B. subtilis* and two distinct signal peptides, SufI (Figure 6) and TorA (Figure 8). The models in these figures present a ribbon shape, which illustrates the alpha helices in a rod-like shape or loops, while the Beta strands are illustrated in ribbons or lines [50].

The docking was performed using the HDOCK web server, which provided a ranking of the top 10 models based on docking scores, ligand root-mean-square deviation (RMSD), and interface residues seen in Figures 7 and 9. The models were evaluated for their docking confidence and interface quality, with results indicating varying degrees of interaction strength and complex stability.

Table 4. *Top ten E. coli str. K-12 and SufI protein-protein top ten docking models.*

Protein Type	LGscore	MaxSub
Receptor (TatC)	2.676	0.162
Ligand (Signal Peptide)	0.180	-0.035

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-265.20	-247.19	-238.20	-236.23	-234.70	-233.82	-233.29	-229.53	-229.94	-227.67
Confidence Score	0.9092	0.8748	0.8537	0.8487	0.8447	0.8424	0.8410	0.8390	0.8319	0.8254
Ligand rmsd (Å)	29.09	49.03	46.60	28.45	46.90	34.82	28.92	28.64	46.00	28.46

The table provides an overview of the top ten protein-protein docking ranks between *E. coli str. K-12* and SufI as assessed by [23]. The leading rank exhibits the highest confidence at 90%, with a docking score of -265 and a ligand RMSD of 29.09. Conversely, the tenth rank displays the lowest assessed metrics, with a docking score of -227, a confidence level of 82%, and a ligand RMSD of 28.46.

Table 4 resulted from the docking, which provided ten ranking scores: rank 1 is the most confident score of 0.90 (90%), a docking score of -265.20, and a Ligand RMSD of 29.09 (29%). The table also provides information on the input protein quality, in which the ligand (SUFI signal peptide) presented a low-quality structure. The docking results from rank 1 were input in PyMOL web server, resulting in the construction of Figure 6.

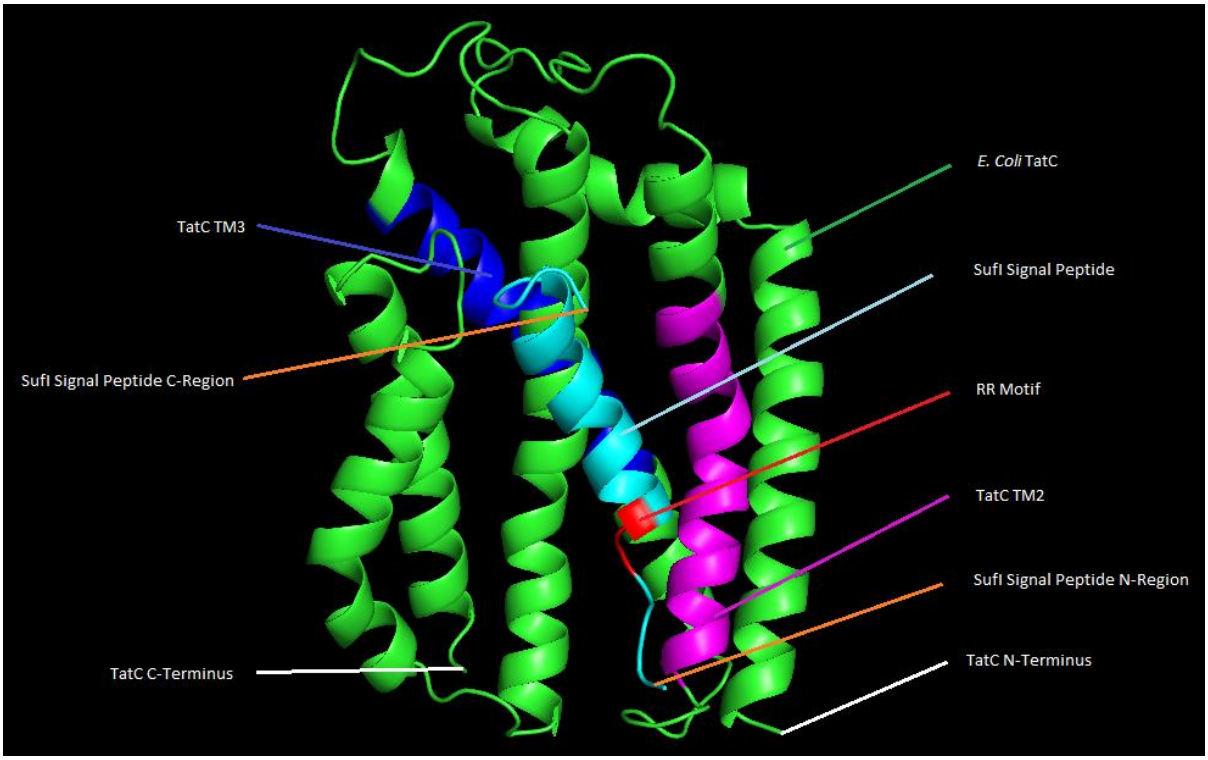

Figure 6. *3D representation of E. coli str. K-12 TatC and SufI protein-protein docking with highlighted features.*

The figure illustrates the secondary structure of the *E. coli str. K-12* TatC illustrated in green, with its transmembrane segments TM2 and TM3 in purple and dark blue, as identified by UniProt (2023) [59]. The SufI signal peptide is shown in blue with its C-region and N-region, and the RR motif highlighted in red.

TorA also underwent the same simulation as SufI, yielding slightly different results (Table 5). This table presented ten ranking scores, where rank 1 provided a docking score of -244.06, a confidence score of 0.86 (86%) and a ligand RMSD of 502 (502%). In the same way as SufI, TorA also presented a low-quality structure. These results were also used to construct its secondary structure in PyMOL.

Table 5. *Top Ten E. coli str. K-12 and TorA protein-protein docking models.*

Protein Type	LGscore	MaxSub
Receptor (TatC)	2.676	0.162
Ligand (Signal Peptide)	0.046	-0.006

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-244.0	-255.5	-224.6	-224.6	-222.7	-221.8	-220.9	-220.2	-216.5	-214.0
Confidence Score	0.867	0.812	0.816	0.816	0.810	0.807	0.805	0.802	0.790	0.782
Ligand rmsd (Å)	502.6	500.1	503.1	486.0	504.5	480.6	501.3	479.0	502.1	479.9

The Table provides an overview of the top ten protein-protein docking ranks between *E. coli str. K-12* and TorA as assessed by HDOCK (2024) [23]. The leading rank exhibits the highest confidence at 86%, with a docking score of -244 and a ligand RMSD of 502.62. Conversely, the tenth rank displays the lowest assessed metrics, with a docking score of -214, a confidence level of 78%, and a ligand RMSD of 479.92. HDOCK also reported a lower structural quality for the ligand in the docking evaluation.

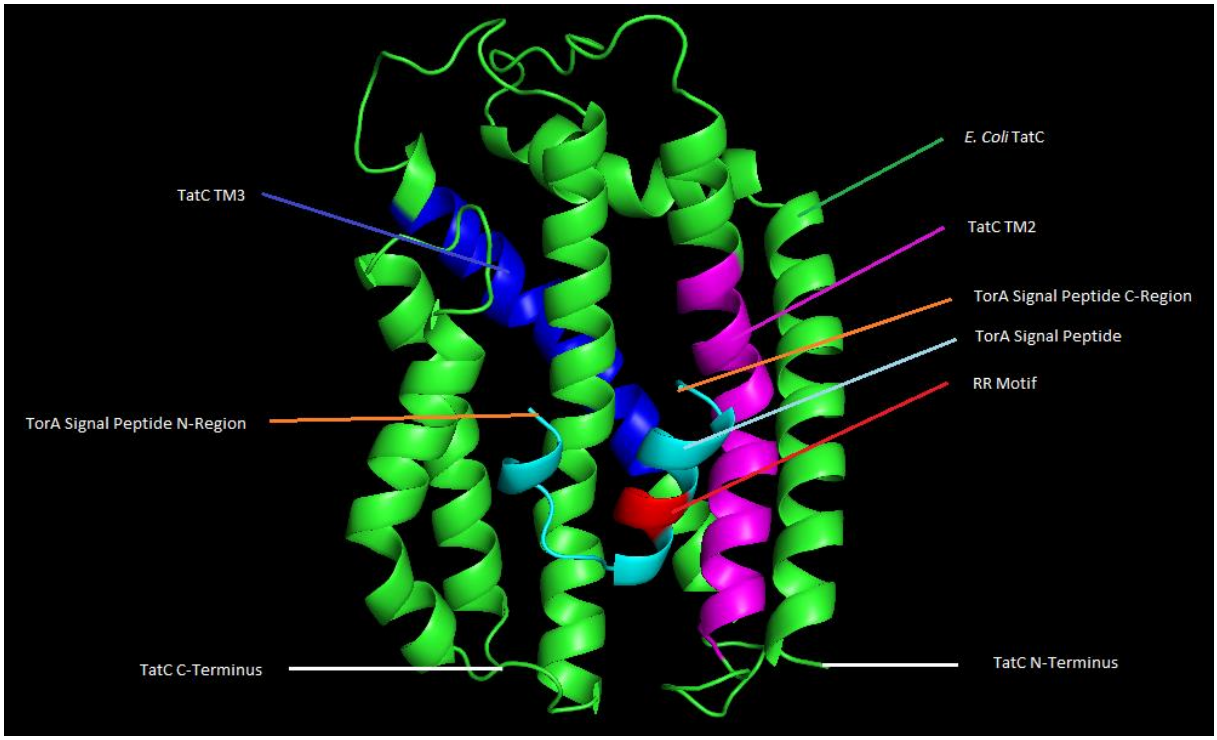

Figure 7. *3D representation of E. coli str. K-12 TatC and TorA protein-protein docking with highlighted features.*

Figure 7 illustrates the secondary structure of the *E. coli str. K-12* TatC illustrated in green highlighting its most confidently scored conformation, with its transmembrane segments TM2 and TM3 in purple and dark blue, as identified by UniProt (2024) [59]. The TorA signal peptide is shown in blue with its C-region and N-region, and the RR motif, essential for the signal peptide's function, is highlighted in red. The image provides insight into the spatial arrangement of the protein's functional domains, including the TatC C-terminus and H-terminus.

Table 6. *Top Ten B. subtilis and TorA protein-protein top ten docking models.*

Protein Type	LGscore	MaxSub
Receptor (TatC)	3.109	0.206
Ligand (Signal Peptide)	0.534	0.056

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-244.7	-227.6	-222.9	-222.2	-219.6	-218.4	-217.1	-213.6	-210.9	-210.9
Confidence Score	0.869	0.825	0.811	0.809	0.801	0.797	0.792	0.781	0.772	0.771
Ligand rmsd (Å)	488	504.4	503.6	488.0	503.9	502.9	503.7	503	502	510.6

The table provides an overview of the top ten protein-protein docking ranks between *B. subtilis* and TorA as assessed by HDOCK (2024) [23]. The leading rank exhibits the highest confidence at 86%, with a docking score of -244 and a ligand RMSD of 488. Conversely, the tenth rank displays the lowest assessed metrics, with a docking score of -210, a confidence level of 77%, and a ligand RMSD of 510. HDOCK also reported a lower structural quality for the ligand in the docking evaluation.

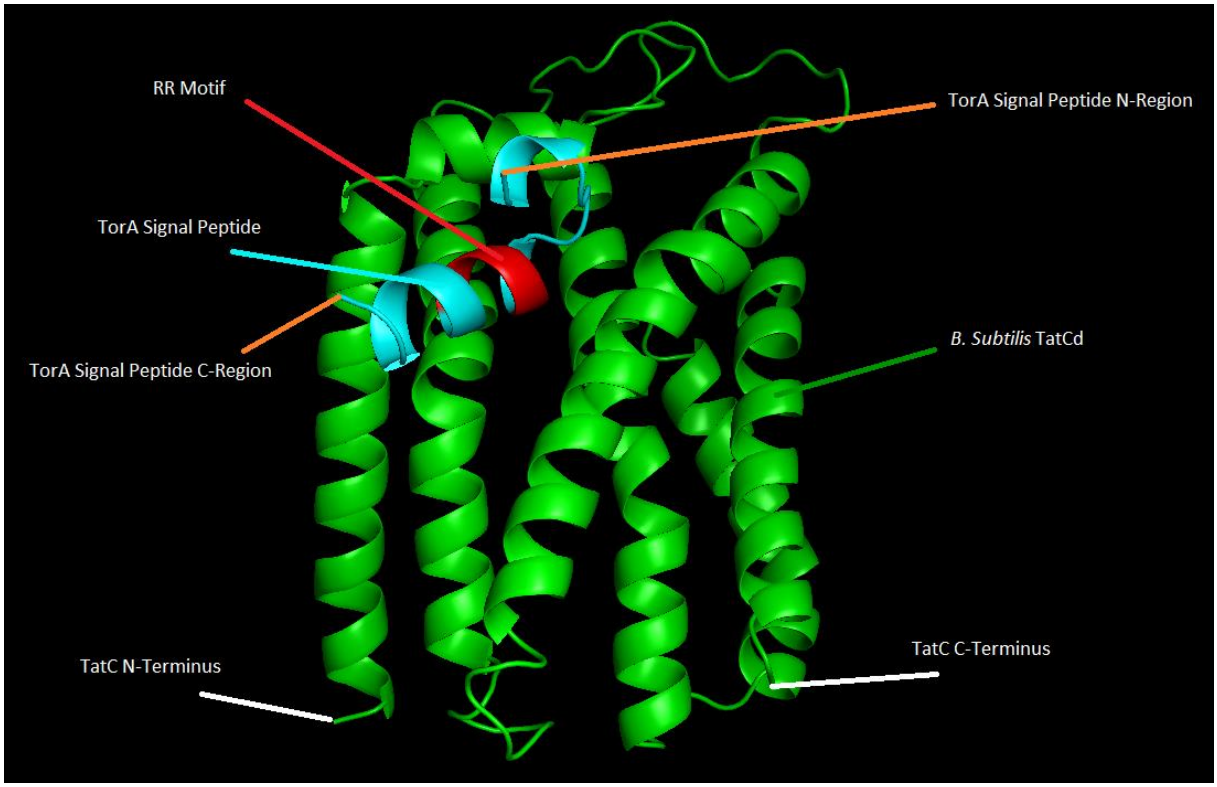

Figure 8. *3D representation of B. subtilis TatCd and TorA protein-protein docking with highlighted features.*

This figure illustrates the secondary structure of the *B. subtilis*. TatCd illustrated in green highlighting its most confidently scored conformation. The TorA signal peptide is shown in blue with its C-region and N-region, and the RR motif, essential for the signal peptide's function, is highlighted in red. The image provides insight into the spatial arrangement of the protein's functional domains, including the TatCd C- terminus and H-terminus.

The examination of models created by HDOCK, analysed using PyMOL, highlights the critical role of the RR motif and

the N- and C-terminal regions of signal peptides in influencing the docking configurations and potential binding efficacy. Despite the initial low quality of input structures suggested by the LGscore, MaxSub, and elevated RMSD values, the preferred models consistently demonstrated interaction patterns between TatC and the signal peptides. Specifically, in the context of SufI binding depicted in Table 4, models 1, 4, 6, 7, 8, and 10 successfully engaged with the TM2 of TatC, whereas the others interacted with TM6. Similarly, for TorA binding shown in Figure 8, models 1, 2, 3, 5, 7, and 9 were observed to bind to TatC TM2, with the rest associating with TM6.

4.4. Binding Site Identification and Ranking

Using POCASA and following Kawabata & Go's (2007) [31] recommendation, a 5Å radius probe was employed to identify potential binding sites, deemed most effective for evaluating peptide interactions. With this approach and default settings, Tables 7 and 8 were constructed, ranking the top five binding sites by pocket number. These tables also detail the pocket volumes, average volume and depth in angstroms, and the total binding sites identified using the 5Å radius probe.

Table 7. *Protein Pocket Binding Characteristics in Gram-Negative.*

Gram-Negative Bacteria	Pocket Binding Rank	Pocket Number	Pocket Volume Å	Average Volume -Depth value Å	Total Binding Sites Probe radius of 5Å
Escherichia coli str. K-12	1	93	407	1108	15
	2	28	231	724	
	3	162	254	720	
	4	265	114	288	
	5	123	99	239	
Pseudomonas aeruginosa	1	114	353	990	12
	2	29	287	858	
	3	142	321	854	
	4	194	153	375	
	5	18	129	337	
Vibrio cholerae	1	147	456	1371	15
	2	157	296	753	
	3	31	170	526	
	4	527	149	368	
	5	137	128	328	
Salmonella enterica	1	116	396	1079	16
	2	159	254	720	
	3	28	216	646	
	4	258	114	288	
	5	124	78	190	
Klebsiella pneumoniae	1	127	596	1648	13
	2	62	122	422	
	3	122	137	333	
	4	367	109	277	
	5	432	88	234	

Binding pocket ranking for Gram-negative bacteria, showing pocket number, volume, average volume, and depth, with a 5 Å probe.

Table 8. *Protein Pocket Binding Characteristics in Gram-Positive.*

Gram-Positive Bacteria	Pocket Binding Rank	Pocket Number	Pocket Volume Å	Average Volume -Depth value Å	Total Binding Sites Probe radius of 5Å
Bacillus subtilis TatCd / TatCy	1	132 / 45	356 / 301	1059 / 866	12 / 17
	2	201 / 193	188 / 297	507 / 846	
	3	52 / 126	109 / 264	371 / 680	
	4	189 / 342	138 / 235	343 / 575	
	5	263 / 93	129 / 62	320 / 161	
Bacillus cereus	1	163	392	1094	12
	2	114	203	564	
	3	61	140	507	
	4	166	194	471	
	5	243	174	423	
Paenibacillus	1	343	656	1726	11
	2	65	250	706	
	3	161	213	651	
	4	17	39	122	
	5	92	20	120	
Streptococcus thermophilus	1	293	635	1652	11
	2	127	229	596	
	3	39	150	464	
	4	213	76	232	
	5	419	71	183	
Streptococcus pneumoniae	1	302	462	1206	10
	2	43	219	717	
	3	113	233	638	
	4	129	77	204	
	5	358	64	184	

Binding pocket ranking for Gram-positive bacteria, including TatCd/TatCy subunit values for *B. subtilis*.

Tables 7 and 8 present the top five ranked binding pockets identified using a 5 Å probe for Gram-negative and Gram-positive bacteria, respectively. The data includes pocket rank, number, volume, average volume and depth (in cubic Å), and total binding sites. For Gram-negative bacteria (Table 7), *Klebsiella pneumoniae* exhibits the largest average pocket volume, while *Salmonella enterica* has the most binding pockets [38]. Among Gram-positive bacteria (Table 8), *Paenibacillus* features the largest average volume [51], and *B. subtilis* displays values for both TatCd and TatCy subunits, with TatCy having the highest pocket count.

4.5. Evidence Overlapping

All collective data was assembled in PyMOL to illustrate the binding results of the signal peptide illustrating the pockets with the most affinity exposed by POCASA, resulting in the construction of Figures 9, 10. This figure presents four models of the same docking TatC protein, presenting the top five docking ranks by hydrogen atoms (white and orange dots), previously mentioned in Tables 7, 8, with the highest affinity in orange, where the signal peptide RR motif stands.

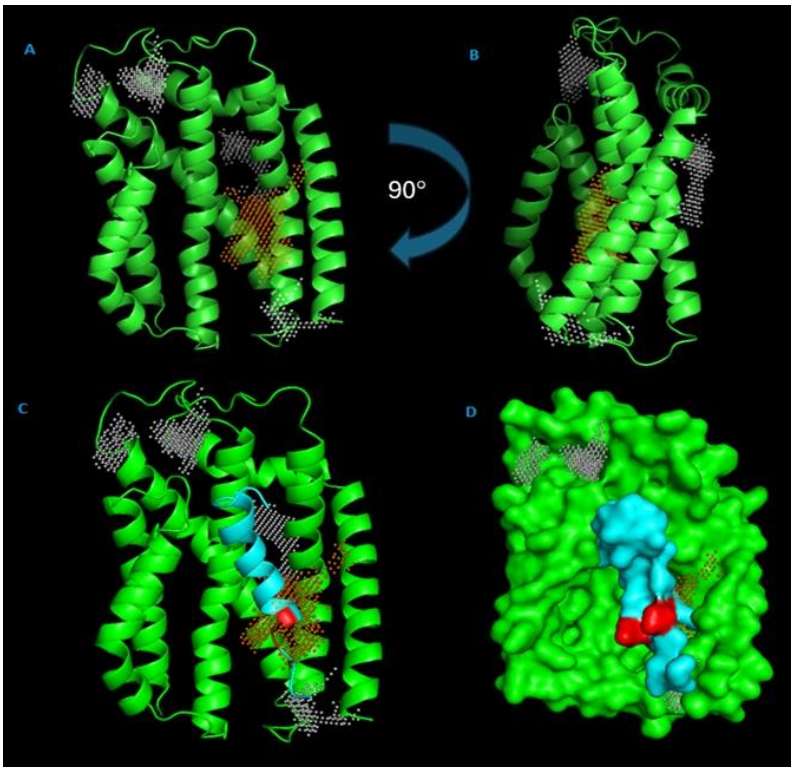

Figure 9. *Structural Analysis of E. coli str. K-12 TatC with Binding Sites and Motifs.*

A and B illustrate the TatC protein in ribbon representation with hydrogens representing the binding sites (white dots). Orange dots indicate the highest ranked (Rank 1) binding site, as identified by POCASA. Model C highlights the SufI signal peptide in blue and the RR motif in red within the TatC protein structure. Model D provides a surface mesh representation of the protein.

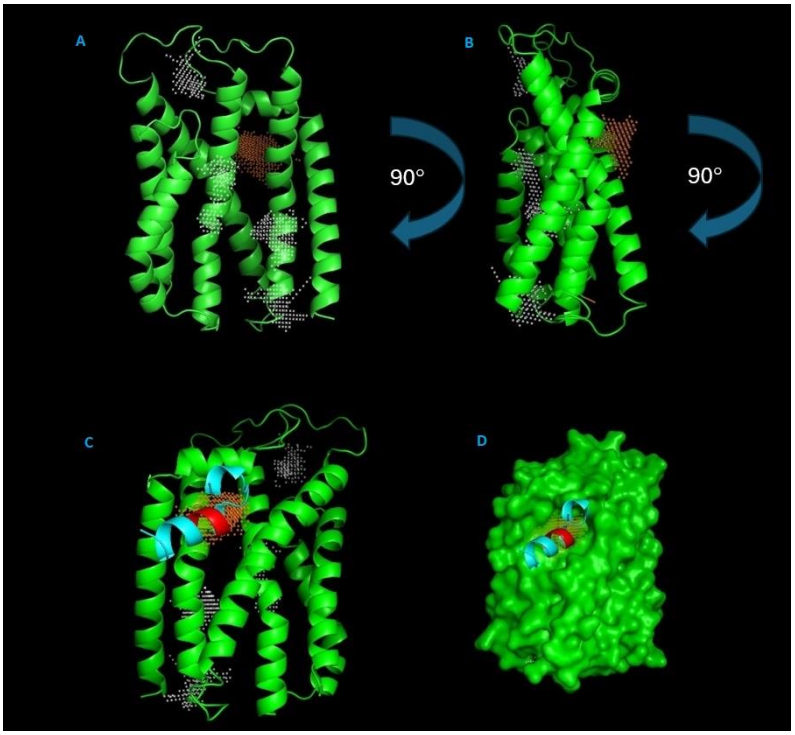

Figure 10. *Structural Analysis of B. subtilis TatCd with Binding Sites and Motifs.*

A, B and C illustrate the TatCd protein in ribbon representation with hydrogens representing the binding sites (white dots). Orange dots indicate the highest ranked (Rank 1) binding site, as identified by POCASA. Model C highlights the TorA signal peptide in blue and the RR motif in red within the TatCd protein structure. Model D provides a surface mesh representation of the protein.

4.6. Hydrophobicity Metrics and Visualisation

Figure 11 illustrates the hydrophobicity of amino acids based on the metrics provided by KVFinder, highlighting their relevance to the TatC hydrophobicity profile.

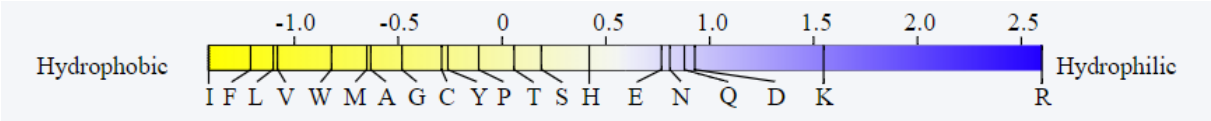

Figure 11. *KVFinder Hydrophobicity Metrics.*

Hydrophobicity metric, with a range from -1.0 for the most hydrophobic amino acids, extending to 2.5 for the most hydrophilic. Below the scores, amino acids are sequenced according to their hydrophobicity rating.

In the docking analysis shown in Figure 12, KVFinder was adjusted to a 5 Å radius, revealing two distinct regions: a highly hydrophilic area in orange and a highly hydrophobic area in pink, with other regions of moderate hydrophobicity marked in blue. Notably, the signaling peptide's RR motif was proximal to the hydrophilic region.

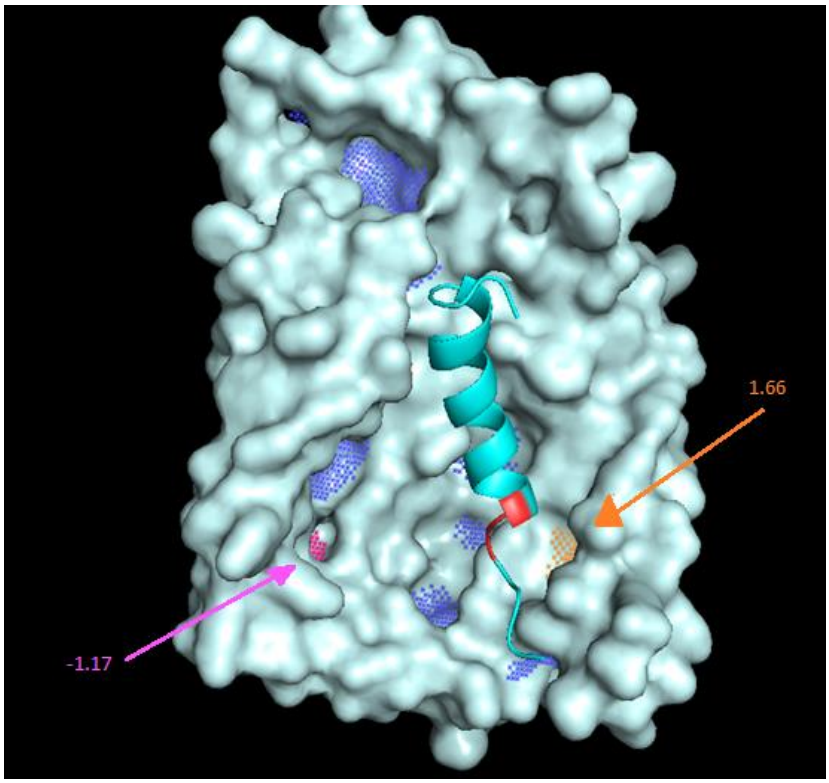

Figure 12. *Hydrophobicity Mapping in E. coli str. K-12 TatC Protein.*

E. coli str. K-12 TatC hydrophobicity analysis. Using a 5 Å radius. Blue dots represent binding regions with negative hydrophobicity, the orange area marks a region with a higher hydrophobicity score of 1.66, and the pink area indicates a region with lower hydrophobicity at -1.17, image assembled with (Pymol, &; KVFinder).

In a newly designed investigative scenario, the wild-type SufI sequence "MSLSRRQFIQASGIALCAGAVPLKASA" from UniProt (2024) [59] was modified by replacing the RR motif with double Isoleucines (II), resulting in "MSLSIIQFIQASGIALCAGAVPLKASA." This change, boosted by Isoleucine's high hydrophobicity on the KVFinder scale, was analysed using HDOCK with parameters consistent with those in Table 5. While docking scores showed similar binding affinities for both sequences, the mutant sequence exhibited a 9% increase in RMSD compared to the wild-type.

Rank	1
Docking Score	-265.83
Confidence Score	0.9102
Ligand rmsd (Å)	39.38

Figure 13. *Results from the modified SufI Sequence that allocates the double Isoleucines.*

The figure presents the Rank 1 result from the Docking of *E. coli str. K-12* TatC protein and the modified signal peptide SufI, in which its docking score was -265, the confidence scores 0.91 and the RMSD 39 Å.

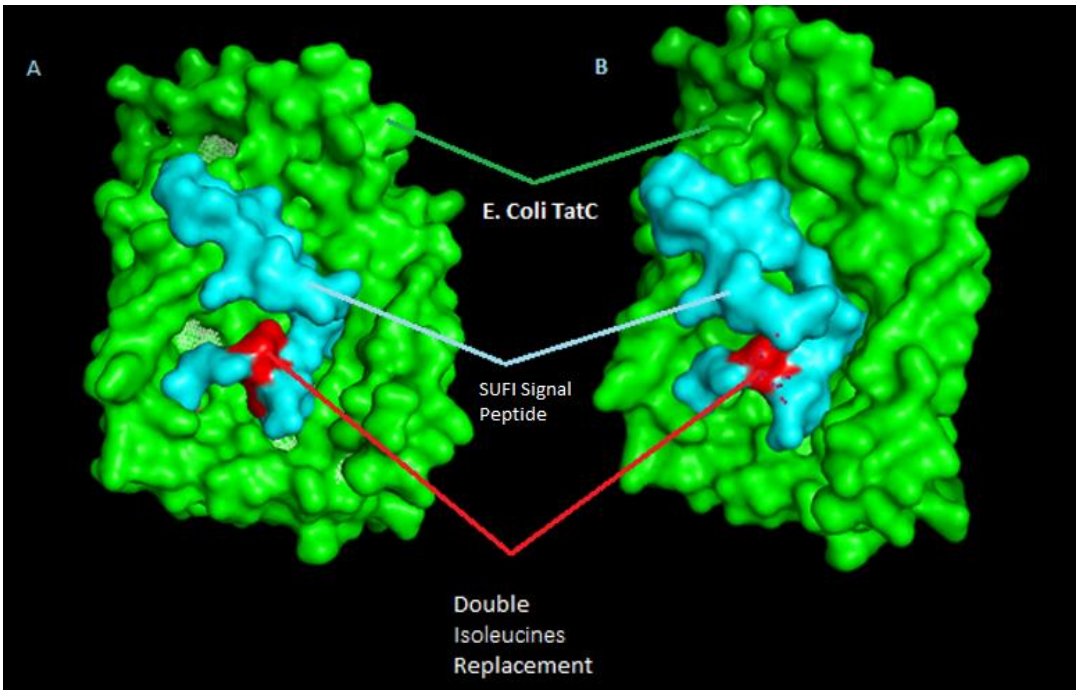

Figure 14. *E. coli str. K-12 TatC docking with SufI mutated sequence.*

The Figure above presents model A and B, where the TatC is highlighted in green, the modified SufI signal peptide in Cyan and the double Isoleucine 'II' change in red. Model B has a 45 degrees rotation to the left based on model A position.

Upon contrasting Model D in Figure 9, which contains the RR motif, a noticeable shift of the signal peptide from TatC's hydrophilic area to a more hydrophobic region is evident, Panel A. Furthermore, this alteration caused the signal peptide to penetrate more deeply into TatC, a change that is clearly illustrated in Panel B of Figure 14.

5. Discussion

This study investigates the critical role of the twin-arginine translocation (Tat) system in mediating the export of folded proteins involved in bacterial metabolic processes and virulence, thereby contributing to pathogenicity. The research primarily focuses on the TatC component, emphasizing the functional significance of conserved amino acid residues in substrate recognition and docking—events that initiate the translocation process. Comparative analysis across phylogenetically diverse bacterial species enabled detailed characterization of TatC's structural conservation and variability.

5.1. TatC and the Importance of Conserved Residues

Research utilizing NCBI, PHYRE2, and Clustal Omega has shed light on the evolutionary relationships among bacterial species based on protein homology. The data reveal that *Salmonella enterica* shares the highest sequence identity with *E. coli str. K-12* among Gram-negative bacteria, suggesting a close evolutionary connection. In contrast, *Pseudomonas aeruginosa* exhibits lower sequence identity, indicating greater evolutionary divergence.

This pattern highlights the dynamic nature of bacterial evolution, where species balance genetic conservation with structural adaptation. Among Gram-positive bacteria, sequence identities related to *Bacillus subtilis* TatCd and TatCy proteins suggest moderate conservation, as seen in *Bacillus cereus*. However, the variation in sequence identities across different species points to significant evolutionary differentiation within this group [62, 22].

5.2. Evolutionary Connections and Protein Homology

The cross-comparison of sequence identities between the two bacterial classifications, as summarized in Table 3, underscores a potential inverse relationship between the degree of pathogenicity and protein sequence similarity—notably, non-pathogenic strains like *E. coli* str. K-12 and *B. subtilis* exhibit higher sequence identity, implying that the evolutionary pressures leading to pathogenicity may drive greater sequence diversity in these proteins. According to (Fitzgerald & Musser, 2001) [13] pathogenic bacteria contain a higher genome diversity than non-pathogenic bacteria, thus presenting the relationship between these non-pathogenic bacteria. Moreover, Dionisio et al. (2023) [12] proposed that gene transfer mechanisms may influence this unexpected evolutionary connection, with *B. cereus* potentially experiencing significant gene alterations that have shaped its evolutionary path. Furthermore, Dionisio et al. (2023) [12] highlighted the potential role of non-pathogenic bacteria in influencing the evolution of pathogenic species through gene transfer processes, thereby possibly augmenting the virulence of pathogenic bacteria despite their non-virulent gene composition.

5.3. Phylogenetic Trees and Residue Analysis

The phylogenetic trees generated using MAFFT web server (Figures 3, 4 and 5) provide an evolutionary perspective, confirming the genetic distances and relationships suggested by sequence identities. The high bootstrap values associated with these trees reflect strong statistical confidence in the phylogenetic classifications, offering a solid foundation for exploring the evolutionary trajectories of these bacterial species [48]. Furthermore, the examination of exact,, and semi-conserved residues (Figure 2) highlights the critical role of these residues in preserving the structural and functional integrity of the TatC protein. The variation in these residues across different taxa may reflect evolutionary adaptations that tailor the protein to the specific ecological niches and pathogenic strategies of the bacteria [6]. Notably, the phylogenetic analysis indicates that each bacterium within the two groups shares 28 identical residues, underscoring a profound evolutionary connection. These residues are likely essential for upholding the protein's structure and function.

Furthermore, 43 residues remain conserved, underscoring their importance in protein function. According to Landau et al. (2005), [34] these residues are critical for the protein, resulting in high evolutionary pressure in resisting amino acid replacements, which correlate with possible roles in catalytic activities and binding processes. In contrast, the 32 semi-conserved residues vary in their amino acid composition among the bacteria. Hsin Liu et al. (2012) [26] explain that these residues do not share chemical properties, suggesting a contribution to protein flexibility and specificity. This variation could be an evolutionary response to each species' unique environmental challenges and selective pressures unimportant to exact and conserved residues. To summarise, exact residues are consistently identical across all sequences examined, conserved residues display similarity in amino acid properties, and semi-conserved residues exhibit a permissible degree of diversity that allows the protein to retain its function despite variations.

5.4. Docking and Pockets Analysis

In this investigation, docking simulations were employed to explore protein-protein interactions, explicitly focusing on evaluating binding affinities and the intricacies of the binding sites. The POCASA web server with a 5Å probe radius highlighted that the most favourable binding sites typically showcased a volume depth over 1000 Å. A noteworthy finding was that the binding site ranked highest, with a volume depth of 1108 Å, achieved a docking score of -265.20 and a confidence level of 0.90 (Figures 9 and 10), according to HDOCK simulations. These figures imply that interactions with docking scores above -200 and confidence levels surpassing 0.7 are reliable (HDOCK, 2024) [23]. However, an exception was identified in the case of *P. aeruginosa*, where the volume depth of the binding site was slightly reduced to 990 Å. The insights gained from the POCASA analysis underlined the significant impact of spatial dimensions on binding site affinity, establishing a clear linkage between greater volume depths and enhanced binding affinities.

Additionally, the research incorporated KVFinder to map the hydrophobicity regions within the *E. coli str. K-12 TatC* amino acids, revealing that the most hydrophilic region corresponded with the highest affinity binding site identified in the docking simulations. This correlation underscores the intricate relationship between hydrophobicity and binding efficiency, highlighting the importance of hydrophilic regions in facilitating protein-protein solid interactions. Despite the high confidence levels indicated by HDOCK, it was noted that the SufI signal peptide, a key component in these interactions, exhibited a structure of low quality. However, *B. subtilis* binding affinity presented a more hydrophobic pocket.

5.5. RR Motif's Role and Alternative Recognition

The analysis of docking results highlights the role of the signal peptide's h-region in facilitating movement within TatC's cavities, which is essential for precise protein-protein interaction alignment [57]. This process notably involves the RR motif's interaction with TatC's hydrophilic areas, observed in docking interactions binding between TM2 and TM3 (Figures 6 and 7), consistent with Frain et al. (2019) [14]. HDOCK simulations, with a 90% confidence level, suggest the signal peptide can rotate significantly, up to 180 degrees, enabling orientation adjustments with the C-region upwards and the N-region downwards, as depicted in Figure 8. This rotational ability supports Palmer and Stansfeld's (2020) [43]

translocation model, which proposes a hairpin conformation driven by the proton motive force, crucial for translocation. These findings enhance understanding of the Tat pathway's molecular mechanics, emphasising the complex coordination necessary for effective protein translocation.

Contrary to the common belief that the RR motif plays a pivotal role in TatC recognition, research by Huang and Palmer (2017) [27] suggests that the RR motif's significance may be overstated. Their findings indicate that increasing the hydrophobicity of the signal peptide can compensate for the absence of the RR motif, challenging the previously understood necessity of this feature for effective TatC recognition. This revelation underscores the complexity of the Tat pathway and the potential for alternative recognition strategies that deviate from established theories. Additionally, Ulfig et al. (2017) [56] report that the h-region of the signal peptide significantly contributes to its positioning within TatBC and its binding affinity, alongside the RR motif. This observation aligns with docking scenarios where the more hydrophobic signal peptide shifted from the conventional TM2/TM3 binding site to the most hydrophobic region within TatC [54]. In which, the signal peptide moved deeper into the binding cavity, bypassing the need for twin-arginine residues while still achieving identical confidence scores. This suggests that mutations within the peptide did not significantly alter overall affinity or docking compatibility according to the in silico scoring algorithm. Notably, in *Bacillus subtilis*, hydrophobic binding plays a particularly important role in facilitating recognition and interaction with TatC, further highlighting the adaptability of the system to alternative binding mechanisms beyond those reliant on the RR motif [40].

5.6. Strengths, Weaknesses, and Areas for Future Work

While HDOCK shows high confidence results, evaluating other factors confirming these docking outcomes' reliability is crucial. A significant point of concern is the ligand's root mean square deviation (RMSD), which was found to be 29.09 Å, indicating a deviation of 29% from the anticipated structure. According to Castro-Alvarez, 2017 [7], an RMSD value below 2 Å is acceptable, and values above this benchmark should be interpreted cautiously. Furthermore, the assessment tools used by HDOCK suggest that RMSD might not be a definitive indicator of docking precision [23], indicating possible issues with the ligand's quality. This implies that while the model might confidently predict binding accuracy, the ligand's spatial arrangement may need adjustment. Additional experiments with TorA also showed confidence levels above 86%. Nevertheless, these tests resulted in high RMSD values, raising questions about attributing peptide rotation solely to the hairpin structure identified by Palmer & Stansfeld (2020) [43], in which Ciemny, 2016 [8] demonstrate that ligands take different spatial conformations according to their RMSD. Considering these observations, future studies should consider using different software or improved docking models to verify these findings.

These results partially echoed findings by Huang & Palmer 2017 [27] and Habersetzer, 2017 [21], which indicated a transition of the signal peptide facilitating a cross-linking interaction with the TatB TM helix and the TM5/TM6 interface of TatC. However, the questioned models were positioned closer to TM6, diverging from TM5, potentially due to previously discussed RMSD values.

Future research avenues include addressing the lack of docking studies involving TatB or TatBC complexes and a comprehensive hydrophobicity analysis, which might have improved the reliability of these excluded models. Enhancing signal peptide representation in docking simulations could yield more precise results. Moreover, examining interactions with a broader variety of signal peptides, especially those of different lengths like the approximately 45 amino acids long DmsA, could deepen insights into protein-protein interactions within the system [54]. Additionally, employing alternative docking software could further refine the research outcomes, offering a broader perspective on the complexities of these interactions.

6. Conclusion

The research confirms the hypothesis that signal peptides transition from the TM2/TM3 binding site to the more hydrophobic regions within the *E. coli str. K-12* and *B. subtilis* TatC, highlighting hydrophobicity's key role in protein interactions. It demonstrates that peptides can bind effectively without twin-arginine motifs, aligning with findings by Huang & Palmer 2017 [27] and suggesting a deeper embedding of signal peptides in TatC's hydrophobic zones. While noteworthy, these findings should be treated as preliminary and interpreted with care, emphasizing the need for further research to confirm and expand upon these early results. Despite the advantages of in silico analysis, it only partially encompasses the entire complexity of protein binding, highlighting the importance of conducting more in vitro studies to deepen our understanding. Furthermore, this study enriches the discourse on hydrophobic interactions and their importance in protein dynamics, aligning with the innovative works of Oba et al., 2023 [42] and Huang & Palmer 2017 [27]. Their investigations demonstrate *the* effects of hydrophobicity and helicity on different antimicrobial and cell-penetrating peptides, which facilitate crossing bacterial cell barriers to interact with TatC, contribute valuable insights into this field of study.

Author Contributions

Micael Sousa Correia: Research, Resource, Writing – review & editing

Sharon Mendel Williams: Project administration, Resources, Supervision, Validation, Writing – review & editing

Conflicts of Interest

The authors declare no conflicts of interest.

Appendix

Sequence retrieval from Uniprot:

MSLSRRQFIQASGIALCAGAVPLKASA AGQQQPL SufI + sequence

MSLSKKQFIQASGIALCAGAVPLKASAAGQ SUFI RR changed to KK (double Lysine) hydrophilic (discarded)

MSLSIIQFIQASGIALCAGAVPLKASAAGQ SUFI RR change to II (double Isoleucine) Hydrophobic

References

[1] Alami, M., Lüke, I., Deitermann, S., Eisner, G., Koch, H.-G., Brunner, J., & Müller, M. (2003). Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli. *Molecular Cell*, *12*(4), 937–946. https://doi.org/10.1016/s1097-2765(03)00398-8

[2] Blümmel, A.-S., Drepper, F., Knapp, B., Eimer, E., Warscheid, B., Müller, M., & Fröbel, J. (2017). Structural features of the TatC membrane protein that determine docking and insertion of a twin-arginine signal peptide. *Journal of Biological Chemistry*, *292*(52), 21320–21329. https://doi.org/10.1074/jbc.M117.812560

[3] Bronstein, P., Marrichi, M., & DeLisa, M. P. (2004). Dissecting the twin-arginine translocation pathway using genome-wide analysis. *Research in Microbiology*, *155*(10), 803–810. https://doi.org/10.1016/j.resmic.2004.06.013

[4] Buck, E. D., Lammertyn, E., & Anné, J. (2008). The importance of the twin-arginine translocation pathway for bacterial virulence. *Trends in Microbiology*, *16*(9), 442–453. https://doi.org/10.1016/j.tim.2008.06.004

[5] Buchanan, G., de Leeuw, E., Stanley, N. R., Wexler, M., Berks, B. C., Sargent, F., and Palmer, T. (2002) 'Functional Complexity of the Twin-Arginine Translocase TatC Component Revealed by Site-Directed Mutagenesis.' *Molecular Microbiology* [online] 43(6), 1457–70. available from http://www.ncbi.nlm.nih.gov/pubmed/11952898

[6] Capra, J. A., & Singh, M. (2007). Predicting functionally important residues from sequence conservation. *Bioinformatics*, *23*(15), 1875–1882. https://doi.org/10.1093/bioinformatics/btm270

[7] Castro-Alvarez, A., Costa, A., & Vilarrasa, J. (2017). The Performance of Several Docking Programs at Reproducing Protein–Macrolide-Like Crystal Structures. *Molecules*, *22*(1), 136. https://doi.org/10.3390/molecules22010136

[8] Ciemny, M. P., Debinski, A., Paczkowska, M., Kolinski, A., Kurcinski, M., & Kmiecik, S. (2016). Protein-peptide molecular docking with large-scale conformational changes: the p53-MDM2 interaction. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep37532

[9] Cline, K. (2015) 'Mechanistic Aspects of Folded Protein Transport by the Twin Arginine Translocase (Tat).' *The Journal of Biological Chemistry* [online] 290(27), 16530–8. http://www.ncbi.nlm.nih.gov/pubmed/25975269

[10] Dalbey, R. E. and Kuhn, A. (2012) 'Protein Traffic in Gram-Negative Bacteria - How Exported and Secreted Proteins Find Their Way'. in *FEMS Microbiology Reviews*. vol. 36 (6). Oxford Academic, 1023–1045.

[11] Diggle, S. P., & Whiteley, M. (2019). Microbe Profile: Pseudomonas aeruginosa: Opportunistic Pathogen and Lab Rat. *Microbiology*, *166*(1), 30–33. https://doi.org/10.1099/mic.0.000860

[12] Dionisio, F., Domingues, C. P. F., Rebelo, J. S., Monteiro, F., & Nogueira, T. (2023). The Impact of Non-Pathogenic Bacteria on the Spread of Virulence and Resistance Genes. *International Journal of Molecular Sciences*, *24*(3), 1967. https://doi.org/10.3390/ijms24031967

[13] Fitzgerald, J. Ross., & Musser, J. M. (2001). Evolutionary genomics of pathogenic bacteria. *Trends in Microbiology*, *9*(11), 547–553. https://doi.org/10.1016/s0966-842x(01)02228-4

[14] Frain, K. M., Robinson, C., & van Dijl, J. M. (2019). Transport of Folded Proteins by the Tat System. *The Protein Journal*, *38*(4), 377–388. https://doi.org/10.1007/s10930-019-09859-y

[15] Fröbel, J., Rose, P., & Müller, M. (2012). Twin-arginine-dependent translocation of folded proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1592), 1029–1046. https://doi.org/10.1098/rstb.2011.0202

[16] Gérard, F., & Cline, K. (2006). Efficient Twin Arginine Translocation (Tat) Pathway Transport of a Precursor Protein Covalently Anchored to Its Initial cpTatC Binding Site. *Journal of Biological Chemistry*, *281*(10), 6130–6135. https://doi.org/10.1074/jbc.m512733200

[17] Ghosh, P., & Sowdhamini, R. (2017). Bioinformatics comparisons of RNA-binding proteins of pathogenic and non-pathogenic Escherichia coli strains reveal novel virulence factors. *BMC Genomics*, *18*(1). https://doi.org/10.1186/s12864-017-4045-3

[18] Guerra, J. V. D. S., Ribeiro Filho, H. V., Bortot, L. O., Honorato, R. V., Pereira, J. G. D. C., & Lopes-de-Oliveira, P. S. (2020). ParKVFinder: a thread-level parallel approach in biomolecular cavity detection. *SoftwareX*, *12*, 100606. https://doi.org/10.1016/j.softx.2020.100606

[19] Guerra, J. V. D. S., Ribeiro-Filho, H. V., Jara, G. E., Bortot, L. O., Pereira, J. G. D. C., & Lopes-de-Oliveira, P. S. (2021). pyKVFinder: an efficient and integrable python package for biomolecular cavity detection and characterization in data science. *BMC Bioinformatics*. https://doi.org/10.1186/s12859-021-04519-4

[20] Guerra, J. V. S., Ribeiro-Filho, H. V., Pereira, J. G. C., & Lopes-de-Oliveira, P. S. (2023). KVFinder-web: a web-based application for detecting and characterizing biomolecular cavities. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkad324

[21] Habersetzer, J., Moore, K., Cherry, J., Buchanan, G., Stansfeld, P. J. and Palmer, T. (2017). Substrate-triggered position switching of TatA and TatB during Tat transport in *Escherichia coli*. *Open Biology*, 7(8), pp. 170091–170091. https://doi.org/10.1098/rsob.170091

[22] Haliloglu, T., & Bahar, I. (2015). Adaptability of protein structures to enable functional interactions and evolutionary implications. *Current Opinion in Structural Biology*, *35*, 17–23. https://doi.org/10.1016/j.sbi.2015.07.007

[23] Halpern, M., & Izhaki, I. (2017). Fish as Hosts of Vibrio cholerae. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.00282

[24] HDOCK Group. (2024). HDOCK server for protein-protein and protein-DNA/RNA docking. Retrieved from http://hdock.phys.hust.edu.cn/

[25] Hong, Y., Wu, Y., Xie, Y., Ben, L., Bu, X., Pan, X., Shao, J., Dong, Q., Qin, X., & Wang, X. (2023). Effects of antibiotic-induced resistance on the growth, survival ability and virulence of Salmonella enterica. *Food Microbiology*, *115*, 104331–104331. https://doi.org/10.1016/j.fm.2023.104331

[26] Hsin Liu, C., Li, K.-C., & Yuan, S. (2012). Human protein–protein interaction prediction by a novel sequence-based co-evolution method: co-evolutionary divergence. *Bioinformatics*, *29*(1), 92–98. https://doi.org/10.1093/bioinformatics/bts620

[27] Huang, Q., & Palmer, T. (2017). Signal Peptide Hydrophobicity Modulates Interaction with the Twin-Arginine Translocase. *MBio*, *8*(4). https://doi.org/10.1128/mbio.00909-17

[28] Kadeřábková, N., Furniss, C. D., Mahmood, A., & Mavridou, D. (2023). Making a chink in their armor: Current and next-generation antimicrobial strategies against the bacterial cell envelope. *Advances in Microbial Physiology*, 221–307. https://doi.org/10.1016/bs.ampbs.2023.05.003

[29] Katoh, K., Rozewicki, J., & Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4). https://doi.org/10.1093/bib/bbx108

[30] Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

[31] Kawabata, T., & Go, N. (2007). Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. *Proteins: Structure, Function, and Bioinformatics*, *68*(2), 516–529. https://doi.org/10.1002/prot.21283

[32] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 Web Portal for Protein modeling, Prediction and Analysis. *Nature Protocols*, *10*(6), 845–858. https://doi.org/10.1038/nprot.2015.053

[33] Kuraku, S., Zmasek, C. M., Nishimura, O., & Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research*, *41* (W1), W22–W28. https://doi.org/10.1093/nar/gkt389

[34] Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., & Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Research*, *33* (Web Server), W299–W302. https://doi.org/10.1093/nar/gki370

[35] Lee, P. A., Tullman-Ercek, D., & Georgiou, G. (2006). The Bacterial Twin-Arginine Translocation Pathway. *Annual Review of Microbiology*, *60*(1), 373–395. https://doi.org/10.1146/annurev.micro.60.080805.142212

[36] Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R., & Shi, J. (2020). Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Signal Transduction and Targeted Therapy*, *5*(1). https://doi.org/10.1038/s41392-020-00315-3

[37] Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, *50* (W1). https://doi.org/10.1093/nar/gkac240

[38] Mannion, J. M., Segal, B. M., McLoughlin, R. M., & Lalor, S. J. (2023). Respiratory tract *Moraxella catarrhalis* and *Klebsiella pneumoniae* can promote pathogenicity of myelin-reactive Th17 cells. *Mucosal Immunology*. https://doi.org/10.1016/j.mucimm.2023.04.003

[39] Mehner-Breitfeld, D., Ringel, M. T., Tichy, D. A., Endter, L. J., Stroh, K. S., Heinrich Lünsdorf, Herre Jelger Risselada, & Brüser, T. (2022). TatA and TatB generate a hydrophobic mismatch important for the function and assembly of the Tat translocon in Escherichia coli. *Journal of Biological Chemistry*, *298*(9), 102236–102236. https://doi.org/10.1016/j.jbc.2022.102236

[40] Mendel, S., McCarthy, A., Barnett, J. P., Eijlander, R. T., Nenninger, A., Kuipers, O. P., & Robinson, C. (2008). The Escherichia coli TatABC System and a Bacillus subtilis TatAC-type System Recognise Three Distinct Targeting Determinants in Twin-arginine Signal Peptides. *Journal of Molecular Biology*, *375*(3), 661–672. https://doi.org/10.1016/j.jmb.2007.09.087

[41] National Center for Biotechnology Information (NCBI). (2024). BLAST: Basic Local Alignment Search Tool. Retrieved [January, 2024], from https://blast.ncbi.nlm.nih.gov/Blast.cgi

[42] Oba, M., Nakajima, S., Kurumi Misao, Hidetomo Yokoo, & Tanaka, M. (2023). Effect of helicity and hydrophobicity on cell-penetrating ability of arginine-rich peptides. *Bioorganic & Medicinal Chemistry*, *91*, 117409–117409. https://doi.org/10.1016/j.bmc.2023.117409

[43] Palmer, T., & Stansfeld, P. J. (2020). Targeting of proteins to the twin‐arginine translocation pathway. *Molecular Microbiology*, *113*(5), 861–871. https://doi.org/10.1111/mmi.14461

[44] Pei, D., & Dalbey, R. E. (2022). Membrane translocation of folded proteins. *Journal of Biological Chemistry*, *298*(7), 102107. https://doi.org/10.1016/j.jbc.2022.102107

[45] Peng, S., Guo, C., Cui, H., & Duan, Z. (2023). Complete genome analysis of *Lactiplantibacillus plantarum* VHProbi P06, a novel probiotic that resists *Streptococcus pneumoniae* in the upper respiratory tract. *International Journal of Biological Macromolecules*, *253*, 127320–127320. https://doi.org/10.1016/j.ijbiomac.2023.127320

[46] Potteth, U. S., Upadhyay, T., Saini, S., & Ishu Saraogi. (2021). Novel Antibacterial Targets in Protein Biogenesis Pathways. *ChemBioChem*, *23*(4). https://doi.org/10.1002/cbic.202100459

[47] Rollauer, S. E., Tarry, M. J., Graham, J. E., Jääskeläinen, M., Jäger, F., Johnson, S., Krehenbrink, M., Liu, S.-M., Lukey, M. J., Marcoux, J., McDowell, M. A., Rodriguez, F., Roversi, P., Stansfeld, P. J., Robinson, C. V., Sansom, M. S. P., Palmer, T., Högbom, M., Berks, B. C., and Lea, S. M. (2012) 'Structure of the TatC Core of the Twin-Arginine Protein Transport System'. *Nature* [online] 492(7428), 210–214. available from http://www.nature.com/doifinder/10.1038/nature11683

[48] Ren, A., Ishida, T., & Akiyama, Y. (2013). Assessing statistical reliability of phylogenetic trees via a speedy double bootstrap method. *Molecular Phylogenetics and Evolution*, *67*(2), 429–435. https://doi.org/10.1016/j.ympev.2013.02.011

[49] Ribeiro, A. A. S. T., Horta, B. A. C., & de Alencastro, R. B. (2021). KVFINDER (v1.2.0) [Software]. Available from https://kvfinder-web.cnpem.br/

[50] Robinson, S. W., Afzal, A. M., & Leader, D. P. (2014). Bioinformatics: Concepts, Methods, and Data. *Handbook of Pharmacogenomics and Stratified Medicine*, 259–287. https://doi.org/10.1016/b978-0-12-386882-4.00013-x

[51] Sáez-Nieto, J. A., Medina-Pascual, M. J., Carrasco, G., Garrido, N., Fernandez-Torres, M. A., Villalón, P., & Valdezate, S. (2017). Paenibacillus spp. isolated from human and environmental samples in Spain: detection of 11 new species. *New Microbes and New Infections*, *19*, 19–27. https://doi.org/10.1016/j.nmni.2017.05.006

[52] Schrödinger, LLC. (2024). PyMOL (Ver. 2.0) [Webserver]. Available from https://pymol.org/

[53] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., & Higgins, D. G. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7(1), 539. https://doi.org/10.1038/msb.2011.75

[54] Tara, Tran, V. A., & Turner, R. J. (2013). The Hydrophobic Region of the DmsA Twin-Arginine Leader Peptide Determines Specificity with Chaperone DmsD. *Biochemistry*, *52*(43), 7532–7541. https://doi.org/10.1021/bi4009374

[55] Tarry, M. J., Schäfer, E., Chen, S., Buchanan, G., Greene, N. P., Lea, S. M., Palmer, T., Saibil, H. R., and Berks, B. C. (2009) 'Structural Analysis of Substrate Binding by the TatBC Component of the Twin-Arginine Protein Transport System.' *Proceedings of the National Academy of Sciences of the United States of America* [online] 106(32), 13284–9. http://www.ncbi.nlm.nih.gov/pubmed/19666509

[56] Taw, M. N., Li, M., Kim, D., Rocco, M. A., DeLisa, M. P., & Zhmayev, D. (2021). Engineering a Supersecreting Strain of *Escherichia coli* by Directed Coevolution of the Multiprotein Tat Translocation Machinery. *ACS Synthetic Biology*, *10*(11), 2947–2958. https://doi.org/10.1021/acssynbio.1c00183

[57] Ulfig, A., Fröbel, J., Lausberg, F., Blümmel, A.-S., Heide, A. K., Müller, M., & Freudl, R. (2017). The h-region of twin-arginine signal peptides supports productive binding of bacterial Tat precursor proteins to the TatBC receptor complex. *Journal of Biological Chemistry*, *292*(26), 10865–10882. https://doi.org/10.1074/jbc.m117.788950

[58] Ulfig, A., & Freudl, R. (2018). The early mature part of bacterial twin-arginine translocation (Tat) precursor proteins contributes to TatBC receptor binding. *Journal of Biological Chemistry*, *293*(19), 7281–7299. https://doi.org/10.1074/jbc.ra118.002576

[59] Uniprot. (2023). *UniProt*. Uniprot.org. https://www.uniprot.org

[60] Voulhoux, R. (2001). Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. *The EMBO Journal*, *20*(23), 6735–6741. https://doi.org/10.1093/emboj/20.23.6735

[61] Widdick, D. A., Dilks, K., Chandra, G., Bottrill, A., Naldrett, M., Pohlschroder, M., & Palmer, T. (2006). The twin-arginine translocation pathway is a major route of protein export in Streptomyces coelicolor. *Proceedings of the National Academy of Sciences*, *103*(47), 17927–17932. https://doi.org/10.1073/pnas.0607025103

[62] Wingert, B., Krieger, J., Li, H., & Bahar, I. (2021). Adaptability and specificity: how do proteins balance opposing needs to achieve function? *Current Opinion in Structural Biology*, *67*, 25–32. https://doi.org/10.1016/j.sbi.2020.08.009

[63] Xu, Y., Hu, J., Liu, D., Tang, J., Liang, M., Wu, J., & Xiong, J. (2023). Assessment of the safety and metabolism characteristics of Streptococcus thermophilus DMST-H2 based on complete genome and phenotype analysis. *LWT*, *184*, 114907–114907. https://doi.org/10.1016/j.lwt.2023.114907

[64] Yan, Y., Tao, H., He, J., & Huang, S.-Y. (2020). The HDOCK server for integrated protein-protein docking. *Nature Protocols*. https://doi.org/10.1038/s41596-020-0312-x

[65] Yu, J., Zhou, Y., Tanaka, I., & Yao, M. (2010). Roll: A new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. Bioinformatics, 26(1), 46-52. http://altair.sci.hokudai.ac.jp/g6/Research/POCASA_e.html

[66] Zhang, Y., Wang, L., Hu, Y., & Jin, C. (2014). Solution structure of the TatB component of the twin-arginine translocation system. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1838*(7), 1881–1888. https://doi.org/10.1016/j.bbamem.2014.03.015