Structural and Biochemical Investigation of Tat components

Thesis Submitted to AcSIR For the Award of
the Degree of

DOCTOR OF PHILOSOPHY
In
BIOLOGICAL SCIENCES

By
DEEPANJAN GHOSH
10BB13A26033

Under the guidance of

Dr. V. Koteswara Rao
(Supervisor)

Dr. Sureshkumar Ramasamy
(Co-Supervisor)

Biochemical Sciences Division
CSIR-National Chemical Laboratory
Pune-411008, India

Dedicated to .....

Maa and Baba

सीएसआईआर - राष्ट्रीय रासायनिक प्रयोगशाला
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)
डॉ. होमी भाभा मार्ग, पुणे - 411 008, भारत
CSIR - NATIONAL CHEMICAL LABORATORY
(Council of Scientific & Industrial Research)
Dr. Homi Bhabha Road, Pune - 411 008, India

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled **Structural and Biochemical Investigation of Tat components** submitted by Mr. **Deepanjan Ghosh** to Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy**, embodies original research work under my/our supervision/guidance. We further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material obtained from other sources has been duly acknowledged in the thesis. Any text, illustration, table etc., used in the thesis from other sources, have been duly cited and acknowledged.

It is also certified that this work done by the student under my supervision, is plagiarism free

Deepanjan Ghosh

Deepanjan Ghosh
(Research Student)

Dr. V. Koteswara Rao
(Research Supervisor)

Sureshkumar Ramasamy
Dr. Sureshkumar Ramasamy
(Research Co-Supervisor)

Place: Pune
Date: 25th May 2019

Communication Channels
NCL Level DID : 2590
NCL Board No. : +91-20-25902000
EPABX +91-20-25893300
: +91-20-25893400
FAX
Director's Office : +91-20-25902601
COA's Office : +91-20-25902660
SPO's Office : +91-20-25902664
WEBSITE
www.ncl-india.org

DECLARATION BY THE CANDIDATE

I hereby declare that the thesis entitled " **Structural and Biochemical Investigation of Tat components** ", submitted for the Degree of **Doctor of Philosophy in Biological Sciences** to Academy of Scientific and Innovative Research (AcSIR) is the record of work carried out by me at Biochemical Sciences Division, CSIR-National Chemical Laboratory, Pune-411008, India under the supervision of **Dr. V. Koteswara Rao (Research guide)** and **Dr. Sureshkumar Ramasamy (Research co-guide)**. The work is original and has not been submitted in part or full by me for any other degree or diploma to any other University. I further declared that the material obtained from other sources has been duly acknowledged in the thesis.

Deepanjan Ghosh

Deepanjan Ghosh
Division of Biochemical Sciences,
CSIR- National Chemical Laboratory,
Pune – 411008, Maharashtra, India

Place: Pune

Date: 25th May 2015

Acknowledgement

The writing of this thesis marks the culmination of my journey towards obtaining Ph.D. It is the outcome of several people's contribution, without their support this wonderful journey, which is a blend of multitude of things, learning, patience, help, encouragement, aspiration, ecstasy and melancholy, would have not been possible. Hence it becomes my responsibility to acknowledge the people who had extended their support to me scientifically or morally to make a smooth running of my way. I, now take this opportunity to thank all who have helped me during these unforgettable years which is a crucial episode of my life.

First and foremost I am extremely thankful to **Dr. V. Koteswara Rao**, my research guide, for giving me complete freedom to execute my work. This work would not have been possible without his guidance, support and encouragement.

My utmost and sincere gratitude is to **Dr. Sureshkumar Ramasamy**, who was my guide my mentor and introduced me to the world of science. He supported me completely and helped me in solving all the queries related to this research topic. Although I am new to this field, I was lucky to have him as my supervisor because of which I could now learn structural biology. He has offered his help in several difficulties and I am thankful to him for being there for me in every moment of need. I am privileged to mention here that I feel very lucky to work with him for my PhD. His teaching not just limits to scientific knowledge but extends to many approaches in life. To this day he is one of the best teachers that I have ever had.

My profound gratitude is to **Dr. Archana V. Pundle**, for allowing me to pursue my degree under supervision and be so kind and warm. I am extremely thankful to **Dr. C G Suresh** who taught us crystallography and always made sure we improved our scientific presentation skills and scientific temperment.

I would also like to thank my research advisory committee Dr. Alok Sen, Dr. Narendra kadoo and Dr. H. V. Thulasiram for evaluating my progress, insightful comments and encouragement. I also acknowledge Director CSIR, NCL and head Biochemical Sciences Division, Dr. Ashok Giri, for providing the lab facilities and infrastructure. I would like to profusely thank Dr. Kiran Kulkarni and Dr. Subashchandrabose Chinnathambi who supported my work by giving access to their instruments and facilities day in and day out. I am also thankful to Dr. Saikrishnan and Dr. Gayatri Mahesh and Jyoti from IISER Pune for allowing me to use the X-ray facility. I am also thankful to Dr. Radha Chauhan, Dr. Janesh Kumar and Ashiwini from NCCS for allowing me to use Mosquito facility and their help. I am grateful to Dr. Ravindra Makde, Dr. Ashwani Kumar and Dr. Biplab Ghosh from PXBL21 beamline, Indus-II, RRCAT, DAE, Indore, for their kind help & support in the crystal data collection.

I also thank **Prof. Bil Clemons** and his lab from Caltech, Pasadena, USA for providing the Mycobacterial Tat constructs and other clones. I am also thankful to **Dr. Manidipa Banerjee** and Kimi Azad from IIT- Delhi for performing negative staining on their HR-TEM. I thoroughly enjoyed working with **Dr. Sushma Gaikwad** and Snakruti Agarwal on plant lectins.

My entire journey here at CSIR-National Chemical Laboratory would not have been a smooth one but for the support of my friends and lab mates. My sincerest regards to **Dr. Manas Sule** and **Dr. Priyabrata Panigrahi** for teaching me bioinformatics, helping me with codes and guiding me each and every step in

computational studies. I am grateful to **Dr. Deepak Chand, Dr. Manu M.S., Dr. Ameya Bendre, Yashpal Yadav, Vijay Rajput, Shiva Shankar, Debjyoti Boral, Shridhar Chougule** with whome I spent so much time in lab, had hours of discussions and troubleshot many scientific problems. They have been like my family in Pune. I would also like to thank my lab-mates Dr. Nishant Varshney, Dr. Tulika Jaokar, Dr. Ranu Sharma, Dr. Ruby Singh, Tejashri Hingmire, Aditi Bhand, Amol Sawant and Vishwambar Navale for their help and support. My sincere regards to my trainees Avani Goswami, Mallika bhattacharya, Venkat Vaithyanathan and Sugavaneshwaran K who immensely helped with my work.

I would like to thank my batchments Bhagyashri Thorat Gadgil, Shakuntala Kolewar, Anand Sukeerthi and Sneha Menon who made coursework and life fun. I would like to thank Shweta, Nalini, Tushar, Abhishek and other members Dr. Subhashchandrabose lab. Sneha, Zenia, Ashwini, Debopriya, Gopal from Dr. Kiran Kilkarni's lab for helping me with my experiments. My gratitude to Dr. Sayali Dalal, Dr. Ekta Shukla, Dr. Ruchira Mukharjee, Hrishikesh Mungi, Parag Maru, Dr. Rahul Slunke, Dr. Reema Banarjee, Dr. Avinash Sundar, Dr. Pushpa Phelim, Nirbhik Acharaya, Prajna Mishra, Dr. Avinash Pandreka, Dr. Selva Rupa and lab members of Dr. Mahesh Kulkarni, Dr. H.V. Thulasiram, Dr. Narendra Kadoo, Dr. Ashok Giri, Dr.Archana Pundle, Dr. Asmita Prbhune, Dr. Dhanasekaran, Dr. Santosh Jha, Indira Madam and every staff member of Biochemical Sciences for their support.

I would like to thank my friends, Dr. Gaurav Bhattacharjee, Sayan Pal, Dr. Akshay Singhan, Dr. Ejaj Pathan, Dr. Manoj Sharma, Bithika Chatterjee, Dilip Pandey, Meghana Thorat, Shikha Thakur, Dushyant Rao, Sayantan Acharya, Dr. Santosh Kumar, Dr. Deepak Kumar, Amit Kumawat, Subhra Jyotsna and everyone in the hostels with whom I had so much fun. I had a lot of good memories with the Bengali community organizing Saraswati puja my drama team and the Marathi community who took me to deepest parts of Maharashtra and gave me an opportunity to give back to the community. Thanks to my friends Aishu, Chetan, Anusha, Deepti, Shreelu, Kishen, Razzu, Bhanu, Chandan, Poornima, Sameer, Dr. TB, Dr. BG, Swapnil, Balu and Kenny for their encouragement and constant support. My sincere thanks to my teachers Dr. Arunalakshmi K, Dr. Khasim Beebi, Dr. Vishwanatha K, Mrs. Leetaha Nandan and Mrs. Seetamahalkshmi for motivating me and having faith in me.

I am also grateful to the SAC and AcSIR office staff of CSIR-NCL especially Komal,Vaishali and Poornima kole madam for the help they provided me time to time. I thank the staff from stores, purchase, accounts, civil, electrical and every other section of NCL.

I would like to thank Director, CSIR-National Chemical Laboratory (CSIR-NCL, Pune) for giving me the opportunity to work in this great institution. I would like to thank Council of Scientific & Industrial Research (CSIR) for my fellowship.

Finally I would like to thank my parents who have supported me all through my journey and have been my pillars of support. I also thank all those who have directly or indirectly supported me.

Deepanjan Ghosh

Deepanjan Ghosh

Table of contents

List of Tables

List of Figures

Abbreviations

Abbreviation	Full form
aa	Amino acid
APH	Ampipathic helix
AU	Absorbance unit
BLAST	Basic local alignment search tool
bp	Base pair
CD	Circular dichroism
CV	Coloumn volume
DAB	3,3'-diaminobenzidine
DDM	N-Dodecyl-β-D-Maltopyranoside
DM	N-Decyl-β-D-Maltopyranoside
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
EcTatC	E. Coli tatc
EDTA	Ethylenediaminetetraacetic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria-Bertani media
LDAO	N-Dodecyl-N,N-Dimethylamine-N-Oxide
MtTatc	M. Tuberculosis tatc
nm	Nanometer
OD	Optical density
PBS	Phosphate buffer saline
PBST	Phosphate buffer saline Tween20
PCR	Polymerase chain reaction
PDB	Protein data bank
PIC	Protease inhibitor cocktail
PMSF	Phenylmethylsulfonyl fluoride
RMSD	Root mean square deviation
rpm	Revolutions per minute
SDS	Sodium dodecyl sulfate
***T*m**	Melting temperature
TM	Transmembrane
TMD	Transmembrane domain
TMH	Transmembrane helix
wt	Wild type
β-ME	Beta-mercaptoethanol
µl	Microliter

Abstract

The twin-arginine translocase (Tat) pathway exclusively transports folded proteins in an ATP-dependent manner across the bacterial cell membrane, in contrast to the Sec pathway which transports proteins in an unfolded form. Although the Sec pathway is generally predominant in protein translocation in many prokaryotes, the Tat pathway transports almost the entire secretome in halophilic Archaea and it is essential for their viability. The Tat machinery in most bacteria comprised of three distinct membrane protein complexes TatA, TatB and TatC. The substrates of the Tat pathway contain a signal motif with a positively charged N-terminal region, a hydrophobic core, and polar region that contains the cleavage site for a signal peptidase. The canonical twin-arginine (RR) signal in the N-terminal region directly interacts with the TatC receptor. Common Tat substrates include periplasmic enzymes in complex multiprotein respiratory redox systems, bacterial virulence factors, lipoproteins, and proteins involved in maintaining cell-wall integrity. Majority of Tat substrates are complexes containing metal cofactors and many forms oligomeric assemblies. Delivery of these proteins is mediated by special cytoplasmic chaperones (redox-enzyme maturation proteins, REMP) that bind and mask the Tat signal, thereby preventing the futile export of immature protein. The *Eschercihia coli* genome encodes around 31 putative signal peptides containing a twin-arginine motif. The Tat Pathway also plays an important role in bacterial pathogenesis, translocating certain virulence factors, and proteins involved in polysaccharide metabolism and biofilm formation. Apart from its significance as a potential drug target to control pathogenic bacteria, the Tat pathway could also find applications in bioengineering, where its strict quality control can be used to produce complex therapeutic proteins in properly folded states or to rapidly screen for mutations that enhance protein solubility.

In this thesis various structural and biochemical aspects of the Tat pathway has been explored. Studies were designed to shed light on the interaction between the Tat signal sequence and its interaction with REMP chaperon and Tat machineries. Further, a Novel High-throughput Twin Arginine Translocase Assay has been developed based on colonic acid quantification which is directly proportional to Tat activity. This assay can be used for Therapeutic Applications. A detailed analysis of the Haloacheal Tat pathway has been done to get insights into its diversity in Haloarchaea which almost exclusively use the tat pathway for protein secretion. In the final part, efforts were made to get solve the structure of the MtTatC and the *E.coli* Tat membrane complex.

Thesis Overview

Chapter 1: Introduction

This chapter introduces the requirement of protein translocation for cells and the two major translocation pathways; the Sec Pathway and the Tat pathway. It highlights the salient features of each pathway and the differences. A comprehensive study of the components of the Tat pathway, their structural and functional significance and mechanism of transport have been provided. The diversity of Tat substrates and their functionality across prokaryotes have been reviewed and the roles of Tat pathway in pathogenesis were explained. Information about potential biomedical and biotechnological applications of the Tat pathway and current status of the research has been discussed.

Chapter 2: Tat Signal chaperon interaction

This chapter deals with the interaction of the Tat signal and chaperon. The Dimethylsulphoxide reductase, a heterotrimeric periplasmic complex that uses dimethylsulphoxide (DMSO) as an electron acceptor during anaerobic respiration is one of the complexes translocated by the Tat pathway. The central protein, DmsA, is synthesized in the cytoplasm, loaded with its catalytic molybdopterin (MoPt) co-factor and the accessory proteins DmsB and DmsC. This enzyme complex is an classical example of Tat substrates that must be fully assembled prior to translocation to the bacterial periplasm. Here signal sequence plays many roles, which is more than a simple targeting. Our specific goals in this study were the biochemical characterization of DmsD and the signal sequence of DmsA (ssDmsA) complex, also the complex with translocase assembly and to map the location of the binding site of TatBC and DmsD in the DmsA signal sequence, thereby deriving the smallest length of sequence essential for a specific interaction. Further, chimeric ssDmsA and DmsD fusion (DDF) was prepared for structural studies using x-ray crystallography to determine the atomic level interactions. It was observed that the minimal DmsA signal peptide consists of 29 amino acids. The DmsD/ssDmsA interaction analysis showed the presence of two forms of the complex. And equilibrium of these forms can be modulated by N-terminal truncation of the signal sequence. The failure of the complex

crystals to diffract and the lowering of melting temperature of equimolar DmsD and ssDmsA attribute the destabilization of the complex in this conformation.

Chapter 3: Development of a Novel High-throughput Twin-Arginine Translocase Assay in Bacteria for Therapeutic Applications

A high-throughput assay was developed to study the Tat activity in *E .coli* based on colanic acid quantification. The Tat Pathway plays an important role in bacterial pathogenesis translocating certain virulence factors, and proteins involved in polysaccharide metabolism and biofilm formation. Apart from its significance as a potential drug target, the Tat pathway could also find applications in bioengineering, where its strict quality control can be used to produce complex therapeutic proteins in properly folded states or to rapidly screen for mutations that enhance protein solubility. In this context, the development of efficient, quantitative and robust screening technique for detecting the protein translocation activity of the pathway was taken up. Colanic acid is an extracellular capsular polysaccharide produced by most *E. coli* strains as well as by other Enterobacteriaceae, and has significant roles in biofilm formation, protection against dessication, and adhesion of pathogenic bacteria to host membranes. It contains units of L-fucose, D-glucuronic acid, D-glucose, D-galactose, and pyruvate. The capsular polysaccharide gene cluster (wca operon) directs colanic acid biosynthesis. Colanic acid biosynthase (WcaM) is an essential enzyme involved in this pathway and is also translocated to the periplasm via the Tat pathway. The activity of the Tat pathway was measured by quantifying L-fucose one of the monomers of Colanic acid using L-Fucose. The spectrometric assay was standardized for large no of sample sizes and made compatible with a 96-well format microtiter plate.

Chapter 4: Comprehensive analysis and Insights Into The Haloarchaeal Twin-Arginine Translocase Pathway

The Haloarchaeal Tat pathway was analyzed to study the evolution of the Tat pathway in the adaptability of these organisms to survive in high salinity conditions. The high intracellular salt concentrations cause rapid folding of proteins and the quality of protein folding is a major concern for these organisms. This is positively facilitated by the sophisticated Tat pathway components and almost the entire secretome is exported by the

Tat pathway. The presence of unique TatC topologies and universal presence of two different homologs is seen with some organisms having multiple homologs of TatA. The Haloarchaeal Tat pathway is a TatAC system. These Tat sequence analysis of TatC and TatA revealed that the 14TM TatC has two copies of TatC and present in haloarchaea along with other homologs. Based on this the substrates of 20 Haloarchaea with complete proteome were analyzed. The Haloarchaeal Tat protein proofreading chaperons were identified.

Chapter 5: Structural Investigation of Tat Membrane complex

This chapter deals with the structural analysis of the Tat membrane receptor complex. Efforts were made to determine the structure of MtTaC using x-ray crystallography. The *Mycobacterium tuberculosis* Tat pathway exported proteins, including virulence factors of periplasmic substrate-binding proteins of ABC transporters and Lactamases but is not essential for survival. Thus developing a MtTat inhibitor would ideally provide a drug that will make the organism non-virulent, at the same time there is no survival pressure diminishing the possibility of resistance. Heterologous expression of MtTatC was achieved in *E.coli* by swapping of the c-terminal cytoplasmic tail with an *Aquifex* TatC tail. Absence of crystal hits due to low quality and limited protein yield is the bottleneck. The MtTatC structure was modelled using the AaTatC structure for analysis. The advancement of Electron Microscopy in studying protein complexes at cryo-temperatures and determination of structures reaching near atomic resolution makes the *E.coli* Tat membrane complex a good target. Formation of EcTatBC complex and EcTatBC with signal and sequence having molecular weight sufficient for cryo-EM studies was achieved. Formation of stable and homogeneous EcTat memberane complex was studied.

Chapter 6: Summary and conclusions

The salient finding of the present work and conclusions were drawn from the data are put forth and future prospects are outlined in this chapter.

Chapter I - Introduction

Cell is the smallest functional unit of life which contains all the fundamental molecules; Nucleic acids, protein, carbohydrates and lipids. They are referred to as building blocks of life as they can either form an organism by itself such as bacteria, archaea and protozoa or organize themselves into more complex organisms such as sponges, fungi and highly complex life forms such as plants and animals. Cells are discrete and easily recognizable as they are surrounded by a membrane which separates the inner parts of the cell from the environment. Based on how the genetic material in a cell is packaged the cells are classified as Prokaryotic and Eukaryotic. In prokaryotes the DNA is in the cytoplasm whereas in eukaryotes it is separated by a membrane bound nucleus. Apart from nucleus there are many other organelles present in eukaryotic cells such as mitochondria, chloroplast, endoplasmic reticulum, lysosome, Golgi body, peroxisomes etc. which are membrane bound and perform a wide variety of function. For the functioning of the cell there is the need to transport different molecules to different parts of the cells for different functions and there are numerous ways of transport of materials in the cell either passively in an energy independent way or actively by utilization of energy either from ATP, other high energy molecules of by generation of ion gradients.

Proteins are the most abundant macromolecules present in the cell and perform diverse functions such as enzymes, transporters, storage molecules, structural components and messengers [1]. They are synthesized by the ribosomes based on the coding sequence on mRNA, which is transcribed from the DNA. This is the central dogma that was proposed by Francis Crick [2]. Protein synthesis is mainly carried out in the cytoplasm and in mitochondrial matrix and Chloroplast. Transport of these synthesized proteins to their target location is very important and multiple protein targeting pathways are operational inside the cell that ensure the precise targeting of newly synthesized proteins.

1.1 Protein Translocation

Protein translocation is a process by which proteins get transferred between various cellular compartments. Transport within the cytoplasm is much easier and straight forward than through the lipid membranes because of their hydrophobic core. Efficient translocation across the membrane is imperative for transport to different organelles and outside the plasma

membrane [3]. This is carried by different protein translocation pathways such as the general secretory pathway or the Sec Translocase pathway and the Twin arginine translocation (Tat) pathway (Fig.1.1). Protein translocation is very crucial for a cell and defects can lead to serious physiological and functional impairment for the cells. A functional Sec is necessary for cell viability and in some Haloarchaea a functional Tat pathway is also obligatory.

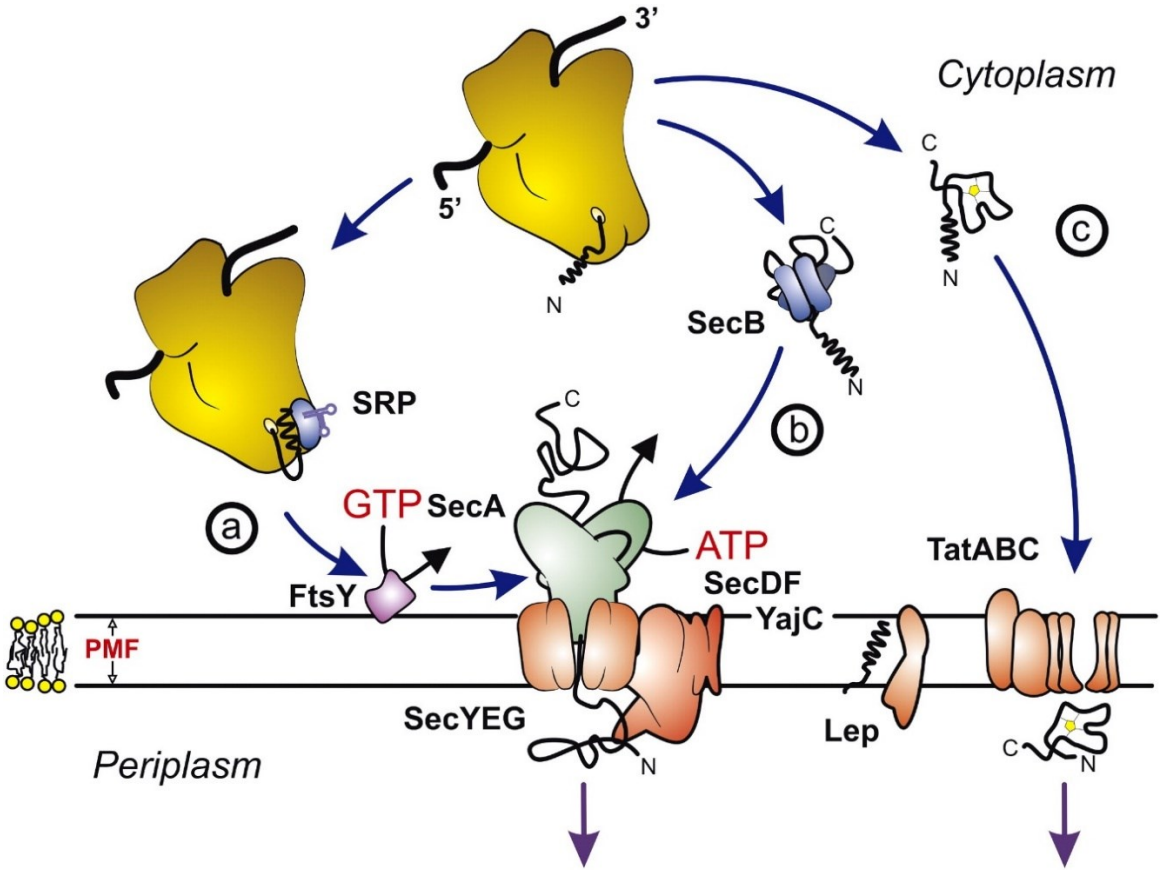


Fig1.1. Schematic overview of the Escherichia coli Sec- and Tat translocases. (a) Co-translational and (b) post-translational targeting routes and translocation of unfolded proteins by the Sec-translocase. (c) Translocation of folded precursor proteins by the Tat translocase.[4]

Proteins are polypeptides of amino acid residues and have vast sequence and structural diversity. It is a great challenge for the Translocase pathways for proper identification and targeting of the proteins to be translocated. A protein targeting signal sequence was discovered by Günter Blobel (awarded Nobel Prize in 1999) that determines the final location of a newly synthesized protein in the cell [5]. This signal peptide is located on the Amino-terminal of the

precursor protein and tripartite in nature (Fig1.2). It has a positively charged N terminal-region followed by a hydrophobic H-region at the center and a polar C-terminal region that has the signal peptidase cleavage site for the release of the mature protein [6].

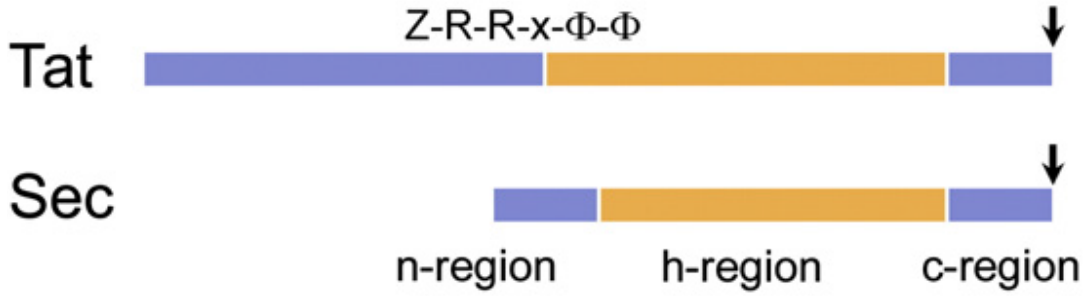


Fig.1.2. Schematic comparison of typical N-terminal signal sequences from substrates of the Tat and Sec systems bearing the polar n-regions with a positive net charge, the uncharged hydrophobic h-regions and short polar c-regions which determine a type I signal peptidase cleavage site (typically AxA).

The major route for protein translocation in the cell is the Sec pathway which transports proteins in unfolded form when compared to the Tat pathway which transports completely folded proteins and protein complexes.

1.2 Sec Translocase pathway

The Sec pathway is the most prevalent universal protein translocation pathway found in the prokaryotes and eukaryotes. It is present in the cytoplasmic membrane of all bacteria, archaea, the thylakoid membrane of plant chloroplasts and Endoplasmic reticulum and is essential for the cell viability [4]. It transports proteins responsible for metabolism, substrate uptake and excretion, membrane fusion, cell envelope structure, sensing and cell communication[7][8]. The SecA is the motor protein which pushes the unfolded polypeptide by ATP hydrolysis [9]. The pore is formed by SecY, SecE and SecG (SecYEG) (Fig1.1) conducts the poly polypeptide to the periplasm or laterally releases the trans-membrane domains into to the phospholipid bilayer[10][11]. The SecA interctswith the SecYEG where it accepts unfolded from the ribosome (Fig1a) or the SecB chaperon.

SecB is utilized in post translational transport where it keeps the preprotein in an unfolded state[12]. Co-translational targeting encompasses the signal recognition particle (SRP) to binds to the amino terminal signal sequence of the nascent secretory protein as it emerges from the

ribosome and the total complex of is targeted to the Sec-translocase[13]. In bacteria, the Co-translocation is extensively used for the integration of integral membrane proteins into the cytoplasmic membrane, while only limited secretory proteins are translocated by this route. In the ER, almost all secretory proteins are targeted co-translationally[10].

The signal sequence (signal peptide) obligatory for accurate protein targeting is present on the N-terminal extension of the secretory protein. The signal has an average length of 20 amino acid residues with a tripartite structure as described above (Fig. 1.2). Programs like SignalP[14] can predict the signal sequence based on a hidden Markov models. The c-region comprises the signal peptidase cleavage site which is recognized by the type I signal peptidase[15]. For recognition positions − 1 and − 3 with respect to the signal peptidase cleavage site are occupied by non-bulky polar amino acids which are recognized by the signal peptidase binding pocket. A lipoprotein box L[AS][GA]C at the − 3 to + 1 position is present in certain substrates. The lipid modified takes place at cysteine prior to translocation and type II lipoprotein signal peptidase cleaves off the signal sequence.

The hydrophobic amino acids of the h-region have a high propensity to form an α-helix and the charged residues of the n-region have been involved in electrostatic interactions with membrane phospholipids. Both the h- and n-regions are critical in recognition by SRP [16] and SecA[17]. SRP and SecA binding affinity is directly proportional hydrophobicity of the h-region and the number of positive charges in the n-region respectively. The bacterial Sec-signal sequences are functionally interchangeable with thylakoid membrane and the ER signals[9].

1.3 Tat Translocase pathway

The twin Arginine Translocase pathway transports folded proteins across the membrane was identified by Ben Breks[18]. The name of the pathway comes from the invariant twin arginines in the consensus S/T-R-R-x-F-L-K motif at the interface of the n-terminal and hydrophobic region [19][20]. Tat pathway is broadly conserved in bacteria (Fig.1.3) [21], archaea [22], chloroplast [23]and mitochondria of certain organisms [24]. A few obligate fermentative or parasitic prokaryotes whose metabolism does not require the transport of cofactor-containing proteins lack the Tat Pathway [25]. Tat pathway normally transports complex redox proteins and protein complexes. Tat substrates are important for cellular processes, including respiratory and photosynthetic energy metabolism, iron and phosphate acquisition, cell

division, cell motility, quorum sensing, organophosphate metabolism, resistance to heavy metals and antimicrobial peptides and symbiotic nitrogen fixation[26]. It has been found to be required for virulence of most animal and plant pathogens[27].

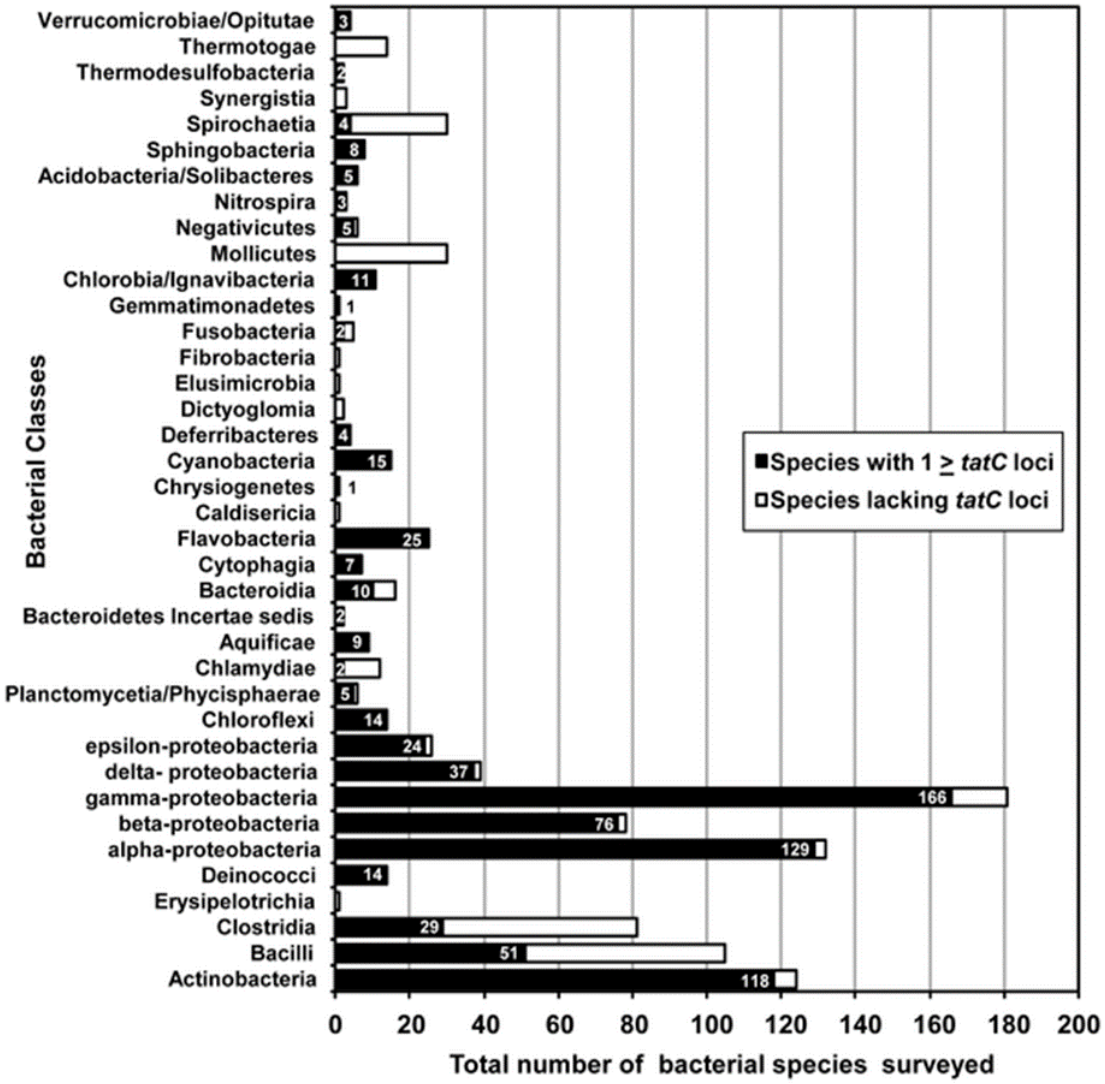


Fig1.3. The bar plot distribution of tatC sequences in sequenced bacterial genomes[21]

Tat transport machinery consists of integral membrane proteins from two structural families, the single transmembrane TatA and the six transmembrane TatC. The minimal Tat systems contain only TatA and TatC i.e TatAC systems which are found in the Firmicutes (low

G + C gram-positive bacteria) and in archaea[28]. The TatABC (Fig 1.4) [29]systems is found in most gram-negative bacteria (*E.coli*) and plants, possess TatB in addition to TatA and TatC [30]. TatB is a functionally distinct member of the TatA family. *E. coli* possesses a TatA paralog called TatE, which has very low abundance and appears to be functionally interchangeable with TatA.

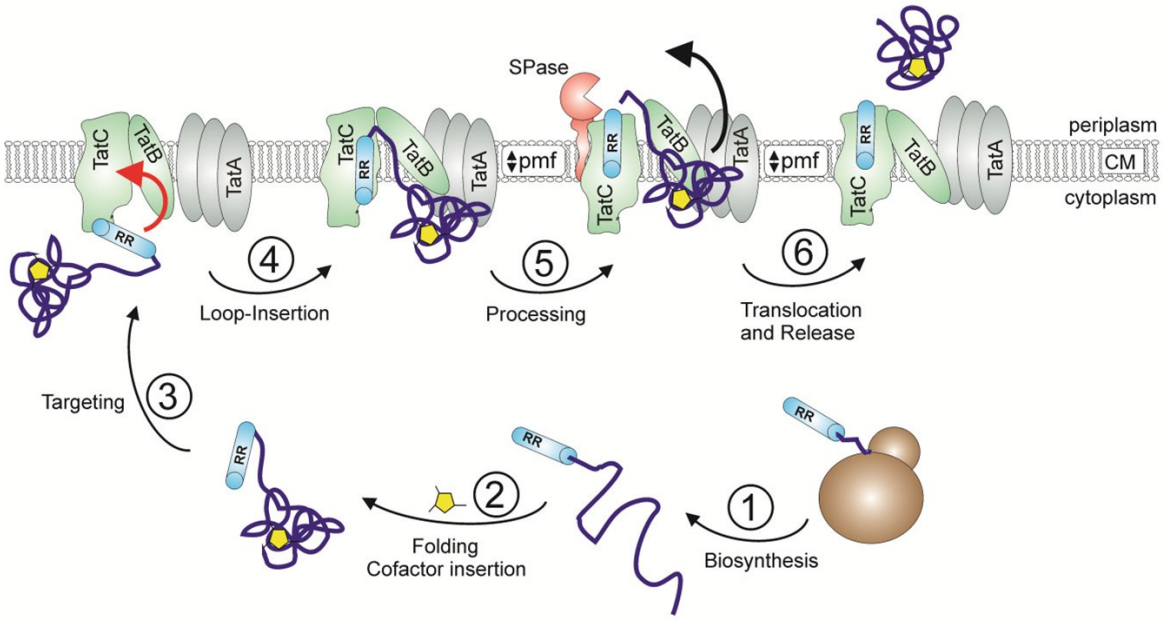


Fig. 1.4. Overview of the Tat-dependent pathway for the protein translocation in bacteria and steps involved in Tat translocation which is a PMF driven transport [29]

TatC is the receptor component that recognizes the characteristic RR-signal [31]. TatC or the TatBC complex interacts with the substrate and recruits TatA which forms an appropriately sized pore through which the substrate is transported across the membrane [32]. The Tat transport uses proton motive force (PMF) for substrate transport [33][34].

Delivery of these proteins is monitored by a special class of cytoplasmic chaperones that bind specifically to the Tat signal of their substrate protein masking the signal sequence thereby ensuring proper maturation and co-factor loading. These chaperones, dubbed redox-enzyme maturation proteins (REMP), prevent the futile export of immature protein [35][36].

There are studies on cross-species complementation which is shows the flexibility of use of Tat components [37][38] and have been listed out in Table1.1.

Table1.1 Interchangeability and functionality of Tat components amongst bacteria[37][38]

S. No	Organism1	Organism2	Function
1	*E.coli* TatBC	*P. syringae* TatA	++
2	*E. coli* TatBC	*S. coelicolor* TatA	++
3	*E. coli* TatBC	*A. aeolicus* TatA	+
4	*E. coli* TatAC	*P. syringae* TatB	+
5	*E. coli* TatAC	*S. coelicolor* TatB	++
6	*E. coli* TatAC	*A. aeolicus* TatB	-
7	*E. coli* TatAB	*P. syringae* TatC	-
8	*E.coli* TatAB	*S. coelicolor* TatC	++
9	*E. coli* TatAB	*A. aeolicus* TatC	-
10	*E.coli* TatABC	*P. syringae* TatABC	+
11	*E. coli* TatABC	*A. aeolicus* TatA1A2BC	-
12	*E. coli* TatBC	*H. pylori* TatA	++
13	*E. coli* TatAC	*H. pylori* TatB	-

1.3.1 Tat Signal Peptide

Tat-targeting signal peptides share over all similarity, but there are noticeable and significant differences that ensure targeting to the correct pathway and avoiding mistargeting (Fig1.5). The Tat signal peptides contain a consensus Ser/Thr-Arg-Arg-X-Phe-Leu-Lys (where X is any polar amino acid) motif. The twin arginines predominantly invariant, greater than 50% frequency is seen in case of other motif residues and a polar amino acid is found at the X position with a few exceptions [39].

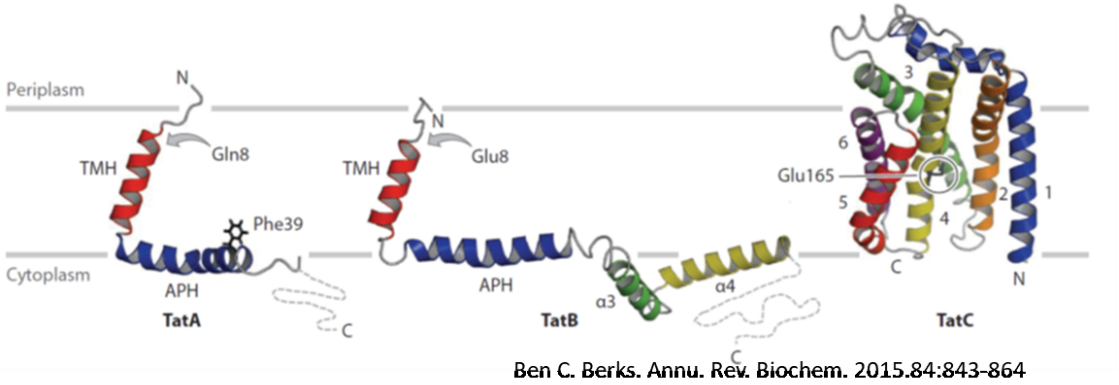
Fig.1.5. Features of a typical Tat signal peptide, ssTorA, from E. coli. The n-region is shown in pink, the h-region in blue, and the c-region in yellow. The consensus Tat motif is boxed. The vertical dashed line indicates the site of cleavage by signal peptidase I. Charged residues in the c-region and mature proteins are marked. [40]

The invariant twin arginines are crucial for Tat signal peptide function, mutation of the other conserved residues generally results in little or no translocation defect[41]. The twin-arginine motif does not prevent Sec-targeting. Some Sec-signal sequences also have twin arginines like in the case of *E.coli* d-alanyl-d-alanine carboxypeptidase (DacD). Sec avoidance is demonstrated by subtle differences between the Tat and sec signal peptide and features of the Tat substrates. The Tat signal peptide have relatively longer h-regions that has lower hydrophobicity than that of Sec-signal [42]. Tat substrates can possibly routed via the sec pathway by incrementally enhancing the hydrophobicity of the h-region. Furthermore there are basic amino acid residues in the c-terminal region of the Tat signal sequence for sec avoidance which do not have any direct role in recognition by the tat machinery [43][19].

Bioinformatics programs such as TATFIND[44], TatP[45] and PredTat[46] can be used to predict Tat signal peptides. A unique characteristic of the Tat pathway is piggyback transport, it can transport proteins without a functional Tat Signal. This phenomenon termed as hitch-hiking is orcastrated by complex formation of these proteins with Tat signal containing proteins [47].

1.3.2 Tat Structure

Structural studies of the Tat components throws light into its mode of function. Atomic resolution structures (Fig1.6) of TatA from *E. coli* [48], TatAd *B. subtilis*[49] and *E. coli* TatB[50] have been solved by NMR . X-ray crystal structure for TatC have been determined for *A. aeolicus* [51][52].



Ben C. Berks. Annu. Rev. Biochem. 2015.84:843-864

Fig.1.6. Structures of the Tat components. Solution NMR structures of the Escherichia coli TatA and TatB proteins and the highestresolution X-ray structure of Aquifex aeolicus TatC are shown as ribbons. The natively unstructured C-terminal tails of TatA and TatB are represented by dashed lines. The transmembrane helices (TMs) and amphipathic helices (APHs) of TatA and TatB are indicated, together with the TM numbering in TatC. Selected *E.coli* residues are depicted. The TatA and TatB TM polar residue is on the far side of the helix in this view [53]

1.3.2.1 TatC Structure

Three identical structures for *Aaquifex(Aa)* TatC were solved in different crystal environments (Fig1.6). This implicates structural rigidity and integrity of TatC. The Six transmembrane domains are arrange with N-terminal and c-terminal of the protein facing the cytoplasmic side and forms a glove like structure[51]. This kind of fold was not previously reported. The "palm" of the hand is formed by the 6TM helices which are kinked due to the presence of conserved proline residues which forms a cavity. This overhangs on the periplasmic side of the membrane by a rigid periplasmic cap. The last two TMHs of TatC are too short to completely span the membrane, resulting in distortion of the membrane and thinning the membrane bilayer. The cytoplasmic loops are very small and molecular dynamic experiments suggest that a very little part of the TatC exposed to the aqueous surroundings [52][51].

The AaTatC TM1 goes in perpendicular to the membrane, and continuing as an amphipathic helix -1A arching under TM2 and constituting a large part of the periplasmic face of TatC. The first periplasmic loop (Per1) continues to TM2, which at the periplasmic end is angled relative to TM1 till the conserved Pro78 residue which generates a kink and the rest of the TM is

parallel to TM1. The short cytoplasmic loop (Cyt1) connects TM2 to TM3. TM3 is angled steeply and contacts TM2, TM4, and TM6 from behind. Periplasmic loop (Per2) connecting TM3 and TM4 is below TM5 and TM6. TM4 is parallel to TM2, is conserved proline kink Pro167 is conserved. The highly conserved glycine161 of TM4 and Gly114 in TM3 forms a tight interface. TM4 is connected to TM5 via a short loop (Cyt2), TM5 moves out from the core of the protein with a steep angle and then contacts to TM4 by kinking sharply. Periplasmic loop (Per3) post TM5 has a highly conserved Pro242 turn that lies in the hydrophobic core of the bilayer. TM6 has is slightly angled and contacts to TM3, TM4, and TM5 on the back. The C terminus of the protein extends into the cytoplasm (Fig1.7).

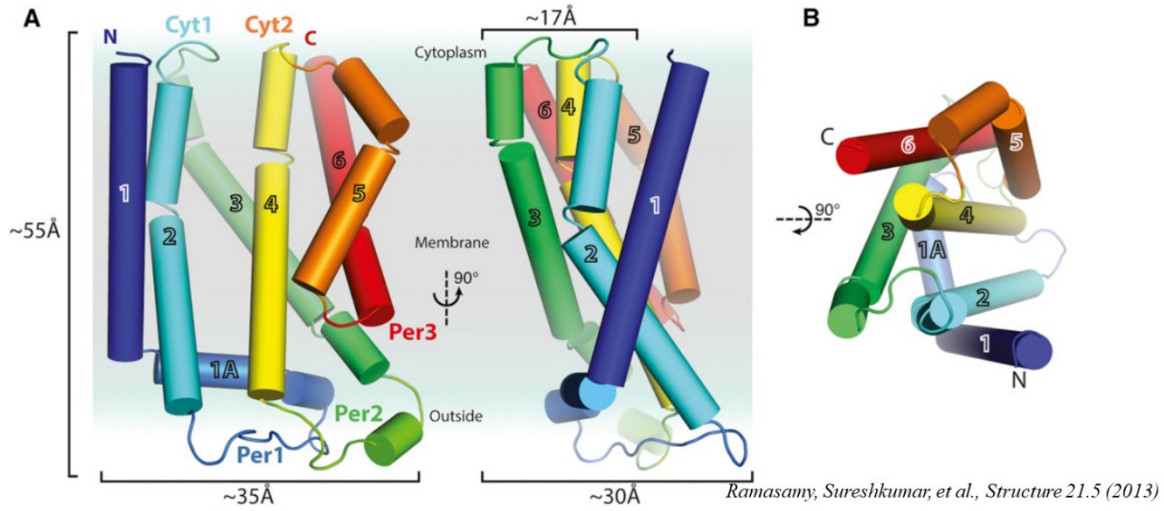
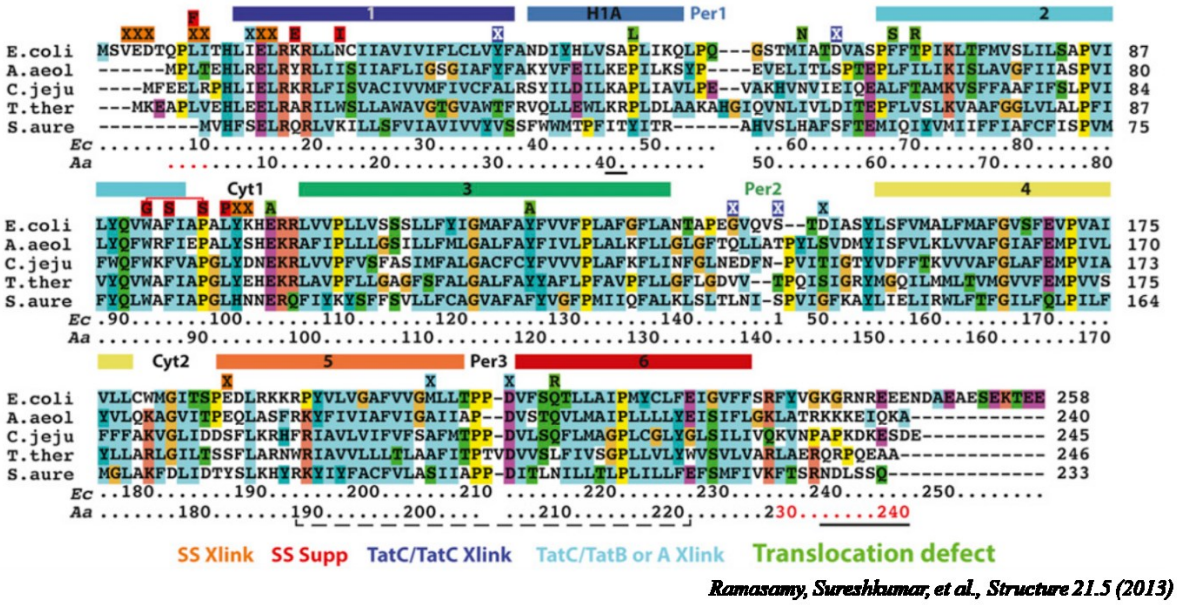


Fig.1.7. Overview of the Structure of AaTatC (A) Cartoon diagram of AaTatC viewed from in the plane of the membrane color-ramped from the N to the C terminus (blue to red) from the front and rotated 90°. TMs and loops are numbered (1–6) according to the text. The amphipathic helix in Per1 is labeled as 1A. The bilayer is represented in the back as a rectangle with the most hydrophobic portions in gray and the hydrophilic head groups in blue. Dimensions are shown for the whole protein and the narrowest point in the glove. (B) As in (A), viewed from the cytoplasm. [51]

The glutamate residue Glu165 in the TM4 of *Aa*TatC lies in the middle of the cavity exposing the polar side chain to the hydrophobic interior of the membrane bilayer. This residue is strictly conserved as either a glutamate or a glutamine in the TatC molecules of other organisms and replacement of this amino acid significantly affects Tat transport in *E. coli* [54]. Extensive site-specific mutants and random mutagenesis studies of TatC has given insights into the residues of TatC that are involved in binding to signal sequence and/or other TAT

components, and transport. These residues have been marked on the sequence alignment (Fig1.8) of TatC homologs of *E. coli, A. aeolicus, Campylobacter jejuni, Thermus thermophilus,* and *Staphylococcus aureus*. These residues have also been marked on the AaTatC structure (Fig1.9).



Ramasamy, Sureshkumar, et al., Structure 21.5 (2013)

Fig.1.8. Sequence alignment of TatC homologs. The species are *E. coli*, *A. aeolicus*, *Campylobacter jejuni*, *Thermus thermophilus*, and *Staphylococcus aureus*. Secondary structure features are highlighted above the sequence with helices as rectangles. Sequence numbering is below the alignment for *E. coli* and *A. aeolicus*. Residues that are disordered in the crystal structure have red numbering. Residues mutated in the AaTatC are underlined in black. X indicates residues involved in crosslinks, and other letters indicate specific mutations. The dashed line marks a salt bridge between K190 and E221. [51]

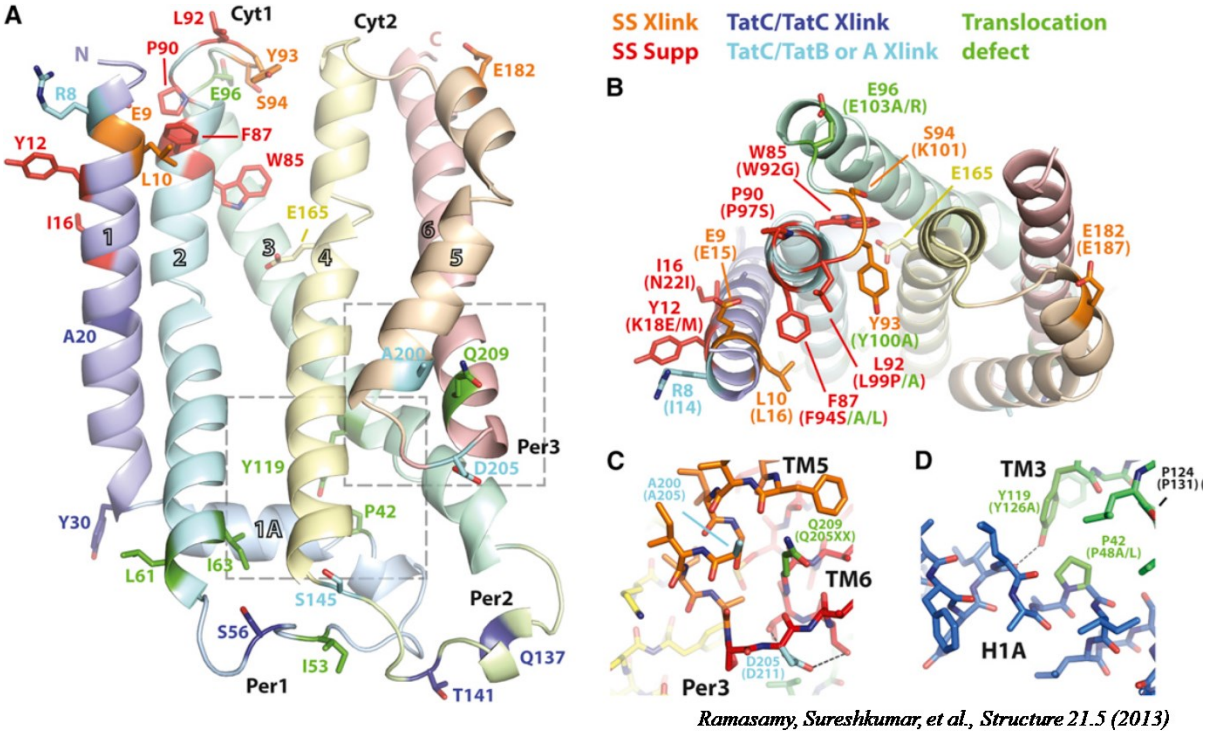


Fig.1.9. Residues Correlated to TatC Function Based on Mutagenesis or Crosslinking (A) Ribbon diagram of AaTatC, with discussed mutants shown as sticks. Residues are colored based on identified function with *A. aeolicus* numbering: signal sequence crosslink (orange), signal sequence suppression mutants (red), TatC to TatC crosslink (blue), TatC to TatA or TatB crosslink (cyan), and mutants that inactivate TatC (green). Dashed boxes indicate regions highlighted in (C) and (D). (B–D) View from the cytoplasm as in (A). Labeling includes *A. aeolicus* numbers with E. coli numbering and mutations in parentheses. (C) and (D) have similar schemes. (C) Per3 in sticks highlighting the conserved hydrogen bond network stabilized by the conserved D205. (D) Sticks view of the hydrogen bond stabilizing the kinks in TM3 and 1A facilitated by P42 and Y119 [51]**.**

1.3.2.2 TatA and TatB Structure

Solution NMR structures have been solved for each subclass of TatA proteins (Fig1.6). The structure of *Bacillus subtilis* TatAd (BsTatAd) [49] is representative of TatAC system and the *E. coli* TatA (EcTatA) [48][49] and TatB (EcTatB) [50] structures provide models for the TatABC systems. All three TatA family proteins have a single transmembrane core structure of a short hydrophobic N-terminal TM helix starting from the periplasmic side followed by an amphipathic helix (APH) approximately at right angle and opening up to the periplasmic side

(Fig1.10) followed by an unstructured loop in case TatA. In TatB there are one or two helices present between the amphipathic helix and the unstructured loop.

Analysis of that the N-terminal hydrophobic helix suggests that the TM is too short to fully span the bilayer and pulling the proximal end APH into the membrane bilayer [55]. The junction between the TM and the APH is called as the hinge and contains an invariant glycine residue. The right angled interhelix orientations for all three structures are due to hydrophobic stacking interactions around the helix junction made up of a short (three- to five-residue) loop. Both TatA and TatB molecules have an unstructured highly charged C-terminal tail. This tail is not important for either Tat transport [56] or substrate induced oligomerization [57]. Its removal only impedes Tat transport activity.

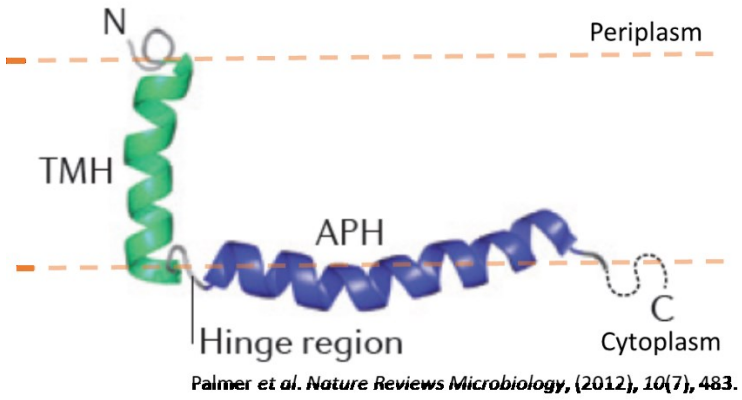


Fig.1.10 Structure of a monomer of the Bacillus subtilis twin-arginine translocation (Tat) pathway protein TatAd in micelles of the detergent dodecyl-phosphocholine (DPC).The transmembrane helix (TM), amphipathic helix (APH), hinge region and protein amino and carboxyl termini are indicated. The 22 amino acid C-terminal tail of the protein is unstructured and represented as a dotted line [26]

The N-terminal of the TM in TatA and TatB is generally a polar residue. Normally a glutamate, glutamine, histidine, or lysine. It is functionally important in chloroplast TatA, chloroplast TatB, and EcTatA[58]. This residue is demonstrated to mediate TatB TatC interaction and substrate-induced oligomerization of TatA pore onto theTatBC complex [59]. But this conserved polar amino acid isn't obligatory for Tat transport as BsTatAd lack it and EcTatB can tolerate this substitution [60].

The distinctive property of TatB compared to TatA is, TatB forms equimolar complexes with TatC, whereas TatA oligomerize and increase association with TatC on substrate binding to TatC. Both proteins have very high sequence similarity until the APH. The most significant TatA-specific sequence feature is the conserved phenylalanine residue at the APH C-terminus which protrudes into the lipid bilayer (Figure 1.6). This functionally essential *E. coli* TatA residue not well conserved in TatB proteins [61]. The periplasmic N-terminal tail before of TatB is important for activity [61]. In *E. coli* the entire TatA APH is sensitive to amino acid substitutions [62], but these substitutions have no effect in TatB functioning [58].

TatB has a longer cytyoplasmic tail and *E.coli* TatB contains two helices between the APH and the natively unstructured tail (Figure 1.6). These helices are linked via highly flexible linkers. These helices have highly polar surfaces and are linked to one another, and to the APH, through flexible linkers. Though deletion of these helices reduces the speed of Tat transport [56], they contain no conserved amino acids.

1.3.3 TatBC and signal binding

The TatBC complex comprise of equimolar ratio of the TatB and TatC proteins and has a size distribution from ~400 kDa and ~700 kDa [63]. In the *B. subtilis* TatAC system a multi subunit TatAC complex has been isolated that is functional analogous to the E.coli TatBC complex[64]. The TatB TM has been predicted to bind near the periplasmic end of TatC TM5 and TM6 (Fig1.11) based on the results of cross-linking studies and genetic analyses [51][65][66]. The structural model places the TatB TM right opposite to the signal peptide n-region-binding site of TatC (Figure 1.11). But there is no interactions between TatB TM and signal peptide have been observed. TatB may exert its functional effects through a different part of the signal peptide, perhaps by controlling access from the TatC central cavity to the periplasm through the adjacent cap/TM5 gap [53].

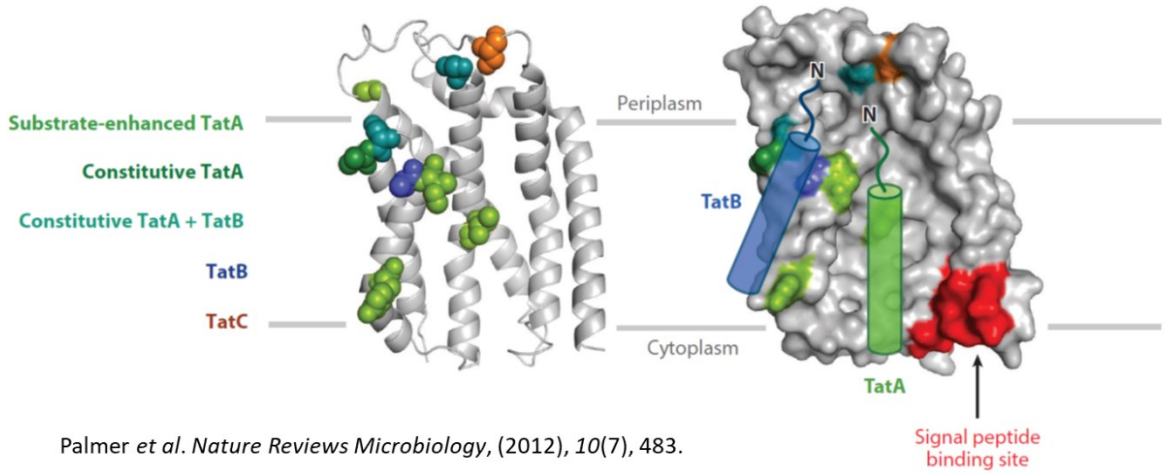


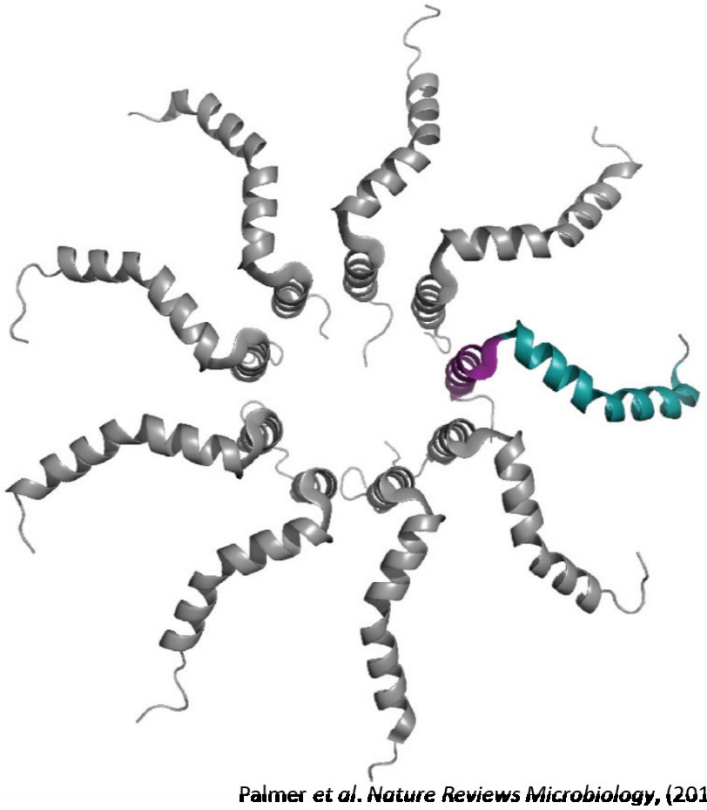
Fig.1.11 TatC contacts with the transmembrane helices of partner proteins and with the n-region of the substrate signal peptide. (Left) Cartoon representation of the Aquifex aeolicus TatC protein. (Right) Surface representation of the same protein. Positions in the transmembrane or periplasmic parts of TatC that cross-link to other Tat components are colored as indicated. Those TatA contacts that are significantly enhanced in the presence of both a substrate and a proton motive force are labeled substrate enhanced. Contact sites for the TatA and TatB amphipathic helices on the cytoplasmic face of TatC are not shown as it is not clear that these represent defined binding sites. In the structure on the right, the inferred positions of the transmembrane helices of TatB and of substrate-enhanced TatA are indicated by cylinders. The signal peptide n-region-binding site is colored red [26]**.**

1.3.4 TatA –TatC interactions

Cross-linking studies suggest that the constitutively bound TatA molecules are positioned close to the TatC-TatB TM binding site (Fig1.11) [67][65]. Additional cross-links form between TatA and TatC during substrate transportation and are possibly due to the contacts between the assembled TatA oligomer and TatC [67]. It starts with more cross-linking or interaction between the TatA TM and TatC near TM5. After this event the interaction of TatA TM close to the conserved glutamate or glutamine residue is the more probable occurrence. Thus, it can be determined that TatA oligomerization takes place on the cavity side of TatC and extends from the TatB TMH-binding site at TM5 into the cavity [53].

1.3.4 TatA oligomer

The final transport of the substrate needs the oligomerization of the dynamic size and number of TatA pore [68]. The experimental observation suggest that each translocation site requires ~25 TatA subunits. This high TatA: TatC ratio requirement is a major rate limiting factor for Tat Transport [69] perse it may be helping for the functional accuracy with less error in translocation. TatA oligomerization requires the TatA TM and APH but C-terminal tail and the functionally essential phenylalanine residue at the end of the APH is not essential. TatA can assemble into large oligomers in native membrane without the requirement of substrate binding [70]. Single-particle electron microscopy of TatA solubilized in different detergents display ring like structure with varying sizes, due to the variation in the number of TatA molecules in these complexes [68]. This variation may also be the different stages of TatA oligomerization and the larger complexes have internal cavities wide enough to accommodate folded protein substrates [68]. By varying the TatA to detergent ratio leads to TatA oligomers small enough for solution NMR methods were produced for determination of individual subunit conformation and inter subunit contact identification (Fig1.12) [48]. Only TM to TM contact is very conspicuous in the small TatA oligomers. The spin labelled oligomers of large detergent extracted TatA oligomers also show similar contacts [71]. So it can be inferred that the subunit arrangement is similar in detergent-solubilized TatA oligomers of different sizes.



Palmer *et al*. *Nature Reviews Microbiology*, (2012), *10*(7), 483.

Fig.1.12 Model for the TatA oligomer in detergent solution, based on NMR and spin-labeling data [26]

MD simulations using TM-TM interaction in the bilayer suggest that oligomers larger than a tetramer will be structurally unstable. The physiological TatA oligomer might be stabilized through interaction with TatC molecules, membrane lipids and inter APH contacts.

A zipper model was proposed by Walther et al. [72] which is based on the appropriate placement of charged residues along the TatA molecule. Basic residues situated in the terminal of APH interact with acidic residues in the proximal C-terminal tail of an adjacent TatA molecule to form a ladder of salt bridges between the TatA subunits (Fig1.13). The proof was based on experimental observation on detergent-solubilized TatA variants but it demands detailed testing in the native membrane environment for the confirmation of this model.

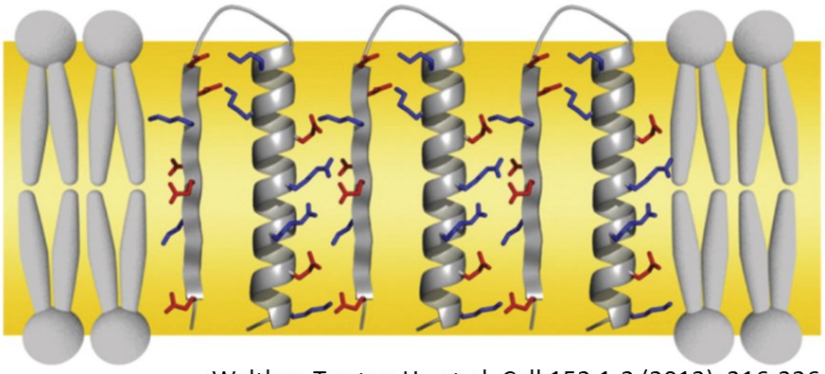


Fig.1.13 Zipper model for TatA oligomerization. Positively charged residues (blue) in the APH of TatA are proposed to form ion pairs with negatively charged residues (red) in the proximal C-terminal tail (PCT) of the same and adjacent TatA subunits [72]**.**

1.3.5 Tat mechanism of transport

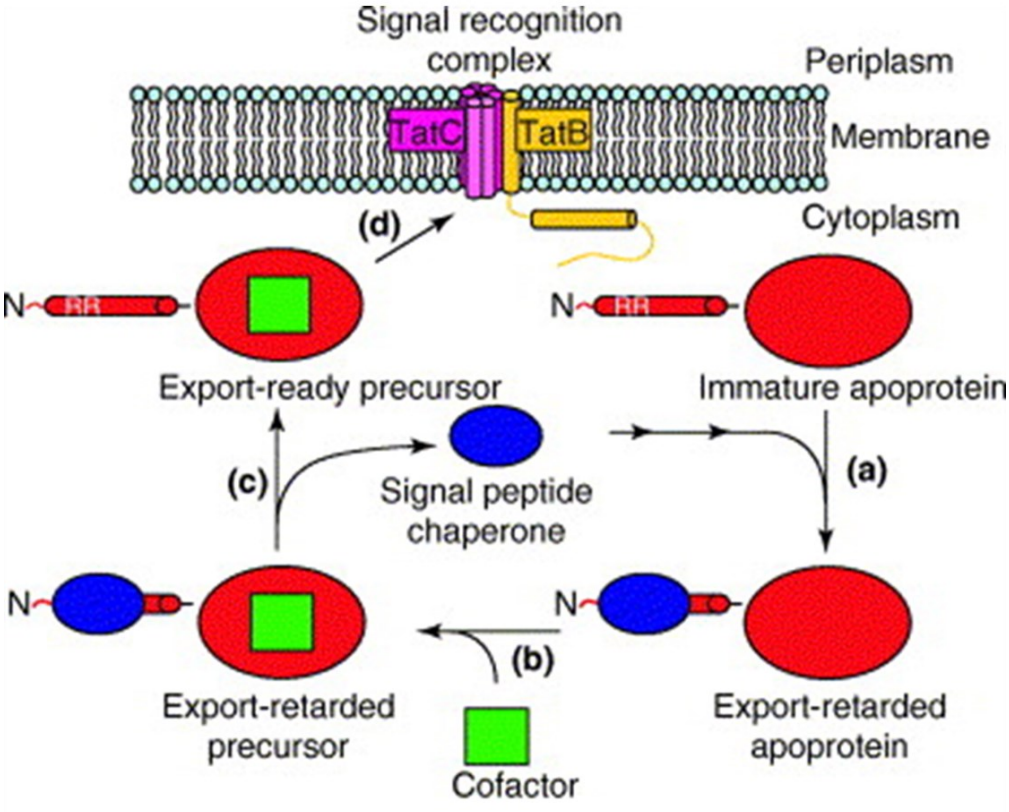
The estimated time taken for the Tat transport cycle is between 1 to 3 min as measured and calculated by kinetic analyses [69] or the time of the first appearance of the transported protein [34][73]. The Tat transport kinetics can be modelled by a two-step mechanism. The first step is TatA oligomerization because its duration is directly proportional to the TatA concentration in the membrane [69] at that time. The second step is the translocation of Tat substrate via the assembled TatABC complex across the membrane bilayer. The TatABC–substrate complex may undergo substantial and rigorous reorganization after assembly of the subunits before transport can take place to transport without any error and with proof reading function [69] and requires about ~1.5 min even if sufficient amount of TatA is present.

A membrane-weakening model proposes that Tat transport may work by local weakening of the membrane due to the presence of large number of TatA [74]. Substrate movement is possibly occurred through transient bilayer rupture or once mechanical force has been exerted on the substrate to conduct it through the weakened bilayer. Here the membrane of the cell remains intact so that it maintains its integrity and impermeable to ions, except during of transport event. Lateral membrane pressure may cause the sealing of the pore by phospholipids packaging around the substrate protein as it moves across the membrane. This mechanism is backed by the observation that transport is strongly influenced by the membrane phospholipid composition, which would be due to the protein–phospholipid interactions [53].

The mechanisms of membrane weakening encompass the TatA APH pulling into the cytoplasmic side of the membrane and the short length of the TatA TM, TM5 and TM6 of TatC. MD simulations of the detergent-solubilized TatA oligomer in a bilayer membrane advocated that the TM length mismatch is capable of locally thinning the bilayer by ~80% and the phospholipids in the thinned membrane show high disorder [48].

1.3.6 Tat Proof reading and quality control

Many of the Tat substrates are cofactor-containing Redox enzymes and these substrates twin-are subject to chaperone-mediated 'proofreading' or error free translocation by Redox enzyme maturation proteins (REMPs) to ensure proper assembled, cofactor loading and any co-exported partner proteins are binding before interaction of the Tat signal sequence with the Tat receptor complex [75]. Here (Fig1.14), the signal peptide of the preprotein is bound by a respective dedicated chaperone, this is followed by cofactor loading on the export retarded apoprotein. One proper folding and cofactor loading is accomplished, the chaperon is released ant the signal peptide is free to interact with the Tat receptor complex. Most Tat substrates have a specific and dedicated chaperon for their export (Table1.2).



Palmer, Tracy, Frank Sargent, and Ben C. Berks., Trends in microbiology 13.4 (2005)

Fig.1.14 A model for 'proofreading' mediated by twin-arginine signal-peptide-binding chaperones. The *E. coli* TatBC signal recognition complex (red and gold) is depicted in its resting state (TatA-dissociated) within the inner membrane. The relative positions of the bacterial periplasm and cytoplasm are shown. (a) When a cofactor-less immature apoprotein (scarlet) is synthesized de novo in the cytoplasm, its exposed twin-arginine signal peptide is recognized and bound by a proofreading chaperone (blue). Binding of the chaperone masks the twin-arginine motif and prevents targeting of the apoprotein to TatBC at this point (i.e. it is an export-retarded apoprotein). (b) The chaperone remains bound to the signal peptide during cofactor loading, preventing any attempted export during this assembly process. (c) Following successful cofactor insertion, the signal-binding chaperone is displaced by an as-yet unknown mechanism, thus revealing the now active twin-arginine signal peptide. (d) The export-ready precursor is now free to associate with TatBC and so enter the Tat transport cycle leading to protein export [76].

Table.1.2 Selected *E.coli* Tat substrates along with Co-exported partner protein and signal chaperon

Protein	Co-exported partner	Signal chaperone
HyaA	HyaB	HyaE
HybO	HybC	HybE
YagT	YagR, YagB	YagQ
YdhX	Unknown	Unknown
TorA	None	TorD
TorZ	None	YcdYb
NapA	None	NapDb
DmsA	DmsB	DmsD
YnfE	YnfG	DmsD
YnfF	YnfG	DmsD
FdnG	FdnH	FdhD, FdhE
FdoG	FdoH	FdhD, FdhE

The Tat pathway has an intrinsic folding quality control mechanism and can differentiate between folded and unfolded substrate proteins. Though smaller unfolded proteins can pass through the Tat machinery [77] [78]. Thus the tat pathway blocks transport very large unstructured proteins or proteins which expose too many hydrophobic stretches.

1.3.7 Tat substrates and function

The Tat pathway normally transports fewer substrates than the Sec system in bacteria [76] but in Haloarchaea almost the entire secretome is transported by the Tat Pathway [79]. Tat substrates have roles in cellular processes such as respiratory and photosynthetic energy metabolism, iron and phosphate acquisition, cell division, cell motility, quorum sensing, organophosphate metabolism, resistance to heavy metals and antimicrobial peptides and symbiotic nitrogen fixation [75].

The Tat system is essential for many redox pathways such as anaerobic respirations with trimethylamine N-oxide, dimethyl sulfoxide, arsenate or chlorinated unsaturated compounds. It is also indispensable for plant nitrogen fixation as nitrous oxide reduction depends on the Tat system. The Rieske subunit of Cytochrome bc1 complex (in plastids the b6f complex) is a cofactor-containing Tat substrate which is essential for aerobic respiratory and photosynthesis. Some Tat substrates which lack cofactors. A large number of Tat substrates in prokaryotes such as *Rhodobacter capsulatus*, *Streptomyces coelicolor* or halophilic archaea includes a high proportion of cofactor-free proteins [4]. About 145 Tat substrates have been predicted in *Streptomyces coelicolor*, which is about 16% of the total exported proteins [80]. Alpha-proteobacteria of the genera *Mesorhizobium* or *Sinorhizobium* have a large number of Tat substrates, *Caulobacter crescentus* has more than 80 predicted Tat substrates though this organism encodes fewer proteins in its genome than *E.coli* [25]. The comprehensive list of Tat substrate function has been provided in the Table 1.3.

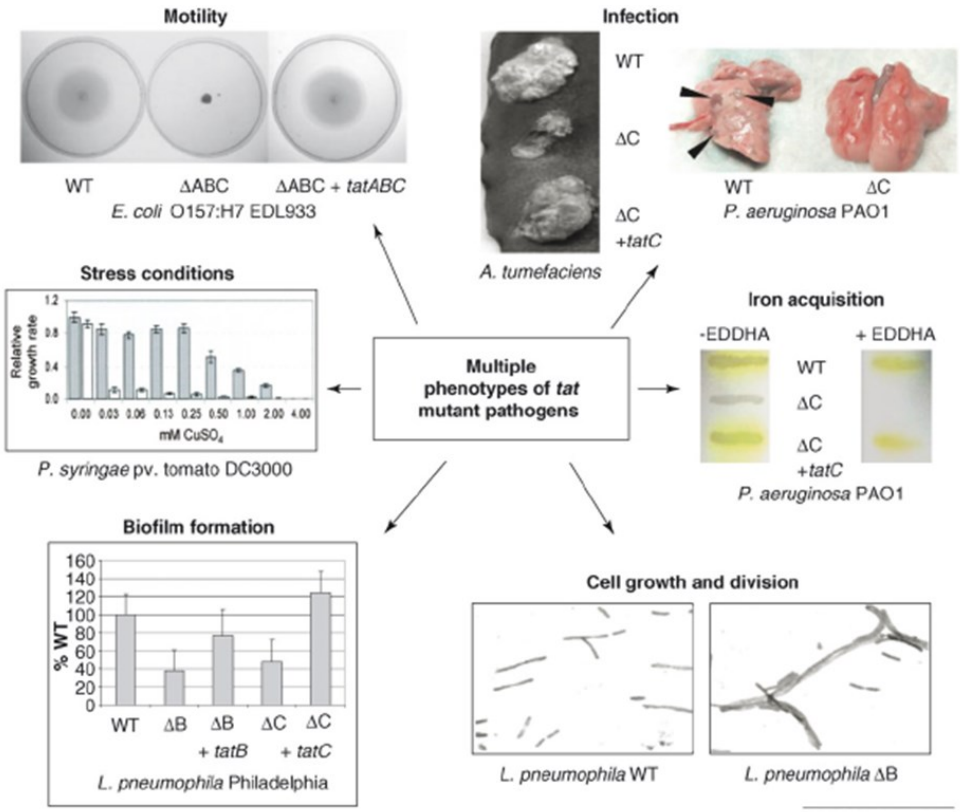
Table 1.3. List of Tat substrate function in different organisms

S.no	*Organism*	Function
1	*Bacillus subtilis*	Salinity stress tolerance
2	*Burkholderia pseudomallei*	B-lactam resistance
3	*Salmonella enterica serovar Typhimurium*	Colonization
4	Halobacterium *sp.* NRC-1	Total secretion
5	*Vibrio alginolyticus*	Biofilm formation, extracellular protease activity, and virulence
6	*Campylobacter jejuni*	Phosphate metabolism
7	*Sodalis glossinidius*	Redox regulation
8	*Xylella fastidiosa*	Virulence
9	*Bdellovibrio bacteriovorus*	Host independent growth
10	*Streptomyces coelicolor*	Lipoprotein biogenesis
11	*Dickeya dadantii 3937*	Virulence and fitness
12	*Streptomyces scabies*	Virulence
13	*Salmonella enterica serovar Typhimurium*	Selenite reduction
14	Staphylococcus	Iron import
15	*Streptomyces sp. Strain K30*	Latex-Clearing protein
16	*Rhizobium leguminosarum bv. Viciae*	Nitrogen fixation
17	*Ralstonia solanacearum*	Virulence
18	*Mycobacterium smegmatis*	B-lactam resistance
19	*Streptomyces lividans*	Xylanase C secretion
20	*Yersinia pseudotuberculosis*	Virulence
21	*Corynebacterium glutamicum*	Detergent tolerance
22	*Legionella pneumophila*	Biofilm formation
23	*Pseudomonas syringae*	Pathogenicity
24	*Rhizobium*	Symbiosis
25	*Pseudomonas stutzeri*	Nitrous oxide reductase
26	*Xanthomonas oryzae*	Flagellar biogenesis and chemotatctic response
28	*Mycobactterium tuberculosis*	B-lactam susceptibility

There is a wide size distribution among Tat Substrates. The smallest known Tat substrates are high potential iron-sulfur proteins (HiPIP) of less than 10 kDa, and the largest are heterodimeric formate dehydrogenases of nearly 150 kDa [81]. This variability shows that the size and shapes of the transported proteins is not very important for Tat transport. However, the Tat system is restricted to substrates that are sufficiently hydrophilic on their surface [82].

1.3.8 Tat pathway of pathogens

The Tat pathway is present in many pathogens and is important for their virulence [27]. The components of the Tat pathway have been identified in the genomes of many bacterial pathogens, including *E. coli O:157*, *Mycobacterium tuberculosis, Vibrio cholerae*, *Agrobacterium tumefaciens*, *Pseudomonas aeruginosa*, *Helicobacter pylori*, *Listeria monocytogenes*, *Haemophilus influenzae*, *Salmonella enterica*, *Neisseria meningitidis*, *Pasteurella multocida*, *Yersinia pestis*, *Xanthomonas campestris***,** *Xanthomonas axonopodis*, *Xyllela fastidiosa*, *Staphylococcus aureus* and *Ralstonia solanacearum* [25]. The tat pathway plays various roles in pathogenesis such as motility, Infection, Iron acquisition, cell growth and division, biofilm formation and stress tolerance (Fig1.15) [27]. In Mycobacterium, the Tat pathway was studied in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. Defective Tat mutants were incapable of exporting the active b-lactamases in both *M. smegmatis* and *M. tuberculosis* [83]. *M. tuberculosis* phospholipase C proteins were found to be Tat-dependent [83][84]. Difficulties in constructing tat mutants of *M. tuberculosis* led to the suggestion that the *M. tuberculosis* tat genes might be essential for viability [84].



De Buck, Emmy, Elke Lammertyn, and Jozef Anné., Trends in microbiology 16.9 (2008)

Fig.1.15 The tat mutants of different pathogens are characterized by multiple phenotypes. Motility: motility assay plates (0.3% soft agar medium) of E. coli O157:H7 EDL933 wild type (WT), tatABCmutant (DABC), and complemented mutant (DABC + tatABC). Infection (left panel): production of tumours by *A. tumefaciens* wild type (WT), tatC mutant (DC) and complemented tatC mutant (DC+ tatC) on susceptible plant tissue. Infection (right panel): pulmonary abscesses were detected on rat lungs that were infected with *P. aeruginosa* PAO1 wild-type cells (arrowheads) but were not observed on lungs infected with DtatC mutant bacteria. Iron acquisition: production of the yellow-green siderophore pyoverdine on low-iron media (casamino acids agar) for *P. aeruginosa* PAO1 wild type, tatC mutant (DC) and complemented mutant (DC+ tatC). In the presence of the iron-chelator EDDHA, the tatC mutant also failed to grow. Cell growth and division: electron microscopic images of stationary-phase-grown *L. pneumophila* Philadelphia wild type (WT) and tatB mutant (DB). Aggregation is observed for the tatB mutant cells. The same image was seen for the tatC mutant. Biofilm formation (the quantification of biofilm formation in *L. pneumophila* Philadelphia): the results given are absorption values at 590 nm after crystal violet staining and corrected for optical density of the cultures. The biofilm formation of the tatB (DB) and the tatC (DC) mutant and the complemented strains (DB+ tatB and DC+ tatC) is given as a percentage of wild-type biofilm formation. Stress conditions: copper susceptibility assays of *P. syringae pv. tomato DC3000* wild type (grey bars) and tatC mutant (white bars) over a range from 0 to 4 mM of CuSO4. All growth data were normalized to growth of wild type in the absence of CuSO4. [27]

The Tat Pathway is a very promising drug target for drug development against pathogenic organisms. A few assays have also been developed to quantify the activity of the Tat pathway to aid in drug development. Tat Signal sequence activity based on the export of a Tat signal bound beta-lactamase was in *M. tuberculosis* [85] and in *Streptomyces coelicolor* agarase assay was developed where Tat signal fused with Streptomyces coelicolor agarase is expressed in *S. lividans* [86]. High-throughput fluorescence assay for screening Tat inhibitors in *E. coli* was developed using a mCherry fused precursor protein substrate spTorA-mCherry-SsrA that gets degraded in the cytoplasm by the ClpXP/ClpAP protease system which recognizes SsrA tag [87]. For the identification of Tat inhibitors in Pseudomonas aeruginosa the PlcH HTS assay was developed that quantified the PlcH(Tat-transported hemolytic phospholipase C) activity [88].

1.3.9 Tat applications

The ability of the Tat Pathway to transport folded proteins across the bacterial membrane and the strict quality control mechanism makes it a tool for heterologous expression of proteins. It is being used widely studied for production of protein in *E.coli* [89], *Bacillus subtilis* [90], *Corynebacterium glutamicum* [91] and *Streptomyces coelicolor* [92] for heterologous expression of protein for industrial applications. The development of a short and universal Tat signal would lead to a significant progress in this aspect and needs proper understanding of the interactions between Tat Signal, Tat receptor complex and Tat chaperons. A phage display technique for Tat pathway [93] is being analyzed in *E.coli* . This could lead to development of a phage display system for oligomeric as well as co-factor-containing proteins, further extending the applicability of filamentous phage display to a wide range of therapeutically interesting proteins.

1.4 Scope of work

The Twin arginine translocase components have been studied in this work. This pathway transports folded proteins across the membrane and its study is of significant interest for therapeutic as well as industrial applications. The main objectives of our study were understanding the interaction between the Tat signal sequence and REMP chaperon, determining a minimal Tat signal peptide necessary for transport, development of a Novel High-throughput Twin Arginine Translocase Assay in *E. coli* for inhibitor screening,

determining the atomic resolution structure of MtTatC and *E. coli* Tat membrane complex. A detailed and comprehensive analysis of the Haloarchaeal Tat pathway was also undertaken to get insights into its diversity in Haloarchaea, which almost exclusively use the tat pathway for protein secretion.

1.5 References

[1] M. Deraitus and K. Freeman, "Essentials of cell biology," *CHI '01 Ext. Abstr. Hum. factors Comput. Syst. - CHI '01*, p. 475, 2001.

[2] F. CRICK, "Central Dogma of Molecular Biology," *Nature*, vol. 227, no. 5258, pp. 561–563, Aug. 1970.

[3] R. Schekman and W. Wickner, "Protein translocation across biological membranes," *Science (80-.).*, vol. 310, no. 5753, pp. 1452–6, 2005.

[4] P. Natale, T. Brüser, and A. J. M. Driessen, "Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms," *Biochim. Biophys. Acta - Biomembr.*, vol. 1778, no. 9, pp. 1735–1756, Sep. 2008.

[5] K. Nałecz, "[The 1999 Nobel Prize for physiology or medicine].," *Neurol. Neurochir. Pol.*, vol. 34, no. 2, pp. 233–242, 2000.

[6] G. von Heijne, "The signal peptide," *J. Membr. Biol.*, vol. 115, no. 3, pp. 195–201, May 1990.

[7] V. T. Lee and O. Schneewind, "Protein secretion and the pathogenesis of bacterial infections.," *Genes Dev.*, vol. 15, no. 14, pp. 1725–52, Jul. 2001.

[8] C. Stathopoulos, D. R. Hendrixson, D. G. Thanassi, S. J. Hultgren, J. W. St. Geme III, and R. Curtiss III, "Secretion of virulence determinants by the general secretory pathway in Gram-negative pathogens: an evolving story," *Microbes Infect.*, vol. 2, no. 9, pp. 1061–1072, Jul. 2000.

[9] W. Wickner, A. J. Driessen, and F. U. Hartl, "The enzymology of protein translocation across the Escherichia coli plasma membrane.," *Annu. Rev. Biochem.*, vol. 60, pp. 101–

24, 1991.

[10] A. R. Osborne, T. A. Rapoport, and B. van den Berg, “PROTEIN TRANSLOCATION BY THE SEC61/SECY CHANNEL,” *Annu. Rev. Cell Dev. Biol.*, vol. 21, no. 1, pp. 529–550, Nov. 2005.

[11] J. de Keyzer, C. van der Does, and A. J. M. Driessen, “The bacterial translocase: a dynamic protein channel complex,” *Cell. Mol. Life Sci.*, vol. 60, no. 10, pp. 2034–2052, Oct. 2003.

[12] A. J. . Driessen, “SecB, a molecular chaperone with two faces,” *Trends Microbiol.*, vol. 9, no. 5, pp. 193–196, May 2001.

[13] J. Luirink, G. von Heijne, E. Houben, and J.-W. de Gier, “BIOGENESIS OF INNER MEMBRANE PROTEINS IN ESCHERICHIA COLI,” *http://dx.doi.org/10.1146/annurev.micro.59.030804.121246*, Sep. 2005.

[14] O. Emanuelsson, S. Brunak, G. von Heijne, and H. Nielsen, “Locating proteins in the cell using TargetP, SignalP and related tools,” *Nat. Protoc.*, vol. 2, no. 4, pp. 953–971, Apr. 2007.

[15] † Mark Paetzel, ‡ Andrew Karla, † and Natalie C. J. Strynadka, and ‡ Ross E. Dalbey*, “Signal Peptidases,” 2002.

[16] J. H. Peterson, C. A. Woolhead, and H. D. Bernstein, “Basic amino acids in a distinct subset of signal peptides promote interaction with the signal recognition particle.,” *J. Biol. Chem.*, vol. 278, no. 46, pp. 46155–62, Nov. 2003.

[17] K. Cunningham and W. Wickner, “Specific recognition of the leader region of precursor proteins is required for the activation of translocation ATPase of Escherichia coli.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 86, no. 22, pp. 8630–4, Nov. 1989.

[18] B. C. Berks, “A common export pathway for proteins binding complex redox cofactors?,” *Mol. Microbiol.*, vol. 22, no. 3, pp. 393–404, Nov. 1996.

[19] E. Bogsch, S. Brink, and C. Robinson, “Pathway specificity for a ΔpH-dependent precursor thylakoid lumen protein is governed by a ‘Sec-avoidance’ motif in the transfer

peptide and a 'Sec-incompatible' mature protein," *EMBO J.*, vol. 16, no. 13, pp. 3851–3859, Jul. 1997.

[20] M. Alami *et al.*, "Differential Interactions between a Twin-Arginine Signal Peptide and Its Translocase in Escherichia coli," *Mol. Cell*, vol. 12, no. 4, pp. 937–946, Oct. 2003.

[21] D. Simone, D. C. Bay, T. Leach, and R. J. Turner, "Diversity and Evolution of Bacterial Twin Arginine Translocase Protein, TatC, Reveals a Protein Secretion System That Is Evolving to Fit Its Environmental Niche," *PLoS One*, vol. 8, no. 11, p. e78742, Nov. 2013.

[22] M. Pohlschröder, K. Dilks, N. J. Hand, and R. W. Rose, "Translocation of proteins across archaeal cytoplasmic membranes," *FEMS Microbiol. Rev.*, vol. 28, no. 1, pp. 3–24, 2004.

[23] J. M. Celedon and K. Cline, "Intra-plastid protein trafficking: How plant cells adapted prokaryotic mechanisms to the eukaryotic condition," *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1833, no. 2, pp. 341–351, Feb. 2013.

[24] W. Pett and D. V. Lavrov, "The Twin-Arginine Subunit C in Oscarella : Origin, Evolution, and Potential Functional Significance," *Integr. Comp. Biol.*, vol. 53, no. 3, pp. 495–502, Sep. 2013.

[25] K. Dilks, R. W. Rose, E. Hartmann, and M. Pohlschröder, "Prokaryotic utilization of the twin-arginine translocation pathway: A genomic survey," *J. Bacteriol.*, vol. 185, no. 4, pp. 1478–1483, 2003.

[26] T. Palmer and B. C. Berks, "The twin-arginine translocation (Tat) protein export pathway," *Nat. Rev. Microbiol.*, vol. 10, no. 7, pp. 483–496, Jul. 2012.

[27] E. De Buck, E. Lammertyn, and J. Anné, "The importance of the twin-arginine translocation pathway for bacterial virulence," *Trends Microbiol.*, vol. 16, no. 9, pp. 442–453, Sep. 2008.

[28] O. V. Nolandt, T. H. Walther, S. Roth, J. Bürck, and A. S. Ulrich, "Structure analysis of the membrane protein TatCd from the Tat system of B. subtilis by circular dichroism," *Biochim. Biophys. Acta - Biomembr.*, vol. 1788, no. 10, pp. 2238–2244,

Oct. 2009.

[29] J. Fröbel, P. Rose, F. Lausberg, A.-S. Blümmel, R. Freudl, and M. Müller, "Transmembrane insertion of twin-arginine signal peptides is driven by TatC and regulated by TatB," *Nat. Commun.*, vol. 3, no. 1, p. 1311, Jan. 2012.

[30] F. Alcock, P. J. Stansfeld, H. Basit, J. Habersetzer, and A. B. Matthew, "Assembling the Tat protein translocase," pp. 1–28, 2016.

[31] J. D. Jongbloed *et al.*, "TatC is a specificity determinant for protein secretion via the twin-arginine translocation pathway.," *J. Biol. Chem.*, vol. 275, no. 52, pp. 41350–7, Dec. 2000.

[32] D. Mangels, J. Mathers, A. Bolhuis, and C. Robinson, "The Core TatABC Complex of the Twin-arginine Translocase in Escherichia coli: TatC Drives Assembly Whereas TatA is Essential for Stability," *J. Mol. Biol.*, vol. 345, no. 2, pp. 415–423, Jan. 2005.

[33] N. A. Braun, A. W. Davis, and S. M. Theg, "The Chloroplast Tat Pathway Utilizes the Transmembrane Electric Potential as an Energy Source," *Biophys. J.*, vol. 93, no. 6, pp. 1993–1998, Sep. 2007.

[34] U. K. Bageshwar and S. M. Musser, "Two electrical potential–dependent steps are required for transport by the Escherichia coli Tat machinery," *J. Cell Biol.*, vol. 179, no. 1, pp. 87–99, Oct. 2007.

[35] R. J. Turner, A. L. Papish, and F. Sargent, "Sequence analysis of bacterial redox enzyme maturation proteins (REMPs)," *Can. J. Microbiol.*, vol. 50, no. 4, pp. 225–238, Apr. 2004.

[36] A. L. Papish, C. L. Ladner, and R. J. Turner, "The twin-arginine leader-binding protein, DmsD, interacts with the TatB and TatC subunits of the Escherichia coli twin-arginine translocase.," *J. Biol. Chem.*, vol. 278, no. 35, pp. 32501–6, Aug. 2003.

[37] M. G. Hicks *et al.*, "Formation of functional Tat translocases from heterologous components," *BMC Microbiol.*, vol. 6, no. 1, p. 64, Jul. 2006.

[38] F. Sargent, N. R. Stanley, B. C. Berks, and T. Palmer, "Sec-independent protein

translocation in Escherichia coli. A distinct and pivotal role for the TatB protein.," *J. Biol. Chem.*, vol. 274, no. 51, pp. 36073–82, Dec. 1999.

[39] N. R. Stanley, T. Palmer, and B. C. Berks, "The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in Escherichia coli.," *J. Biol. Chem.*, vol. 275, no. 16, pp. 11591–6, Apr. 2000.

[40] P. A. Lee, D. Tullman-Ercek, and G. Georgiou, "The Bacterial Twin-Arginine Translocation Pathway," *Annu. Rev. Microbiol.*, vol. 60, no. 1, pp. 373–395, 2006.

[41] F. Sargent, B. C. Berks, and T. Palmer, "The Tat Protein Export Pathway," *EcoSal Plus*, vol. 4, no. 1, Sep. 2010.

[42] S. Cristóbal, J. W. De Gier, H. Nielsen, and G. Von Heijne, "Competition between Sec- and TAT-dependent protein translocation in Escherichia coli," *EMBO J.*, vol. 18, no. 11, pp. 2982–2990, 1999.

[43] N. Blaudeck, P. Kreutzenbeck, R. Freudl, and G. A. Sprenger, "Genetic analysis of pathway specificity during posttranslational protein translocation across the Escherichia coli plasma membrane.," *J. Bacteriol.*, vol. 185, no. 9, pp. 2811–9, May 2003.

[44] R. W. Rose, T. Bruser, J. C. Kissinger, and M. Pohlschroder, "Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway," *Mol. Microbiol.*, vol. 45, no. 4, pp. 943–950, Aug. 2002.

[45] J. Bendtsen, H. Nielsen, D. Widdick, T. Palmer, and S. Brunak, "Prediction of twin-arginine signal peptides," *BMC Bioinformatics*, vol. 6, no. 1, p. 167, Jul. 2005.

[46] P. G. Bagos, E. P. Nikolaou, T. D. Liakopoulos, and K. D. Tsirigos, "Combined prediction of Tat and Sec signal peptides with hidden Markov models," *Bioinformatics*, vol. 26, no. 22, pp. 2811–2817, Nov. 2010.

[47] A. Rodrigue, A. Chanal, K. Beck, M. Müller, and L. F. Wu, "Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial tat pathway.," *J. Biol. Chem.*, vol. 274, no. 19, pp. 13223–8, May 1999.

[48] F. Rodriguez *et al.*, "Structural model for the protein-translocating element of the twin-

arginine transport system.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 12, pp. E1092-101, 2013.

[49] Y. Hu, E. Zhao, H. Li, B. Xia, and C. Jin, "Solution NMR Structure of the TatA Component of the Twin-Arginine Protein Transport System from Gram-Positive Bacterium *Bacillus subtilis*," *J. Am. Chem. Soc.*, vol. 132, no. 45, pp. 15942–15944, Nov. 2010.

[50] Y. Zhang, L. Wang, Y. Hu, and C. Jin, "Solution structure of the TatB component of the twin-arginine translocation system," *Biochim. Biophys. Acta - Biomembr.*, vol. 1838, no. 7, pp. 1881–1888, Jul. 2014.

[51] S. Ramasamy, R. Abrol, C. J. M. Suloway, and W. M. Clemons, "The glove-like structure of the conserved membrane protein TatC provides insight into signal sequence recognition in twin-arginine translocation.," *Structure*, vol. 21, no. 5, pp. 777–88, May 2013.

[52] S. Johnson and S. M. Lea, "Structure of the TatC core of the twin-arginine protein transport system," pp. 1–6, 2012.

[53] B. C. Berks, "The Twin-Arginine Protein Translocation Pathway," *Annu. Rev. Biochem.*, vol. 84, no. 1, pp. 843–864, 2014.

[54] G. Buchanan *et al.*, "Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis," *Mol. Microbiol.*, vol. 43, no. 6, pp. 1457–1470, Mar. 2002.

[55] T. H. Walther, S. L. Grage, N. Roth, and A. S. Ulrich, "Membrane Alignment of the Pore-Forming Component TatA $_{d}$ of the Twin-Arginine Translocase from *Bacillus subtilis* Resolved by Solid-State NMR Spectroscopy," *J. Am. Chem. Soc.*, vol. 132, no. 45, pp. 15945–15956, Nov. 2010.

[56] P. A. Lee, G. Buchanan, N. R. Stanley, B. C. Berks, and T. Palmer, "Truncation analysis of TatA and TatB defines the minimal functional units required for protein translocation.," *J. Bacteriol.*, vol. 184, no. 21, pp. 5871–9, Nov. 2002.

[57] C. Dabney-Smith and K. Cline, "Clustering of C-Terminal Stromal Domains of Tha4

Homo-oligomers during Translocation by the Tat Protein Transport System," *Mol. Biol. Cell*, vol. 20, no. 7, pp. 2060–2069, Apr. 2009.

[58] P. A. Lee *et al.*, "Cysteine-scanning mutagenesis and disulfide mapping studies of the conserved domain of the twin-arginine translocase TatB component.," *J. Biol. Chem.*, vol. 281, no. 45, pp. 34072–85, Nov. 2006.

[59] V. Fincher, C. Dabney-Smith, and K. Cline, "Functional assembly of thylakoid ΔpH-dependent/Tat protein transport pathway components in vitro," *Eur. J. Biochem.*, vol. 270, no. 24, pp. 4930–4941, 2003.

[60] M. G. Hicks, E. de Leeuw, I. Porcelli, G. Buchanan, B. C. Berks, and T. Palmer, "The *Escherichia coli* twin-arginine translocase: conserved residues of TatA and TatB family components involved in protein transport," *FEBS Lett.*, vol. 539, no. 1–3, pp. 61–67, Mar. 2003.

[61] M. G. Hicks, P. A. Lee, G. Georgiou, B. C. Berks, and T. Palmer, "Positive selection for loss-of-function tat mutations identifies critical residues required for TatA activity," *J. Bacteriol.*, vol. 187, no. 8, pp. 2920–2925, 2005.

[62] N. P. Greene *et al.*, "Cysteine scanning mutagenesis and disulfide mapping studies of the TatA component of the bacterial twin arginine translocase.," *J. Biol. Chem.*, vol. 282, no. 33, pp. 23937–45, Aug. 2007.

[63] A. Bolhuis, J. E. Mathers, J. D. Thomas, M. Claire, L. Barrett, and C. Robinson, "TatB and TatC Form a Functional and Structural Unit of the Twin-arginine Translocase from Escherichia coli," *J. Biol. Chem.*, vol. 276, no. 23, pp. 20213–20219, 2001.

[64] J. P. Barnett, R. T. Eijlander, O. P. Kuipers, and C. Robinson, "A Minimal Tat System from a Gram-positive Organism A BIFUNCTIONAL TatA SUBUNIT PARTICIPATES IN DISCRETE TatAC AND TatA COMPLEXES," *J. Biol. Chem.*, vol. 283, no. 5, pp. 2534–2542, 2007.

[65] S. Zoufaly *et al.*, "Mapping precursor-binding site on TatC subunit of twin arginine-specific protein translocase by site-specific photo cross-linking.," *J. Biol. Chem.*, vol. 287, no. 16, pp. 13430–41, Apr. 2012.

[66] H. Kneuper *et al.*, "Molecular dissection of TatC defines critical regions essential for protein transport and a TatB-TatC contact site," *Mol. Microbiol.*, vol. 85, no. 5, pp. 945–961, Sep. 2012.

[67] C. Aldridge, X. Ma, F. Gerard, and K. Cline, "Substrate-gated docking of pore subunit Tha4 in the TatC cavity initiates Tat translocase assembly," *J. Cell Biol.*, vol. 205, no. 1, pp. 51–65, 2014.

[68] U. Gohlke *et al.*, "The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 30, pp. 10482–6, Jul. 2005.

[69] J. M. Celedon and K. Cline, "Stoichiometry for binding and transport by the twin Arginine translocation system," *J. Cell Biol.*, vol. 197, no. 4, pp. 523–534, 2012.

[70] C. Dabney-Smith, H. Mori, and K. Cline, "Oligomers of Tha4 organize at the thylakoid Tat translocase during protein transport," *J. Biol. Chem.*, vol. 281, no. 9, pp. 5476–5483, 2006.

[71] G. F. White *et al.*, "Subunit Organization in the TatA Complex of the Twin Arginine Protein Translocase A SITE-DIRECTED EPR SPIN LABELING STUDY," *J. Biol. Chem.*, vol. 285, no. 4, pp. 2294–2301, 2009.

[72] T. H. Walther *et al.*, "Folding and Self-Assembly of the TatA Translocation Pore Based on a Charge Zipper Mechanism," *Cell*, vol. 152, no. 1–2, pp. 316–326, Jan. 2013.

[73] H. Mori and K. Cline, "A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid ΔpH/Tat translocase," *J. Cell Biol.*, vol. 157, no. 2, pp. 205–210, 2002.

[74] T. Brüser and C. Sanders, "An alternative model of the twin arginine translocation system.," *Microbiol. Res.*, vol. 158, no. 1, pp. 7–17, Jan. 2003.

[75] T. Palmer, F. Sargent, and B. C. Berks, "Export of complex cofactor-containing proteins by the bacterial Tat pathway," *Trends Microbiol.*, vol. 13, no. 4, pp. 175–180, 2005.

[76] B. C. Berks, T. Palmer, and F. Sargent, "Protein targeting by the bacterial twin-arginine

translocation (Tat) pathway," *Curr. Opin. Microbiol.*, vol. 8, no. 2, pp. 174–181, 2005.

[77] U. Lindenstrauß and T. Brüser, "Tat transport of linker-containing proteins in Escherichia coli," *FEMS Microbiol. Lett.*, vol. 295, no. 1, pp. 135–140, 2009.

[78] S. Richter and T. Brüser, "Targeting of unfolded PhoA to the TAT translocon of Escherichia coli," *J. Biol. Chem.*, vol. 280, no. 52, pp. 42723–42730, 2005.

[79] K. Dilks and M. I. Giménez, "Genetic and Biochemical Analysis of the Twin-Arginine Translocation Pathway in Halophilic Archaea Genetic and Biochemical Analysis of the Twin-Arginine Translocation Pathway in Halophilic Archaea," 2005.

[80] D. a Widdick *et al.*, "The twin-arginine translocation pathway is a major route of protein export in Streptomyces coelicolor.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 47, pp. 17927–17932, 2006.

[81] B. Hou and T. Brüser, "The Tat-dependent protein translocation pathway," *Biomol. Concepts*, vol. 2, no. 6, pp. 507–523, 2011.

[82] S. Richter, U. Lindenstrauss, C. Lücke, R. Bayliss, and T. Brüser, "Functional tat transport of unstructured, small, hydrophilic proteins," *J. Biol. Chem.*, vol. 282, no. 46, pp. 33257–33264, 2007.

[83] J. A. McDonough, K. E. Hacker, A. R. Flores, M. S. Pavelka, and M. Braunstein, "The twin-arginine translocation pathway of Mycobacterium smegmatis is functional and required for the export of mycobacterial β-lactamases," *J. Bacteriol.*, vol. 187, no. 22, pp. 7667–7679, 2005.

[84] J. E. Posey, T. M. Shinnick, and F. D. Quinn, "Characterization of the twin-arginine translocase secretion system of Mycobacterium smegmatis," *J. Bacteriol.*, vol. 188, no. 4, pp. 1332–1340, 2006.

[85] J. R. McCann, J. A. McDonough, M. S. Pavelka, and M. Braunstein, "β-Lactamase can function as a reporter of bacterial protein export during Mycobacterium tuberculosis infection of host cells," *Microbiology*, vol. 153, no. 10, pp. 3350–3359, 2007.

[86] D. A. Widdick, R. T. Eijlander, J. M. van Dijl, O. P. Kuipers, and T. Palmer, "A Facile

Reporter System for the Experimental Identification of Twin-Arginine Translocation (Tat) Signal Peptides from All Kingdoms of Life," *J. Mol. Biol.*, vol. 375, no. 3, pp. 595–603, 2008.

[87] U. K. Bageshwar *et al.*, "High throughput screen for Escherichia coli twin arginine translocation (Tat) inhibitors," *PLoS One*, vol. 11, no. 2, pp. 1–25, 2016.

[88] M. L. Vasil, A. P. Tomaras, and A. E. Pritchard, "Identification and Evaluation of Twin-Arginine Translocase Inhibitors," *Antimicrob. Agents Chemother.*, vol. 56, no. 12, pp. 6223–6234, 2012.

[89] C. F. R. O. Matos *et al.*, "High-yield export of a native heterologous protein to the periplasm by the tat translocation pathway in Escherichia coli," *Biotechnol. Bioeng.*, vol. 109, no. 10, pp. 2533–2542, 2012.

[90] Ling Lin Fu, Zi Rong Xu, Wei Fen Li, Jiang Bing Shuai, Ping Lu, and Chun Xia Hu, "Protein secretion pathways in Bacillus subtilis: Implication for optimization of heterologous protein secretion," *Biotechnol. Adv.*, vol. 25, no. 1, pp. 1–12, 2007.

[91] H. Teramoto, K. Watanabe, N. Suzuki, M. Inui, and H. Yukawa, "High yield secretion of heterologous proteins in Corynebacterium glutamicum using its own Tat-type signal sequence," *Appl. Microbiol. Biotechnol.*, vol. 91, no. 3, pp. 677–687, 2011.

[92] J. Anné, B. Maldonado, J. Van Impe, L. Van Mellaert, and K. Bernaerts, "Recombinant protein production and streptomycetes," *J. Biotechnol.*, vol. 158, no. 4, pp. 159–167, Apr. 2012.

[93] J. Speck, K. M. Arndt, and K. M. Muller, "Efficient phage display of intracellularly folded proteins mediated by the TAT pathway," *Protein Eng. Des. Sel.*, vol. 24, no. 6, pp. 473–484, Jun. 2011.

Chapter II - Tat Signal chaperon interaction

2.1 Introduction

The twin-arginine translocation (Tat) pathway is responsible for the post-translational transport of folded proteins across the cytoplasmic membrane of bacteria, archaea and the thylakoid membrane of the plant chloroplast. The minimal Tat system, here described for *E. coli*, is composed of the integral membrane proteins TatA, TatB and TatC. Substrates of this pathway contain a characteristic Tat specific Arg-Arg motif in the N-terminal signal sequence [1]. The Tat signal localizes the protein to the membrane where it is recognized by TatC [2] and then the protein is translocated across the cytoplasmic membrane though a pore formed by TatA (reviewed in [3-5]). In most cases, Tat substrates are either a component of the respiratory electron transport chain or are bacterial virulence factors [6, 7]. These substrates often incorporate large co-factors or assemble into complexes before getting translocated [8].

Dimethylsulphoxide reductase is a heterotrimeric periplasmic complex that uses dimethylsulphoxide (DMSO) as an electron acceptor during anaerobic respiration. The central protein, DmsA, is synthesized in the cytoplasm, loaded with its catalytic molybdopterin (MoPt) co-factor and the accessory proteins DmsB and DmsC. This enzyme complex is an example of Tat substrates that must be fully assembled prior to translocation to the bacterial periplasm [9]. Delivery of these proteins is monitored by a special class of cytoplasmic chaperones that bind specifically to the Tat signal of their substrate protein masking the signal sequence thereby ensuring proper maturation and co-factor loading (Fig.1.14) [10, 11]. These chaperones, dubbed redox-enzyme maturation proteins (REMP), prevent the futile export of immature protein [12]. Each REMP generally has a specific binding partner and there are a number of verified examples. DmsD, a well characterized REMP, binds the DmsA N-terminal Tat signal sequence which contains the Tat consensus motif (S/TRRxLVK) [13] [14].

Generally signal peptides have tripartite structure with a positively charged N-region, a middle hydrophobic stretch followed by a polar C-region that contains a signal peptidase cleavage site [15, 16]. The *E. coli* genome encodes at least 29 putative signal sequences that contain the Tat motif. [17]. Many studies have emphasized the significance of the presence of

RR in the signal sequence during translocation [2, 18, 19]. The Tat signal is targeted to the pre-formed complex between TatB and TatC, the TatBC recognition complex [20, 21]. Replacing the RR of the signal sequence with a pair of lysines retains the ability to co-purify with the TatBC complex. This suggests that the essential RR residues are not necessary for binding to the TatBC recognition complex but are required for sucessful entry into the transport cycle [22]. Additional evidence revealed a major role for TatB in initial binding of the precursor protein [22]. The interaction between TatBC and the Tat signal of DmsA has been captured by bimolecular fluorescence complementation (BiFC) [23]. The residues essential for the TatC recognition has been studies extensively [19] Random mutational analysis in the Tat signal peptide indicated that there is some of gain of function mutation with respect to rate of translocation occurred within or near the Tat motif, highlighting the significance of this region [24].

Other characterized REMP/signal peptides are TorD/TorA [8] NapD/NapA [25], NarJ/NarG, NarW/NarZ and HybE/HybO [26]. Other chaperones, such as DnaK and SlyD, were shown to bind broad range of different Tat signal sequences [27]. Bindings of these chaperones are not essential for the Tat depended transport of substrate across the membrane [28, 29].

The role of the signal sequence is more than a simple targeting factor. This is exemplified by the fact that substitution of a DmsA signal with a TorA signal sequence peptide showed poor growth in anaerobic condition despite the targeting of DmsABC to the membrane [9]. Without a correct signal sequence the maturation and production of functional enzyme prior to targeting is lost indirectly implying the significance of DmsD binding for the production of functional enzyme [9]. DmsD can form a complex with the N-terminus of DmsA independent of the globular domain of the protein [13]. *In vivo* studies suggest that DmsD is not required for targeting [29]. Also there is evidence that signal peptide is extensively crosslinked to ribosomal components and the trigger-factor chaperone during synthesis but does not play a critical role in the export of Tat dependent protein [30]

Despite structural information, the molecular characteristics of REMP/signal sequence binding have not been fully elucidated [31] [32]. REMP chaperones contain two specific motifs

Y/F/W-X-X-L-F and E-(P-X/X-P)-D-H/Y that are conserved among all members of this family and mutations in these regions lower the binding affinity to substrates [33].

The binding of REMPs to signal sequences has been localized to the Tat signal and the possible protease site [23]. Further studies using DmsD and the signal peptide indicate binding via a hydrophobic interaction with micromolar affinity in an equimolar ratio [34]. More extensive work has been done on the interaction between TorD and the signal peptide of TorA. Here, recognition has been localized to the hydrophobic h-region [35]. Glutamine-scanning mutagenesis through this region demonstrated that an L31Q variant of the TorA signal peptide impaired binding of TorD [36]. Synthetic peptide truncation variants of the TorA signal showed weaker or no binding of TorD with C-terminal truncations while N-terminal truncations, including the twin arginine motif, did not dramatically affect TorD binding [35]. This suggests that the twin-arginine motif is not essential for chaperone recognition.

Many mechanistic questions about the role of REMPs remain. The coordination between co-factor loading and the subsequent transfer of the mature protein to the translocase assembly is unknown. Our specific goals in this study were the biochemical characterization of DmsD and the signal sequence of DmsA (ssDmsA) complex, also the complex with translocase assembly and to map the location of the binding site of TatBC and DmsD in the DmsA signal sequence, thereby deriving the smallest length of sequence essential for a specific interaction.

2.2 Materials and Methods

2.2.1 Materials

LB media, 2xYT, TB and LB agar used for all the bacterial culture were purchased from Hi-media. Antibiotics used for selection of transformants such as kanamycin, Ampicillin, and chloramphenicol were obtained from Sigma-Aldrich, USA. Chemicals used for the purification of proteins such as Trizma, Sodium chloride, Imidazole, β- mercaptoethanol, glycerol, DTT, Nickel sulphate, Bromophenol-blue (BPB), Acrylamide, N,N'-methylene bisacrylamide, Sodium dodecyl sulfate (SDS), Acetic acid, Methanol, TEMED (N,N,N',N'-Tetramethylethylenediamine), Ammonium persulfate (APS), were purchased from Sigma-Aldrich, USA. Molecular weight marker for SDS-PAGE was purchased from Bio-Rad Laboratories, USA. Ni-NTA used for affinity purification were purchased from Qiagen,

Germany. Size exclusion columns were obtained from GE, USA. Protein samples were concentrated using Amicon® ultra centrifugal filters procured from Merck-Millipore, USA.

Commercial screens procured from Hampton research, USA and Qiagen, Germany were used for initial crystallization screening. Sodium cacodylate, lithium sulfate, PEG 4000, Ethylene glycol, glycerol, 2-methyl-2,4-pentanediol, propan-2-ol etc used in crystallization trials were obtained from Sigma-Aldrich, USA. Two-well sitting-drop plates were obtained from Hampton research, USA. 24 well plates and coverslips were obtained from Corning® (Sigma- Aldrich, USA) and Blue Star, India respectively. Other specialized chemicals and instruments used in the experiments are mentioned in the appropriate places. The constructs used in this study have been described in Table2.1.

Table 2.1. Constructs used in the study with details of vector, Tag, Study and sequence details

S. No.	Construct	vector	Tag	Study	Remarks
1	ssDmsA	pET33b	6-His	Signal chaperon interaction, SPR, minimal signal , Ala scanning	DmsA signal sequence 1-50 amino acids
2	ssDmsA-GFP	pET33b	6-His	Signal chaperon interaction, SPR, minimal signal , Ala scanning	DmsA signal sequence 1-50 amino acids +GFP
3	DmsD	pACYC	None	Signal chaperon interaction, SPR, minimal signal , Ala scanning	DmsD without tag
4	MBP-DmsD	pACYC	MBP	Signal chaperon interaction, SPR, minimal signal , Ala scanning	Maltose binding protein + DmsD
5	EcTatBC	pACYC	None	Signal chaperon interaction	E. coli TatB and TatC without Tag
6	EcTatBC-His	pET33b	6-His	SPR, DSF	E. coli TatB and TatC + 6-His Tag
7	ss45DmsA-GFP	pET33b	6-His	SPR, minimal signal	DmsA signal sequence 1-45 amino acids + GFP
8	ss40DmsA-GFP	pET33b	6-His	SPR, minimal signal	DmsA signal sequence 1-40 amino acids + GFP
9	ss35DmsA-GFP	pET33b	6-His	SPR, minimal signal	DmsA signal sequence 1-35 amino acids + GFP

10	ss30DmsA-GFP	pET33b	6-His	SPR, minimal signal	DmsA signal sequence 1-30 amino acids + GFP
11	ss25DmsA-GFP	pET33b	6-His	SPR, minimal signal	DmsA signal sequence 1-50 amino acids + GFP
12	ssNtrun5DmsA-GFP	pET33b	6-His	SPR, minimal signal	DmsA signal sequence 5-50 amino acids + GFP
13	ssNtrun10DmsA-GFP	pET33b	6-His	SPR, minimal signal	DmsA signal sequence 10-50 amino acids + GFP
14	ssNtrun15DmsA-GFP	pET33b	6-His	SPR, minimal signal	DmsA signal sequence 15-50 amino acids + GFP
15	ssDmsD_Minimum peptideDMSA-GFP	pET33b	6-His	SPR, minimal signal	DmsA signal sequence 13-41 amino acids + GFP
16	E13A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
17	V14A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
18	S15A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
19	R16A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
20	R17A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
21	G18A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
22	L19A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
23	V20A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
24	K21A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
25	T22A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
26	T23A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
27	A24S ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
28	I25A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
29	G26A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
30	G27A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
31	L28A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
32	A29S ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
33	M30A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
34	A31S ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
35	S32A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
36	S33A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
37	A34S ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
38	L35A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
39	T36A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
40	L37A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
41	P38A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
42	F39A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
43	S40A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA

44	R41A ssDmsA	pET33b	6-His	Ala scanning	Ala mutant ssDmsA
45	E13A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
46	V14A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
47	S15A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
48	R16A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
49	R17A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
50	G18A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
51	L19A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
52	V20A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
53	K21A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
54	T22A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
55	T23A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
56	A24S ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
57	I25A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
58	G26A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
59	G27A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
60	L28A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
61	A29S ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
62	M30A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
63	A31S ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
64	S32A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
65	S33A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
66	A34S ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
67	L35A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
68	T36A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
69	L37A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
70	P38A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
71	F39A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
72	S40A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
73	R41A ssDmsA-GFP	pET33b	6-His	Ala scanning	Ala mutant ssDmsA+GFP
74	DDF	pET33b	6-His	Crystallization,DSF	DmsA 1-54+DmsD
75	DDFΔ2-10	pET33b	6-His	Crystallization ,DSF	DmsA 11-54+DmsD
76	DDFΔ33-53	pET33b	6-His	Crystallization	DmsA 1-33+DmsD
77	DmsDp33	pET33b	6-His	DSF	DmsD with 6-His Tag
78	DN2	pET33b	6-His	DSF	DmsA signal sequence 10-50 amino acids + GFP

2.2.2 Over-expression of DDF

To obtain the DDF, the expression host *E. coli* BL21G DE3 DTat was transformed with DDF /pET33b construct by calcium chloride treatment. Selection of transformed colonies was performed on LB agar plate containing Kanamycin. 10ml of the overnight culture was used to inoculate 1000 ml of fresh LB/TB broth with kanamycin. When the absorbance at 600nm reached above ~1, the culture was induced with 0.5 mM IPTG and grown for 4hours. Cells were harvested by centrifugation at 4500 rpm for 10 minutes.

2.2.3 Purification of DDF, DDFΔ2-10 and DDFΔ33-53

DDF overexpressed cells were resuspended in lysis buffer (50 mM Tris pH 7.5, 100 mM NaCl, 30 mM imidazole, 2 mM β-mercaptoethanol,) and sonicated for 10 minutes. After the cell disruption, the lysate was centrifuged for 30 minutes at 13000 rcf and the supernatant was collected. This supernatant was allowed to pass through the pre-equilibrated Ni-NTA column. The column was washed with 10 times the column volume of wash buffer (50 mM Tris pH 7.5, 100 mM NaCl, 30 mM imidazole, 2 mM β-mercaptoethanol (β-ME)). The elution of the protein was carried out with elution buffer containing 300 mM imidazole (50 mM Tris pH 7.5, 100 mM NaCl, 300 mM imidazole, 2 mM β-ME). Eluted protein fractions were pooled and concentrated using amicon 10K concentrator. The concentrated protein was further pass through pre-equilibrated (50 mM HEPES pH 8.0, 2 mM DTT) S200 size exclusion column. The eluted fractions were pooled and further concentrated before crystallization. Concentration of DDF was measured using Bradford method and the purity was checked by SDS-PAGE.

2.2.4 Site Directed Mutagenesis for DDFΔ2-10 and DDFΔ33-53

DDF /pET33b plasmid was extracted from DH5alpha cells were grown overnight in 10ml LB media in Kanamycin. The Site Directed mutagenesis primers were designed according to the quick-change method. Here the primers were complementary having the mutations and amplified the whole plasmid by extending in the opposite directions in 18-20 cycles of PCR amplification using Takara Ex-Taq DNA polymerase. The PCR product was

treated with Dpn I for 3hours which cleaved the DAM methylated plasmids leaving only the complete mutants. The mutants were screened by transformation into DH5alpha cells and plating on Kanamycin LB agar plates. Mutations were confirmed by sequencing to ensure that no undesired mutation has been introduced.

2.2.5 Crystallization

Purified the DDF and DDFΔ2-10 was purified to the homogeneity and concentrated up to 15 mg/ml using Amicon ultra 10kDa and 30 kDa cutoff centrifugal filters (Millipore, USA) at 4 ºC at 4000 rpm for Monomer and dimer fractions respectively. The proteins were screened against several commercially available crystallization screens including, Index (Hampton Research Corp.) Nextal PACT (Qiagen), Nextal Protein Complex Suite (Qiagen) and JCSG plus (Molecular Dimensions). Screenings were performed using the Mosquito Robot (TTP labtech) by vapour diffusion method in 96 well MRC2 sitting drop plates Hampton MRC-SD2 and SD3 plates with multiple protein to screen ratios of 1:1, 1:2 and 2:1 in 400ul sitting drops and incubated at 20 ºC in a RuMed vibration free incubator.

2.2.6 Cryo-protection and X-ray Diffraction

In general, crystals are exposed to high-intensity ionizing radiation during X-ray diffraction. In order to minimize the radiation damage during this exposure, crystals were diffracted in a cryogenic temperature at 100K. Ethylene Glycol, Glycerol, 2-methyl-2,4-pentanediol, sucrose, PEG 200, perfluoropolyether oil etc are commonly used cryoprotectants. Different cryo-protecting agents were tried but 25% PEG200 was found to be the most suitable for DDF crystals.

2.2.7 Co-expression and Pull down

DmsD (w/o tag) and DmsA signal mutant plasmids were co transformed into Bl21Gold DTat cells by chemical transformation. The co-transformed cells were selected on dual antibiotic LBagar plates containing Kanamycin and chloramphenicol. Only co-transformed

cells will survive and were used for pull down analysis. 1ml of the overnight culture was used to inoculate 100 ml of fresh LB broth with kanamycin and chloramphenicol. When the absorbance at 600 nm reached ~ 0.8, the culture was induced with 0.5 mM IPTG and grown for 4hours. After induction, cells were harvested by centrifugation at 4500 rpm for 10minutes. The interaction of DmsD (w/o tag) and DmsA signal mutant plasmids was studied by affinity chromatography. Harvested bacterial cells were resuspended in lysis buffer (50 mM Tris pH 8, 100 mM NaCl, 2mM DTT), followed by sonication for 5 minutes with 5 sec on/off pulse and 45% amplitude. After the cell disruption, the lysate was centrifuged for 30 minutes at 12000 rpm. The supernatant was collected and subjected to Ni-NTA affinity chromatography as used above in DDF purification. Fractions eluted from Ni-NTA were for all co-expressed samples were run on 12% SDS-PAGE. To check the interaction between the DmsD/ssDmsA complex and the EcTatBC assembly, the over expressed cells of EcTatBC without 6xHis tag were lysed by passing four times though a microfludizer and the total membrane fraction was isolated by centrifugation at 45,000g at 4 ºC for 1 hr. The membrane pellet was solublized in 1% digitonin, centrifuged at 35,000g at 4 ºC for 30 min. The supernatant was used for the pull-down experiments; the supernatant was diluted 10 fold with the buffer (50 mM Tris pH 7.5, 100 mM NaCl, 2 mM β-ME), to minimize the cause of excess detergent in complex formation.

2.2.8 Differential Scanning Fluorimetry (DSF)

Measurements of thermal unfolding was performed with the Prometheus NT.48 nanoDSF (NanoTemper Technologies, Germany). DSF experiments were based on changes in intrinsic fluorescence of the protein where the shift in fluorescence emission is plotted as the ratio between 350 and 330nm. The capillaries were filled with mixture of 20µl of protein placed onto the capillary tray of the Prometheus NT.48. Start to end temperatures were set as 20°C to 95°C and the heating rate was defined as 1°C/min. The inflection point of the resulting sigmoidal curve provided the melting temperature (*T*m) of the protein. The *T*m of the various combinations of proteins were studied in native condition.

2.2.9 SPR analysis

All SPR analysis was performed at 25°C on a Biacore T100 system using Biacore CM5 research grade sensor chips (GE Healthcare). Purified fractions of DmsD were immobilized on sensor chips using amine coupling at 400 µg/ml of total protein. The dextran matrix on the sensor chip surface is first activated with 1:1 mixture of 0.4 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and 0.1 M N-hydroxysuccinimide (NHS) to create reactive succinimide esters. The maximum immobilization level of the DmsD was achieved after three injections and the sensogram baseline was increased to 2000 RU. The surfaces were deactivated by passage of ethanolamine-HCl (1.0 M at pH 8.5). Experiments were carried out in running buffer containing 10 mM HEPES pH 7.4, 150 mM NaCl, and 0.005% v/v Surfactant P20 (GE Healthcare). Each of the ssDmsA mutants was analyzed for binding by injecting 200 µl of a serially diluted (10, 20, 40, 80, 160 and 320 nM) purified sample at a flow rate of 10 µl/min for 10 minutes. The dissociation step was performed at the same flow rate with same buffer for 15 minutes. Regeneration solution (1 M NaCl, 100 mM glycine pH 9.5) for 2 minute at 5 µl/min were used to regenerate the DmsD surface after each injection. For the *Ec*TatBC complex, purified fractions of *Ec*TatBC with 6xHis tag was used directly for immobilization on sensor chips using the amine coupling in presence of 0.1% (w/v) digitonin at 250 µg/ml of total protein. The dextran matrix on the sensor chip surface was prepared as mentioned above. The maximum immobilization level of the *Ec*TatBC was achieved after three injections and the sensogram baseline was increased to 2000 RU. The surfaces were then deactivated by passage of ethanolamine-HCl (1.0 M at pH 8.5). Experiments were carried out in running buffer containing Sodium acetate buffer pH 5.0 with NaCl 100mM, 10% v/v glycerol and 0.1% (w/v) digitonin. Each of the ssDmsA mutants, DmsD and ssDmsA/DmsD complexes were analyzed for binding by injecting 200 µl of a serially diluted (100, 200, 400, 800, 1600 and 3200 nM) purified sample in PBS buffer contain 10% v/v glycerol and 0.1% (w/v) digitonin at a flow rate of 10 µl/min for 10 minutes. The dissociation step was performed at the same flow rate with same buffer for 15 minutes. Regeneration solution [1 M NaCl, 100 mM glycine pH 9.5 contains 10% (v/v) glycerol and 0.1% (w/v) digitonin] for 2 minutes at 5 µl/min were used to regenerate the *Ec*TatBC surface after each injection.

One flow cell of the CM5 chip was used as the blank control. The signal of each binding experiment was corrected for non-specific binding by subtracting the signal obtained for a blank surface. The kinetic of the association and dissociation phase were measured at several flow rates from 5 to 40 µL. the kinetic rates measured were not affected by flow rate, demonstrating that the system is not mass-transfer limited. Kinetic data analysis was done in BIAevaluation software v.4.1 for multiple concentrations simultaneously using a 1:1 Langmiur model. Some data curves were eliminated to reduce the $\chi 2$ value for a better fitting; A more detailed description of the parameters and equation is provided in the Biacore (GE Healthcare) T100 software Handbook (BR-1006-48, edn AE, 186-187)

2.2.10 CD analysis

All CD measurements were performed on a JASCO J-715 spectropolarimeter, using circular quartz cells with path lengths of 0.02 cm. The response time setting for the spectrometer was 2 s with a data acquisition time of 5 s. The CD spectra are reported in terms of [e], molar ellipticity, deg cm^2 dmol. Each measurement was the average of three repeated scans in steps. The temperature of the sample was controlled by a circulating water bath (Lauda, type K2R) linked to the outer jacket of the cuvette (Helmann).

For CD measurements of the complex of DmsD and ssDmsA peptides, a molar ratio of 1.5:1 (peptide:protein), at a peptide concentration of 50 µm, was used, which is comparable to the K_D, value of DmsD to the peptides. The differential spectrum of DmsD was obtained by subtracting the CD spectrum of the peptides from the CD spectrum of the complex at the same concentration as in the complex before the conversion to the ellipticity values.

2.2.11 Multi-angle light-scattering analysis

Purified *Ec* ssDmsA and DmsD complex (15 mg ml^{-1}) was loaded onto a Shodex protein KW-803 size-exclusion column equilibrated with 50 mM Tris-HCl pH 7.0, 150 mM NaCl, 10mM β-mercaptoethanol and connected in-line with a Dawn 18-angle light-scattering detector coupled to an Optilab interferometric refractometer and a WyattQELS Quasi-Elastic Light-Scattering instrument (Wyatt Technologies). Data analysis was performed with the

ASTRA V5.3.4.14 software (Wyatt Technologies) and molecular weights were calculated using the Zimm fit method.

2.3 Results

2.3.1 Characterization of DmsD/ssDmsA complex

The signal sequence of DmsA (first 50 amino acids) construct followed by a protease site (Tev), green fluorescent protein (GFP) and a six histidine-tag for purification (ssDmsA-GFP). An expression construct of full length maltose binding protein (MBP) tag followed by a second protease site (Thr) and DmsD (MBP-DmsD). Both native proteins and their mutants were purified to homogeneity by chromotagraphy. Wild type and mutant variants of ssDmsA were purified and tested for their binding capability with wild type DmsD by incubating purified MBP-DmsD with ssDmsA-GFP fusion protein and separating on a size-exclusion column (SEC). The two different purification tags were added for ease of purification of homogenous complex (Fig 2.1).

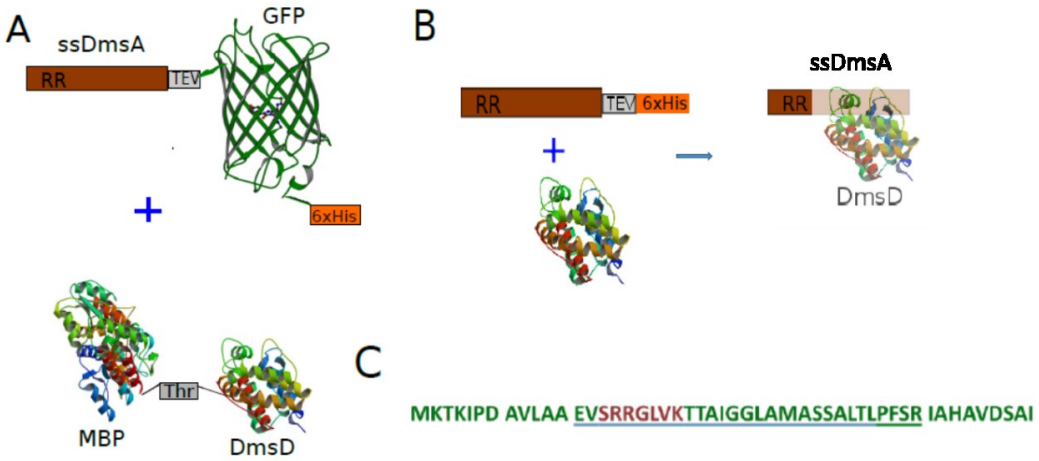


Fig. 2.1 The constructs used for DmsA signal sequence and DmsD binding studies A) DmsA (first 50 amino acids), Tev protease site and GFP and a 6-his tag for purification (ssDmsA-GFP) complexed with full length MBP- tag followed Thr protease site and DmsD (MBP-DmsD) B) DmsA (first 50 amino acids) Tev protease and a 6-his tag for purification (ssDmsA) complexed to Wild type DmsD (DmsD) C) The DmsD signal sequence.

The observed results are presented in the Fig 2.2. The SEC profiles clearly show DmsD/ssDmsA complex formation. The formation of complex was further analyzed by mixing different molar ratios of each component. A two-fold excess of ssDmsA-GFP saturates DmsD and forms homogeneous complex DmsD.

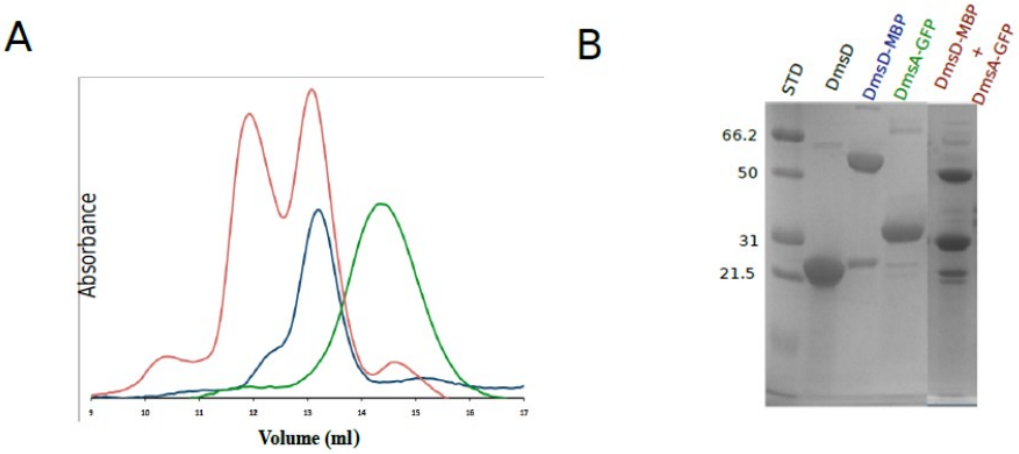


Fig 2.2 Complex formation between the co-expressed ssDmsA-GFP with MBP-DmsD A) Size exclusion chromatography shows greem peak for ssDmsA-GFP, blue peak for DmsD-MBP and red peaks corresponding to free MBP-DmsD and the ssDmsA-GFP MBP-DmsD complex B) SDS PAGE of DmsD, ssDmsA-GFP, MBP-DmsD and the peak corresponding to ssDmsA-GFP with MBP-DmsD complex

The complex formed between the co-expressed ssDmsA (6xHis tag) with DmsD (Figure 2.1.B) had two distinct populations' i.e. monomeric complex and dimeric co-complex (Fig. 2.3A), these complexes were confirmed by running the samples on SDS PAGE (Fig. 2.3.B).

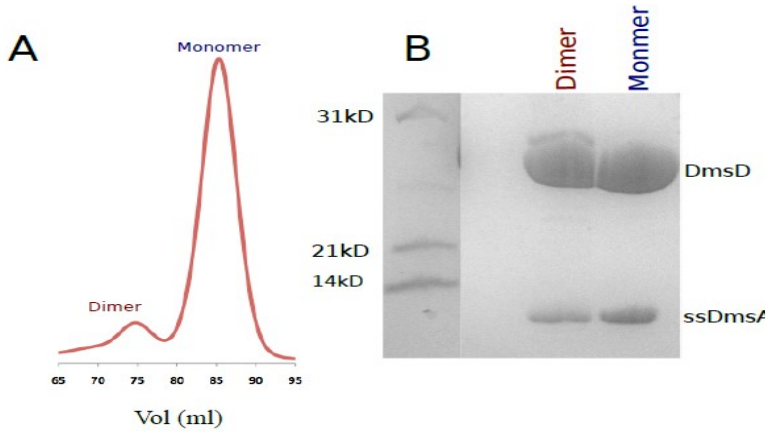


Fig.2.3 Complex formation between the co-expressed ssDmsA (6xHis tag) with DmsD A) Size exclusion chromatography shows two separate peaks corresponding to monomeric and dimeric complex B) SDS PAGE analysis of the Monomeric and Dimeric fractions show the presence of both ssDmsA (6xHis tag) and DmsD

A comparison of the elution profiles of monomeric ssDmsA/DmsD complex and DmsD alone (Fig2.4) is shown which shows the difference in elution volumes between them.

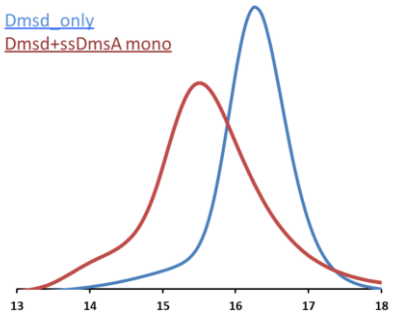


Fig 2.4 Size Exclusion Chromatography peak comparison between the elution profiles of monomeric ssDmsA/DmsD complex (Red) and DmsD (Blue).

To determine the stoichiometry of the ssDmsA/DmsD complex we performed multi-angle light scattering (MALS) (Fig 2.5). First DmsD was injected and its molecular weight corresponded 23 kDa (Fig 2.5.A). Then the DmsD/ssDmsAcomplex was injected. The results suggest that DmsD/ssDmsA complex exists either as a single copy of ssDmsA and DmsD (monomeric) having a molecular weight of 28kDa (23kDa-DmsD+6kDa-ssDmsA) or one copy of ssDmsA and two copies of DmsD (dimeric) having a molecular weight of 53kDa (2x23kDa-DmsD+6kDa-ssDmsA).

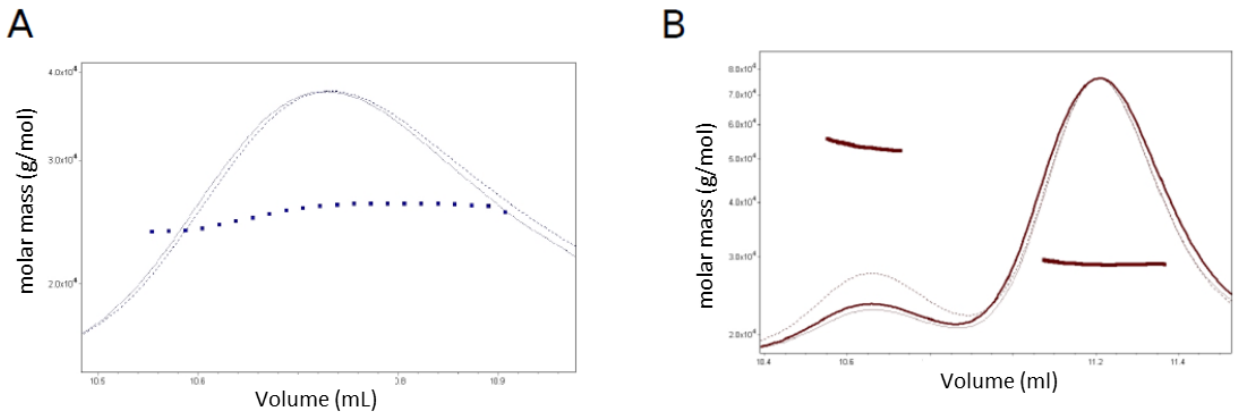


Fig 2.5 Molecular weight determination by MALS A) DmsD - molecular weight 23 kDa B) DmsD/ssDmsA complex with monomeric peak having 28kDa(23kDa-DmsD+6kDa-ssDmsA) and dimeric peak having a molecular weight of 53kDa (2x23kDa-DmsD+6kDa-ssDmsA).

We suspected that a signal sequence bound to DmsD would have a stabilizing effect on DmsD. We tested this using circular dichroism (CD) where we generated melting curves for DmsD alone and bound to ssDmsA (monomeric form) (Figure 2.6.A). The difference in the thermal stability were shown in the (Figure 2.6.B). The thermal stability analysis showed that DmsD is less stable in the presence of ssDmsA. This instability can be attributed to the intrinsic disorder of the hydrophobic region (h-region) of the ssDmsA.

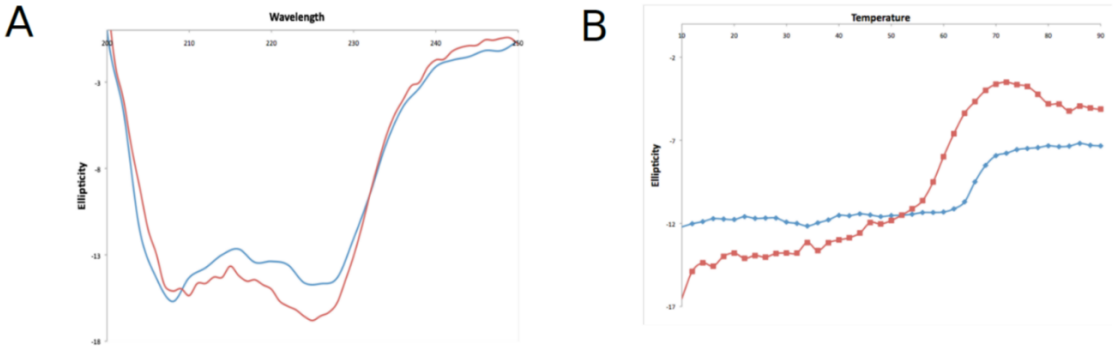


Fig 2.6 CD analysis of DmsD and ssDmsA-DmsD monomeric complex molar reported in terms of [e], ellipticity, deg cm2 dmol A) DmsD (Blue) vs monomeric ssDmsA-DmsD complex(Red) scan at 25°C B) Temperature vs ellipticity for DmsD (Blue) vs monomeric ssDmsA-DmsD complex(Red) fom 10°C to 90°C at 5°C intervals

2.3.2 Tetrameric complex formation

Interaction between TatBC with DmsA, DmsD and DmsD/ssDmsA complex was observed after incubating each component for 2 hrs at 4°C and then affinity capture using the 6xHis tag of ssDmsA or DmsD. As a control, TatBC was overexpressed without a tag and ssDmsA/DmsD complex was also observed to be intact in the presence of detergents. Complex formation was confirmed by a shift in the SEC elution profile (Fig 2.7). The tetrameric complex of DmsA/ssDmsA/TatBC only formed if the monomeric complex DmsD/ssDmsA was used. The dimeric form ssDmsA/DmsD was unable to capture TatBC.

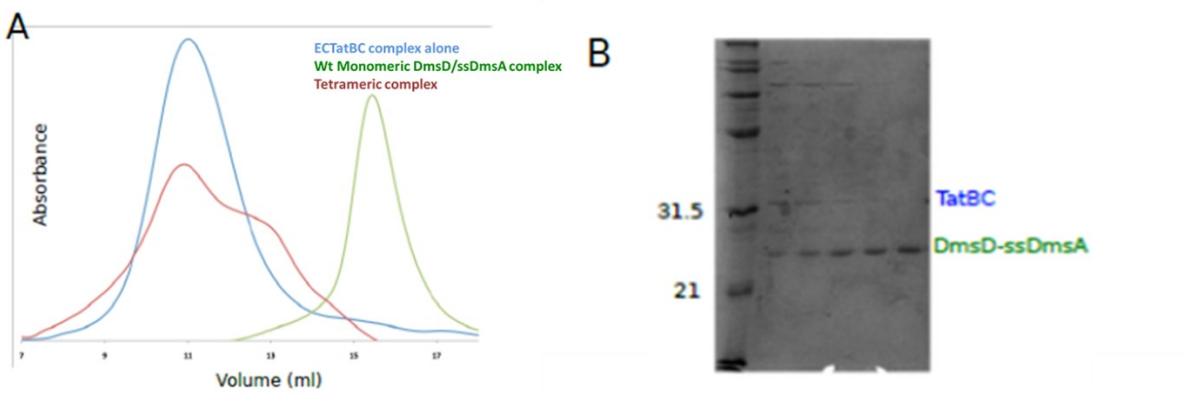


Fig 2.7 Tetrameric complex formation of TatBC and DmsD/ssDmsA A) Size exclusion chromatography of monomeric DmsD-ssDmsA (green), TatBC (blue) and tetrameric comlex of DmsD-ssDmsA and TatBC (red) B) SDS PAGE of tetrameric complex showing TatBC and DmsD-ssDmsA

2.3.3 Sequence comparison of Tat signals with reference to DmsA

Tat signal sequences are typically 30-50 amino acid long [36]. The signal sequences of the closely related DmsA and TorA were aligned (Fig 2.8A). DmsA signal sequences are ~50 amino acid long while TorA signal sequences are relatively shorter, ~35 amino acid. An alignment of all Tat signals are recognized by REMP chaperones from *E. coli* is shown in Fig 2.8.B. The alignment shows the conservation of Tat signal motif, located in the border of n- and h- regions, between the various Tat substrate from *E. coli*. The h-region is relatively longer in the case of DmsA signal peptide in and Ala-29, Ala-31, Leu-37, Pro-38 and Phe-39 in h-regions are conserved at least among DmsA signal peptides.

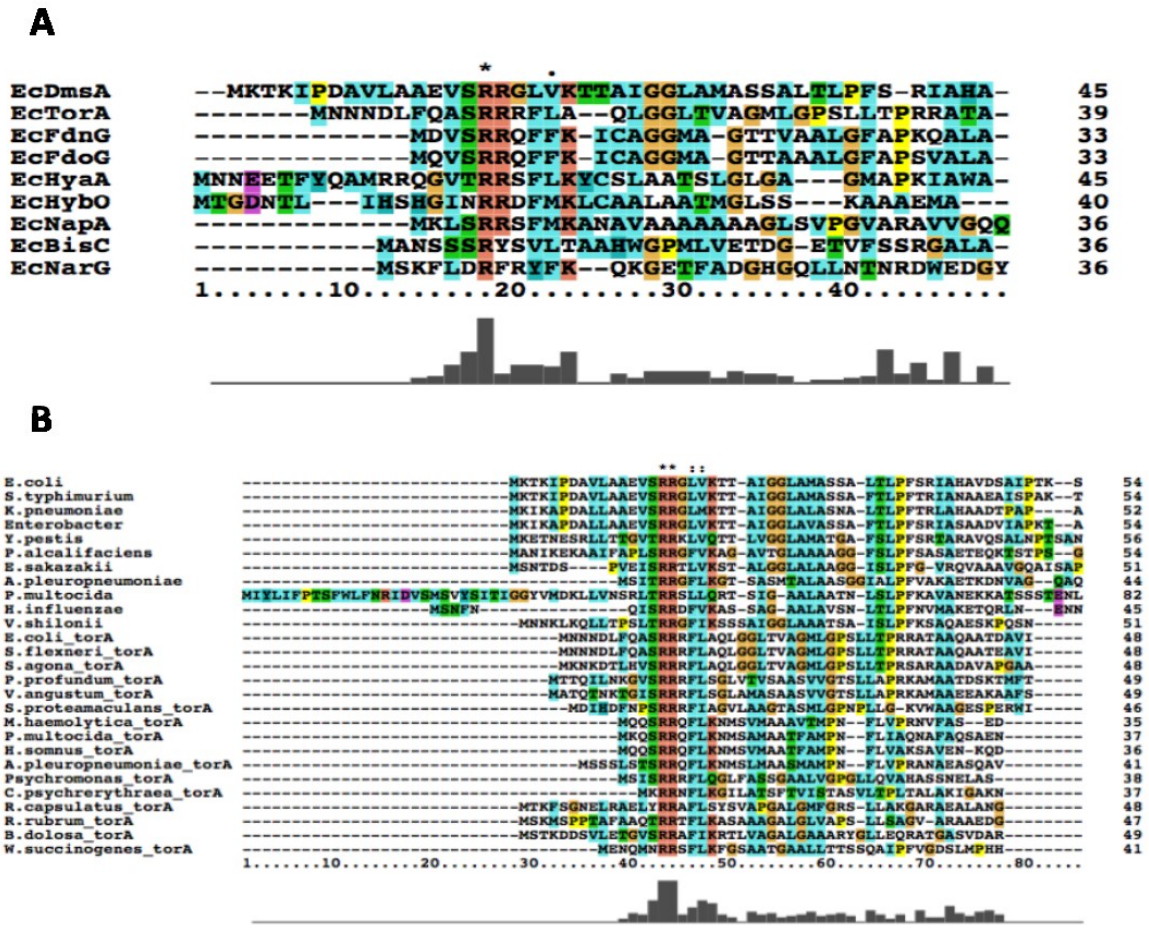


Fig 2.8 Sequence analysis of Tat signals A) Multiple sequence alignment of Tat signal sequences of the closely related DmsA and TorA B) Multiple sequence alignment of all Tat signals are recognized by REMP chaperones from *E. coli*.

2.3.4 Effect of truncation of ssDmsA on DmsD and TatC binding

To check for the efficacy of the signal sequence to bind to DmsD and TatC, a series of truncations were made from the N- and C- terminal. These truncation mutants of ssDmsA were tested for binding to DmsD. Truncation constructs removed five amino acids at a time from either end of the signal sequence (Figure 2.1.C). In the first round of truncation only C-terminal or N-terminal truncation were tested. Based on these results, double mutants were tested in subsequent cycle. They were tested for binding to DmsD in both *in vivo* and *in vitro* conditions. The pull down of DmsD with these truncations on ssDmsA-GFP by co expression provided *in vivo* binding (Fig2.9).

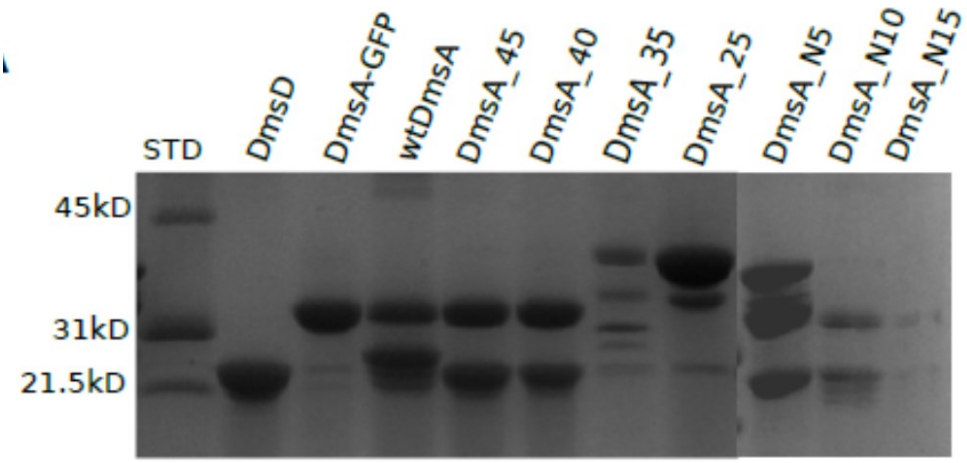


Fig 2.9. SDS-PAGE of DmsD pull down with ssDmsA-GFP truncations by co-expression *in vivo*

The deletion of first 10 residues in N-terminal had no significant effect on DmsD binding, except the change in the ratios of the monomeric versus dimeric complex formation compared to wild type ssDmsA (Fig 2.10). Ten amino acids could be deleted from the C-terminus before a significant loss of binding occurred. Altogether, binding to DmsD required a minimum of 29 amino acids in the signal sequence from residue 13 to 41 (Figure 2.1.C).

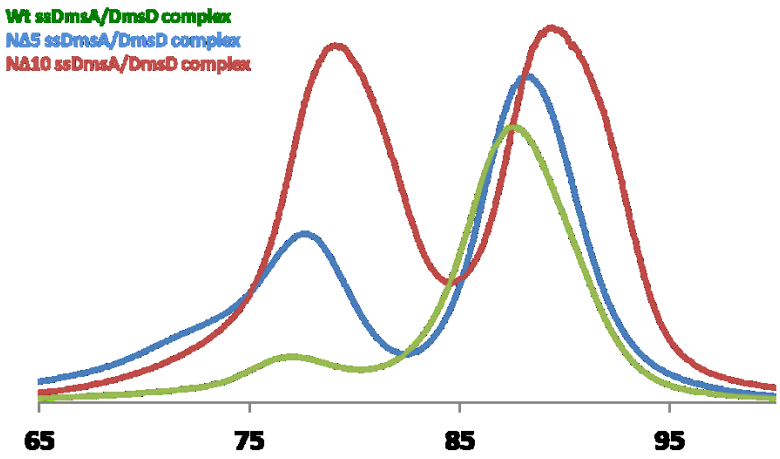


Fig 2.10. Size exclusion chromatography of ssDmsA-GFP truncations and DmsD pull down. Shift in distribution among monomeric and dimeric complex in ssDmsA-GFP and DmsD (green), NΔ5ssDmsA-GFP and DmsD (blue) and NΔ10ssDmsA-GFP and DmsD (red)

Then all the truncated constructs of ssDmsA-GFP have been tested for its ability to form complex with EcTatBC. Pull down the EcTatBC complex with purified tagged protein of ssDmsA-GFP to determine and map the region in signal peptide responsible for this tetrameric complex formation. From the figure 8 we can clearly observe that there is different forms of complex formation between the EcTatBC and NΔssDmsA/DmsD. The residue deletion from C-terminal of ssDmsA-GFP signal peptide has no effect on the complex formation.

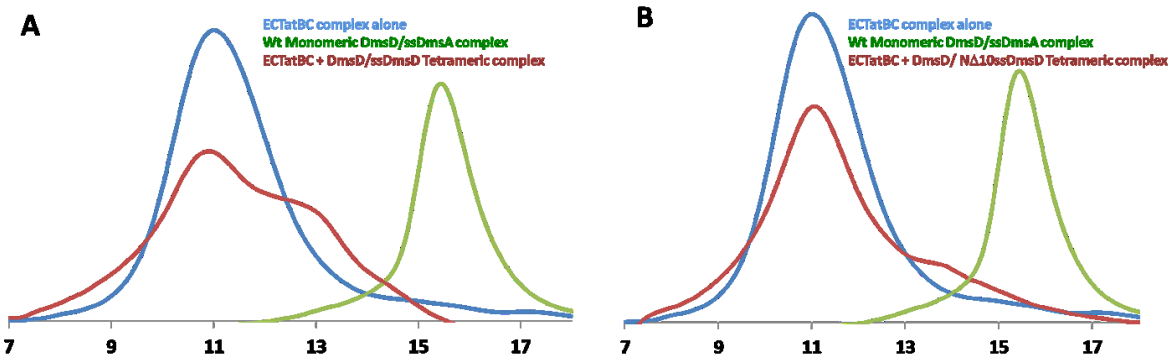


Fig 2.11 Difference between tetrameric complex formation of TatBC - DmsD/ssDmsA and TatBC - DmsD/ NΔ10ssDmsA A) Size exclusion chromatography of monomeric DmsD-ssDmsA (green), TatBC (blue) and tetrameric comlex of DmsD-ssDmsA and TatBC (red) B) Size exclusion chromatography of monomeric DmsD-ssDmsA (green), TatBC (blue) and tetrameric comlex of DmsD- NΔ10ssDmsA and TatBC (red)

2.3.5 Sequence determinants of the minimal signal sequence

Alanine scanning of the minimal 29 amino acid long DmsA signal sequence was used to establish the important recognition elements for DmsD binding. Each amino acid was individually replaced with either Ala or Ser (in the case of Ala). We assayed each mutant for the ability to form complex by co-expression and then affinity capture on a Ni-resin. ssDmsA-GFP-6xHis construct co-expressed with MBP-DmsD. Using this construct, the visualization of GFP fluorescence allowed for quantification of the expression level of each mutant indicating that they was no significant effect on expression (Fig 2.12).

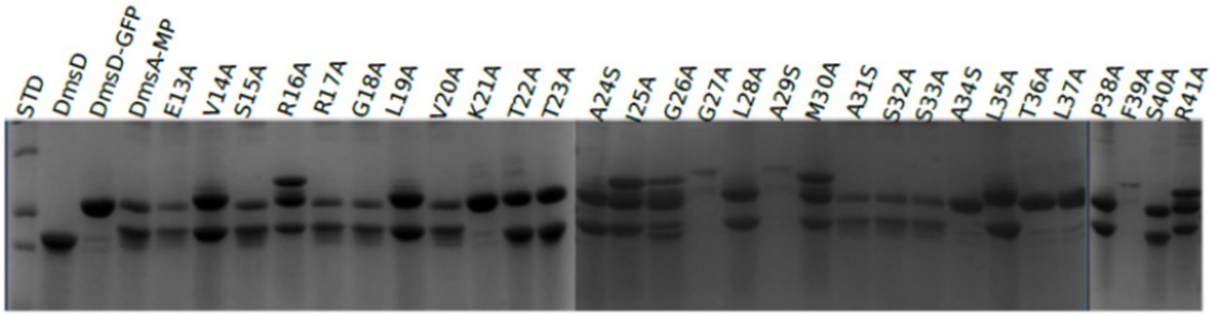


Fig 2.12 SDS PAGE of Ni-NTA affinity capture of MBP-DmsD by alanine scanning mutation on the minimal signal region of ssDmsA-GFP

Additionally, the effect was confirmed *in vitro* by mixing saturated amount of purified DmsD with mutants of the ssDmsA-6XHis and followed by affinity purification. The presence of a stable complex is indicated by the ability to capture DmsD visualized by SDS-PAGE (Fig.2.13). Based on this analysis, Lys 21, Gly27, Ala29, Ala34, Thr36, Leu37 and Phe 39 were all required for complex formation. Mutation of the invariant RR motif completely abolishes the Tat-specific export [37] but had no effect on DmsD binding

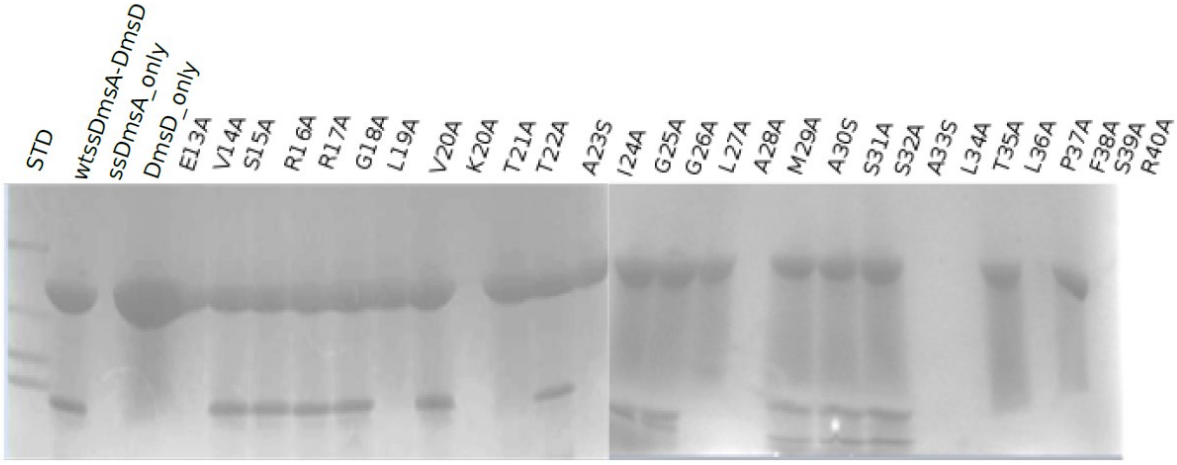


Fig 2.13 SDS PAGE of Ni-NTA affinity capture of purified DmsD with mutants of the ssDmsA-6XHis by alanine scanning mutation on the minimal signal region of ssDmsA-GFP

2.3.6 Kinetic analysis of DmsD/ssDmsA complex formation and interactions with TatBC

We used surface plasmon resonance (SPR) to obtain association and dissociation rate constants during complex formation between DmsD and ssDmsA along with the higher order TatBC interaction. Kinetic rate constants for the binding of the Tat signal peptide of DmsA to DmsD were determined by flowing ssDmsA-GFP to a chip containing immobilized DmsD. As a control, free GFP showed no interaction with immobilized DmsD indicating that any interaction was due solely to the signal sequence.

The key to a successful SPR experiment is the ability to saturate the surface of the chip with the Ligand. We were unable to immobilize the DmsD effectively by affinity tag; therefore, DmsD was attached to the chip by direct amine coupling. We tested a variety of conditions and were able to regenerate the chip using 150mM Nacl and 100 mM glycin buffer pH 9.5. This fully removed the ssDmsA-GFP without disrupting the native conformation of DmsD. These conditions allowed for reproducible binding of ssDmsA to DmsD.

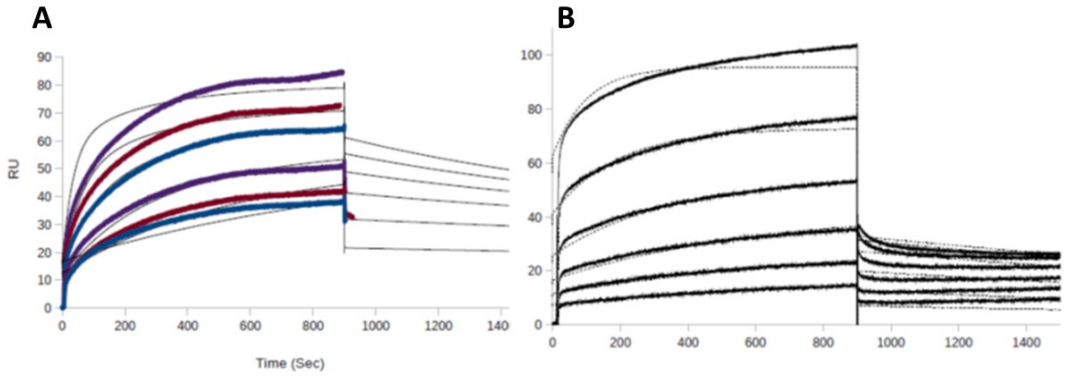


Fig 2.14 SPR sensogram to obtain association and dissociation rate constants of ssDmsA constructs A) ssDmsA-GFP to immobilized DmsD B) ssDmsA and DmsD with TatBC by immobilized TatBC

The sensogram of wild type peptide and other truncation mutants were shown (Figure 2.14). Figure 2.14 shows the experimental sonograms obtained, association of analyte ssDmsA (wt or mutants) with the DmsD proceeded for 10 minutes and dissociation in analyte free buffer was typically 15 minutes. For each titration, a zero analyte surface as a control was subtracted from test sensogram. The kinetic rate constants (K_a and K_d), as well as equilibrium dissociation constant (K_D) were estimated by global fitting analysis of the titration curve to the 1:1 [38] . The results are summarized in Table 2.2.

Table 2.2 Kinetic rate constants (K_a and K_d), as well as equilibrium dissociation constant (K_D) for ssDmsA-GFP and DmsD interaction and Pull down interactions.

ssDmsA	Pull down with DmsD	K_a X 10^4 (1/Msec)	K_d X 10^{-4} (1/sec)	K_D (nM)
wt	++	16.4 ±0.8	2.2±0.08	56.25±3.8
45	++	11.0 ±1.3	2.5±0.3	55.88±4.9
40	++	13.0 ±1.1	2.7±0.09	64.5±7.6
35	+	13.8 ±2.0	18.0±0.8	59.1±5.7
30	-	4.5±0.3	51.0±1.4	82.0±5.6
25	-	5.6±0.7	57.7±1.9	195.4±12
Ntrun5	++	16.3 ± 0.4	3.4 ± 0.6	58.30± 2.8
Ntrun10	++	12.8 ± 0.8	4.2 ± 0.2	55.98 ± 3.4
Ntrun15	+	3.8±0.3	23.0±1.1	166.9±9.7
DmsD_Minimum peptide	-	10.7±0.5	2.4±0.1	56.51±2.9

Good fitting of experimental data to the calculated curves has been observed, suggesting the correctness of the used fitting model. It provided the insight into the association and dissociation kinetics of the interaction between either wide type or truncated signal peptide of ssDmsA and DmsD. The slower association rate for ssDmsA_wt (K_a = 16.4 X 10^4 Ms^{-1}) and resultant complex between the ssDmsA_wt and DmsD were stable as illustrated by slow dissociation rate (K_d =2.2 X $10^{-4}s^{-1}$). Kinetic analysis of the sensogram curve suggests that ssDmsD_45 is more preferred (K_D = 55.88 nM) over other truncated forms. But the stable associations were observed upto 15 amino acid deletions, with considerable decrease in affinity. N-terminal truncations had no significant effect on DmsD binding and consistent with

the earlier results. The derived minimal peptide has similar K_a and K_d value suggesting that the essential regions of the signal sequence for the interaction with DmsD.

To analyze the interaction of ssDmsA and DmsD with TatBC by SPR, we immobilized TatBC from *E.coli*. (Table 2.3). These interactions were also observed in affinity capture experiments (Fig 2.15).

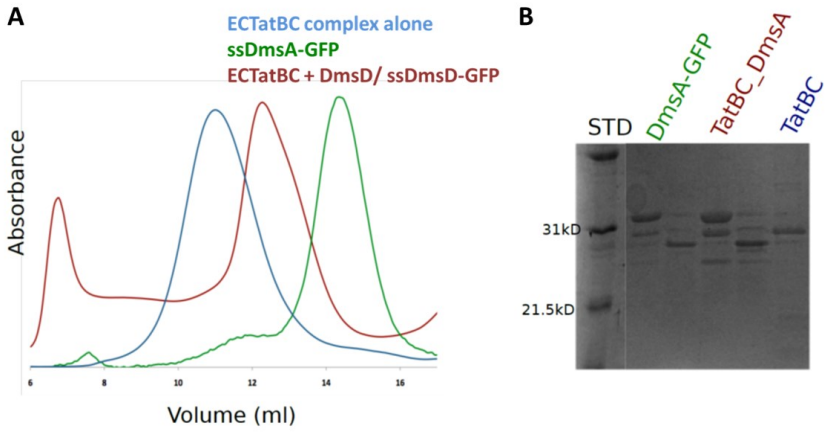


Fig 2.15 Complex formation of TatBC, DmsD and ssDmsA-GFP used in SPR analysis A) Size exclusion chromatography of ssDmsA-DmsA (green), TatBC (blue) and tetrameric comlex of DmsD-ssDmsA-GFP and TatBC (red) B) SDS PAGE of tetrameric complex showing ssDmsA-GFP, TatBC and TatBC + DmsD + ssDmsA-GFP

The affinity constants were higher for the DmsD/ssDmsA complex compared to the individual components. Interestingly, although the N-terminal truncation had no effect on the DmsD/ssDmsA complex formation, truncations significantly lowered the affinity towards TatBC suggesting a role in the higher order complex formation. Moreover, this effect was even seen in measurements using ssDmsA alone. It is worth to note that the derived minimal peptide has significantly less affinity towards the TatBC complex. Even though minimal peptide has showed no difference compared to wild type ssDmsA in the formation of ssDmsA/DmsD complex. This result denotes that N-terminal of signal peptide may be involved in the interaction with translocon complex. Another striking observation is that dimeric form of DmsD/ssDmsA complex has less affinity towards TatBC compared to monomeric form which is on par with the result of affinity capture experiments.

Table 2.3 kinetic rate constants (K_a and K_d), as well as equilibrium dissociation constant (K_D) for ssDmsA and DmsD with TatBC by immobilized TatBC

ssDmsA or and DmsD	K_a X 10^4 (1/Msec)	K_d X 10^{-4} (1/sec)	K_D (µM)
ssDmsA_wt	9.0	5.2	5.8
ssDmsA_45	12.3	3.4	5.3
ssDmsA_40	15.2	4.2	6.8
ssDmsA_35	20.8	1.6	8.2
ssDmsA_30	ND	ND	11.4
ssDmsA_25	ND	ND	10.4
ssDmsA_Ntrun5	ND	ND	7.3
ssDmsA_Ntrun10	30.7	2.3	15.6
ssDmsA_Ntrun15	ND	ND	18.7
DmsA_Minimum peptide	26.5	1.7	19.8
ssDmsAwt RR mutants	ND	ND	8.8
DmsD alone	1.1	3.1	2.8
ssDmsD/DmsD monomer	3.0	5.1	1.7
ssDmsD/DmsD Dimer	ND	ND	13.2

2.3.7 Characterization of DDF and mutants

As the interaction between both was known, a chimeric protein DDF having the *E. coli* signal sequence of DmsA (first 53 amino acids) construct followed by DMSD was designed with an hypothesis that the Tat Signal sequence and the chaperon would interact either by folding itself into the chaperon or in a dimeric manner where the signal of one construct interacts with the chaperon subunit of a second construct. The clone of signal sequence of DmsA (first 53 amino acids) construct followed by DmsD followed by a 6-Histine tag in a pET 33b expression vector was used for overexpression (Fig. 2.16)

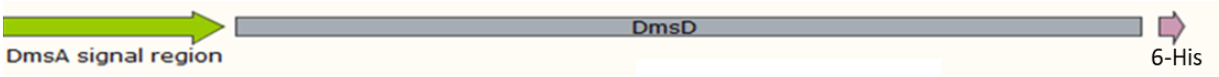


Fig 2.16 DDF construct by fusion of ssDmsA with DmsD followed by 6-His tag

The protein expression was optimised and overexpressed in Bl21G DTat cells by inducing culture with 0.5 mM IPTG and incubation at 37°C for 4hours to obtain the cytoplasmic soluble fraction. The overexpressed protein was purified to homogeneity using a Ni–NTA affinity column (Fig2.17).

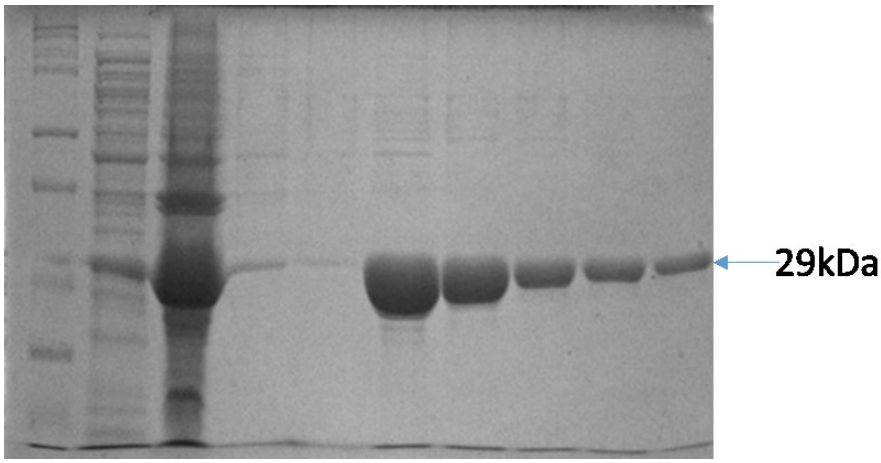


Fig. 2.17 Ni-NTA Purification of DDF on SDS PAGE

Following Ni-NTA purification gel filtration (column: Superdex S200 10/300 increase) was carried out to remove any contaminant proteins, soluble aggregates (Fig2.18). The molecular weight of peak at 16.2ml approximately is ~29kDa which is the monomer and that of the 14.7ml peak is ~60kDa which is the dimeric form of DDF.

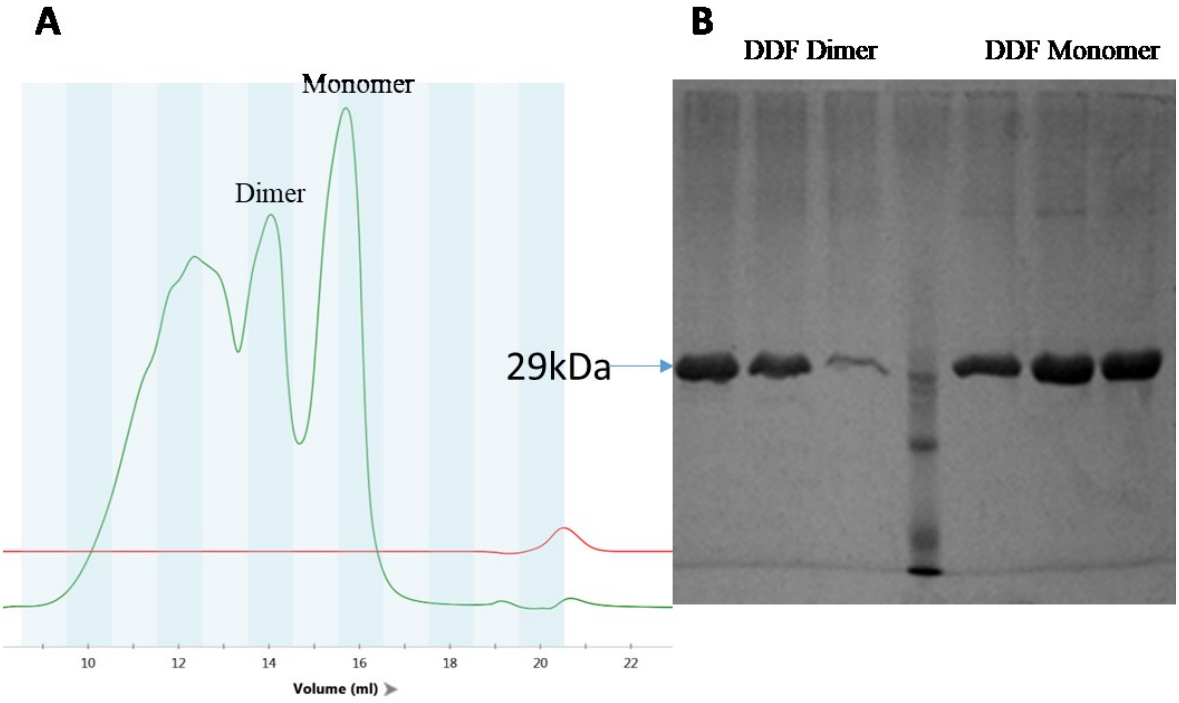


Fig 2.18 Size exclusion chromatography of DDF A) The molecular weight of peak at 16.2ml approximately is ~29kDa which is the monomer and that of the 14.7ml peak is ~60kDa which is the dimeric form of DDF B) SDS PAGE of DDF monomer and dimer

Two new constructs DDF Δ2-10 and DDFΔ33-53 were created by site directed mutagenesis based on the ssDMSA signal truncation studies (Fig2.19). The mutants were made using quick change deletion primers and confirmed by sequencing.

Fig 2.19 DDF mutants by truncation of ssDmsA 2-10 and C-terminal 33-53 fused with DmsD followed by 6-His tag DDFΔ2-10 and DDFΔ33-53 respectively

The protein expression for both the mutants were optimised and overexpressed in Bl21G DTat cells by inducing culture with 0.5 mM IPTG and incubation at 37°C for 4hours. Ni-NTA affinity chromatography was used for purification of protein from lysed supernatant (Fig 2.20).

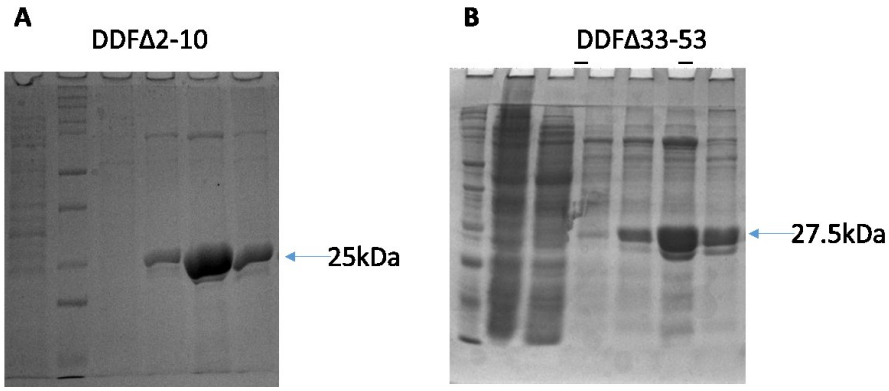


Fig 2.20 Ni-NTA Purification of DDF mutants on SDS PAGE A) DDFΔ2-10 B) DDFΔ33-53

Ni-NTA purified DDFΔ33-53 was subjected Size exclusion chromatography using GE S200 Sephacryl 16/60 column in buffer containing 50mM HEPES and 2mM DTT to separate out monomer and dimer (Fig 2.21). Only monomer was obtained after size exclusion chromatography this was probably due to the loss of flexibility of the signal region.

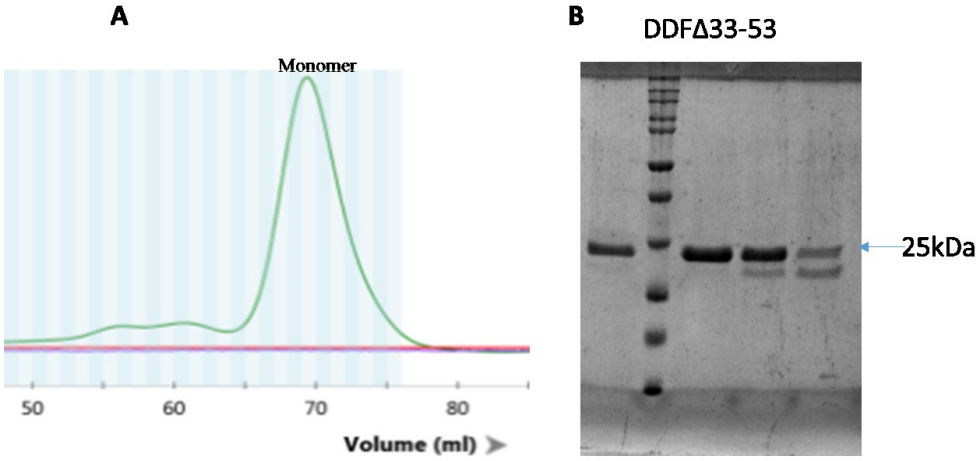


Fig 2.21. Size exclusion chromatography of DDFΔ33-53 A) The molecular weight of peak at 71.2ml approximately is ~25kDa which is the monomer DDFΔ33-53 B) SDS PAGE of DDFΔ33-53 monomer

DDFΔ2-10 was subjected Size exclusion chromatography using GE S200 superdex10/300 column in buffer containing 50mM HEPES and 2mM DTT to separate out monomer and dimer (Fig 2.22). The molecular weight of peak at 16.3ml approximately is ~27.5kDa which is the monomer and that of the 14.8ml peak is ~55kDa which is the dimeric form of DDFΔ2-10.

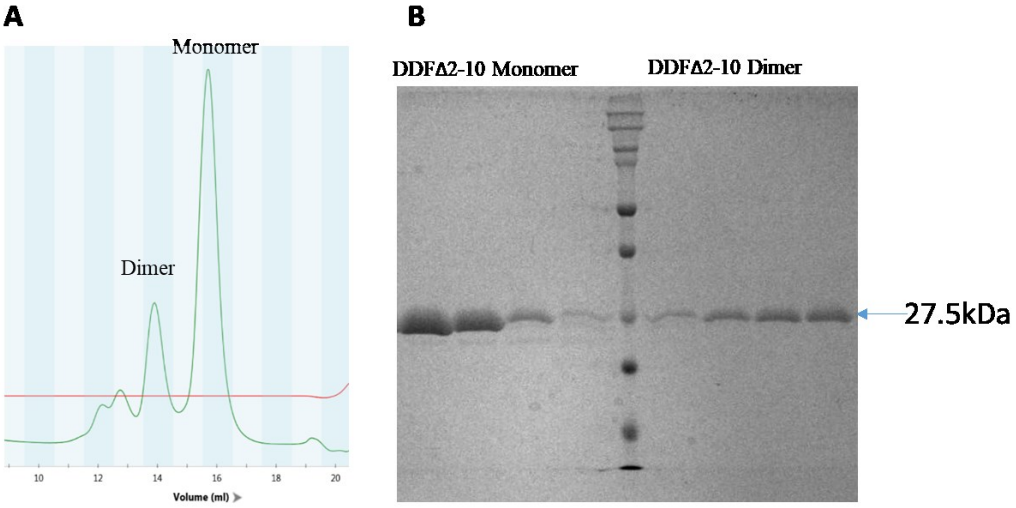


Fig 2.22 Size exclusion chromatography of DDFΔ2-10 A) The molecular weight of peak at 16.3ml approximately is ~27.5kDa which is the monomer and that of the 14.8ml peak is ~55kDa which is the dimeric form of DDFΔ2-10 B) SDS PAGE of DDFΔ2-10 monomer and dimer

2.3.8 Crystallization of DDF and mutants

Purified protein from Size exclusion chromatography was concentrated up to 15 mg/ml using Amicon ultra 10kDa and 30 kDa cutoff centrifugal filters (Millipore, USA) at 4 ºC at 4000 rpm for Monomer and dimer fractions respectively. The proteins were screened against several commercially available crystallization screens including, Index (Hampton Research Corp.) Nextal PACT (Qiagen), Nextal Protein Complex Suite (Qiagen) and JCSG plus (Molecular Dimensions) by vapour diffusion method in 96 well MRC2 sitting drop plates Hampton MRC-SD2 and SD3 plates with multiple protein to screen ratios of 1:1, 1:2 and 2:1 in 400ul sitting drops and incubated at 20°C. One batch of DDF was also set up with 1:100 thermolysin: DDF for chopping out protruding hydrophobic region of the protein. DDF Dimer

crystals were obtained in 0.1M- Phosphate/Citrate pH4.2; 40%v/v Ethanol; 5%w/v PEG1000 and 0.1M Sodium- Citrate pH5; 8%w/v PEG8000 after 3-4weeks at 20°C (Fig2.23).

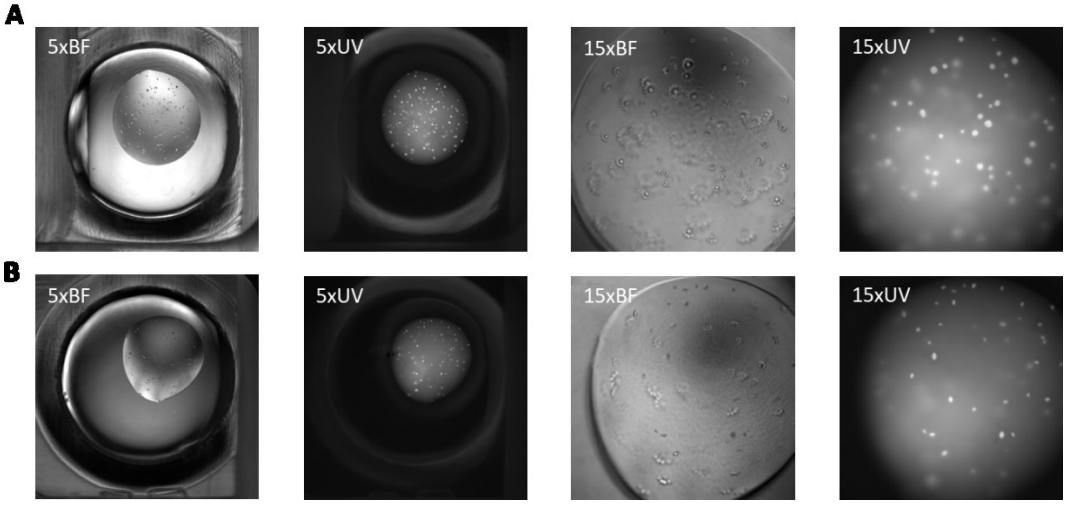


Fig.2.23 Crystal images of DDF Dimer in Bright field and UV A) 0.1M- Phosphate/Citrate pH4.2; 40%v/v Ethanol; 5%w/v PEG1000 and B)0.1M Sodium- Citrate pH5; 8%w/v PEG8000 after 3-4weeks at 20°C

DDF Monomer crystals were obtained in thermolysin treated protein in 0.2M Lithium Sulphate monohydrate, 0.1M BisTris, pH5.5; 25%w/v; PEG335 and 0.1M BisTris pH6.5; 20%w/v; PEGmmE 5000 (Fig2.24).

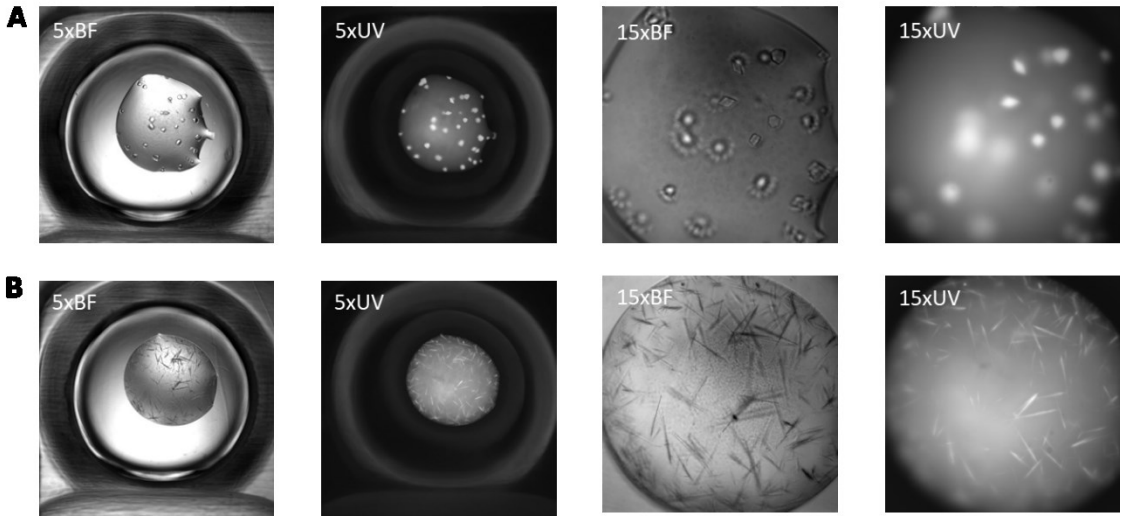


Fig.2.24 Crystal images of Thermolysin treated DDF Monomer in Bright field and UV A) 0.2M Lithium Sulphate monohydrate, 0.1M BisTris, pH5.5; 25%w/v; PEG335 and B)0.1M BisTris pH6.5; 20%w/v; PEGmmE 5000

Additional additive screening and manual crystal screening was performed for DDF monomer and dimer but there was no improvement in crystals. No crystals were obtained for both DDFΔ2-10 and DDFΔ33-53 after extensive screening. The original crystals were screened at BL21px beamline at RRCAT Indore but failed to diffract failed to diffract.

2.3.9 Differential Scanning Fluorimetry (DSF)

DSF analysis was done for different combinations of DmsA signal Sequence, WtDmsD, DDF and DDFΔ2-10 to find out the complex with greater stability for proceeding with crystal trials for improved complex. The combinations were analysed in the Prometheus NT.48 nanoDSF (NanoTemper Technologies, Germany). Intrinsic fluorescence of the protein where the shift in fluorescence emission is plotted as the ratio between 350 and 330nm. The *T*m of the various combinations of proteins were studied in native condition. Tm was calculated from the first derivate of absorbance at 350nm (Fig2.25). Highest melting temperature was of DMSDp33 + DDF Dimer at 66.03°C and the lowest melting temperature was of DDF Monomer at 57.91°C (Table 2.4).

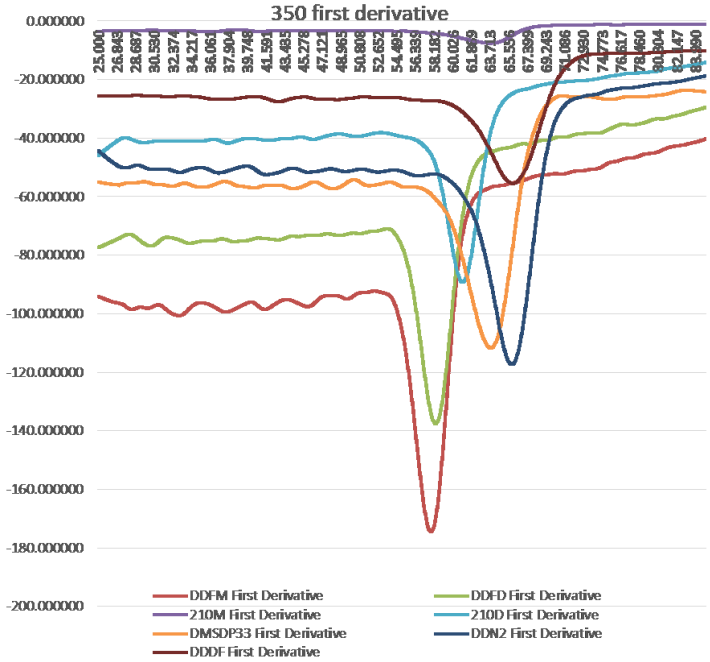


Fig.2.25 Melting temperature(T_m) analysis for different constructs using DSF by measurement of first derivative of absorbance at 350nm

Table 2.4 Melting temperature(T_m) analysis for different constructs using DSF by measurement of first derivative of absorbance at 350nm

Sr.No	Proteins	*T*m
1	DDF Dimer	58.35°C
2	DDF Monomer	57.91°C
3	DDFΔ2-10 Dimer	61.03°C
4	DDFdel2-10 Monomer	63.84°C
5	DMSDp33	63.84°C
6	DMSDp33 + DN2	65.9°C
7	DMSDp33 + DDF Dimer	66.03°C

2.4 Discussions

This work reveals that DmsD binds to the N-terminal signal peptide of ssDmsA and forms two distinct populations. The results from the Biacore analysis and the SEC data or affinity capture were well correlated. Based on this data we have designed the minimum peptide for formation of complex, which consist of 29 amino acids. And the kinetic analysis from SPR suggest that this peptide is having faster association rate (K_a = 10.7 X 10^4 Ms^{-1}) compare to all other mutant and wild type, and the affinity is very close to the wild type peptide K_D = 56.5 nM.

The sequence comparisons of all Tat signal, which are having binding chaperons reveals some region are more conserved (Fig 2.8). In spite of the divergence in Sec and Tat transport systems, they do have features in common. The signal peptides of both pathways are evolutionarily well conserved and display the same tripartite organization. Even though overall conservation signal peptide found in Tat substrate do have several exclusive features as compared to those found in Sec substrates. Tat substrate signal n-region is much longer, and they have twin arginine motif in the n- and h- region boundary. The h-region in Tat substrate usually longer and also has less number of hydrophobic residues compare to Sec signal. An additional lysine or arginine is often present in the c-region of Tat signal peptide, which serve as sec avoidance motif [39].

The sequence alignment provides some inferences with regard to the mechanism of substrate specificity by REMPs. Homology of the entire REMP sequences and the hydrophobic regions of signal peptide imply recognition specificity, with importance on the position and length of the continuous hydrophobic stretch following the Tat motif. The varying architecture of this hydrophobic stretch is likely to adapt to the binding pocket of different REMP structural classes such as those recently described [25]. The higher percentage of sequence conservation of DmsD/HybE than for observed NapD/HybE pairs suggest that homology of REMPs is not the only factor consider whether two REMPs will interact with the same signal peptide [40].

The truncation mutations of ssDmsA reveals that the C-terminal part of the ssDmsA is more directly involved in DmsD binding. In case TorA/TorD system truncated signal peptide of 10-36 of TorA essentially identical to full length and showed tighter binding [35]. Also showed that there is no preference over twin-lysine peptide, which suggest that twin arginine motif is not essential for chaperone recognition [35]. This is on par with our Ala scanning results where the replacement of RR with Ala has no effect on the DmsD binding (Figure 2.11). The another study suggested that Ser in -1 position and Leu in +2 position in the peptide were essential for the translocation [41].

A leucine rich region within the signal peptide h-region is shown to be involved in TorD binding both *in vivo* and *in vitro* [36], which correlates well with our alanine scanning results in ssDmsA/DmsD system (Fig 2.12). Before the pre-protein interacts with Tat translocon it interact with inner membrane that is stabilized by both electrostatic and hydrophobic contribution [42]. The studies are revealed that the positively charged signal peptide region and anionic lipid head groups play major role in the association of preprotein with the lipid bilayer. And h-region of the signal peptide interacts with the apolar environment of phospolipid [42]. It has been suggested that increasing hydrophobicity of h-region favors the membrane binding rather chaperone binding.

Here we report for the first time that the DmsD/ssDmsA interaction analysis showed undoubtedly the presence of two forms of complex. And equilibrium of these forms can be modulated by N-terminal truncation. Further *in vitro* analysis showed that the concentration of DmsD also influence the architecture of the complex.

The NarJ/NarG system, which is similar to DmsA/DmsD has been investigated thoroughly in structural level by NMR. This result suggests that NarG signal peptide is in helical conformation and also there is change in conformation in NarJ upon the signal peptide binding. pH depended modulation of the peptide binding affinity has also been observed [43].

Translocation involves a very intimate association of signal peptide with the receptor complex binding site [44]. Previous reports suggested that DmsD alone interact with Tat apparatus especially TatB and TatC [45]. Here we could able isolate and identify the regions responsible for this interaction between TatBC with DmsD, ssDmsA and ssDmsA/DmsD complex. The N-terminal of the signal peptide has been mainly involved in the TatBC recognition, which includes the Tat motif. The effect of mutation on the Tat motif has been well studied on the translocation previously [2]. The N- terminal of Tat motif has seems to play some role in the complex formation. Even though the deletion of the N-terminal upstream of the Tat motif still forms complex with TatBC, the association is weak and forms different type of complex. The C-terminal of the signal peptide doesn't seem affect binding.

TatB is found in contact with the entire signal sequence and adjacent parts of mature part of the protein, but in TatC the interaction restricted to a discrete area around the consensus motif [2].

Among monomeric and dimeric types of ssDmsA/DmsD complex, only the monomeric population interacts with TatBC and forms the tetrameric complex. In case of monomeric complex of DmsD/ssDmsA, both interaction site of signal peptide and the DmsD may be still accessible for the Tat translocase assembly. The dimeric complex could be formed due to stable and very strong interaction of DmsD molecules. The DmsD concentration depended binding towards signal peptide has been observed previously [34] and in this work.

It is possible that both of these types of complex may be needed and serve as mechanism for the stringent incorporation of co-factor and quality control of DmsA maturation. The model that emerged from this study has been represented in Figure 2.26.

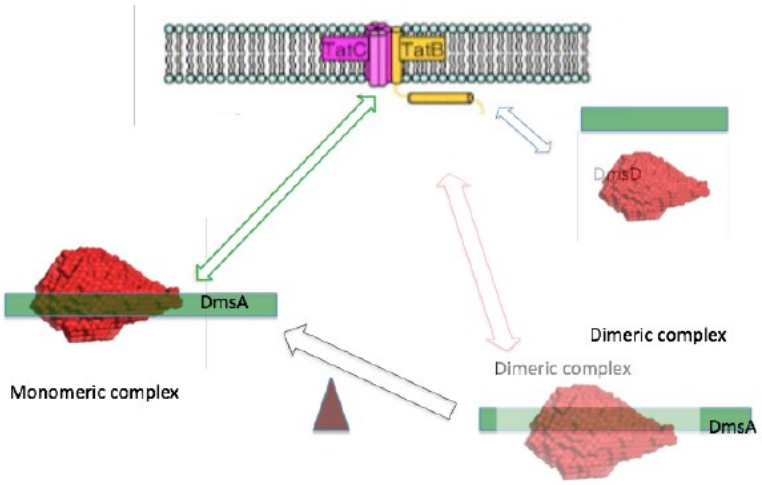


Fig 2.26 Postulated quality control mechanism of DmsA. The Preprotein is bound to a DmsD dimeric complex, after co-factor loading one of the DmsD subunits dissociates allowing the movement and targeting to the TatBC receptor complex

After the nascent peptide of DmsA emerge from ribosome DmsD form dimeric complex. So the DmsA signal has been complexly masked and it is inaccessible for the translocon and also this dimeric complex obstructs the DmsD interaction with the receptor complex. Once the cofactor has been inserted and after completion of sub-unit assembly the DmsD dissociate to monomeric form which has now exposed all its interaction sites over signal peptide as well as in DmsD and this monomeric complex readily interact with translocase assembly. This is supported by the DSF experiments where the highest melting temperature is for the DMSDp33 + DDF Dimer at 66.03°C which shows it is a stable complex and the lowest melting temperature is that of DDF Monomer at 57.91°C which resembles a DmsD monomer bound to the DmsA signal sequence. There is enough evidence that the mature part of DmsA interact with DmsD [46]. The other possible scenario is DmsD may form monomeric or dimeric complex in normal circumstance but once the degradation increases due to improper folding and assembly [47] it may trigger the DmsD synthesis that eventually increases the relative concentration of DmsD with respect to DmsA and equilibrium shifted towards the dimeric form. Dimeric form restricts the access to the translocon prematurely and leads to stringent quality control of DmsA maturity as a functional assembly. One interesting and supporting fact to this model has been observed in this studies was once the N-terminal of the

signal peptide truncated there is shift in monomeric and dimeric form of the complex which also indirectly tells us the steric hindrance may be playing role in the complex regulation.

The crystallization trials suggest that the disorder of the DmsA signal sequence is a major factor of non-diffraction. The DMSDp33 + DDF Dimer with a high Melting temperature is a good candidate for further crystallization trials.

2.5 References

[1] A. M. Chaddock, A. Mant, I. Karnauchov, S. Brink, R. G. Herrmann, R. B. Klosgen, *et al.*, "A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the delta pH-dependent thylakoidal protein translocase," *EMBO J,* vol. 14, pp. 2715-22, Jun 15 1995.

[2] M. Alami, I. Luke, S. Deitermann, G. Eisner, H. G. Koch, J. Brunner, *et al.*, "Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli," *Mol Cell,* vol. 12, pp. 937-46, Oct 2003.

[3] M. Muller and R. B. Klosgen, "The Tat pathway in bacteria and chloroplasts (review)," *Mol Membr Biol,* vol. 22, pp. 113-21, Jan-Apr 2005.

[4] P. A. Lee, D. Tullman-Ercek, and G. Georgiou, "The bacterial twin-arginine translocation pathway," *Annu Rev Microbiol,* vol. 60, pp. 373-95, 2006.

[5] B. C. Berks, F. Sargent, and T. Palmer, "The Tat protein export pathway," *Mol Microbiol,* vol. 35, pp. 260-74, Jan 2000.

[6] U. A. Ochsner, A. Snyder, A. I. Vasil, and M. L. Vasil, "Effects of the twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis," *Proc Natl Acad Sci U S A,* vol. 99, pp. 8312-7, Jun 11 2002.

[7] N. Pradel, C. Ye, V. Livrelli, J. Xu, B. Joly, and L. F. Wu, "Contribution of the twin arginine translocation system to the virulence of enterohemorrhagic Escherichia coli O157:H7," *Infect Immun,* vol. 71, pp. 4908-16, Sep 2003.

[8] R. L. Jack, G. Buchanan, A. Dubini, K. Hatzixanthis, T. Palmer, and F. Sargent, "Coordinating assembly and export of complex bacterial proteins," *EMBO J,* vol. 23, pp. 3962-72, Oct 13 2004.

[9] D. Sambasivarao, R. J. Turner, J. L. Simala-Grant, G. Shaw, J. Hu, and J. H. Weiner, "Multiple roles for the twin arginine leader sequence of dimethyl sulfoxide reductase of Escherichia coli," *J Biol Chem,* vol. 275, pp. 22526-31, Jul 21 2000.

[10] B. C. Berks, F. Sargent, E. De Leeuw, A. P. Hinsley, N. R. Stanley, R. L. Jack*, et al.*, "A novel protein transport system involved in the biogenesis of bacterial electron transfer chains," *Biochim Biophys Acta,* vol. 1459, pp. 325-30, Aug 15 2000.

[11] F. Sargent, "The twin-arginine transport system: moving folded proteins across membranes," *Biochem Soc Trans,* vol. 35, pp. 835-47, Nov 2007.

[12] R. J. Turner, A. L. Papish, and F. Sargent, "Sequence analysis of bacterial redox enzyme maturation proteins (REMPs)," *Can J Microbiol,* vol. 50, pp. 225-38, Apr 2004.

[13] I. J. Oresnik, C. L. Ladner, and R. J. Turner, "Identification of a twin-arginine leader-binding protein," *Mol Microbiol,* vol. 40, pp. 323-31, Apr 2001.

[14] M. Muller, "Twin-arginine-specific protein export in Escherichia coli," *Res Microbiol,* vol. 156, pp. 131-6, Mar 2005.

[15] B. C. Berks, "A common export pathway for proteins binding complex redox cofactors?," *Mol Microbiol,* vol. 22, pp. 393-404, Nov 1996.

[16] S. Cristobal, J. W. de Gier, H. Nielsen, and G. von Heijne, "Competition between Sec- and TAT-dependent protein translocation in Escherichia coli," *EMBO J,* vol. 18, pp. 2982-90, Jun 1 1999.

[17] D. Tullman-Ercek, M. P. DeLisa, Y. Kawarasaki, P. Iranpour, B. Ribnicky, T. Palmer*, et al.*, "Export pathway selectivity of Escherichia coli twin arginine translocation signal peptides," *J Biol Chem,* vol. 282, pp. 8309-16, Mar 16 2007.

[18] C. A. McDevitt, G. Buchanan, F. Sargent, T. Palmer, and B. C. Berks, "Subunit composition and in vivo substrate-binding characteristics of Escherichia coli Tat protein complexes expressed at native levels," *FEBS J,* vol. 273, pp. 5656-68, Dec 2006.

[19] N. R. Stanley, T. Palmer, and B. C. Berks, "The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in Escherichia coli," *J Biol Chem,* vol. 275, pp. 11591-6, Apr 21 2000.

[20] E. Holzapfel, G. Eisner, M. Alami, C. M. Barrett, G. Buchanan, I. Luke*, et al.*, "The entire N-terminal half of TatC is involved in twin-arginine precursor binding," *Biochemistry,* vol. 46, pp. 2892-8, Mar 13 2007.

[21] E. M. Strauch and G. Georgiou, "Escherichia coli tatC mutations that suppress defective twin-arginine transporter signal peptides," *J Mol Biol,* vol. 374, pp. 283-91, Nov 23 2007.

[22] C. Maurer, S. Panahandeh, A. C. Jungkamp, M. Moser, and M. Muller, "TatB functions as an oligomeric binding site for folded Tat precursor proteins," *Mol Biol Cell,* vol. 21, pp. 4151-61, Dec 2010.

[23] J. S. Kostecki, H. Li, R. J. Turner, and M. P. DeLisa, "Visualizing interactions along the Escherichia coli twin-arginine translocation pathway using protein fragment complementation," *PLoS One,* vol. 5, p. e9225, 2010.

[24] M. P. DeLisa, P. Samuelson, T. Palmer, and G. Georgiou, "Genetic analysis of the twin arginine translocator secretion pathway in bacteria," *J Biol Chem,* vol. 277, pp. 29825-31, Aug 16 2002.

[25] J. Maillard, C. A. Spronk, G. Buchanan, V. Lyall, D. J. Richardson, T. Palmer*, et al.*, "Structural diversity in twin-arginine signal peptide-binding proteins," *Proc Natl Acad Sci U S A,* vol. 104, pp. 15641-6, Oct 2 2007.

[26] A. Dubini and F. Sargent, "Assembly of Tat-dependent [NiFe] hydrogenases: identification of precursor-binding accessory proteins," *FEBS Lett,* vol. 549, pp. 141-6, Aug 14 2003.

[27] W. Graubner, A. Schierhorn, and T. Bruser, "DnaK plays a pivotal role in Tat targeting of CueO and functions beside SlyD as a general Tat signal binding chaperone," *J Biol Chem,* vol. 282, pp. 7116-24, Mar 9 2007.

[28] E. Holzapfel, M. Moser, E. Schiltz, T. Ueda, J. M. Betton, and M. Muller, "Twin-arginine-dependent translocation of SufI in the absence of cytosolic helper proteins," *Biochemistry,* vol. 48, pp. 5096-105, Jun 16 2009.

[29] N. Ray, J. Oates, R. J. Turner, and C. Robinson, "DmsD is required for the biogenesis of DMSO reductase in Escherichia coli but not for the interaction of the DmsA signal peptide with the Tat apparatus," *FEBS Lett,* vol. 534, pp. 156-60, Jan 16 2003.

[30] W. S. Jong, C. M. ten Hagen-Jongman, P. Genevaux, J. Brunner, B. Oudega, and J. Luirink, "Trigger factor interacts with the signal peptide of nascent Tat substrates but does not play a critical role in Tat-mediated export," *Eur J Biochem,* vol. 271, pp. 4779-87, Dec 2004.

[31] S. K. Ramasamy and W. M. Clemons, Jr., "Structure of the twin-arginine signal-binding protein DmsD from Escherichia coli," *Acta Crystallogr Sect F Struct Biol Cryst Commun,* vol. 65, pp. 746-50, Aug 1 2009.

[32] C. M. Stevens, T. M. Winstone, R. J. Turner, and M. Paetzel, "Structural analysis of a monomeric form of the twin-arginine leader peptide binding chaperone Escherichia coli DmsD," *J Mol Biol,* vol. 389, pp. 124-33, May 29 2009.

[33] C. S. Chan, T. M. Winstone, L. Chang, C. M. Stevens, M. L. Workentine, H. Li*, et al.*, "Identification of residues in DmsD for twin-arginine leader peptide binding, defined through random and bioinformatics-directed mutagenesis," *Biochemistry,* vol. 47, pp. 2749-59, Mar 4 2008.

[34] T. L. Winstone, M. L. Workentine, K. J. Sarfo, A. J. Binding, B. D. Haslam, and R. J. Turner, "Physical nature of signal peptide binding to DmsD," *Arch Biochem Biophys,* vol. 455, pp. 89-97, Nov 1 2006.

[35] K. Hatzixanthis, T. A. Clarke, A. Oubrie, D. J. Richardson, R. J. Turner, and F. Sargent, "Signal peptide-chaperone interactions on the twin-arginine protein transport pathway," *Proc Natl Acad Sci U S A,* vol. 102, pp. 8460-5, Jun 14 2005.

[36] G. Buchanan, J. Maillard, S. B. Nabuurs, D. J. Richardson, T. Palmer, and F. Sargent, "Features of a twin-arginine signal peptide required for recognition by a Tat proofreading chaperone," *FEBS Lett,* vol. 582, pp. 3979-84, Dec 10 2008.

[37] M. Alami, D. Trescher, L. F. Wu, and M. Muller, "Separate analysis of twin-arginine translocation (Tat)-specific membrane binding and translocation in Escherichia coli," *J Biol Chem,* vol. 277, pp. 20499-503, Jun 7 2002.

[38] R. I. Masel, *Principles of adsorption and reaction on solid surfaces*. New York: Wiley, 1996.

[39] E. Bogsch, S. Brink, and C. Robinson, "Pathway specificity for a delta pH-dependent precursor thylakoid lumen protein is governed by a 'Sec-avoidance' motif in the transfer

peptide and a 'Sec-incompatible' mature protein," *EMBO J,* vol. 16, pp. 3851-9, Jul 1 1997.

[40] C. S. Chan, L. Chang, K. L. Rommens, and R. J. Turner, "Differential Interactions between Tat-specific redox enzyme peptides and their chaperones," *J Bacteriol,* vol. 191, pp. 2091-101, Apr 2009.

[41] S. Mendel, A. McCarthy, J. P. Barnett, R. T. Eijlander, A. Nenninger, O. P. Kuipers*, et al.*, "The Escherichia coli TatABC system and a Bacillus subtilis TatAC-type system recognise three distinct targeting determinants in twin-arginine signal peptides," *J Mol Biol,* vol. 375, pp. 661-72, Jan 18 2008.

[42] A. Shanmugham, H. W. Wong Fong Sang, Y. J. Bollen, and H. Lill, "Membrane binding of twin arginine preproteins as an early step in translocation," *Biochemistry,* vol. 45, pp. 2243-9, Feb 21 2006.

[43] S. Zakian, D. Lafitte, A. Vergnes, C. Pimentel, C. Sebban-Kreuzer, R. Toci*, et al.*, "Basis of recognition between the NarJ chaperone and the N-terminus of the NarG subunit from Escherichia coli nitrate reductase," *FEBS J,* vol. 277, pp. 1886-95, Apr 2010.

[44] F. Gerard and K. Cline, "The thylakoid proton gradient promotes an advanced stage of signal peptide binding deep within the Tat pathway receptor complex," *J Biol Chem,* vol. 282, pp. 5263-72, Feb 23 2007.

[45] A. L. Papish, C. L. Ladner, and R. J. Turner, "The twin-arginine leader-binding protein, DmsD, interacts with the TatB and TatC subunits of the Escherichia coli twin-arginine translocase," *J Biol Chem,* vol. 278, pp. 32501-6, Aug 29 2003.

[46] C. S. Chan, L. Chang, T. M. Winstone, and R. J. Turner, "Comparing system-specific chaperone interactions with their Tat dependent redox enzyme substrates," *FEBS Lett,* vol. 584, pp. 4553-8, Nov 19 2010.

[47] C. F. Matos, C. Robinson, and A. Di Cola, "The Tat system proofreads FeS protein substrates and directly initiates the disposal of rejected molecules," *EMBO J,* vol. 27, pp. 2055-63, Aug 6 2008.

Chapter III - Development of a Novel High-throughput Twin-Arginine Translocase Assay in Bacteria for Therapeutic Applications

3.1 Introduction

The twin-arginine translocase (Tat) pathway exclusively transports folded proteins in an ATP-independent manner across the bacterial cell membrane [1]. Common Tat substrates [2] include periplasmic enzymes in complex multiprotein respiratory redox systems, bacterial virulence factors, lipoproteins, and proteins involved in maintaining cell-wall integrity. These are mostly complexes containing cofactors, and many form oligomeric assemblies. Delivery of these proteins is mediated by special cytoplasmic chaperones (redox-enzyme maturation proteins, REMP) that bind to and mask the Tat signal, thereby preventing the futile export of immature protein [3]. The *Escherichia coli* genome encodes around 31 putative signal peptides [4] containing a twin-arginine motif. The Tat Pathway also plays an important role in bacterial pathogenesis [5], translocating certain virulence factors, and proteins involved in polysaccharide metabolism and biofilm formation. Apart from its significance as a potential drug target, the Tat pathway could also find applications in bioengineering, where its strict quality control can be used to produce complex therapeutic proteins in properly folded states or to rapidly screen for mutations that enhance protein solubility [6]. In this context, it would be highly beneficial to develop efficient and robust screening techniques for detecting the protein translocation activity of the pathway.

Currently, a few assays have been deployed to measure Tat activity. The list of *in vivo* and *in vitro* assay developed till now for Tat pathways are listed in Table 3.1.

Table 3.1. The currently available Tat Assays for identifying signal activity or inhibitor screening

Sr. No	Assay	Function	organism	Protein	Method
1	BlaTEM-1	Tat signal peptide Identification	*M. tuberculosis*	*E. coli* BlaTEM-1	β-lactam sensitivity of ΔblaC mutant of *M. tuberculosis*, lacking the chromosomally encoded β-lactamase BlaC [7]
2	*S. coelicolor* agarase	Tat signal peptide Identification	*S. lividans.*	*S. coelicolor* Agarase	Tat signal fused with Streptomyces coelicolor agarase expressed in S. lividans. Agarase secreted degrades Agar into smaller oligosaccharides which are visualized by staining the plates with iodine [8]
3	spTorA-mCherry-SsrA	Tat inhibitor screening	*E. coli*	spTorA-mCherry -SsrA	Quantification of periplasmic mCherry emission. Untranslocated spTorA-mCherry-SsrA in the cytoplasm is degraded by the ClpXP/ ClpAP protease system [9]
4	PlcH HTS	Tat inhibitor screening	*P. aeruginosa*	PlcH	Tat pathway secreted PlcH activity is estimated with NPPC (O-(4-Nitrophenylphosphoryl)choline) at A410 [10]
5	Mobility	Motility and chemotaxis assay	*X. oryzae*		Motility assays were performed in NYGB 0.3% soft agar plate [11]
6	EPS aasay	EPS production	*X. oryzae*		EPS was precipitated from the culture supernatant with ethanol, then was dried at 80°C to constant weight. [11]
7	Hydrogenasase assay	Hydrogenase activity	*Rhizoium leguminosarum bv viciae*	*R. leguminosarum* hydrogenase	Detection of hydrogenase subunits was carried out by immunobloting [12].
8	Cytochrome C assay	Cytochrome C-dependent O_2 respiration	*Rhizoium leguminosarum bv viciae*		Reduction of cytochrome *c* was monitored for membrane using Ubiquinone50 spectrophotometrically [12].
9	PehC assay	Galacturonic acid measurement	*Ralstonia solanacearum*	PehC of *R. solanacearum*	Release of galacturonic acid from polygalacturonate by the exopolygalacturonase PehC was measured by separation and visualization of reaction products on thin-layer chromatography plates [13]
10	Xylanase assay	Xylanase activity	*Steptomyces lividens*	*S. lividens* Xylanase C	Dinitrosalicylic acid assay to quantify amount of reducing sugar [14]
11	Tyrosinase assay	Tyrosinase acticity	*Steptomyces lividens*	*S. lividens* Tyrosinase	Dopachrome assay procedure with l-dihydroxyphenylalanine (l-DOPA) as a substrate [14]
12	TMAO assat	TMAO : benzl viologen	*E. coli*	*E.coli* TorA	TMAO and Me2SO reductase were assayed as the substrate-dependent oxidation of reduced benzyl viologen [15]
13	GFP	Tat pathway folding	*E. coli*	*E.coli* TorA – GFP fusion protein	Green fluorescent protein export by *E.coli* TorA–GFP fusion protein export visualized by con-focal microscope [16]

Tat Signal sequence activity based on the export of a Tat signal bound beta-lactamase [7] was developed where a ΔblaC mutant of *M. tuberculosis*, lacking the chromosomally encoded β-lactamase BlaC is substituted by a BlaTEM-1 from *E. coli* preceded with a Tat or Sec Signal sequence to identify the active Tat and Sec Signal sequences in *Mycobacterium.* The inactive signal leads to susceptibility to B-lactam antibiotics. In *Streptomyces coelicolor* agarase assay [8] Tat signal fused with Streptomyces coelicolor agarase is expressed in *S. lividans*, active Tat peptides caused formation as a zone of digested agar having around the colony. The broken down sugars from agar can be visualized by staining the plates with Iodine. Recently, Bageshwar et al. [9] have developed a high-throughput fluorescence assay for screening Tat inhibitors, using a mCherry fused precursor protein substrate spTorA-mCherry-SsrA that gets degraded in the cytoplasm by the ClpXP/ClpAP protease system which recognizes SsrA. When the Tat pathway is active extracellular fluorescence is observed but when Tat pathway is inhibited there is no fluorescence due to degradation in the cytoplasm. This can be used to screen for Tat inhibitors in *E.coli*. For the identification of Tat inhibitors in *Pseudomonas aeruginosa* the PlcH HTS [10] assay was developed that quantified the PlcH(Tat-transported hemolytic phospholipase C) activity. NPPC was used to quantify the PlcH and was estimated by absorbance at A410nm.

However, these assays are time consuming, and are also limited by the fact that they depend on proteins not native to the assay organism. Heterologous expression of these proteins might cause undue stress and introduce new variables that can affect the sturdiness of the assay. In this context, we have evaluated an E. coli-native Tat-substrate colonic acid biosynthase enzyme, involved in bacterial polysaccharide metabolism, as a viable alternative to develop a robust high-throughput assay for more specific and direct analysis of Tat pathway function.

3.2 Materials and Methods

3.2.1 Materials

Type I water (autoclaved MilliQ) was used for buffer preparation and dilution. Sulphuric acid AR grade was obtained from Qualigens, India. L-Fucose and Cysteine HCl were procured from Sigma Aldrich, USA. All bacterial media components were from HiMedia,

India. Bacterial Strains and Growth Conditions. The Tat pathway deletion mutant E. coli BL21 (DE3) Gold DTat has the Tat A, B, C components deleted [5]. E. coli BL21 (DE3) Gold strain was used as control. The DTat strain was routinely maintained on Luria–Bertani (LB) agar. Both the strains were cultivated (1% inoculum) in a 250 ml Erlenmeyer flask containing 100 ml LB at 37°C and 180 rpm. Another set of Tat-functional wt cells were grown in media containing 0.75% DMSO. After A_{600} reached 0.1, aliquots were removed at regular intervals (15 min for Bl21G cells and Bl21G (DMSO) cells, and10 min for DTat) to monitor the growth curve till A_{600} ~1.2.

3.2.2 Colanic Acid Estimation

Extracellular polysaccharides were extracted using acid hydrolysis, and spectrophotometric quantification of colonic acid was performed as described earlier [17] by measuring non-dialyzable methylpentose (L-fucose). Briefly, cells from 1 ml culture (A_{600} = 1) were resuspended in 1 ml water and mixed with 4.5 ml of H2SO4:H2O (6:1; v/v). The samples were boiled at 100°C for 20 min and cooled to room temperature. For each sample, absorbance at 396 and 427 nm was measured (A396-co and A427-co, respectively), and the measurements were repeated after the addition of 100 µl L-cysteine hydrochloride (A396-cyand A427-cy, respectively). The absorption due to unspecific reaction of cellular material with acid was subtracted from the total absorption of the sample: A396-co and A427-co were subtracted from A396-cy and A427-cy, respectively, to obtain DA396 and DA427. Values of (DA396–DA427) were directly correlated to methylpentose concentration using a standard curve.

3.2.3 High Throughput assay

For a lower volume high-throughput assay, cells from 0.5 ml overnight grown culture ($A_{600} \leq 1$) were centrifuged at 4000 rpm for 10 min and resuspended in 100 µl water. To 60 µl cell suspension taken in a 96-well microtitre plate (clear polystyrene, Corning USA), 150 µl H2SO4:H2O (6:1) was added and mixed at 180 rpm for 1 min (Fig1). The plate was incubated in a water bath at 75°C for 45 min and cooled to room temperature. Absorbance was measured at 401 nm and 459 nm before (co) and 5 min after (cy) the addition of 3 µl of 5% L-cysteine HCl (using Thermo Scientific VarioskanFlashTM microplate reader).

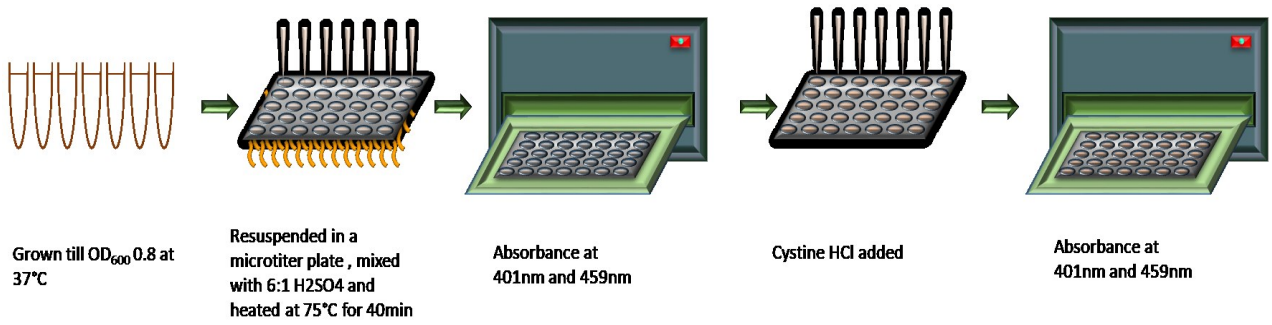


Fig. 3.1. Schematics of the high throughput assay. E.coli was grown till A_{600} 0.8 at 37 °C - 0.5ml culture was pelleted down at 4000 rpm- the cells were resuspended in 100 µl Milliq – 60 µl of cell solution was taken into a 96 well microtiter plate – 150 µl of 6:1 (H_2SO_4:H_20) was added – the plate was agitated at 180rpm for 60 sec and heated at 75°C for 45min- Absorbance was measured at 401 and 459nm – 3 µl of 5% L-Cystine HCl was added- Absorbance was again measured at 401 and 459nm

3.3 Results

3.3.1 Development of the colonic acid based assay

Colanic acid biosynthase (WcaM) is an essential enzyme involved in this pathway and is also translocated to the periplasm via the Tat pathway [18]. The WcaM activity is positively correlated with the pathway, making it an ideal reporter to help quantify Tat-mediated protein export (Fig. 3.2). Based on this reasoning, we evaluated the suitability of employing colanic acid production as a direct indicator of Tat pathway function. Extracellular polysaccharides were extracted using acid hydrolysis and tested for their L-fucose (methylpentose) content, which relates to the amount of colanic acid.

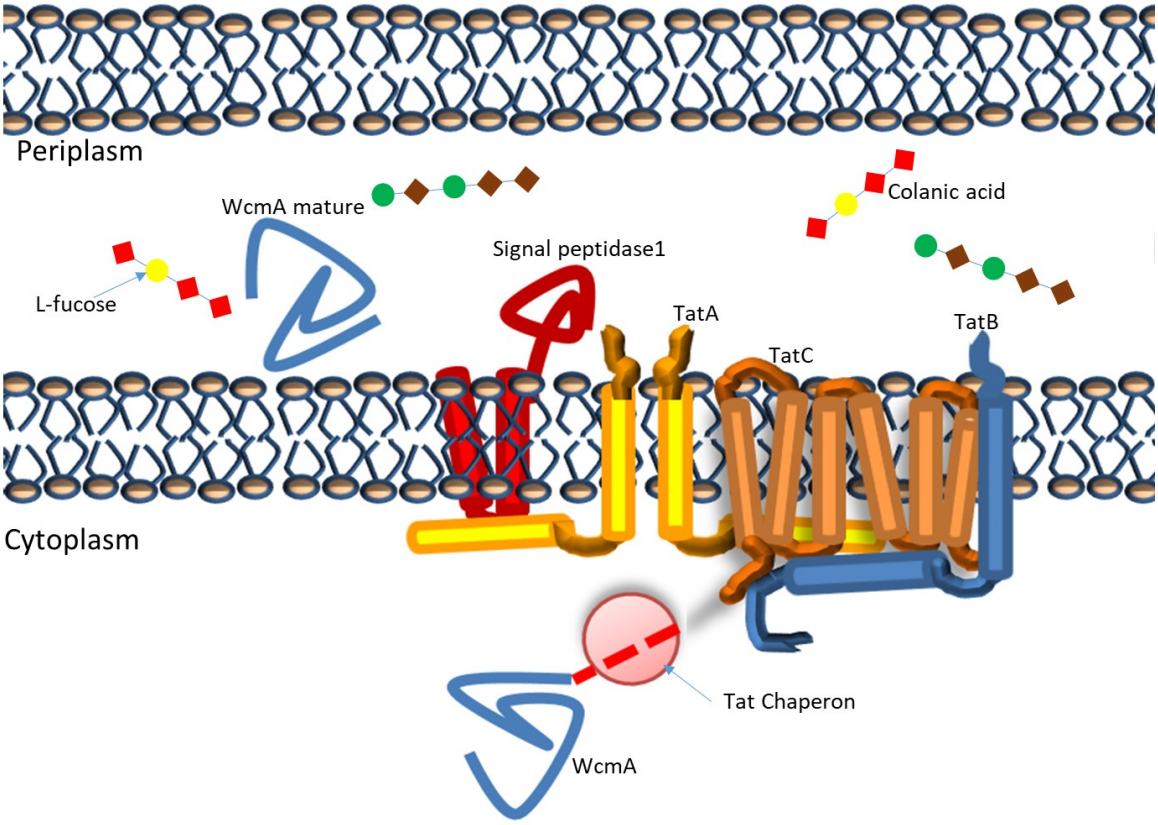


Fig. 3.2. WcaM transport through the Tat pathway for colanic acid biosynthesis

The L-fucose concentration from the Tat deletion (DTat) strain was 8 µM (Fig.3.3, lane2) when compared to 42 µM (Fig. 3.3, lane1) of wt BL21 DE3 (Gold) cells in which Tat export is functional. Moreover, complementation of the Tat component proteins (pET33b-TatABC) in the DTat mutant effectively restored L-fucose concentration to normal levels (Fig. 3.3), confirming that colanic acid production is directly linked to protein export via the Tat pathway. Importantly, it serves as proof of concept for adapting the method for determination of L-fucose (colanic acid) concentration as an efficient assay with high specificity for the analysis of Tat function.

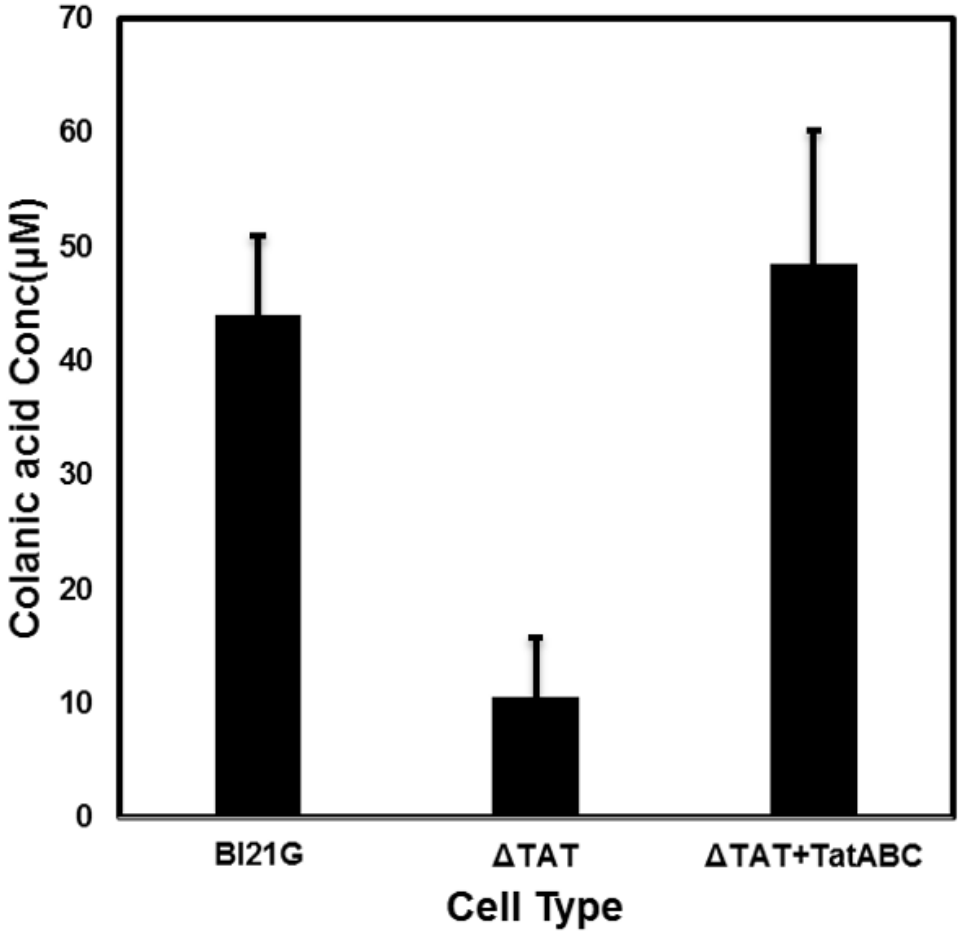


Fig. 3.3. Colanic acid concentrations in E. coli BL21(DE3) Gold wildtype (BL21G), Tat mutant (DTAT) , and TatABC complemented (DTAT + TatABC) strains

3.3.2 Adaptation to High-throughput assay

For adaptation to high-throughput analysis and possible automation, the assay was performed in 96-well microtitre plates in a total volume of 210 µl. Polysaccharide extraction was carried out at 75 °C for 45 min to prevent damage to integrity of the plates. The reaction of L-cysteine HCl with the L-fucose component of the colonic acid in acidic environment is an important step in the assay, and a concentration of 5% cysteine HCl was optimized for detecting L-fucose levels up to 200 µM. New absorbance maxima were identified as 401 nm (instead of 396 nm) and 459 nm (instead of 427 nm) for the microplate reader by performing full spectra absorbance scans (Fig.3.4).

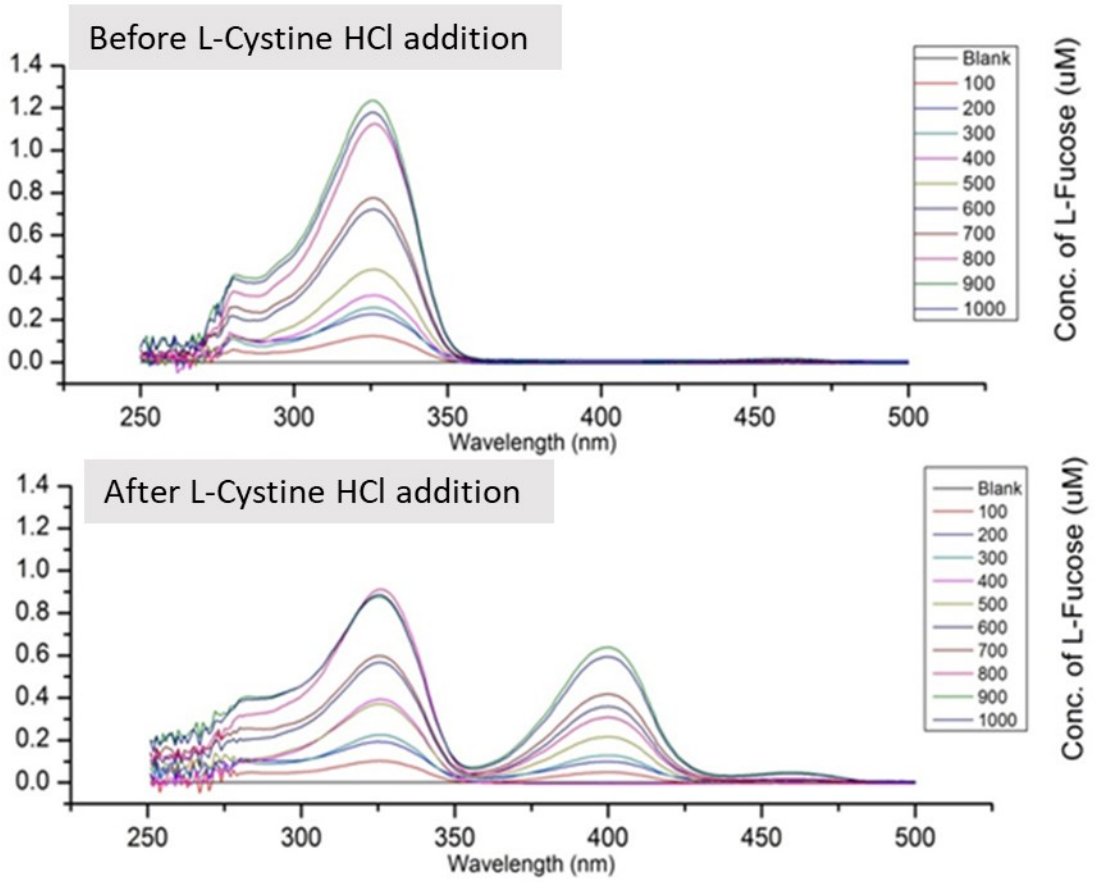


Fig. 3.4. Absorbance vs wavelength scan to determine the absorbance max at different concentrations of L-fucose was determined in Thermo Varioskan microplate reader by full spectra scan from 250-500 nm. The absorbance max were determined to be 401 nm and 459 nm.

3.3.3 L-Fucose standard curve

For detection of L-Fucose, the crucial step is where L-Cysteine HCl reacts with L-Fucose in acidic environment. To ensure that L-Cysteine HCl concentration is appropriate for detecting any possible concentration of L-Fucose. Different concentrations of L-Cysteine HCl from 3% to 10% were checked,by applying against concentration gradient of L-Fucose. 5 % L-Cysteine HCl was found out to be most suitable. L-Fucose standard was performed. Gradient of L-Fucose was prepared using 1mM working stock from 25 µM to 200 µM in 25 µM steps. The values of (DA401(cy-co) –DA459(cy-co)) were plotted to get the L-Fucose standard curve(Fig. 3.5) and directly correlated to L-fucose concentration (y = 0.008x + 0.0078, R2 =

0.9844). Results are reported as the mean of three independent experiments with standard deviation.

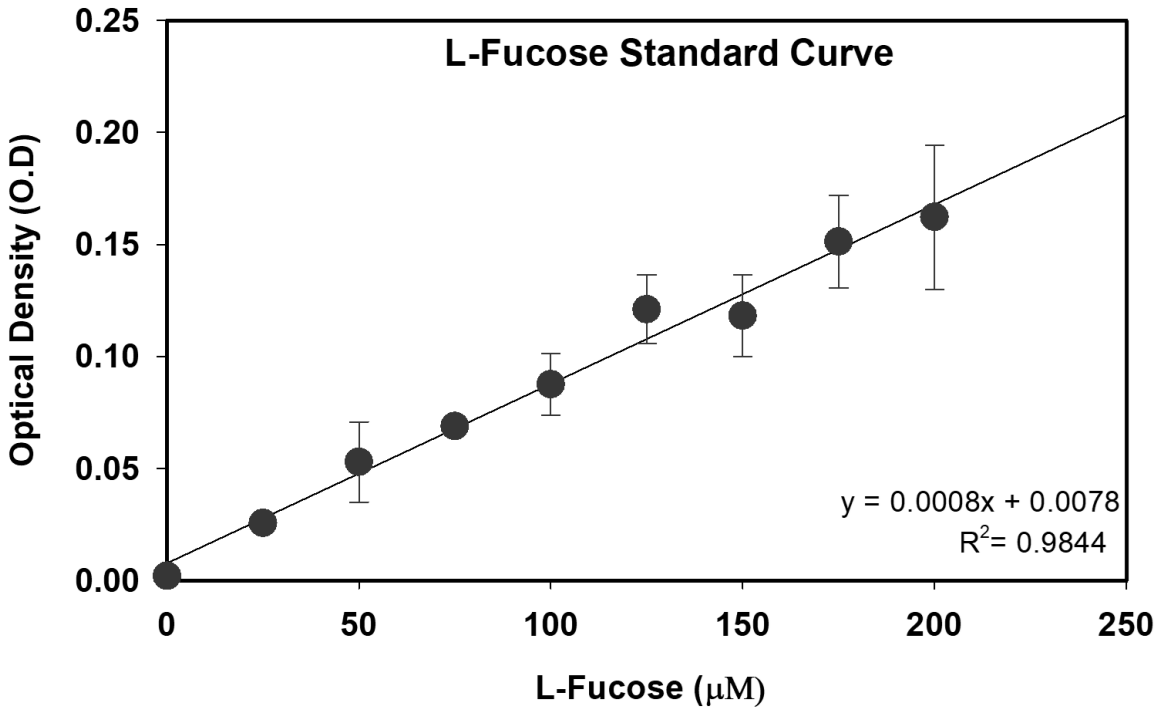


Fig. 3.5. L-Fucose standard curve - O.D. vs L-Fucose concentration plotted from 25 µM to 200 µM in 25 µM steps and the standard curve equation is obtained by drawing a linear fit line.

3.3.4 Growth vs Colanic acid

L-Fucose concentration was also monitored at different stages of bacterial growth to establish the optimum measurement window. Though the Tat machinery is not absolutely essential for E. coli growth, deletion strains have been reported to exhibit significant growth defects [19]. In agreement, the Tat cells used in this study exhibited a longer initial lag period of growth, taking 120 min to reach A_{600} = 0.1 against 95 min for Tat-functional BL21G wt cells. Further growth appeared to be normal, although distinctly slower than wt cells (Fig. 3.6.A). Colanic acid levels were subject to some fluctuations, but it could be discerned that colanic acid production in wt cells during the exponential growth phase was at least 2-3fold higher than in Tat mutant. The maximum colanic acid concentration for wt cells (38 µM) was

estimated at A_{600} = 0.8; we therefore suggest that measurements be made after this growth point. However, it should be noted that over-growth (A_{600}~1.5) could also potentially lead to accumulation of cellular products, resulting in a significant background on acid hydrolysis [20]. A basal concentration (10–15 μM) of L-fucose was still observed in the Tat mutant, which might be attributed to free pentose in the cell. However, the colanic acid production was markedly inhibited in the Tat strain throughout the course of growth in comparison to wt cells, making it apparent that colanic acid production is governed by the Tat pathway. Colanic acid levels were also checked in the presence of DMSO, as it is present in many available drug libraries that might be used for screening Tat inhibitors. The presence of DMSO caused an initial increase in colanic acid production (till A_{600} = 0.4), but CA levels in later growth stages were observed to be similar to the wildtype (Fig. 3.6.B).

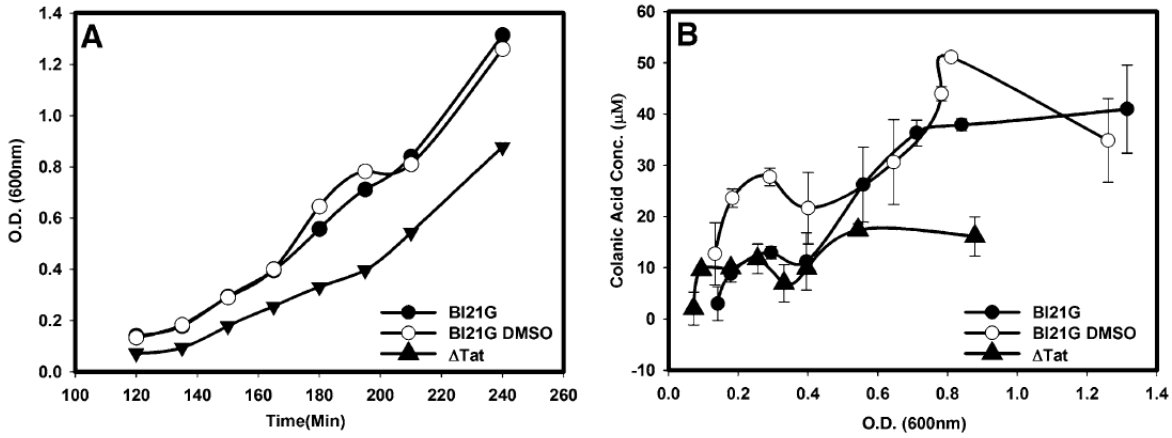


Fig. 3.6. A) Change in colanic acid concentration across different growth stages. B) Bacterial cell growth (A_{600}) monitored over time.

3.4 Discussions

The Tat pathway exports 6% of all proteins in *E. coli* [1], and is very significant for viability in some halophilic archaea [10] and for virulence in *Mycobacterium tuberculosis* [21]. A large volume of research has been done in recent years to understand the functions of different components of the Tat pathway [22], and the finer details of the mechanism of protein translocation across the bacterial membrane [23]. Moreover, the Tat pathway is specific to bacteria and transports virulence factors in pathogenic bacteria [24], thus making it an attractive target to screen for inhibitors to design novel drugs in human healthcare. In this

context, it can be deemed necessary to develop an efficient assay that could enable the specific analysis of Tat function and mode of protein translocation. Colanic acid (CA) is an extracellular capsular polysaccharide produced by most E. coli strains as well as by other Enterobacteriaceae [25], and has significant roles in biofilm formation, protection against dessication, and adhesion of pathogenic bacteria to host membranes. It contains units of L-fucose, D-glucuronic acid, D-glucose, D-galactose, and pyruvate, and forms a mucoid matrix on the cell surface. Fucose is quite widely present in bacterial polysaccharides and eukaryote glycoproteins, and it is known [26] that GDP–L-fucose is the precursor of such fucose residues in both bacteria and eukaryotes. The capsular polysaccharide gene cluster (wca operon) directs colanic acid biosynthesis [27], comprising 19 different proteins. Colanic acid biosynthase (WcaM) is an essential enzyme involved in this pathway and is also translocated to the periplasm via the Tat pathway [18].

Moreover, the use of colonic acid biosynthesis native to *E. coli* as a reporter of Tat export makes available a quantitative and Tat-specific method that can be used to probe Tat pathway function, and scout for new Tat inhibitors. The extraction of colanic acid after cell growth can be performed using a very simple procedure as there is no need of heterologous expression of a reporter protein, enabling easy and swift quantification of Tat export in the different organisms where heterologous expression is not straight forward.

This colanic acid-based assay could essentially serve as a simple yet efficient indicator of Tat pathway export and function, including checking the effect of mutations in the Tat component proteins. Screening of Tat inhibitors can also be performed, but care has to be taken to differentiate inhibitors specific for the Tat pathway against general growth inhibitors, by comparing the cell growth with colonic acid concentration. Compounds that effect an acute reduction in colanic acid production combined with only a modest retardation in cell growth can be developed as potential Tat inhibitor candidates, while a general growth inhibitor could result in a more severe growth impediment. It is also possible that inhibition of any enzyme of the colanic biosynthesis pathway could result in decreased colanic acid production, unrelated to the Tat pathway; however, inhibition of growth is not expected in such cases.

The colanic acid biosynthase (WcaM) and glucan biosynthesis enzyme (MdoD) are substrates of the Tat pathway important for polysaccharide biosynthesis [18]. An analysis of their

distribution in pathogenic bacteria (Table3.2) revealed that they are mostly restricted to Enterobacteriaceae, including many strains of *E. coli*. Polysaccharide production in these bacteria, most of which are gastrointestinal pathogens, could play a significant role in biofilm formation and adhesion to the host cells [28]. Disruption of colanic acid production through inhibition of the Tat pathway appears to be an attractive option for designing novel drugs. In conclusion, the assay described in this study offers an easy and robust analysis system in *E. coli* for experiments concerning Tat pathway function and a screen for identification of new Tat substrates and inhibitors in gastrointestinal pathogens. It combines the advantages of a single-step extraction procedure and the specificity for Tat pathway, paving the way for the development of an economical highthroughput assay for Tat screening.

Table 3.2. Distribution of colanic acid biosynthase (WcaM) and glucan biosynthesis enzyme (MdoD) in pathogenic bacteria. Organisms with conserved WcaM and MdoD in bold

Organism	MdoD	WcaM
Klebsiella pneumoniae	+	+
Salmonella enterica	+	+
Salmonella typhi	+	+
Salmonella typhimurium	+	+
Shigella boydii	+	+
Shigella flexneri	+	+
Klebsiella oxytoca	+	-
Shigella dysenteriae	+	-
Vibrio cholerae	+	-
Yersinia pestis	+	-
Yersinia enterocolitica	+	-
Serratia marcescens	+	-
Stenotrophomonas maltophilia	+	-
Xanthomonas citri	+	-
Pseudomonas aeruginosa	+	-
Bacillus anthracis	-	-
Campylobacter jejuni	-	-
Campylobacter lari	-	-
Corynebacterium diphtheriae	-	-
Haemophilus influenzae	-	-
Helicobacter pylori	-	-
Listeria monocytogenes	-	-
Mycobacterium leprae	-	-
Mycobacterium tuberculosis	-	-
Proteus mirabilis	-	-
Rickettsia prowazekii	-	-
Rickettsia rickettsii	-	-
Brucella abortus	-	-
Brucella ovis	-	-

3.5 References

[1] T. Palmer and B. C. Berks, "The twin-arginine translocation (Tat) protein export pathway," *Nat. Rev. Microbiol.*, vol. 10, no. 7, pp. 483–496, Jul. 2012.

[2] H. Shruthi, P. Anand, V. Murugan, and K. Sankaran, "Twin arginine translocase pathway and fast-folding lipoprotein biosynthesis in E. coli: interesting implications and applications," *Mol. Biosyst.*, vol. 6, no. 6, p. 999, May 2010.

[3] S. K. Ramasamy, W. M. Clemons, and IUCr, "Structure of the twin-arginine signal-binding protein DmsD from *Escherichia coli*," *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, vol. 65, no. 8, pp. 746–750, Aug. 2009.

[4] G. Orfanoudaki and A. Economou, " Proteome-wide Subcellular Topologies of E. coli Polypeptides Database (STEPdb) ," *Mol. Cell. Proteomics*, vol. 13, no. 12, pp. 3674–3687, 2014.

[5] P. A. Lee, D. Tullman-Ercek, and G. Georgiou, "The Bacterial Twin-Arginine Translocation Pathway," *Annu. Rev. Microbiol.*, vol. 60, no. 1, pp. 373–395, 2006.

[6] S. D. Branston, C. F. R. O. Matos, R. B. Freedman, C. Robinson, and E. Keshavarz-Moore, "Investigation of the impact of Tat export pathway enhancement on E. coli culture, protein production and early stage recovery," *Biotechnol. Bioeng.*, vol. 109, no. 4, pp. 983–991, 2012.

[7] J. R. McCann, J. A. McDonough, M. S. Pavelka, and M. Braunstein, "β-Lactamase can function as a reporter of bacterial protein export during Mycobacterium tuberculosis infection of host cells," *Microbiology*, vol. 153, no. 10, pp. 3350–3359, 2007.

[8] D. A. Widdick, R. T. Eijlander, J. M. van Dijl, O. P. Kuipers, and T. Palmer, "A Facile Reporter System for the Experimental Identification of Twin-Arginine Translocation (Tat) Signal Peptides from All Kingdoms of Life," *J. Mol. Biol.*, vol. 375, no. 3, pp. 595–603, 2008.

[9] U. K. Bageshwar *et al.*, "High throughput screen for Escherichia coli twin arginine translocation (Tat) inhibitors," *PLoS One*, vol. 11, no. 2, pp. 1–25, 2016.

[10] M. L. Vasil, A. P. Tomaras, and A. E. Pritchard, "Identification and Evaluation of Twin-Arginine Translocase Inhibitors," *Antimicrob. Agents Chemother.*, vol. 56, no. 12, pp. 6223–6234, 2012.

[11] L. Chen *et al.*, "Identification and molecular characterization of twin-arginine translocation system (Tat) in Xanthomonas oryzae pv. oryzae strain PXO99," *Arch. Microbiol.*, vol. 191, no. 2, pp. 163–170, Feb. 2009.

[12] S. Meloni, L. Rey, S. Sidler, J. Imperial, T. Ruiz-Argüeso, and J. M. Palacios, "The twin-arginine translocation (Tat) system is essential for Rhizobium-legume symbiosis," *Mol. Microbiol.*, vol. 48, no. 5, pp. 1195–1207, May 2003.

[13] E. T. Tans-Kersten, D. G. Guan, J. K. Allen, and C. Allen, "Ralstonia solanacearum pectin methylesterase is required for growth on methylated pectin but not for bacterial wilt virulence," *Appl. Environ. Microbiol.*, vol. 64, no. 12, pp. 4918–23, Dec. 1998.

[14] S. De Keersmaeker, L. Van Mellaert, E. Lammertyn, K. Vrancken, J. Anné, and N. Geukens, "Functional analysis of TatA and TatB in Streptomyces lividans," *Biochem. Biophys. Res. Commun.*, vol. 335, no. 3, pp. 973–982, Sep. 2005.

[15] A. Silvestro, J. Pommier, and G. Giordano, "The inducible trimethylamine-N-oxide reductase of Escherichia coli K12: biochemical and immunological studies," *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.*, vol. 954, pp. 1–13, Jan. 1988.

[16] J. D. Thomas, R. A. Daniel, J. Errington, and C. Robinson, "Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in Escherichia coli," *Mol. Microbiol.*, vol. 39, no. 1, pp. 47–53, Jan. 2001.

[17] and L. B. S. Dische, Zacharias, "A new spectrophotometric test for the detection of methylpentose," *J. Biol. Chem.*, vol. 92(2), pp. 579–582, 1951.

[18] M. M. Reynolds *et al.*, "Abrogation of the Twin Arginine Transport System in Salmonella enterica Serovar Typhimurium Leads to Colonization Defects during Infection," vol. 6, no. 1, 2011.

[19] N. R. Stanley, K. Findlay, B. C. Berks, and T. Palmer, "Escherichia coli strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope.," *J.*

Bacteriol., vol. 183, no. 1, pp. 139–44, Jan. 2001.

[20] B. Obadia, S. Lacour, P. Doublet, H. Baubichon-cortay, A. J. Cozzone, and C. Grangeasse, "Influence of Tyrosine-Kinase Wzc Activity on Colanic Acid Production in Escherichia coli K12 Cells," pp. 42–53, 2007.

[21] J. A. McDonough, K. E. Hacker, A. R. Flores, M. S. Pavelka, and M. Braunstein, "The twin-arginine translocation pathway of Mycobacterium smegmatis is functional and required for the export of mycobacterial β-lactamases," *J. Bacteriol.*, vol. 187, no. 22, pp. 7667–7679, 2005.

[22] R. Patel, S. M. Smith, and C. Robinson, "Protein transport by the bacterial Tat pathway," *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1843, no. 8, pp. 1620–1628, Aug. 2014.

[23] B. C. Berks, "The Twin-Arginine Protein Translocation Pathway," *Annu. Rev. Biochem.*, vol. 84, no. 1, pp. 843–864, 2014.

[24] E. De Buck, E. Lammertyn, and J. Anné, "The importance of the twin-arginine translocation pathway for bacterial virulence," *Trends Microbiol.*, vol. 16, no. 9, pp. 442–453, Sep. 2008.

[25] C. Whitfield, " Biosynthesis and Assembly of Capsular Polysaccharides in Escherichia coli ," *Annu. Rev. Biochem.*, vol. 75, no. 1, pp. 39–68, 2006.

[26] K. Andrianopoulos, L. E. I. Wang, and P. R. Reeves, "Identification of the Fucose Synthetase Gene in the Colanic Acid Gene Cluster of Escherichia coli K-12," vol. 180, no. 4, pp. 998–1001, 2006.

[27] G. Stevenson, K. Andrianopoulos, M. Hobbs, and P. R. Reeves, "Organization of the Escherichia coli K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid.," *J. Bacteriol.*, vol. 178, no. 16, pp. 4885–93, Aug. 1996.

[28] A. Hanna, M. Berg, V. Stout, and A. Razatos, "Role of capsular colanic acid in adhesion of uropathogenic Escherichia coli," *Appl. Environ. Microbiol.*, vol. 69, no. 8, pp. 4474–4481, 2003.

Chapter IV - Comprehensive analysis and Insights Into The Haloarchaeal Twin-Arginine Translocase Pathway

4.1 Introduction

Archaea are known to be evolutionarily distinct from both bacteria and eukarya in many of their cellular characteristics. In the case of halophilic archaea, the Tat pathway has a decisive role in transporting almost 50% of secretome [1][2][3]. This might be because of the high cytoplasmic salt concentration in halophilic archaea, which these organisms maintain in order to balance the high sodium concentration of the environment. The high cytoplasmic salt concentration results in faster folding of proteins to prevent aggregation, so relatively few unfolded proteins are transported. The transport of folded proteins also ensures that the proteins are not folded in the extracellular environment, where chaperones might not be available and the chance of misfolding is therefore much higher [4].

Tat translocation is carried out by integral membrane proteins in the TatA and TatC families [5][6]. In some organisms, such as archaea and Gram-positive bacteria with low guanine-cytosine (GC) genomic content, single TatA and TatC components are sufficient to mediate transport through the Tat pathway (the TatAC system). For example, *Bacillus subtilis* contains two TatC genes (denoted as TatCd and TatCy), which coordinate functions with their respective TatA partners (TatAd and TatAy). However, in other cases, Tat-mediated translocation involves another member of the TatA family, termed TatB. The Tat pathway in Escherichia coli consists of three distinct membrane-localized proteins (TatA, TatB, and TatC) and is the most studied model in which all the three components are essential for Tat functioning. TatABC systems are present in Gram-negative bacteria (including *E. coli*) [7][8], in Gram-positive bacteria with high GC-content genomes, and in plant chloroplasts [9]. TatC is the receptor component that recognizes the characteristic RR-signal in the substrate N-terminal [10][11].TatC or the TatBC complex interacts with the substrate and recruits TatA. TatA forms an appropriately sized pore through which the substrate is transported across the

membrane [12]. In plants, the thylakoid Tat system is required for the assembly of many essential components, such as photosystem II and the cytochrome b6f complex [13][14]. The recently solved crystal structures of TatC from the thermophilic bacteria *Aquifex aeolicus* [15][16] confirmed the presence of six transmembrane (TM) domains and the orientation of the amino and carboxy termini towards the cytoplasmic side of the membrane [7][17][17]. The structure of TatAd from *Bacillus subtilis* has been determined by NMR to be composed of a single TM domain and a cytosolically located amphipathic helix [18]. Common Tat substrates include respiratory redox enzymes, bacterial virulence factors, lipoproteins [19][20][21], and proteins involved in maintaining cell-wall integrity. These are mostly complexes containing cofactors, and many form oligomeric assemblies. Delivery of these proteins is monitored by a special class of cytoplasmic chaperones that bind specifically to the Tat signal of their substrate protein masking the signal sequence thereby ensuring proper maturation and co-factor loading. These chaperones, dubbed redox-enzyme maturation proteins (REMP), prevent the futile export of immature protein [22][23][24].

In this study, we examined 20 halophilic archaea whose complete proteomes were available in the Universal Protein Resource (UniProt) database. One or two TatA homologs were found in each organism. In contrast, many of the organisms contained two to four TatC homologs. We investigated the implications of the presence of these TatC homologs in these organisms. Experiments have shown that an atypical TatC homolog with fourteen TM domains (TatCt) is present in halophilic archaea [25]. Unlike other organisms, the archaeal cell membrane phospholipid is composed of branched isoprene units linked by ether groups to glycerol. It may have some advantages to possess the novel protein transport components such as TatCt may be the result of the unusual membrane phospholipid structure. We believe that the variety in TatC homologs would have some significance for the substrates being transported [20].To assess this, we analyzed the topology, horizontal gene transfer, and substrate diversity for the different TatC homologs. We also analyzed the TatA components with respect to E. coli and *A. aeolicus* TatA.

4.2 Materials and Methods

4.2.1 Distribution of Tat components and topology assignment

All data on Tat pathway components was obtained via a query search in the UniProt database and was tabulated by organism. The amino acid sequences for TatC in these organisms were retrieved and were analyzed with the TMHMM Server v. 2.0 [26], which segregated the TatC homologs into different topologies. Different TM prediction servers were used to reinforce these results: the DAS-TM filter server, the HMMTOP server, and SCAMPI [27][28][29].

4.2.2 Sequence analysis and GC-content analysis

All TatC homologs were aligned using ClustalW and a neighbor-joining tree with a bootstrap value of 10,000 was generated. The sequence analysis was performed by Clustal-Omega-generated multiple sequence alignment and the relative positions of secondary structures were marked according to the TMHMM data. The GC content of the TatC gene along with fifty upstream and fifty downstream genes were plotted as box plots. For comparison of TatC versus genomic GC content, the entire gene chromosome sequence was used to obtain the GC content. The two datasets were compared using a two tailed z-test.

4.2.3 Substrate analysis

The complete proteomes for all twenty haloarchaea were downloaded from the UniProt database. The Tat substrates were identified using the TatFind server. To segregate the substrates, a bidirectional BLAST was performed against each other using standalone BLAST version 2.2.29. An E-value cutoff of 0.5 was imposed; all hits with less than 50% query coverage were removed. From these hits only the bidirectional hits (BETs) were considered for analysis (Fig. 4.8).

4.2.4 Signal sequence analysis

Multiple sequence alignment files for the RR signal in each dataset were generated from the TatFind data with ClustalW. TheWebLogo was generated at http://weblogo.berkeley.edu/logo.cgi.

4.2.5 Pfam domain analysis

The Pfam hidden Markov models (HMMs) present in each substrate were identified through a Pfam batch sequence search. The unique and common HMMs were then identified in the different groups. Venn diagrams were generated with http://bioinfogp.cnb.csic.es/tools/venny/.

4.2.6 Chaperon Identification

Sequences for the known tat related chaperons were obtained from UniProt and blasted against the proteomes of these organisms using standalone BLAST (blast-2.2.29+) with an E-value cutoff of 1. Sequences were blasted and homologs for each of them were obtained. Standard BLASTp output provided pairwise sequence alignment using which the signature sequence conservation was checked.

4.3 Results

The archaeal Tat system generally consists of TatA and TatC, but is devoid of TatB. The Tat pathway is present in about half of the Euryarchaeota members and several of the Crenarchaeota species that have been sequenced. The number of substrates transported by the Tat system in archaea is generally comparable to that for bacteria, with many archaea, such as *Sulfolobus tokodaii* and *Archaeoglobus fulgidus*, encoding only a few known Tat substrates, primarily cofactor-containing redox proteins [30][31]. But, in the case of haloarchaea, a substantially greater number of substrates are translocated via the Tat pathway. Sequenced and annotated data for twenty haloarchaeal species available in UniProt were used for this analysis.

4.3.1 Membrane topological variants of TatC

The amino acid sequences for TatC in haloarcheal were retrieved and using TMHMM server v. 2.0 (32) the TMs were assigned. The homologs were segregated into different topologies. This analysis segregated the TatC homologs into four classes TatC with 6-TMhelices (TatCo), TatC with 6-TM helices and possessing a long N-terminal cytoplasmic loop (TatCn), TatC with 10-TMhelices (TatCx) and TatC with 14-TMhelices (TatCt) based on the major differences in membrane topology, specifically the number of TM helices and the length of the N-terminal cytoplasmic region. Different TM-prediction servers were used to validate these results: the DAS-TMfilter server [28], the HMMTOP server [27], and SCAMPI [29]. The results of the analyses from these servers confirmed the results we obtained from the TMHMM server, except in few cases and on an average all the proteins fell into one of the topological categories described above (Fig.4.1 and Table 4.1) [17].

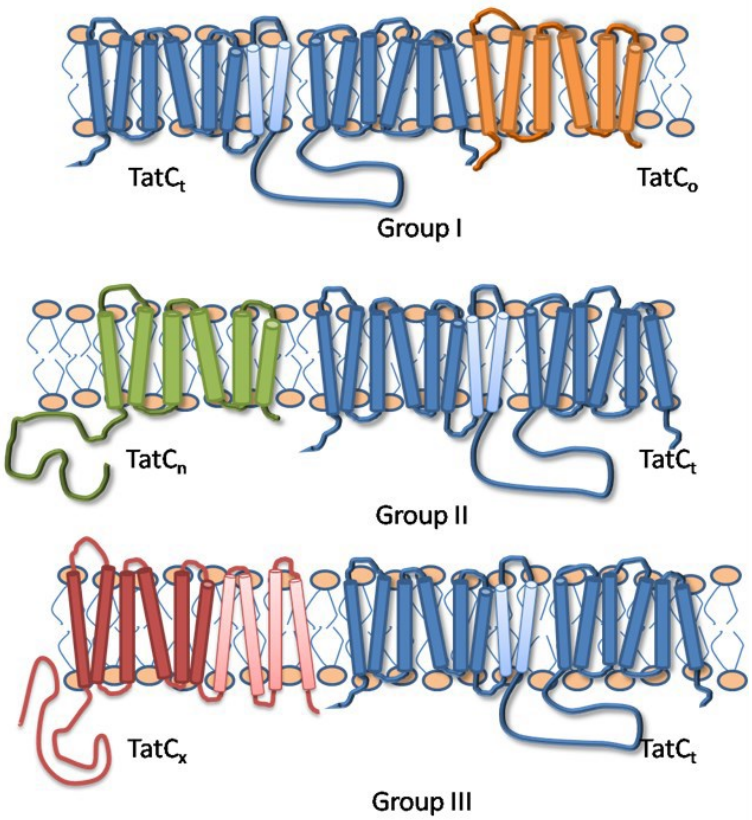


Fig. 4.1. Schematic representation of different TatC topology and grouped in to different groups based on combinations. TatCt+TatCo, TatCt+TatCn and TatCt+TatCx which are denoted a group1, Group2 and group3, respectively.

The expected 6-TM topology of TatC has been confirmed through analysis of its crystal structure [15][16] and the existence of a 14-TM homolog has been demonstrated experimentally [20]. Multiple sequence alignment additionally confirmed that the TM predictions were appropriate. The consistent prediction by all tools of a TatC homolog with 10-TM helices strongly suggests the existence of a 10-TM TatC homolog. The existence of this TatCx topological class is further supported by previous studies in H. marismortui [20]. Topology similar to that of the TatCn category has been commonly observed in TatC from chloroplasts [32].

Table 4.1: Analysis of Tat Pathway receptor component and number of substrates present in Haloarchaea that were sequenced and annotated from Uniprot.

Organism	No. of tat substrates	No. of tatC	TatCt	$TatC_x$	$TatC_n$	TatC0	tatC(id)
Halophilic							
Halalkalicoccusjeotgali	103/4212	2	1			1	D8J5Y4(6),D8J5Y5(14)
Haloarculahispanica	122/3860	2	1	1			G0HQX6(10),G0HQX5(14)
Haloarculamarismortui	148/4237	2	1	1			Q5UYP5(10),Q5UYP6(14)
Halobacteriumsalinarum (strain ATCC 29341)	72/2578	2	1			1	B0R7G7(6),B0R7G6(14)
Halobacteriumsalinarum (strain ATCC 700922)	63/2426	2	1			1	B0R7G7(6),B0R7G6(14)
Haloferaxmediterranei	171/4712	3	1		2		I3R116(6L),I3R115(14),M0IS38(6L)
Haloferaxvolcanii	143/4813	3	2		2*		L9UI27(6L),D4GZD0(6L), D4GZC9(14),L9UGM9(14)
Halogeometricumborinquense	161/4438	4	1		2	1	L9UQ35(6L),E4NRH6(14), E4NRH7(6L),E4NPB6(6)
Halomicrobiummukohataei	125/3343	2	1	1			C7P1B7(14),C7P1B8(10)
Halopigerxanaduensis	194/4221	2	1		1		F8D3F7(6L),F8D3F6(14)
Haloquadratumwalsbyi (strain DSM 16790)	64/2546	2	1		1		Q18E63(6L),Q18E62(14)
Haloquadratumwalsbyi (strain DSM 16854)	59/2638	2	1		1		G0LFQ9(14),G0LFQ8(6L)
Halorhabdusutahensis	100/3001	2	1	1			C7NVD1(10),C7NVD2(14)
Halorubrumlacusprofundi	98/3497	2	1		1		B9LTY5(14),B9LTY6(6L)
Haloterrigenaturkmenica	230/5116	2	1		1		D2RTM5(14),D2RTM6(6L)
Halovivaxruber	132/3099	2	1		1		L0IGT4(6L),L0ID04(14)
Natrialbamagadii	228/4936	2	1		1		D3SVL3(6L),D3SVL2(14)
Natrinemapellirubrum	171/5026	2	1		1		L0JIB8(6L),L0JEW5(14)
Natrinema sp.	142/4296	2	1		1		I7CEK6(6L),I7CPH1(14)
Natronobacteriumgregoryi	138/3624	2	1		1		L0AHR8(6L),L0AFL9(14)
Natronomonasmoolapensis	69/2721	2	1			1	M1XNN2(6),M1XZE0(14)
Natronomonaspharaonis	106/2764	2	1			1	Q3ISX0(6),Q3ISW9(14)
Others							
Aquifexaeolicus	15/1553	1				1	O67305(6)
Acidianushospitalis	2/2329	2				2	F4B4X0(6),F4B6B8(6)
Desulfurococcuskamchatkensis	0/1470	0					
sulfobusilandicus	3/2661	0					
Archaeoglobusveneficus	7/2065	2				2	F2KP54(6),F2KQ26(6)
Methanobrevibactersmithii	0/1783	0					
Picrophilustorridus	2/1535	1				1	Q6KZY6(6)

*Same type

The TatCt topological class is noteworthy because it is unique to haloarchaea. Further investigation via BLASTp analysis indicated that TatCt comprises two putative TatC domains. Multiple sequence analysis and TMHMM data confirmed that the first six TM helices at the N-terminal and the last six helices at the C-terminal had significant homology with the normal TatC domains (Fig. 4.2 and Table 4.2). Two TM helices and a large cytoplasmic loop separate these two homologous domains. No homologs could be identified for the middle two TM helices. For TatCx, the first six N-terminal TM helices form a characteristic TatC domain. No sequence homology was observed for the last four TM helices with any of the sequences. We then examined the distribution of these four TatC topology classes among the haloarchaea species. It was observed that all the species encompasses at least one copy of TatCt and combination with TatCo, TatCn or TatCx (Table 4.1). Thus, we defined three main groups (Fig. 4.1): TatCt+TatCo (Group I), TatCt+TatCn (Group II) and TatCt+TatCx (Group III).

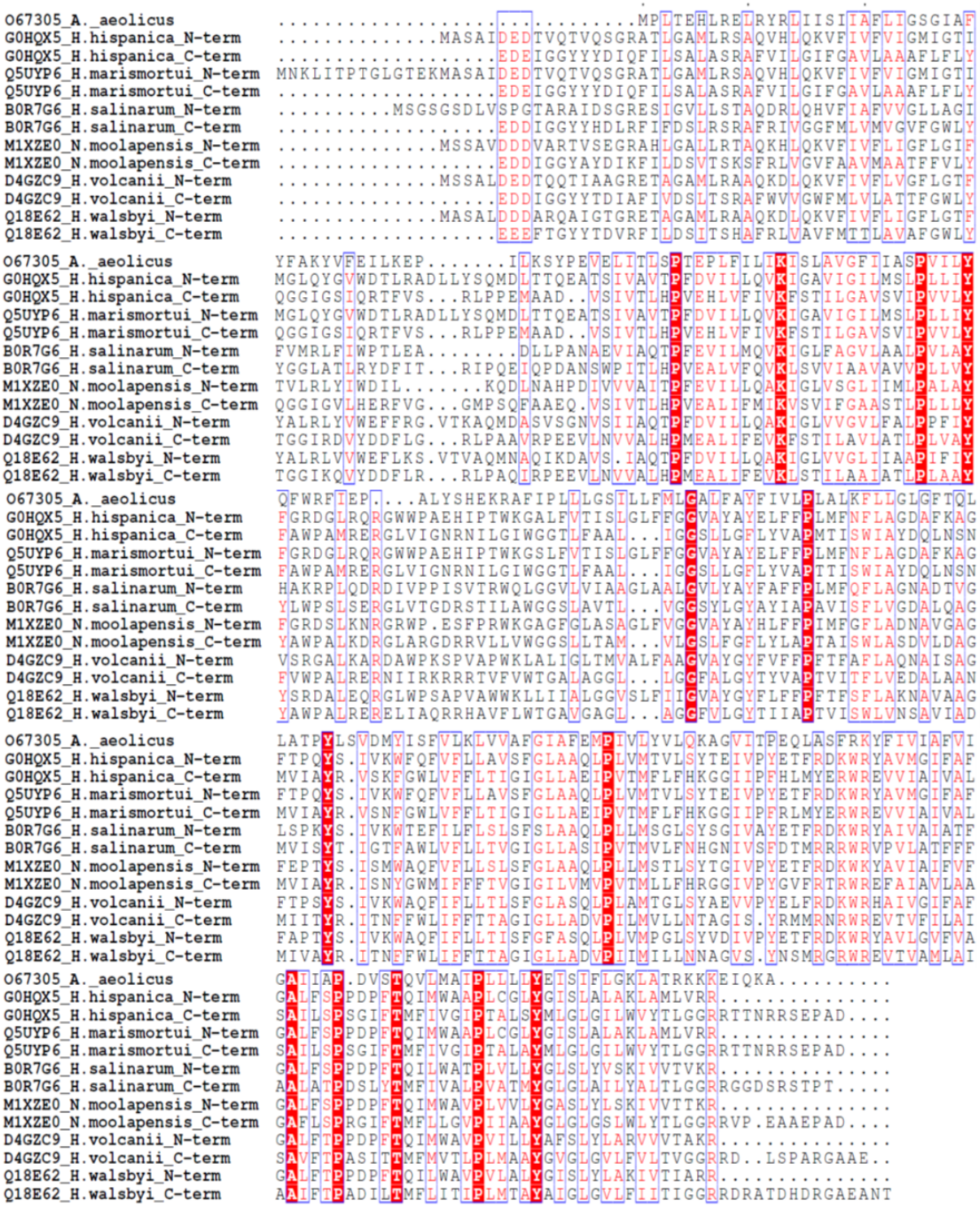


Fig. 4.2. The Multiple sequenc alignment between N and C terminal TatC domain of $TatC_t$ showing high similarity between the two domains

Table 4.2: Percentage identity between N-terminal and C-terminal TatC domains of $TatC_t$ of Holoarchaea

Organism	Percentage Identity
Halalkalicoccus jeotgali_D8J5Y5	20
Haloarcula hispanica_G0HQX5	21
Haloarcula marismortui_Q5UYP6	21
Halobacterium salinarum_B0R7G6	23
Haloferax mediterranei_I3R115	23
Haloferax volcanii_D4GZC9	22
Haloferax volcanii_L9UGM9	26
Halogeometricum borinquense_E4NRH6	22
Halomicrobium mukohataei_C7P1B7	23
Halopiger xanaduensis_F8D3F6	21
Haloquadratum walsbyi_Q18E62	22
Haloquadratum walsbyi_G0LFQ9	22
Halorhabdus utahensis_C7NVD2	22
Halorubrum lacusprofundi_B9LTY5	23
Haloterrigena turkmenica_D2RTM5	22
Halovivax ruber_L0ID04	22
Natrialba magadi_D3SVL2	22
Natrinema pellirubrum_L0JEW5	20
Natrinema sp._I7CPH1	22
Natronobacterium gregoryi_L0AFL9	24
Natronomonas moolapensis_M1XZE0	23
Natronomonas pharaonis_Q3ISW9	25

4.3.2 Sequence and phylogenetic analysis of TatC

A neighbor-joining tree generated from aligned TatC homologs was analyzed for all unique TatCt homologs, to assess the clustering of organisms in the three main groups outlined above. We found that the organisms with TatCt+TatCx and TatCt+TatCo were formed the distinct cluster (Fig 4.3). The organisms with TatCt+TatCn did not cluster together but were widely distributed.

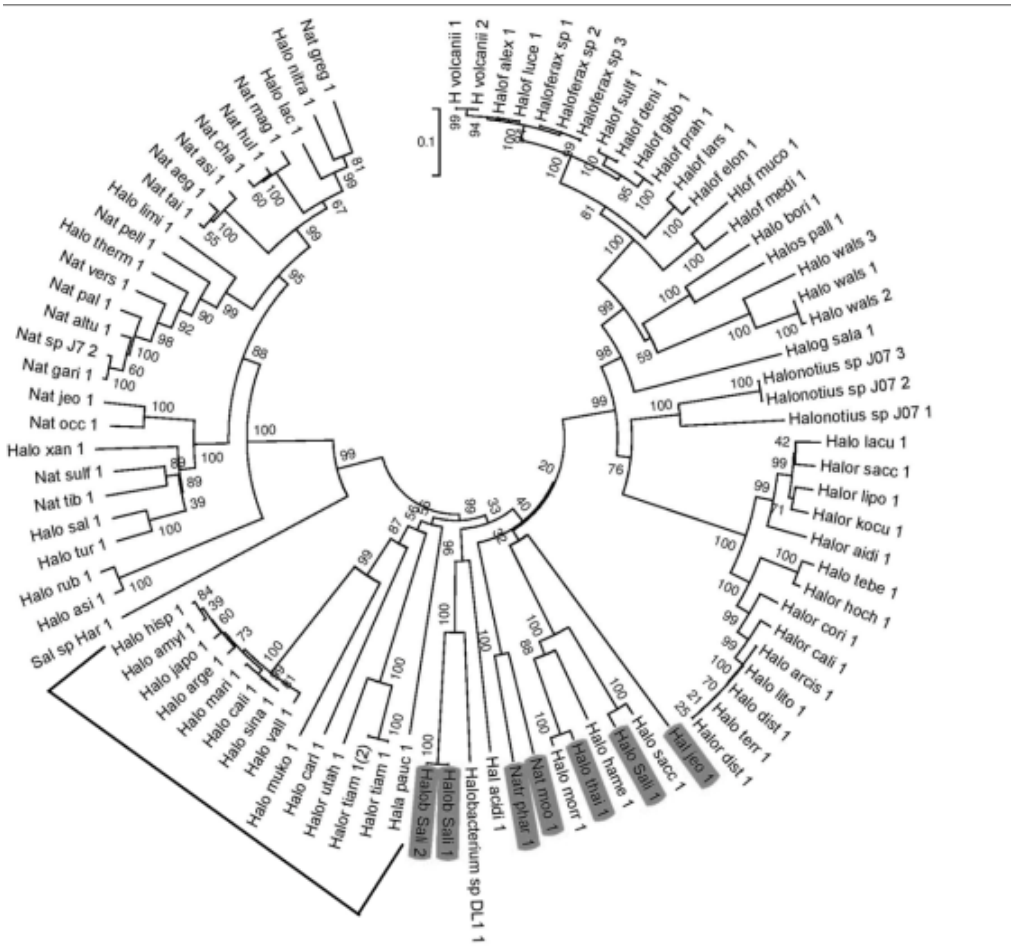
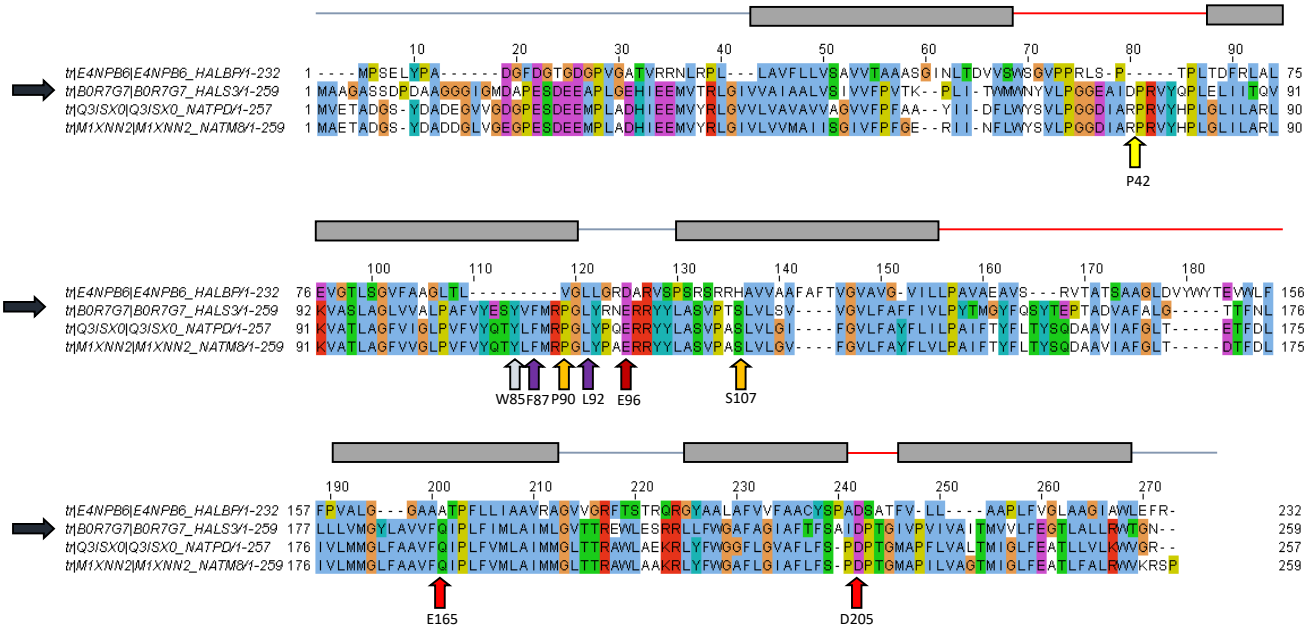


Fig. 4.3. Phylogenetic analysis of all TatC$_t$ from haloarchaea. Group 3 TatCt shown in bracket and group1 members are highlighted cluster together; the remaining members from group2 group.

The sequences were further analyzed by multiple sequence alignment against E. coli and A. aeolicus, for which crystal structure is available, to determine the crucial residues that were actively involved in Tat signal interaction and dimer formation, and explore whether these residues were conserved across sequences. This analysis provides an insight into the probable oligomerization state of the different classes of TatC in archaea (Fig. 4.4. And Table 4.3). The TM regions were marked to ensure that the sequences conserved in the analysis were in proximity to the reported regions of the protein.

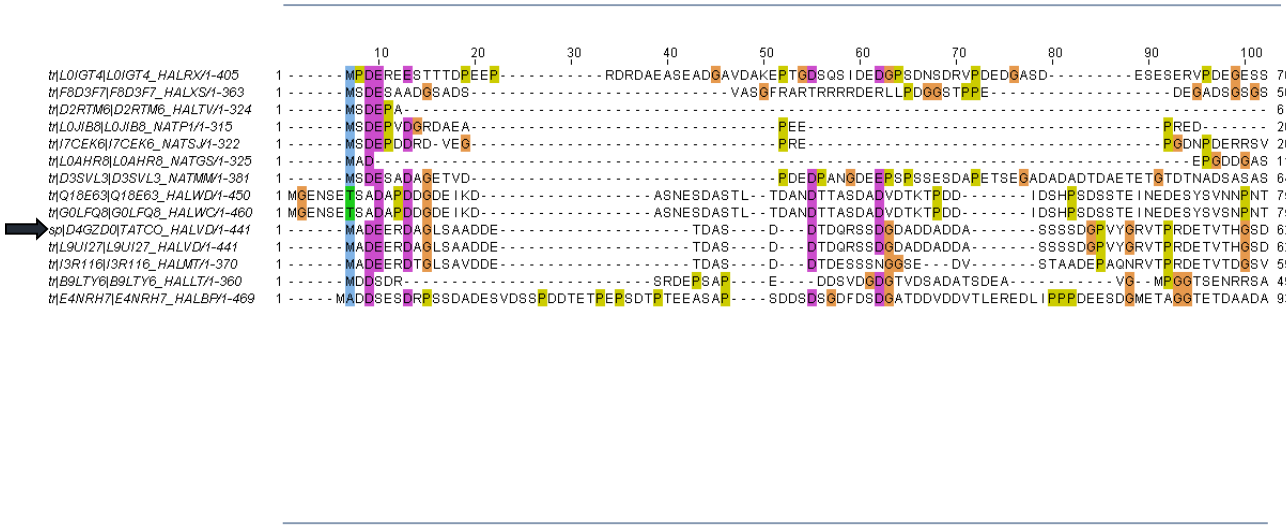
A

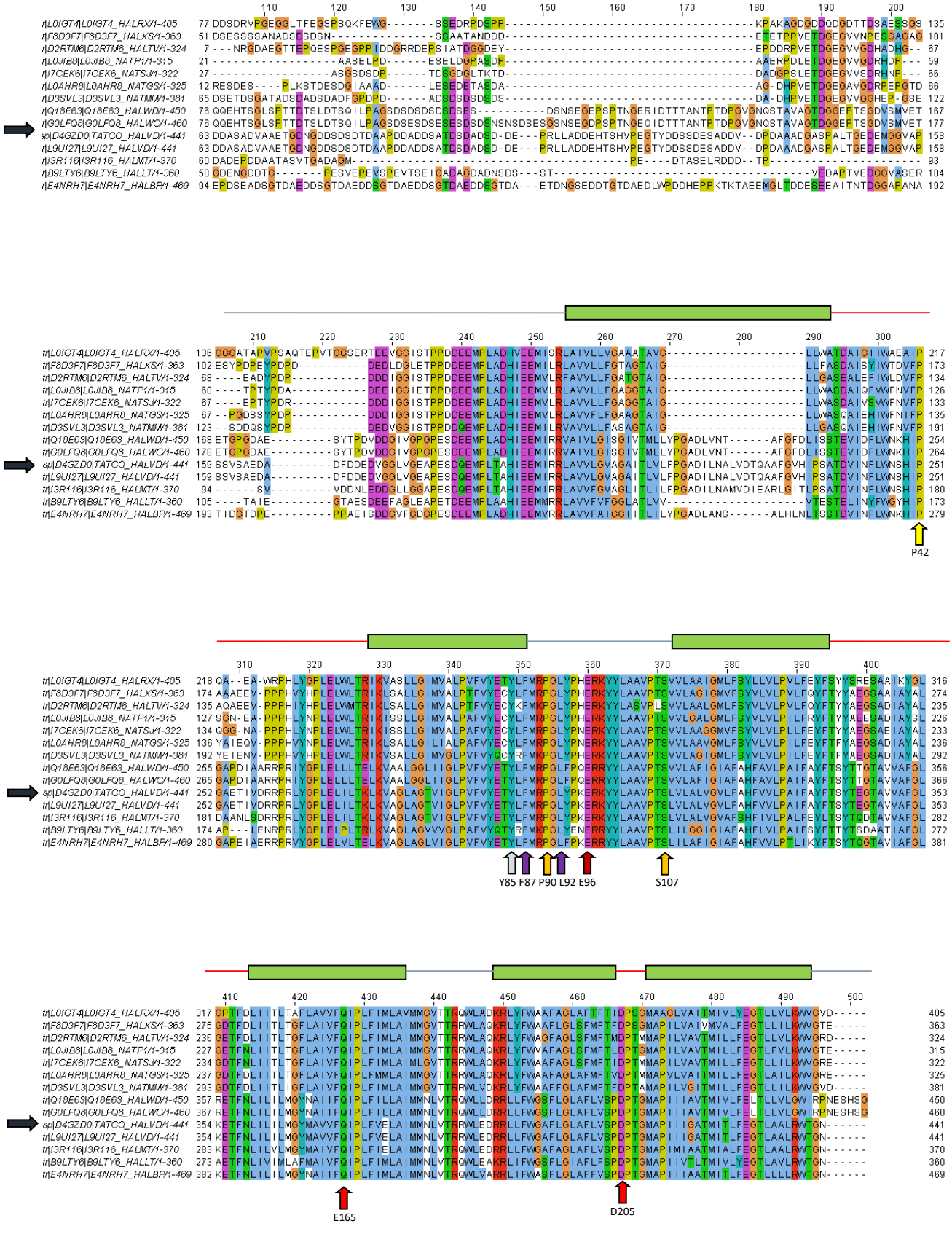
TatCo



B

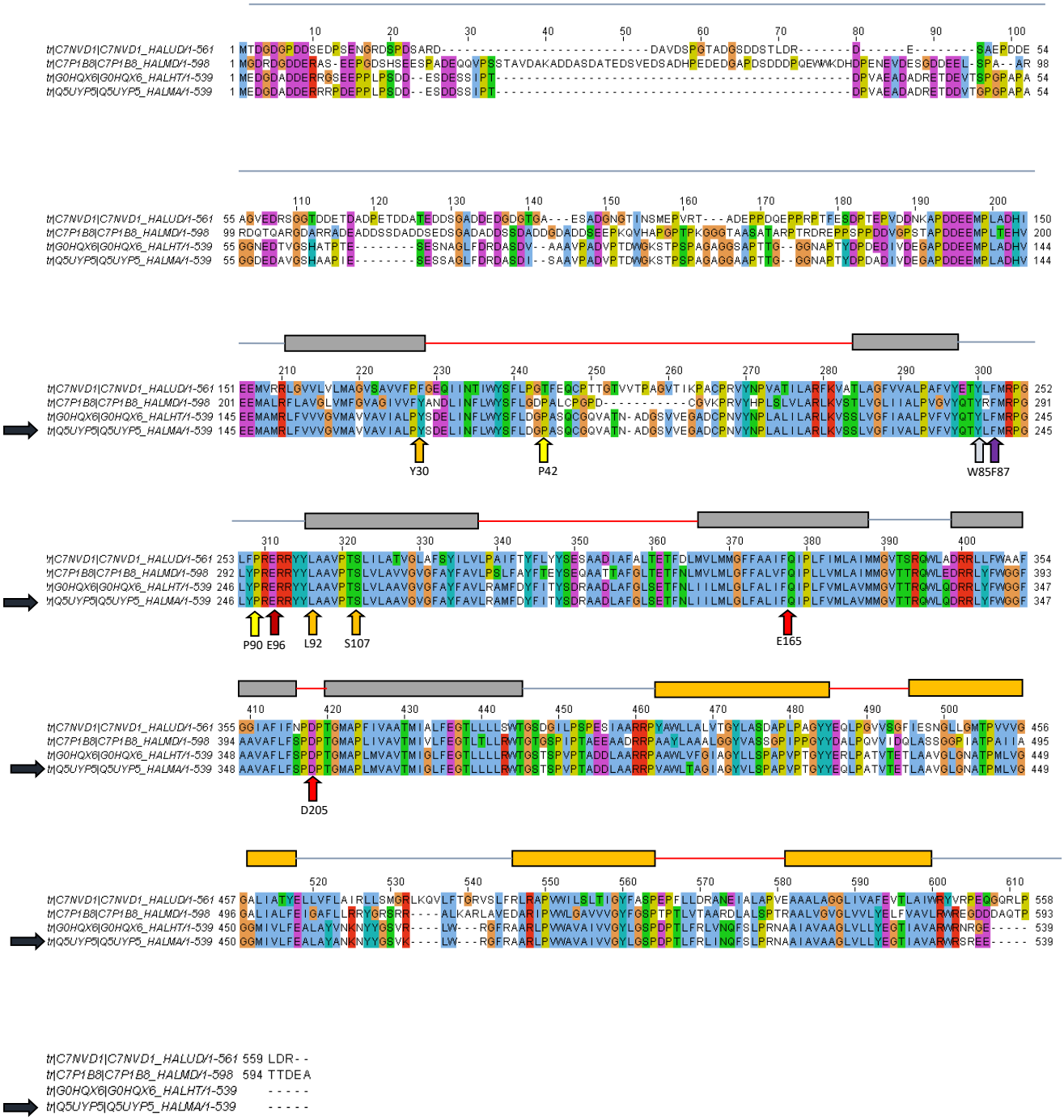
TatCn




P42
Y85 F87 P90 L92 E96
S107
E165
D205

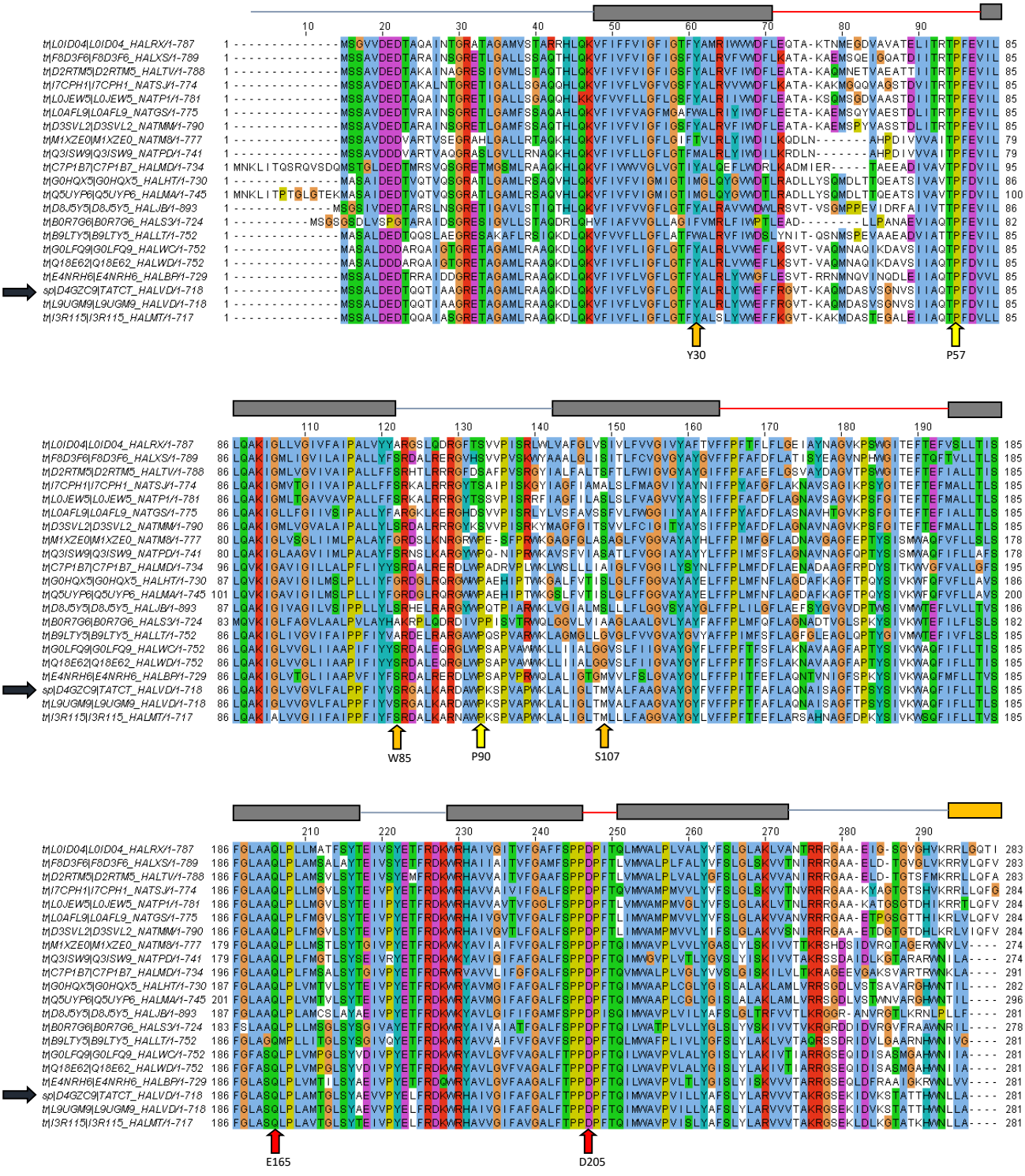
C

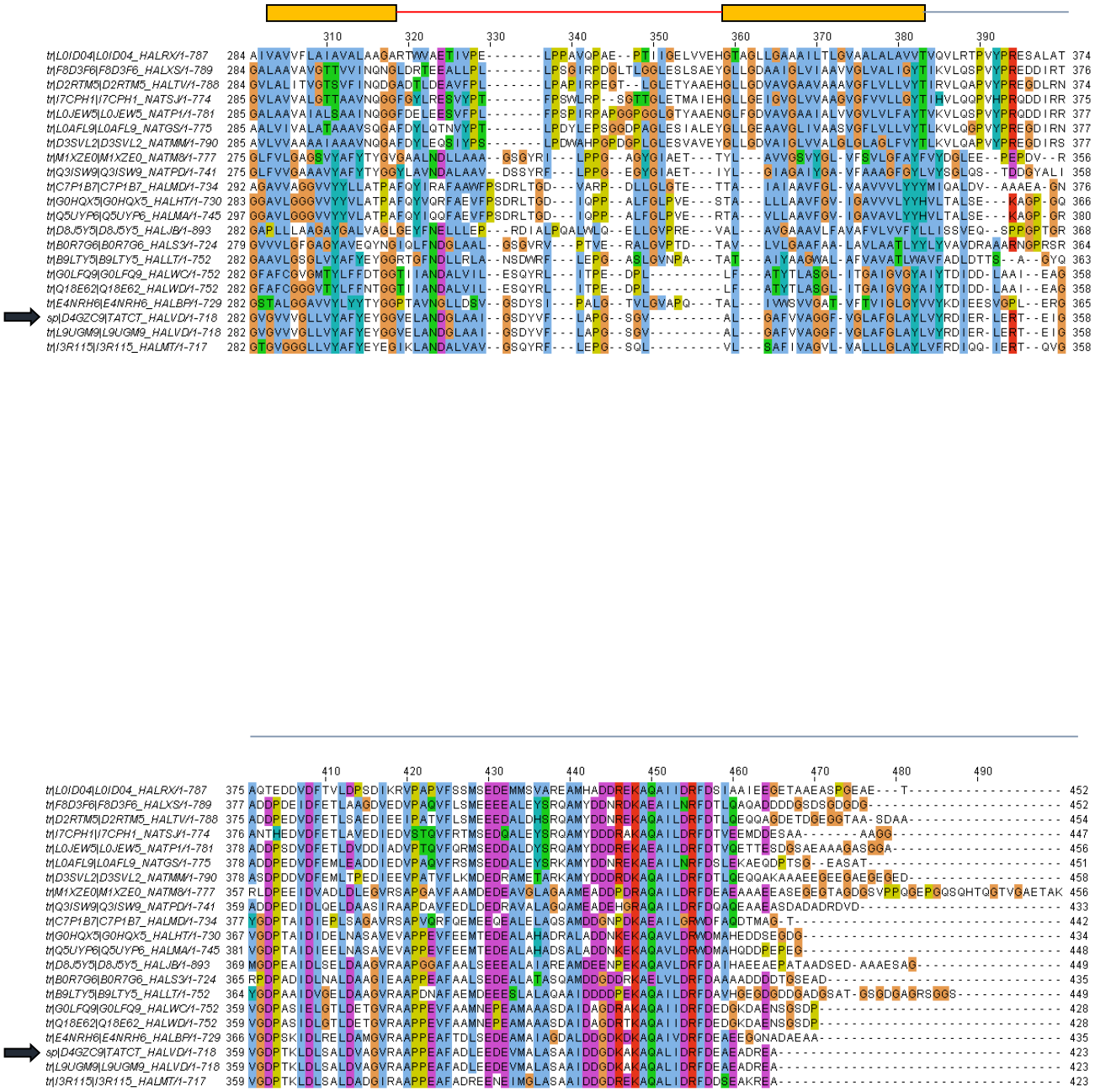
TatCx

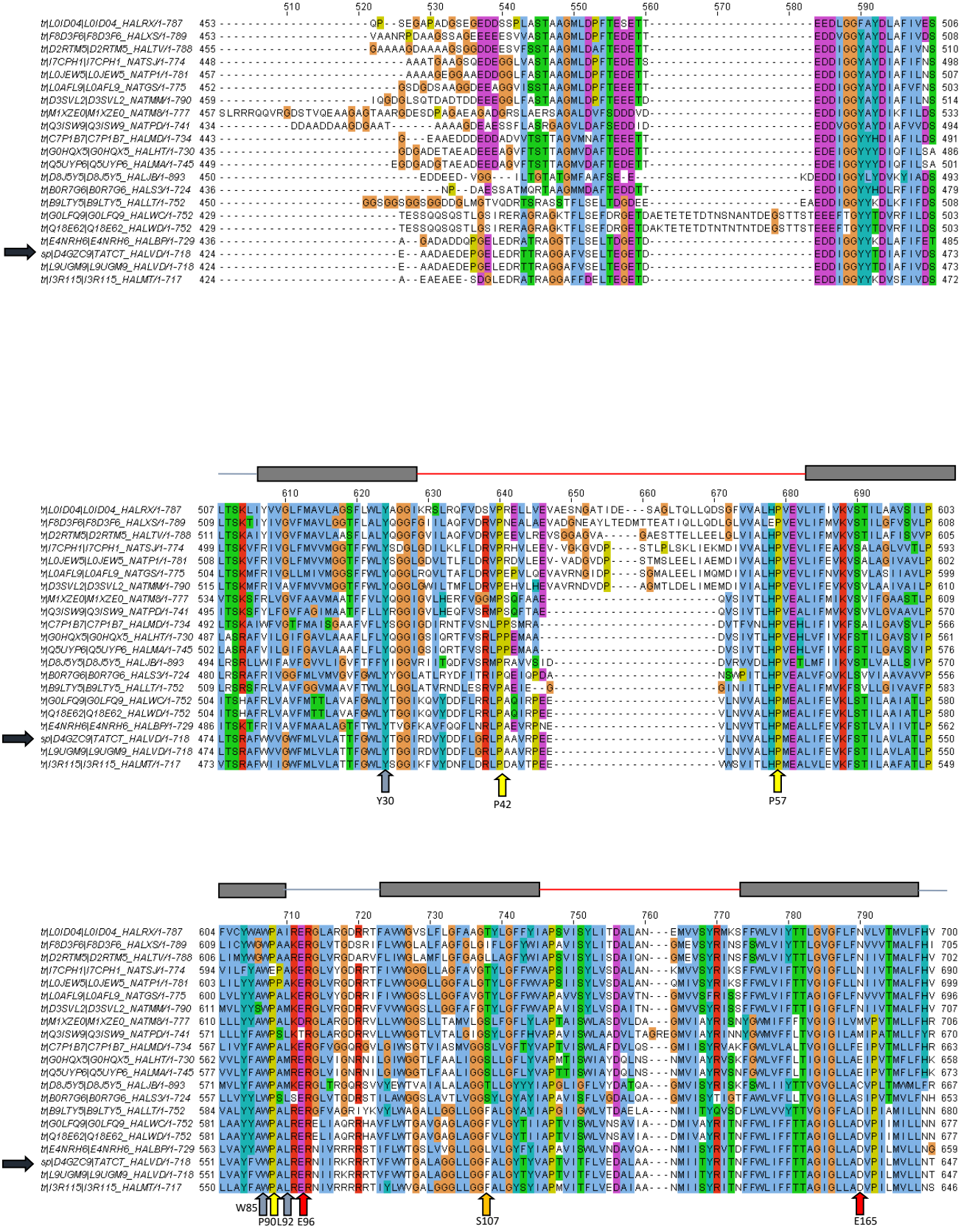

t|C7NVD1|C7NVD1_HALUD/1-561
t|C7P1B8|C7P1B8_HALMD/1-598
t|G0HQX6|G0HQX6_HALHT/1-539
t|Q5UYP5|Q5UYP5_HALMA/1-539
Y30
P42
W85F87
P90 E96
L92
S107
E165
D205

D

TatCt




310 320 330 340 350 360 370 380 390
410 420 430 440 450 460 470 480 490


Y30
P42
P57
W85
P90L92 E96
S107
E165

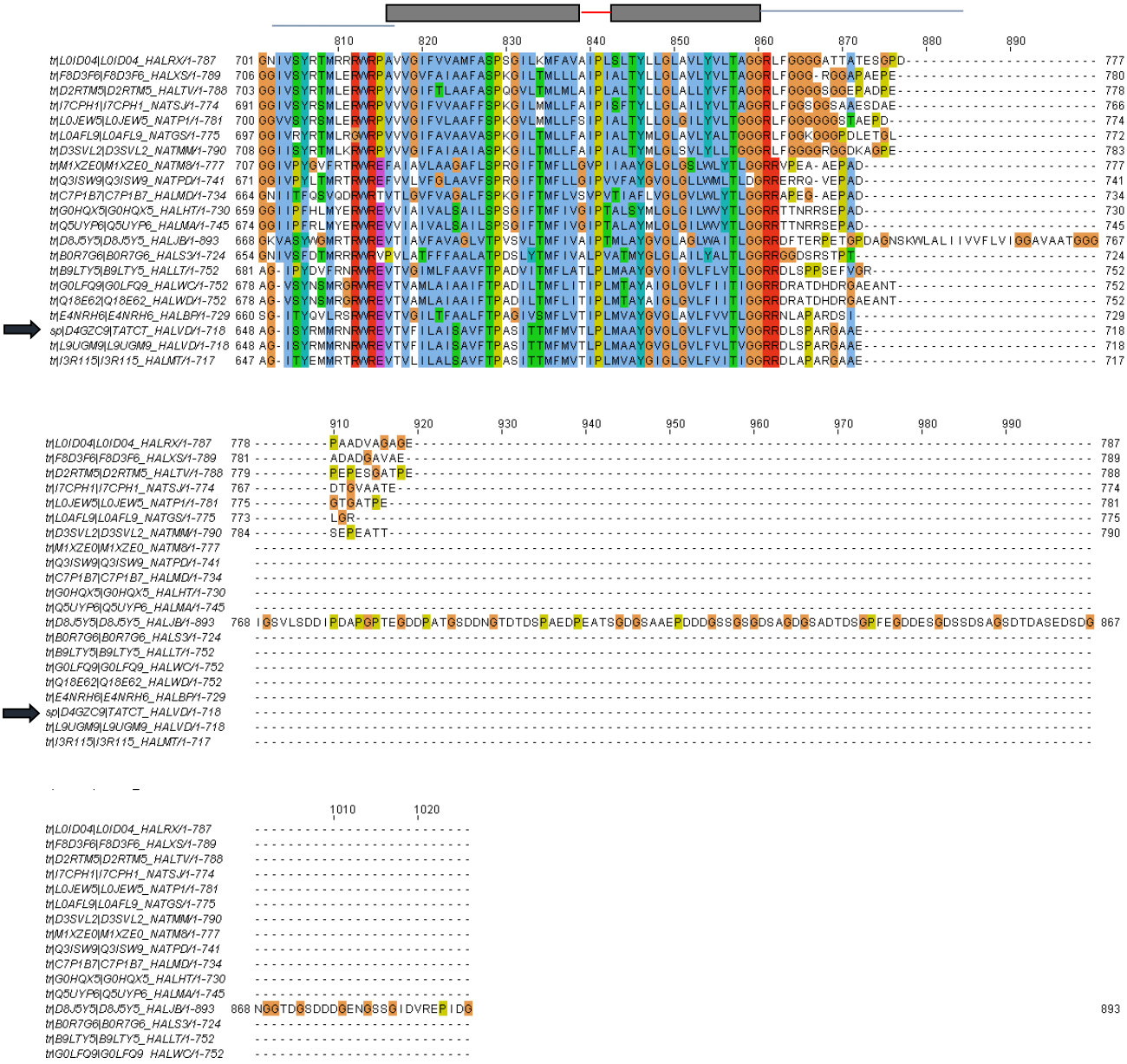


Fig. 4.4. Multiple Sequence alignment of the TatC proteins of different topologies. The important *aquifex* residues corresponding in the haloarchaea were marked with arrows having *aquifex* numbering. The black arrow corresponding to the species is the one for which transmembrane helices have been marked base on TMHMM predictions. Rectangles depicting TM domains, grey lines and red lines depicting cytoplasmic and extracellular loops respectively are marked on the top of the alignment. A) TatCo B) TatCn C)TatCn and D) TatCt

A. aeolicus Glu165, which is exposed at the center of the concave face and thus places an ionizable group in the hydrophobic interior of the bilayer [15][33], was replaced by glutamine and was conserved across TatCo, TatCn, TatCx and the N-terminal domain of TatCt. However, in the C-terminal TatCt domain, Glu165 was replaced with asparagine or aspartic acid. This forms hydrogen-bond networks with Ser107 and Tyr85 [34]. Both these residues were conserved across TatCo, TatCn and TatCx; they were only partially conserved in the two TatC domains of TatCt - Ser107 in the N-terminal TatCt domain and Tyr85 in the C-terminal TatCt domain. This evidence indicates that the difference in membrane properties in the haloarchaea Tat pathway might be due to the conservation of glutamine instead of glutamic acid. In addition, the differential conservation of the three residues Glu165, Ser107, and Tyr85 in the two Tat domains of TatCt might enable complementary functioning of the two domains for active transport.

Table 4.3 Conservation of important sequences across different TatC based on the multiple sequence alignments of the different TatC topologies

Residue (*Aquifex* numbering)	TatC0	TatCn	TatCx	TatCt-Nterm	TatCt-Cterm
Glu165	Asp/Asn	Asp/Asn	Asp/Asn	Asp/Asn	-
Ser107	+	+	+	+	-
Tyr85	+	+	+	-	+
Glu96	+	+	+	+	+
Pro42	+	+	+	-	+
Phe87	+	+	+	-	-
Pro90	+	+	+	+	+
Leu92	+	+	+	-	-
D205	+	+	+	+	-
Y30	-	-	+	-	+

Glu96 is another critical residue for Tat transport. This residue was conserved across TatCo, TatCn, TatCx and the C-terminal Tat domain of TatCt. A similar distribution was found even in case of the residues involved in signal binding in *Aquifex* [35]. The Pro42, Phe87, Pro90 and Leu92 residues were conserved in TatCo, TatCn, TatCx and C-terminal Tat domain of TatCt but only Pro90 was conserved in TatCt. Pro57 and Pro48 were conserved in the N-terminal Tat domain of TatCt instead of Pro42. We have also investigated the residues involved in TatC dimerization and interaction with other Tat pathway membrane components. While the Asp205 was conserved across all TatC classes except for the C-terminal TatC domain of TatCt, the Tyr30 residue was not conserved in TatCo, TatCn or the N-terminal TatC domain of TatCt.

We further surveyed a GC-content for these TatC homologs in order to test the possibility of horizontal gene transfer to haloarchaea. The GC content of the TatC gene and 50 upstream and 50 downstream genes were plotted. For comparison of TatC versus genomic GC content, the entire gene chromosome sequence was used to obtain the GC content. The two datasets were compared using a two tailed z-test. It is clear from Fig. 4.5 that there is no horizontal gene transfer. No significant difference was observed between the individual GC content versus the genomic GC content for any of the TatC homologs ($z = 1.044$; two-tailed $p = 0.2965$), thereby indicating that these TatC gene homologs may have evolved within the haloarchaea and consequently the presence of a correlation between the different TatC homologs and the types of substrate transported by them.

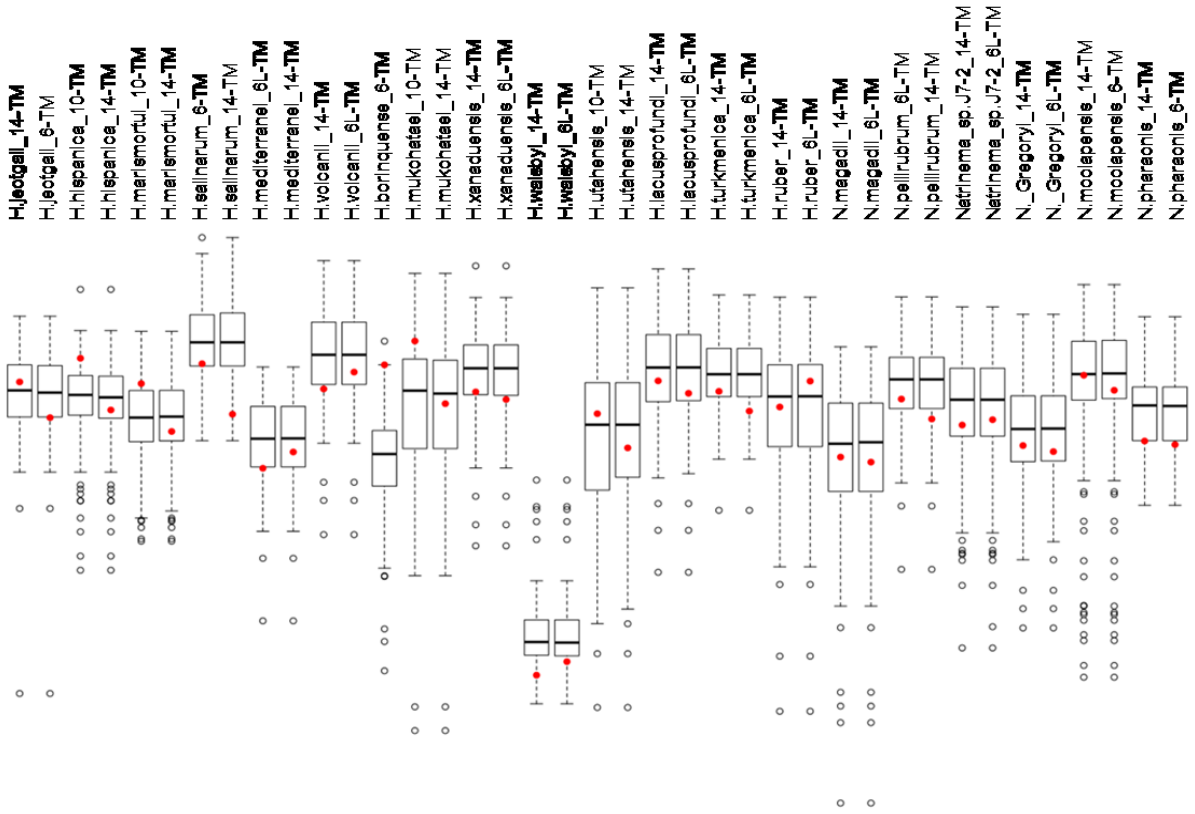


Fig. 4.5. The bar diagram representation of GC content of fifty genes flanking either side of the TatC (represented as filled red dot) of respective organism. The box corresponding to the percentage between first and the third quartile with the middle line being the median. The outliers are marked a hollow circles.

4.3.3 Haloarchaeal TatA

The size of TatA components in the selected Haloarchaea ranges from 75-145 amino acids. Phylogenetic analysis of unique TatA sequences (aligned with E. coli TatA and TatE and *A. aeolicus* TatA1 and TatA2) established that unlike TatCt, the TatA do not segregate according to the groups based on TatC combinations (Fig. 4.6). The alignment of E.coli TatA structure with haloarchaeal showed that the first TM region and the amphipathic helix region is highly conserved across all TatA components, but there was very little similarity in the n-terminal region (Fig. 4.7). TM residue Gln8 of *E. coli* TatA points inward, resulting in a short hydrophobic pore in the center of the complex. Different views were proposed based on simulations of the TatA complex in lipid bilayers indicate that the short TM domain distorts

the membrane [36]. This residue is replaced with Glu in all haloarchaea. The *E.coli* Gly21 remains conserved in haloarchaea that keeps TM helix and amphipathic helix at right angles except in the case of TatA1 of *Halomicrobiu mmukohataei* (UniProt id C7P2I0). The movement of C-terminal portion of the amphipathic helix is considered important for TatA function and Phe39, mutation (F39A) of which causes TatA inactivation [37], is conserved across haloarchaea. *Halorubrum lacusprofundi* TatA (UniProt id B9LS70) is an exception which has a non-aromatic amino acid leucine. Phylogenetic analysis also showed that these two TatA homologs (C7P2I0 and B9LS70) segregate out, which may suggest different substrate interaction or inactive TatA homologs as these organisms contain another TatA homolog. Presence of a predicted coiled coil motif in the C-terminal of *Natronomonas moolapensis* TatA homolog (UniProt id M1XKM7) might play additional function in recognition.

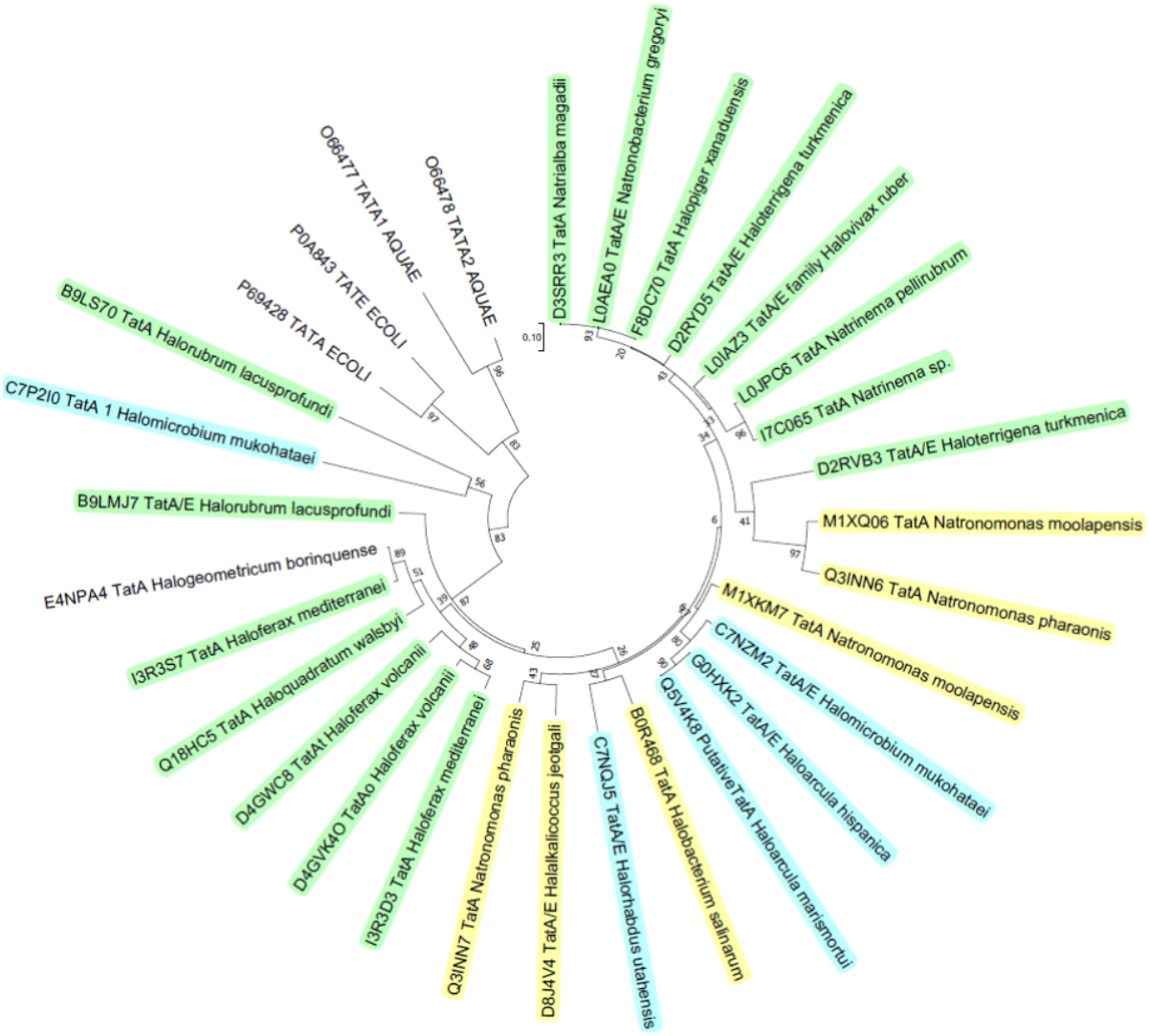


Fig. 4.6. Phylogenetic analysis of TatA/E from 20 haloarcheal species studies. Group 1 members are highlighted in yellow, group 2 in green and group 3 in blue.

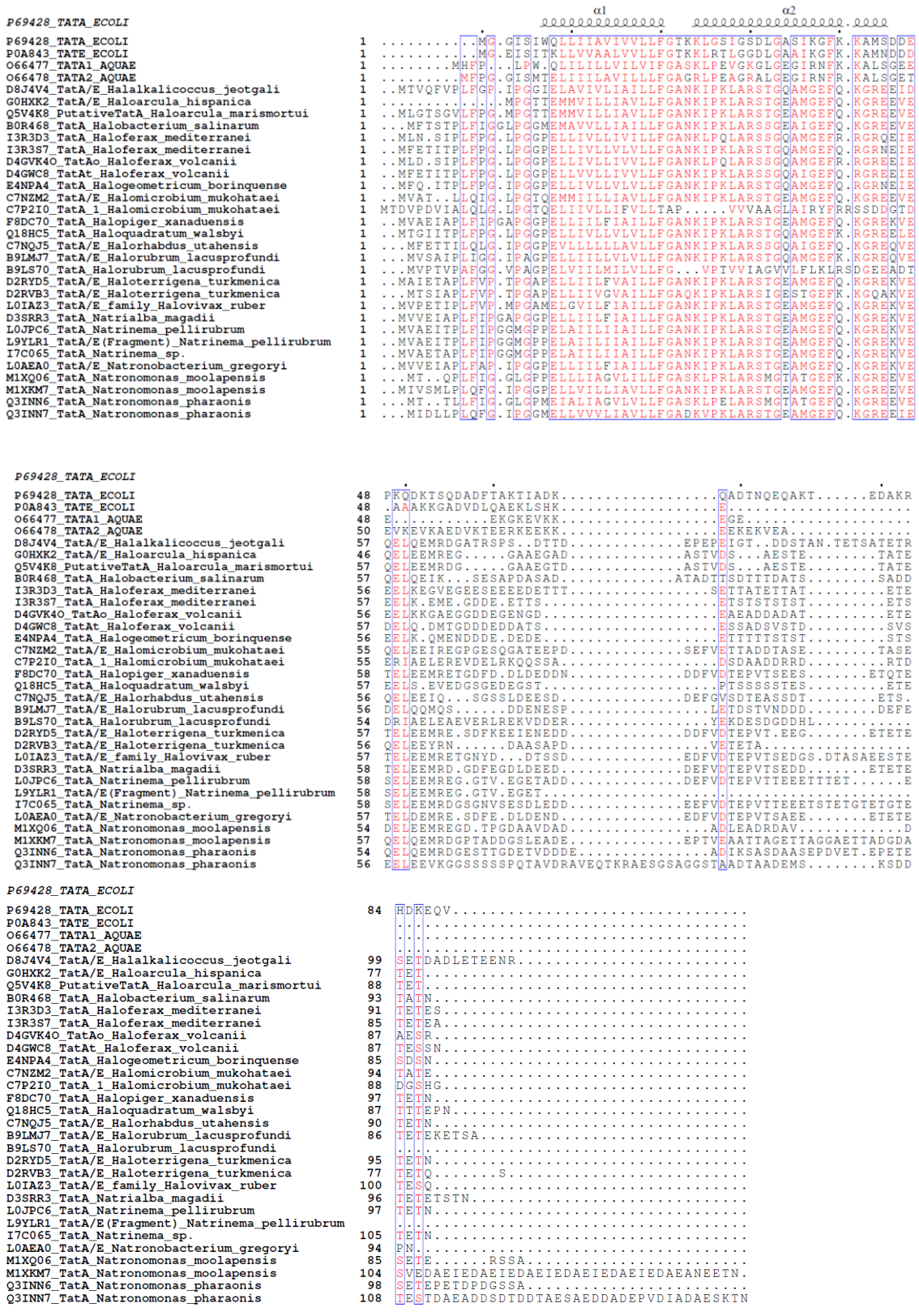


Fig.4.7 The alignment of Haloarchaea TatA/E with *E. coli* and *A. aeolicus* TatA/E

4.3.4 Tat substrate distribution among the different groups of haloarchaea

As discussed above, the three groups of species had TatCt in common; thus we hypothesized that the Tat substrates in each group would be translocated either by TatCt or by TatCo, TatCn or TatCx. The Tat substrates were identified from proteome of all twenty haloarchaea using the TATFIND 1.4 server [1]. To segregate the substrates, a bidirectional BLAST was performed against each other using standalone BLAST version 2.2.29. From these hits, only the bidirectional hits (BETs) were considered for analysis. The BETs were considered as shared sets and the substrates which had no hits were termed as unique sets. Furthermore, the substrates translocated by TatCt might exhibit common characteristics across all three groups in terms of signal and functional domains. We conducted extensive sequence analysis of the Tat substrates for these haloarchaea to check for any unique features exhibited by each group of substrates: those potentially transported by TatCt and those transported by the other TatC homologs [20]. From a total of 1,287,495 proteins, 2670 substrates were identified for the Group II (TatCt+TatCn), Group III (TatCt+TatCx), and Group I datasets (TatCt+TatCo) (Fig. 4.8).

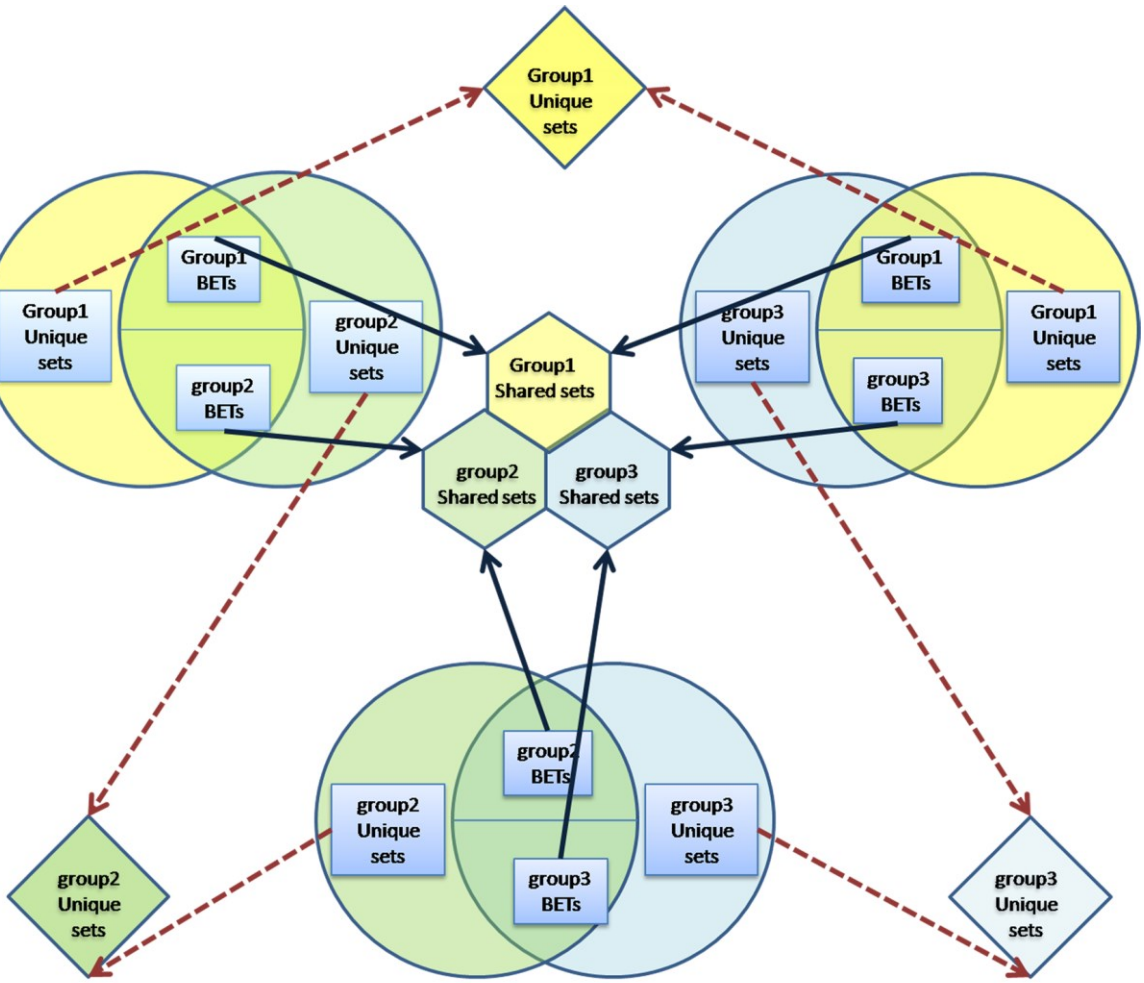


Fig. 4.8. Segregation of Substrates by biderectional BLAST analysis using stand alone BLAST, E-value cutoff of 0.5 was imposed and all hits more than 50% query coverage were considered as hits. The Bidirectional Best hits(BETs) were grouped as the shared sets and the non hits were grouped as unique sets.

4.3.5 Tat signal motif analysis

The Tat signal motif consists of three basic domains: a positively charged region at the N-terminal, a hydrophobic core and a more polar region that contains the cleavage site for a signal peptidase [1][30][38][39]. The TATFIND server identifies Tat substrates and also provides information about the N-terminal signature RR signal and the middle hydrophobic region. The canonical RR signal is the region, which directly interacts with the TatC receptor [40][41]; thorough analysis of this region was conducted for all Tat substrates in our dataset.

It was clear that the phenylalanine in the fifth position of the signal sequence was highly conserved in almost all the Tat substrates examined (Fig.4.9), however, there were also differences in the mean hydrophobicity. These results indicate that the signal region clearly plays a role in recognition, but that recognition is not solely due to either the RR signal or the hydrophobicity score of the signal region alone.

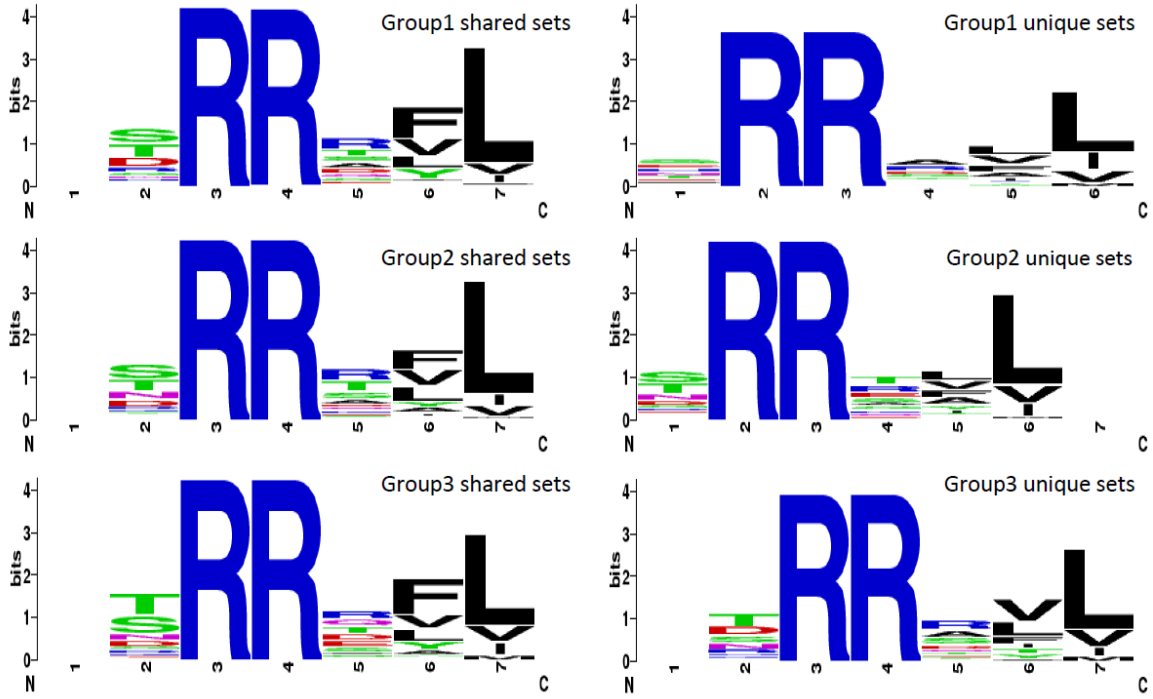


Fig. 4.9. Weblogo representation of Tat signal motifs in each class of substrates. WebLogo was generated at http://weblogo.berkeley.edu/logo.cgi.

The multiple sequence alignments of fifty amino acids of the N-terminal clearly showed a high occurrence of AGC sequences at the C-terminal of the hydrophobic region in a shared set of Tat substrates, but this was not observed in a unique set of substrates. This AGC sequence forms a distinctive recognition site called the 'lipobox' [25] and was depicted via the WebLogos (Fig.4.10). In the mature lipoprotein, this cysteine residue is attached to a membrane-associated lipid anchor. Analysis using the TatLipo server [2] confirmed that most of the substrates in the shared set were lipobox positive and were probably lipoproteins. This

feature therefore clearly distinguishes between the shared (exported by more than one class of TatC) substrate datasets and the unique (exported by one type of TatC) substrate datasets. It is also probable that TatCt mainly transports proteins from the shared set, which are probably lipoproteins, whereas the other TatC homolog classes may transport the remaining unique set of substrates, which may be secreted proteins.

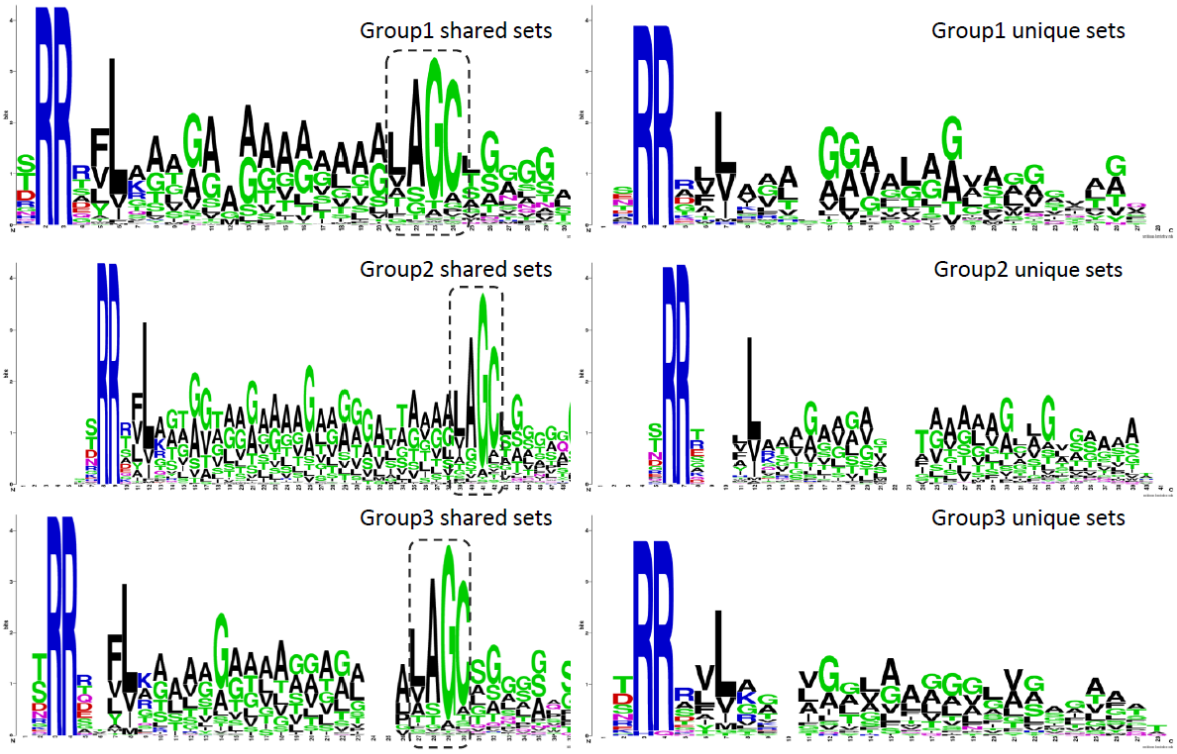


Fig. 4.10. Weblogo representation of Tat signal peptide regions from different classes of substrates showing the occurrence of lipoboxmotif (AGC) in the shared substrates. WebLogo was generated at http://weblogo.berkeley.edu/logo.cgi.

4.3.6 Functional analysis of the Tat pathway of haloarchaea

We assessed the substrates in our dataset for specific functional roles, but in haloarchaea only a few have been experimentally tested and most are uncharacterized. We analyzed each substrate using the Pfam database to identify the different domains, repeats and signature motifs. The unique and common HMMs were identified in the different Tat groups (Table 4.4). A total of 891 unique families were identified among the 2670 haloarchaea Tat

substrates. From the analysis, it can be inferred that there is more domain diversity across the unique set and that the shared set has many domains in common (Fig. 4.11 and Table 4.4).

Table 4.4. Distribution of Pfam families of the different datasets as identified from Pfam analysis showing the number of Pfam families identified versus the number of proteins with Pfam families predicted in each data set

Datasets	Number of pfam families	Number of proteins with pfam families predicted
Group1 sharedsets	109	231
Group1 uniquesets	47	29
Group2 sharedsets	237	555
Group2 uniquesets	208	144
Group3 sharedsets	136	276
Group3 uniquesets	47	31

The Pfam families for the shared set of substrates were then compared across datasets. Fifty eight families were common to all the groups for the shared set, but only two families were common to all the groups for the unique set (Fig.4.11). To assess the significance of the shared and unique sets of families, the distribution of these domains across the groups were analyzed by re-mapping the list of Pfam families to the substrates in each set and then mapping all the families present in these proteins. Although there were many unique Pfam families in both sets, most of the shared substrates contained one or more common Pfam families.

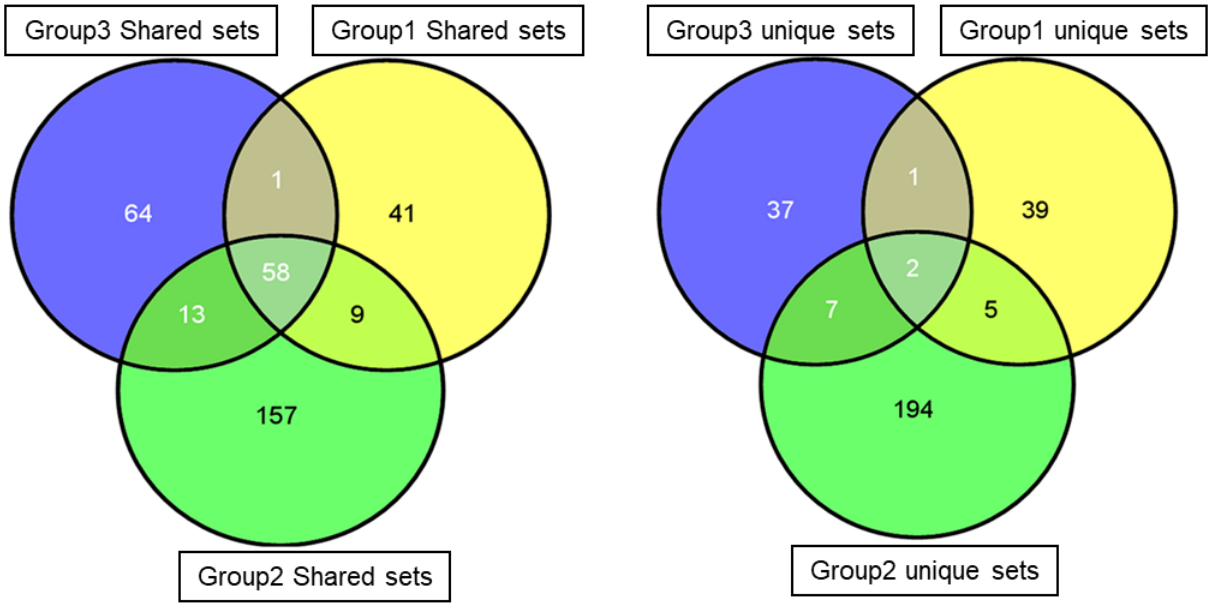


Fig. 4.11. Venn diagrammatic representation of the Pfam family distribution in shared and unique group of substrates. Venn diagrams were generated with http://bioinfogp.cnb.csic.es/tools/venny/.

The 58 frequently occurring Pfam families found in all the shared sets were mostly metalbinding domains, periplasmic substrate-binding domains that are responsible for ion transport, glycosidases, dehydrogenases and domains responsible for the biosynthesis of pyrimidine and trypsin-2. There were also four domains of unknown function and four from PfamB. It was quite clear that the proteins in the shared substrates set were mostly membrane-associated proteins, which correlates with the presence of lipobox in the signal region. The unique substrates set generally had lyase or hydrolase-type domains. Interestingly, a group of unique substrates in the group II were identified as 7TM-GPCRs, PAZ domains, Cox II, and a few metal-binding domains. However, most of the Tat substrates present in the unique substrates set were secreted proteins.

4.3.7 Protein-folding quality control

The quality-control mechanism for substrates transported by the Tat machinery is well established. In the bacterial system, this process is taken care of by chaperones that referred as REMP [42], which mask the twin-arginine signal and ensure proper folding or substrate

maturation with appropriate cofactor loading [43]. In the case of halophilic archaea, the presence of such a control mechanism may be even more vital because the Tat pathway transports almost the entire secretome. We therefore identified chaperone homologs from the proteomes of these organisms. Homologs for E. coli REMP such as DmsD, HyaE, HybE, NapD, NarJ, NarW, TorD and YcdY were identified. Pairwise sequence analysis showed that the key sequence motif shown in Fig.4.12, Table. 4.5 was partially conserved and the protein sizes were comparable in most of the cases, indicating that the haloarchaeal Tat pathway uses chaperones similar to their bacterial counterparts.

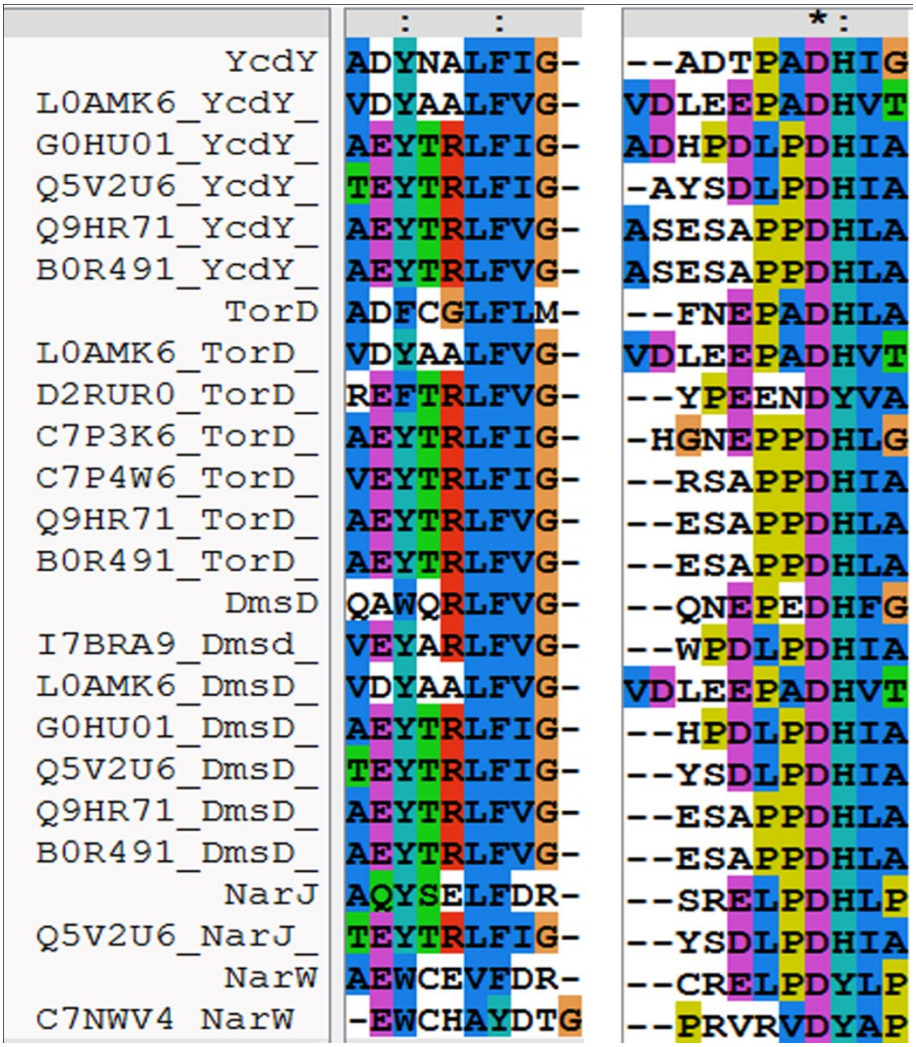


Fig. 4.12. REMP motifs *E. coli* Chaperons aligned to identified Haloarchaeal chaperons

Table 4.5. BLASTp pairwise sequence alignment using REMP motif signature sequence of *E. coli* from which the Haloarchaea Tat Chaperons were identified

Protein	*E .coli* protein ID	Hit ID	%identity	q start	q end	h start	h end	e value	bitscore	q coverage
DmsD	E0IWW9	I7BRA9	32.23	61	181	46	165	3.00E-13	64.7	59
DmsD	E0IWW9	L0AMK6	29.6	72	182	68	190	4.00E-10	57.4	54
DmsD	E0IWW9	L0AFH1	31.67	63	175	72	189	5.00E-04	39.7	55
DmsD	E0IWW9	F8D6S0	39.66	131	181	144	201	6.00E-04	39.7	25
DmsD	E0IWW9	D2RRH6	32.76	131	181	144	201	7.20E-02	33.1	25
DmsD	E0IWW9	Q5V2U6	30.77	51	188	53	195	2.00E-16	73.2	68
DmsD	E0IWW9	G0HU01	31.97	72	188	74	195	1.00E-13	65.1	57
DmsD	E0IWW9	C7P4W6	30.77	72	180	66	182	2.00E-11	58.9	53
DmsD	E0IWW9	Q9HR71	33.86	72	193	66	192	6.00E-15	68.9	60
DmsD	E0IWW9	B0R491	33.86	72	193	66	192	6.00E-15	68.9	60
DmsD	E0IWW9	M1XZ62	40.74	105	131	197	223	0.28	30	13
HyaE	E0IZ67	Q18DX2	38.46	52	88	39	77	0.049	31.2	28
HyaE	E0IZ67	G0LGF3	38.46	52	88	39	77	4.90E-02	31.2	28
HyaE	E0IZ67	L0AH44	27.21	1	121	89	213	6.60E-01	29.3	92
HyaE	E0IZ67	C7NUT4	34.04	22	67	251	297	0.33	28.9	35
HyaE	E0IZ67	M1XTM8	37.04	58	111	164	216	0.084	30.4	41
HyaE	E0IZ67	M1XTG4	42.86	79	106	169	196	0.49	28.5	21
HybE	P0AAN1	C7NUU4	29.17	23	69	18	65	0.6	28.5	29
NapD	P0A9I5	G0HVU2	32.56	27	65	207	248	0.81	26.2	45
NapD	P0A9I5	D8J682	35.29	16	63	102	150	0.12	28.5	55
NarJ	E0IV30	Q5UZZ6	32.76	36	85	928	984	0.46	30	21
NarJ	E0IV30	G0HXZ2	27.27	36	115	914	995	0.49	30	34
NarJ	E0IV30	Q5V2U6	24.37	16	124	36	145	0.5	29.3	46
NarJ	E0IV30	G0HS22	32.76	36	85	928	984	0.6	29.6	21
NarW	E0J6C1	L9YDB2	33.82	110	176	19	86	0.52	30.8	29
NarW	E0J6C1	L0JRK8	33.82	110	176	26	93	0.52	30.8	29
NarW	E0J6C1	C7NWV4	32.94	54	129	211	292	0.045	32.7	33
TorD	A7ZKA2	L0AMK6	27.89	12	186	20	188	0.009	35.8	88
TorD	A7ZKA2	F8D6S0	35.29	117	195	126	209	5.00E-02	33.5	40
TorD	A7ZKA2	D2RRH6	32.22	116	195	125	209	6.00E-02	33.5	40
TorD	A7ZKA2	D2RUR0	24.53	49	188	43	195	7.50E-02	32.7	70
TorD	A7ZKA2	F8DCA1	26.39	57	188	58	195	2.20E-01	31.6	66
TorD	A7ZKA2	L0ICG2	35.29	99	149	34	84	3.40E-01	29.6	26
TorD	A7ZKA2	Q18KR0	40.63	15	46	104	135	7.80E-01	30	16
TorD	A7ZKA2	G0LHK7	40.63	15	46	104	135	8.00E-01	30	16
TorD	A7ZKA2	C7P3K6	26.54	5	197	23	220	1.00E-05	42.7	97
TorD	A7ZKA2	C7P4W6	23.4	68	192	60	188	5.00E-04	38.1	63

TorD	A7ZKA2	Q5V7U6	36.51	62	124	26	84	8.80E-02	31.2	32
TorD	A7ZKA2	Q5V855	38.1	62	124	26	84	9.20E-02	31.2	32
TorD	A7ZKA2	G0HU01	23.2	74	185	70	186	1.70E-01	30.4	56
TorD	A7ZKA2	G0HY93	30.88	100	167	3	70	4.20E-01	28.5	34
TorD	A7ZKA2	Q5V0U6	30.88	100	167	3	70	5.40E-01	28.1	34
TorD	A7ZKA2	C7NNI8	33.77	124	198	154	226	5.70E-01	28.9	38
TorD	A7ZKA2	G0HZ34	36.54	91	133	90	141	6.60E-01	28.9	22
TorD	A7ZKA2	G0HUZ8	32.05	80	151	518	589	6.70E-01	29.3	36
TorD	A7ZKA2	Q9HR71	25.81	74	184	62	177	3.00E-05	41.6	56
TorD	A7ZKA2	B0R491	25.81	74	184	62	177	3.00E-05	41.6	56
TorD	A7ZKA2	Q3IR60	28	112	195	149	238	7.00E-03	35	42
TorD	A7ZKA2	Q3IMT7	29.03	104	195	106	188	6.00E-02	32	46
TorD	A7ZKA2	O52027	41.67	108	154	367	414	9.80E-02	31.6	24
TorD	A7ZKA2	B0RA87	41.67	108	154	367	414	0.098	31.6	24
TorD	A7ZKA2	M1XL07	42.22	108	151	366	410	0.47	29.6	22
YcdY	P75915	L0AMK6	32.03	56	164	60	187	1.00E-06	47	59
YcdY	P75915	G0HU01	33.58	54	172	64	194	6.00E-11	57	65
YcdY	P75915	Q5V2U6	32.09	54	171	64	193	2.00E-09	52.8	64
YcdY	P75915	C7P4W6	27.4	48	172	50	189	8.00E-07	45.8	68
YcdY	P75915	Q9HR71	27.08	44	172	46	186	7.00E-08	48.5	70
YcdY	P75915	B0R491	27.08	44	172	46	186	7.00E-08	48.5	70
YcdY	P75915	Q3INV6	34.09	136	179	19	62	3.40E-01	29.6	24

4.4 Discussion

Overall, these analyses provide insight into the diversity of the indispensable Tat pathway in the haloarchaeal system. The evolution of the Tat pathway itself suggests that there may be a fundamental conflict between the substrates transported and the Sec export mechanism in organisms such as haloarchaea that live in extreme habitats. The diversity and multiple unique topologies of the TatC receptor in haloarchaea indicate that they may exist specifically for transport of large numbers and a wide variety of Tat substrates. Conservation in the N-terminal region of TatA homologs to that of *E.coli* implies similar mechanism of transport but diversity in the C-terminal region could be useful in interaction with the wide range of substrates. This is a requirement for haloarchaeal proteins; they would otherwise aggregate after exiting the ribosome because haloarchaea have a high intracellular concentration of positive ions in order to maintain osmotic balance against the high

extracellular sodium concentration. The quality of protein folding is a major concern for these organisms, and is positively facilitated by the sophisticated Tat pathway components. Unlike in other systems, many substrate groups are present here based on the class of receptor required for the translocation. This may be necessary for differential translocation dynamics and to ensure the stringent folding quality. Further invivo and invitro studies are required for dissecting the dynamics of the Tat pathway in haloarchaea.

4.5 References

[1] R. W. Rose, T. Bruser, J. C. Kissinger, and M. Pohlschroder, "Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway," *Mol. Microbiol.*, vol. 45, no. 4, pp. 943–950, Aug. 2002.

[2] M. Pohlschröder, S. Storf, F. Pfeiffer, K. Dilks, Z. Q. Chen, and S. Imam, "Mutational and bioinformatic analysis of haloarchaeal lipobox-containing proteins," *Archaea*, vol. 2010, 2010.

[3] K. Dilks, R. W. Rose, E. Hartmann, and M. Pohlschröder, "Prokaryotic utilization of the twin-arginine translocation pathway: A genomic survey," *J. Bacteriol.*, vol. 185, no. 4, pp. 1478–1483, 2003.

[4] J. R. Bolhuis, A., Kwan, D. and Thomas, *Halophilic adaptations of proteins. Protein adaptation in extremophiles*. 2008.

[5] T. Palmer and B. C. Berks, "The twin-arginine translocation (Tat) protein export pathway," *Nat. Rev. Microbiol.*, vol. 10, no. 7, pp. 483–496, Jul. 2012.

[6] F. Alcock *et al.*, "Assembling the Tat protein translocase.," *Elife*, vol. 5, 2016.

[7] F. Sargent *et al.*, "Overlapping functions of components of a bacterial Sec-independent protein export pathway," *EMBO J.*, vol. 17, no. 13, pp. 3640–3650, Jul. 1998.

[8] J. H. Weiner *et al.*, "A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins," *Cell*, vol. 93, no. 1, pp. 93–101, 1998.

[9] B. C. Berks, T. Palmer, C. Robinson, N. R. Stanley, E. G. Bogsch, and F. Sargent, "An Essential Component of a Novel Bacterial Protein Export System with Homologues in

Plastids and Mitochondria," *J. Biol. Chem.*, vol. 273, no. 29, pp. 18003–18006, 2002.

[10] J. D. Jongbloed *et al.*, "TatC is a specificity determinant for protein secretion via the twin-arginine translocation pathway.," *J. Biol. Chem.*, vol. 275, no. 52, pp. 41350–7, Dec. 2000.

[11] Q. Huang and T. Palmer, "Signal Peptide Hydrophobicity Modulates Interaction with the Twin-Arginine Translocase," *MBio*, vol. 8, no. 4, pp. 1–17, 2017.

[12] D. Mangels, J. Mathers, A. Bolhuis, and C. Robinson, "The Core TatABC Complex of the Twin-arginine Translocase in Escherichia coli: TatC Drives Assembly Whereas TatA is Essential for Stability," *J. Mol. Biol.*, vol. 345, no. 2, pp. 415–423, Jan. 2005.

[13] S. Molik, I. Karnauchov, C. Weidlich, R. G. Herrmann, and R. B. Klösgen, "The Rieske Fe/S protein of the cytochrome b6/f complex in chloroplasts: Missing link in the evolution of protein transport pathways in chloroplasts?," *J. Biol. Chem.*, vol. 276, no. 46, pp. 42761–42766, 2001.

[14] C. P. New, Q. Ma, and C. Dabney-Smith, "Routing of thylakoid lumen proteins by the chloroplast twin arginine transport pathway," *Photosynth. Res.*, vol. 138, no. 3, pp. 289–301, 2018.

[15] S. Ramasamy, R. Abrol, C. J. M. Suloway, and W. M. Clemons, "The glove-like structure of the conserved membrane protein TatC provides insight into signal sequence recognition in twin-arginine translocation.," *Structure*, vol. 21, no. 5, pp. 777–88, May 2013.

[16] S. E. Rollauer *et al.*, "Structure of the TatC core of the twin-arginine protein transport system.," *Nature*, vol. 492, no. 7428, pp. 210–4, Dec. 2012.

[17] C. Punginelli *et al.*, "Cysteine scanning mutagenesis and topological mapping of the Escherichia coli twin-arginine translocase TatC component," *J. Bacteriol.*, vol. 189, no. 15, pp. 5482–5494, 2007.

[18] Y. Hu, E. Zhao, H. Li, B. Xia, and C. Jin, "Solution NMR Structure of the TatA Component of the Twin-Arginine Protein Transport System from Gram-Positive Bacterium *Bacillus subtilis*," *J. Am. Chem. Soc.*, vol. 132, no. 45, pp. 15942–15944,

Nov. 2010.

[19] H. Shruthi, P. Anand, V. Murugan, and K. Sankaran, "Twin arginine translocase pathway and fast-folding lipoprotein biosynthesis in E. coli: interesting implications and applications," *Mol. Biosyst.*, vol. 6, no. 6, p. 999, May 2010.

[20] K. Dilks and M. I. Giménez, "Genetic and Biochemical Analysis of the Twin-Arginine Translocation Pathway in Halophilic Archaea Genetic and Biochemical Analysis of the Twin-Arginine Translocation Pathway in Halophilic Archaea," 2005.

[21] M. Moser, S. Panahandeh, E. Holzapfel, and M. Müller, "In Vitro Analysis of the Bacterial Twin-Arginine-Dependent Protein Export," in *Protein Targeting Protocols*, Totowa, NJ: Humana Press, 2007, pp. 63–79.

[22] R. J. Turner, A. L. Papish, and F. Sargent, "Sequence analysis of bacterial redox enzyme maturation proteins (REMPs)," *Can. J. Microbiol.*, vol. 50, no. 4, pp. 225–238, Apr. 2004.

[23] S. K. Ramasamy, W. M. Clemons, and IUCr, "Structure of the twin-arginine signal-binding protein DmsD from *Escherichia coli*," *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, vol. 65, no. 8, pp. 746–750, Aug. 2009.

[24] A. L. Papish, C. L. Ladner, and R. J. Turner, "The twin-arginine leader-binding protein, DmsD, interacts with the TatB and TatC subunits of the Escherichia coli twin-arginine translocase.," *J. Biol. Chem.*, vol. 278, no. 35, pp. 32501–6, Aug. 2003.

[25] A. Bolhuis, "REVIEW ARTICLE Protein transport in the halophilic archaeon Halobacterium sp . NRC-1 : a major role for the twin-arginine translocation pathway ?," pp. 3335–3346, 2002.

[26] A. Krogh, B. Larsson, G. Von Heijne, and E. L. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes," *J. Mol. Biol.*, vol. 305, no. 3, pp. 567–580, 2001.

[27] G. E. Tusnády and I. Simon, "The HMMTOP transmembrane topology prediction server," *Bioinformatics*, vol. 17, no. 9, pp. 849–850, 2001.

[28] M. Cserzö, F. Eisenhaber, B. Eisenhaber, and I. Simon, "On filtering false positive transmembrane protein predictions," *Protein Eng. Des. Sel.*, vol. 15, no. 9, pp. 745–752, 2002.

[29] A. Bernsel, H. Viklund, J. Falk, E. Lindahl, G. von Heijne, and A. Elofsson, "Prediction of membrane-protein topology from first principles," *Proc. Natl. Acad. Sci.*, vol. 105, no. 20, pp. 7177–7181, 2008.

[30] M. Pohlschröder, M. I. Giménez, and K. F. Jarrell, "Protein transport in Archaea: Sec and twin arginine translocation pathways.," *Curr. Opin. Microbiol.*, vol. 8, no. 6, pp. 713–9, Dec. 2005.

[31] D. Kwan and A. Bolhuis, "Analysis of the twin-arginine motif of a haloarchaeal Tat substrate," *FEMS Microbiol. Lett.*, vol. 308, no. 2, pp. 138–143, 2010.

[32] H. Mori, E. J. Summer, and K. Cline, "Chloroplast TatC plays a direct role in thylakoid ΔpH-dependent protein transport," *FEBS Lett.*, vol. 501, no. 1, pp. 65–68, 2001.

[33] S. E. Rollauer *et al.*, "Structure of the TatC core of the twin-arginine protein transport system," *Nature*, vol. 492, no. 7428, pp. 210–214, Dec. 2012.

[34] E. Holzapfel *et al.*, "The entire N-terminal half of TatC is involved in twin-arginine precursor binding," *Biochemistry*, vol. 46, no. 10, pp. 2892–2898, 2007.

[35] G. Buchanan *et al.*, "Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis," *Mol. Microbiol.*, vol. 43, no. 6, pp. 1457–1470, Mar. 2002.

[36] F. Rodriguez *et al.*, "Structural model for the protein-translocating element of the twin-arginine transport system," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 12, pp. E1092-101, 2013.

[37] M. G. Hicks, P. A. Lee, G. Georgiou, B. C. Berks, and T. Palmer, "Positive selection for loss-of-function tat mutations identifies critical residues required for TatA activity," *J. Bacteriol.*, vol. 187, no. 8, pp. 2920–2925, 2005.

[38] B. C. Berks, "A common export pathway for proteins binding complex redox

cofactors?," *Mol. Microbiol.*, vol. 22, no. 3, pp. 393–404, Nov. 1996.

[39] A. M. Chaddock *et al.*, "A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the delta pH-dependent thylakoidal protein translocase.," *EMBO J.*, vol. 14, no. 12, pp. 2715–22, 1995.

[40] M. Alami *et al.*, "Differential Interactions between a Twin-Arginine Signal Peptide and Its Translocase in Escherichia coli," *Mol. Cell*, vol. 12, no. 4, pp. 937–946, Oct. 2003.

[41] X. Ma and K. Cline, "Mapping the Signal Peptide Binding and Oligomer Contact Sites of the Core Subunit of the Pea Twin Arginine Protein Translocase," *Plant Cell*, vol. 25, no. 3, pp. 999–1015, 2013.

[42] C. S. Chan, L. Chang, K. L. Rommens, and R. J. Turner, "Differential interactions between tat-specific redox enzyme peptides and their chaperones," *J. Bacteriol.*, vol. 191, no. 7, pp. 2091–2101, 2009.

[43] C. M. Stevens, T. M. L. Winstone, R. J. Turner, and M. Paetzel, "Structural Analysis of a Monomeric Form of the Twin-Arginine Leader Peptide Binding Chaperone Escherichia coli DmsD," *J. Mol. Biol.*, vol. 389, no. 1, pp. 124–133, 2009.

Chapter V - Structural Investigation of Tat Membrane complex

5.1 Introduction

In *Mycobacterium tuberculosis* the Tat pathway exported proteins, include virulence factors of periplasmic substrate-binding proteins of ABC transporters and Lactamases [1]. The Tat system is required for optimal growth of *M. smegmatis* in-vitro [2][1]. Thus the pathway can be targeted to control the virulence of *M. tuberculosis*. The strategies to develop the antimicrobial or antibiotic from the studies are very fascinating. By given the fact that elimination or hampering the Tat pathway which is not present in the human or in higher mammals [3]are very advantageous in many aspects. Also, it is very sustainable idea as it won't put pressure on the organism to develop the resistant or alternative pathway by not hampering the survival of the organism but the pathogenicity only. The main components of the Tat pathway are TatA/E, TatB and TatC [4]. Where TatB and TatC form the receptor complex and TatA forms the pore for translocation [3]. The six transmembrane (Tm) TatC is the main receptor for tat substrate recognition and extensive studies in *E.coli* have shown that mutations in the first and second periplasmic loops of TatC disrupt Tat export. We aim to find the crystal structure of *Mycobacterium* TatC but the biggest challenge is no or very low expression of MtTatC in E.coli in its native form or in codon optimized form.

The *E. coli* Tat pathway has been studied extensively but still there is lack of complete understanding of interaction between the components of the Tat machinery and how the variable size TatA pore is formed to efficiently transport Substrates of different size. The advancement of Electron Microscopy in studying protein complexes and determination of structures reaching near atomic resolution makes the *E.coli* Tat membrane complex a good target. Achieving homogenous Tat Membrane complex comprising of either TatBC or TatABC or TatABC + Substrate with good homogeneity should be a step forward in determination of atomic structure of Tat membrane complex.

5.2 Materials and methods

5.2.1 Materials

LB media, 2xYT, TB and LB agar used for all the bacterial culture were purchased from Hi-media. Antibiotics used for selection of transformants such as kanamycin, Ampicillin, and chloramphenicol were obtained from Sigma-Aldrich, USA. Chemicals used for the purification of proteins such as Trizma, Sodium chloride, Imidazole, β-mercaptoethanol, glycerol, DTT, Nickel sulphate, Bromophenol-blue (BPB), Acrylamide, N,N'-methylene bisacrylamide, Sodium dodecyl sulfate(SDS), Acetic acid, Methanol, TEMED (N,N,N',N'-Tetramethylethylenediamine), Ammonium persulfate (APS), were purchased from Sigma-Aldrich, USA. Molecular weight marker for SDS-PAGE was purchased from Bio-Rad Laboratories, USA. Ni-NTA used for affinity purification were purchased from Qiagen, Germany. Size exclusion columns were obtained from GE, USA. All detergents were procured from Anatrace USA. Protein samples were concentrated using Amicon® ultra centrifugal filters procured from Merck-Millipore, USA.

Commercial screens procured from Hampton research, USA and Qiagen, Germany were used for initial crystallization screening. Sodium cacodylate, lithium sulfate, PEG 4000, Ethylene glycol, glycerol, 2-methyl-2,4-pentanediol, propan-2-ol etc used in crystallization trials were obtained from Sigma-Aldrich, USA. Two-well sitting-drop plates were obtained from Hampton research, USA. 24 well plates and coverslips were obtained from Corning® (Sigma- Aldrich, USA) and Blue Star, India respectively. Other specialized chemicals and instruments used in the experiments are mentioned in the appropriate places. The constructs used are tabulated in Table5.1.

Table 5.1. MtTatC and EcTat constructs used in current study

S.No.	Construct	protein	Tag	Vector	Study	Remarks
1	N10M	MtTatC	N-term 10His	pET33b	MtTatC Structure determination	*E. Coli* codon bias MtTatC sequence
2	N10MA	MtTatC - AaTail	N-term 10His	pET33b	MtTatC Structure determination	*E. Coli* codon bias MtTatC sequence with Aquifex Tail
3	DM	GFP-MtTatC	C-term 10His	pET33b	MtTatC Structure determination	*E. Coli* codon bias MtTatC sequence +GFP
4	DMA	GFP- MtTatC - AaTail	C-term 10His	pET33b	MtTatC Structure determination	*E. Coli* codon bias MtTatC sequence +GFP with Aquifex Tail
5	BMTA	Bril - MtTatC - AaTail	N-term 6His	pET28a	MtTatC Structure determination	N-terminal Bril with *E. Coli* codon bias MtTatC sequence with Aquifex Tail to avoid free GFP
6	EcBC	EcTatC + EcTatB	C-term 6-His EcTatC	pACYC	*E.coli* Tat membrane complex structure determination	*E.coli* TatB and TatC receptor complex
7	DmsD	EcDmsD	No-tag	pACYC	*E.coli* Tat Tetrameric complex	Tat chaperon to be complexed with DmsA signal sequence
8	DN1	EcDmsASignal + GFP	C-term 6His	pET33b	*E.coli* Tat Tetrameric complex	Tat signal sequence ss5-50DmsA+GFP

5.2.2 Expression of Mt TatC constructs

Many variations of MtTatC had been tested for expression by shuffling various transmembrane domains and loops and alternately by forming chimeric fusion of MtTatC and AaTatC in various combinations (Fig 5.1). Finally, it was found out that the swapping of the C-terminal region with a stable tail i.e. of *Aquifex* stabilized MtTatC [5].

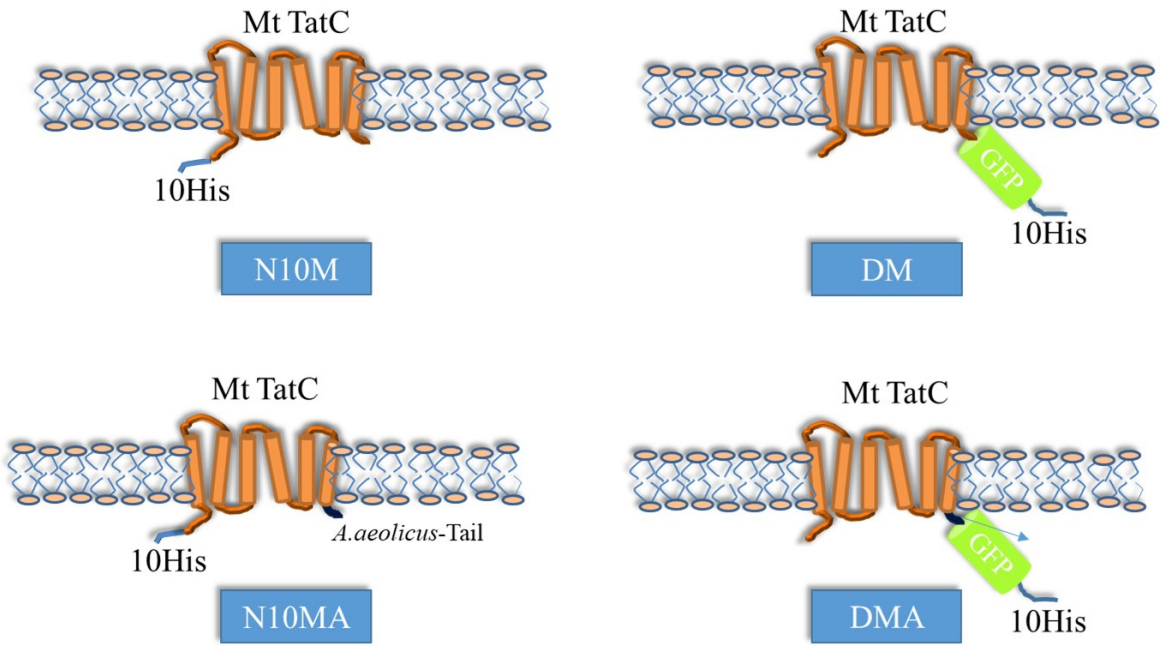


Fig 5.1 Constructs tested for expression of *Mt*TatC N10MT (n-terminal 10-histag + MtTatC); N10MTA (n-terminal 10histag + MtTatCwith AaTail); DM(C-terminal His tagged DrewGFP + MtTatC) and DMA (C-terminal His tagged DrewGFP + MtTatC with AaTail).

Expression analysis was performed four different constructs of MtTatC namely N10M (n-terminal 10histag+MtTatC); N10MA (n-terminal 10histag+ MtTatCwith AaTail); DM(C-terminal His tagged DrewGFP+MtTatC) and DMA(C-terminal His tagged DrewGFP + MtTatC with AaTail). These four constructs were tested for expression using different cell types (Bl21*, pLEMO, C41 and BL21G DTat), growth media (2xYT, LB and TB) and induction temperatures (16°C and 37°C).

5.2.3 Expression, Purification and detergent testing of DMA

DMA was expressed in BL21* and Bl21gold DTat cells in 1L cultures of 2xYT using 0.5mM IPTG induction after growth at 37 °C till 0.4 OD(600) and then grown overnight at 16°C. After that the cells were pelleted down at 4500 rpm (Bekman Coulter JLA8.1000 rotor) for 10min at 4 °C and all further steps were carried out in 4 °C. Then the pellets were re-

suspended in resuspension buffer (300 mM NaCl, 50 mM TRIS, 10% glycerol, 2 mM PMSF and PIC; pH 7.4) in 10% w/v and then cells were ruptured using a microfluidizer at 15000 psi for 15min. after that the lysate was pelleted down at 180000Xg for 1hour. The pellet obtained was resuspended for detergent exchange after resuspension using a hand homogenizer in TSG300 buffer (300 mM NaCl, 50 mM TRIS, 10% glycerol, 2 mM PMSF and PIC; pH 7.4 and 1% w/v detergent). Initially DDM was used but later on LDAO was used in this step for stabling DMA post Ni-NTA. After overnight exchange the solution was again spun down at 180000xg for 1hour during this step the lipids get pelleted down and the extracted membrane proteins stay in the supernatant. The supernatant was incubated for binding with Invitrogen Ni-NTA resin for 1hour. Then the mixture was passed through a gravity flow column and washed with 10CV of wash buffer (100 mM NaCl, 50 mM TRIS, 10% v/v glycerol, 30 mM Imidazole, 2 mM PMSF and PIC; pH 7.4 and 0.03% w/v DDM/ 0.1% w/v DM/ 0.1% w/v LDAO). After that the protein was eluted using 300 mM Imidazole containing elution buffer (100 mM NaCl, 50 mM TRIS, 10% v/v glycerol, 300 mM Imidazole; pH 7.4 and 0.03% w/v DDM/ 0.1% w/v DM / 0.1% w/v LDAO). The elution fractions were concentrated using a 30kDa cutoff filter and concentrated until 250µl. This was loaded into a size exclusion column (BioradSec650) equilibrated with TSG100 buffer (100 mM NaCl, 50 mM TRIS, 10% v/v glycerol; pH 7.4) containing 0.1% w/v LDAO. 250 µl of protein was injected and passed at 0.35ml/min.

5.2.4 Crystallization of DMA

The DMA was concentrated using a 30kDa concentrator about 5 mg/ml protein and set up for crystallization. It was screened against several commercially available crystallization screens including, MemFac Hampton Research Corp.), Index (Hampton Research Corp.), Nextal PACT (Qiagen), Nextal Protein Complex Suite (Qiagen) and JCSG plus (Molecular Dimensions) by vapour diffusion method in 96 well MRC2 sitting drop plates Hampton MRC-SD2 plates with multiple protein to screen ratios of 1:1 and 2:1 in 400ul sitting drops using the Mosquito robot(TTP Labtech) and incubated at 20°C.

5.2.5 Cloning and expression of BrilMtA (BMTA)

5.2.5.1 Amplification of MTA

MtTatC construct with AaTtail protein (MTA) with an N-terminal Bril was designed The MTA part of the protein was extracted from the DMA construct and inserted into a pET28a Bril-NDC3 vector such that there is an N-terminal his-tag followed by Bril and MtTatC with AaTail (BMTA) (Fig5.2).

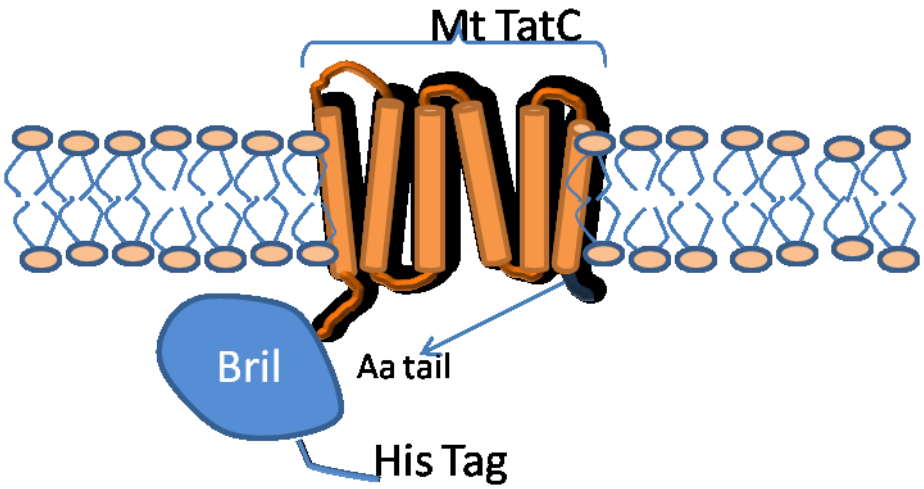


Fig5.2. BMTA construct, N-terminal 6-His tag followed by Bril and *Mt*TatC with *Aa*Tail C-Terminal Tail

MtTatC with Aquifex tail (MTA) was amplified from the DMA plasmid by gene specific primers. Forward primer NtBRIL_MTtatC_Fwd_BamHI 5'-CAGTC**GGATCC**ATGGGGTCTCTCGTAGACC-3' and reverse primer NtBRIL_MTtatC_Rev_HindIII 5'-CAGTC**AAGCTT**TTAAGCCTTCTGAATCTCCTTCTTTTTGC-3'. The restriction sites are highlighted. The PCR was carried out in a thermal cycler (AB-life technology, USA) under the following conditions: 1 cycle of pre-denaturation at 94 °C (2 min) followed by 28 cycles of denaturation at 94 °C (25 s), primer annealing at 64 °C (3 s) and extension at 72 °C (1 min), a final extension at 72 °C (10 min), and then cooled to 4 °C using Amplitaq Gold360 (Invitrogen). The PCR amplified products were analysed on 1% w/v agarose gel.

5.2.5.2 Restriction digestion of plasmid (pET28BrNDC3) vector and PCR product

The vector pET28BrNDC3 with Thrombin (Fig. 2.1) and the amplified PCR product were digested with appropriate restriction enzymes (BamHI /Hind III). Both reactions were carried out at 37 ºC with NEB CutSmart buffer for 4 hr. After restriction digestion the products were loaded on to a 1.0 % agarose gel. The digested vector and PCR product bands were carefully sliced from the gel and purified using Takara gel extraction kit. The double digested DNA products were quantified by a Nanodrop (Thermo-scientific).

5.2.5.3 Ligation

Approximately 1:10 ratio of vector to insert was incubated with T4 DNA Ligase from NEB with suitable buffer at 16 ºC overnight.

2.2.5.4 Transformation of cloned plasmid (pET28aBrNDC3+MTA) and colony PCR

Cloned plasmid was transformed in DH5α and plated on the LB agar plates containing 50mg/ml Kanymicin and incubated at 37 °C overnight. Single colonies were picked up and colony PCR was performed using T7 promoter and terminator primers to confirm the MTA gene insert in the cloned plasmid. Positive colonies were sequenced for confirmation of sequence and to check for unwanted mutations.

2.2.5.5 Confirmation of BMTA expression by western blotting

DMA was expressed in BL21* and BL21goldDTat cells in 1L cultures of 2xYT using 0.5 mM IPTG induction after growth at 37 °C till 0.4 OD(600) and then grown overnight at 16°C. Then the pellets were re-suspended in resuspension buffer TSG300 (300 mM NaCl, 50 mM TRIS, 10% glycerol, 2 mM PMSF and PIC; pH 7.4) in 10%w/v and then cells were ruptured using a microfluidizer at 15000Psi for 15min. The lysate was pelleted down at 180000Xg for 1hour. Pellet obtained was resuspended for detergent exchange using a hand homogenizer in TSG100 buffer (100 mM NaCl, 50 mM TRIS, 10% glycerol, 2 mM PMSF and PIC; pH 7.4 and 1% LDAO). After overnight exchange the solution was again spun down at 180000xg for 1hour. The supernatant was kept for binding with Invitrogen Ni-NTA resin for 1hour. Then the mixture was passed through a gravity flow column and washed with 10CV of

Wash buffer (100mM NaCl, 50mM TRIS, 10% glycerol, 30mM Imidazole, 2mM PMSF and PIC; pH7.4 and 0.1%LDAO). After that the protein was eluted using 300 mM Imidazole containing elution buffer (100 mM NaCl, 50 mM TRIS, 10% glycerol, 300 mM Imidazole; pH 7.4 and 0.1%LDAO).

Western blot was performed to confirm the expression of the protein by electro-blotting on to a PVDF membrane. PVDF membrane was blocked with 5% bovine skim milk in 1X PBS for overnight followed by 3 washes of PBST (PBS+ Tween 20) buffer in the 5 minute interval. The PVDF membrane was incubated with primary anti-penta-his (Quiagen) mouse antibody at 1:3000 dilutions for 2 hours. Three washes with PBST were given to remove any unbound antibody. Further the membrane was incubated with secondary anti-mouse IgG (Fc-region specific) hrp conjugated antibody at 1:2000 dilutions for the 1 hour followed by three PBST washes. For visualizing the bands, the membrane was incubated with substrate (DAB).

5.2.6 Sequence analysis and Modelling of MtTatC

M. tuberculosis (strain ATCC 25618 / H37Rv) TatC sequence P9WG97 used for sequence analysis and structure prediction was downloaded from Uniprot. Other TatC sequences were also downloaded from Uniprot. Multiple sequence analysis was done in ClustalOmega using ClustalW. MtTatC was modelled in MODELLER ver.9.17 [6] using *A. aelocious* TatC (4HTS) [7] as template. The model with the lowest discrete optimized protein energy (DOEP) scores was selected. The generated model was refined in the GalaxyWEB web server (Computational Biology Lab, Department of Chemistry, Seoul National University) [8].

5.2.7 Expression of E.coli Tat membrane complex

EcTatBC assembly in pACYC was over expressed in BL21gold DTat cells in 1L cultures of 2xYT using 0.5 mM IPTG induction after growth at 37 °C till 0.4 OD(600) and then grown overnight at 16 °C. Purification was similar to DMA purification as in section 5.2.3.The elution fractions were concentrated using a 100 kDa cutoff filter and concentrated. 250 µL of protein was injected into a GE superdexS200 increase10/300 size exclusion column

equilibrated with TSG100 buffer (100 mM NaCl, 50 mM TRIS, 10% glycerol, pH7.4 and 0.03%DDM). The EcTatBC complex obtained was used for negative staining. DmsD without 6-His tag (pACYC) and DN1 (DmsD signal sequence+GFP+6-His pET33b+) were also co-expressed in Bl21gold DTat cells. They were grown in 1L culture of LB using 0.5 mM IPTG induction after growth at 37 °C till 0.4 OD (600) and then grown for 4 hours at 37 °C. After induction, cells were harvested by centrifugation at 4500 rpm for 10minutes. Harvested bacterial cells were resuspended in lysis buffer (50 mM Tris pH 8, 100 mM NaCl, 2 mM β-ME), followed by sonication for 5 minutes with 5 sec on/off pulse and 45% amplitude. After the cell disruption, the lysate was centrifuged for 30 minutes at 12000 rpm. The supernatant was collected and subjected to Ni-NTA affinity chromatography as in section 5.3.2 in the absence of any detergent. The purified EcTatBC complex from Size exclusion chromatography was mixed in in 1:1 ratio for complex formation and injected into a GE superdexS200 increase10/300 size exclusion column equilibrated with a buffer containing 100 mM NaCl, 50 mM TRIS, 10% v/v glycerol, pH7.4 and 0.03% w/v DDM. The tetrameric protein complex peak was used for negative staining.

5.2.8 Negative staining of *E.coli* Tat Membrane complex

The negative staining of the EcTatBC complex and EcTatBC+DmsD+DN1 complex was performed by loading 5 µl of protein sample onto a Formavar-coated 300-mesh copper grids (Electron Microscopy Sciences) and allowed to adsorb for 2 min. Then the sample was blotted off using a filter paper. 5 µl of negative stain solution (2% uranyl acetate) was loaded onto the grid immediately and incubated for 2 min, the rest of the stain was blotted off with a filter paper and the grid was air-dried [9]. The grids were visualized in FEI Tecnai TF20 in a high resolution Transmission Electron Microscope (TEM) equipped with a FEG source at an accelerating voltage of 200 KV. The images were recorded with a 4K x 4K Eagle CCD Camera.

5.3 Results

5.3.1 Expression of Mt TatC constructs

Many variations of MtTatC had were tested for expression forming chimeric fusion of MtTatC, GFP and AaTatC cytoplasmic Tail in various combinations. Finally it was found out that the swapping of the C-terminal tail with a stable tail i.e. of *Aquifex* TatC C-terminal cytoplasmic tail stabilized MtTatC. These four constructs were tested for expression using different cell types, media and induction temperatures as shown in Table5.2.

Table5.2. Expression studies of MtTaC constructs: - indicates no expression, +indicates low levels of expression and ++ indicates good expression

Expression cell	Temperature	37°C			16°C		
	Media	2xYT	LB	TB	2xYT	LB	TB
N10M	BL21*	-	-	-	-	-	-
	PLEMO	-	-	-	-	-	-
	C41	-	-	-	-	-	-
	BL21G DTat	-	-	-	-	-	-
N10MA	BL21*	-	-	-	-	-	-
	PLEMO	-	-	-	-	-	-
	C41	-	-	-	-	-	-
	BL21G DTat	-	-	-	-	-	-
DM	BL21*	-	-	-	-	-	-
	PLEMO	-	-	-	-	-	-
	C41	-	-	-	-	-	-
	BL21G DTat	-	-	-	-	-	-
DMA	BL21*	+	+	+	++	+	+
	PLEMO	-	-	-	-	-	-
	C41	-	-	-	-	-	-
	BL21G DTat	+	+	+	+	+	+

Expression was obtained only in case of DMA. The expressed DMA was loaded beside native GFP on 12% w/v SDS-PAGE to confirm expression of full length DMA. This was done since no band was observed in the gel when the DMA sample was heated and loaded, this may be protein aggregation and did not move into resolving gel. It was observed that there was some free GFP along with DMA. The protein was confirmed by western blot (Fig 5.3).

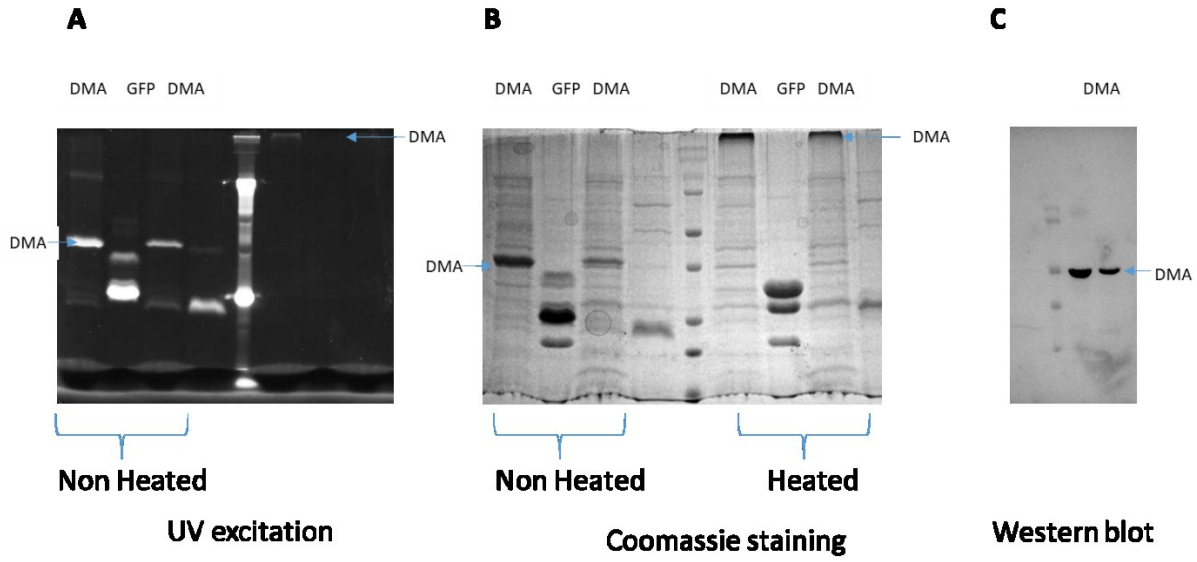


Fig.5.3. Confirmation of DMA expression and comparison with free GFP in heated and non-heated condition SDS-PAGE (A) Visualization under UV (B) Comassie Stained SDS-PAGE (C) Western blot using Anti-His Antibody

5.3.2 Purification of DMA and detergent testing

Further expression and purification was carried out in Bl21* which gave the maximum yield. Initially DDM was used for solubilization of membrane and protein extraction, but it was observed that the purified protein after Ni-NTA would precipitate within a few hours of purification. Later DM and LDAO were used for extraction and purification (Fig5.4). It was observed that the protein stayed stable in solution when extraction was done with LDAO and Ni-NTA purification was carried out with 0.1% LDAO.

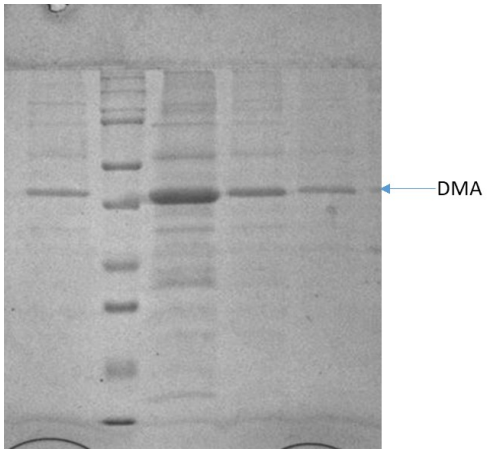


Fig5.4. Ni-NTA Purification of DMA in LDAO

The Ni-NTA elution fractions were concentrated using a 30KDa cutoff filter. 250µl protein was injected into a Bio-rad Sec650 size exclusion column equilibrated with a buffer containing 100mM NaCl, 50mM TRIS, 10% v/v glycerol, pH7.4 and 0.1% w/v LDAO. The eluted protein fractions were run on 12% w/v SDS-PAGE (Fig5.5).

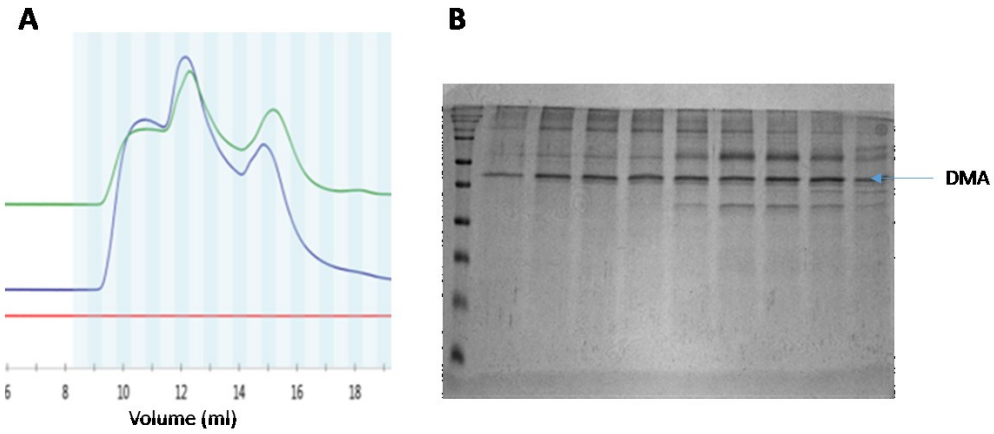


Fig5.5. (A) Size exclusion chromatogram of DMA in SEC650 (B) Different fractions obtained after size exclusion in 12% SDS-PAGE

5.3.3 Crystallization of DMA

The DMA peak was obtained at 12.5ml Fig11(b) and the fractions were concentrated using a 30KDa concentrator till about 5mg/ml and set up for crystallization. Several commercially available crystallization screens including, MemFac Hampton Research Corp.), Index (Hampton Research Corp.), Nextal PACT (Qiagen), Nextal Protein Complex Suite (Qiagen) and JCSG plus (Molecular Dimensions) by vapour diffusion method in 400nl sitting drops using the Mosquito robot (TTP Labtech). No Crystal hits were obtained.

5.3.4 Cloning of BrilMtA(BMTA)

The inbuilt problem of GFP attached MtTatC (DMA) is the presence of free GFP by cleavage of GFP from the MtTatC portion. This creates problem in proper purification and also the presence of GFP in crystal setup may cause the formation of just GFP crystals; thus not providing any details of MtTatC structure. Also the absence of any crystal hit urged us to produce another MtTatC with AaTtail protein (MTA) with another tag namely Bril. The MTA part of the protein was extracted from the DMA construct and inserted into a pET28a Bril-NDC3 a Bril-NDC vector such that there is an N-terminal his-tag followed by Bril and MtTatC with AaTail (BMTA). The positive colonies were screened by colony PCR using T7 Promoter and T7Terminator Primers. Out of 20 colonies screened on Kanamycin plates 9 were positive for the insert. Sequencing was done for confirmation of sequence. (Fig5.6).

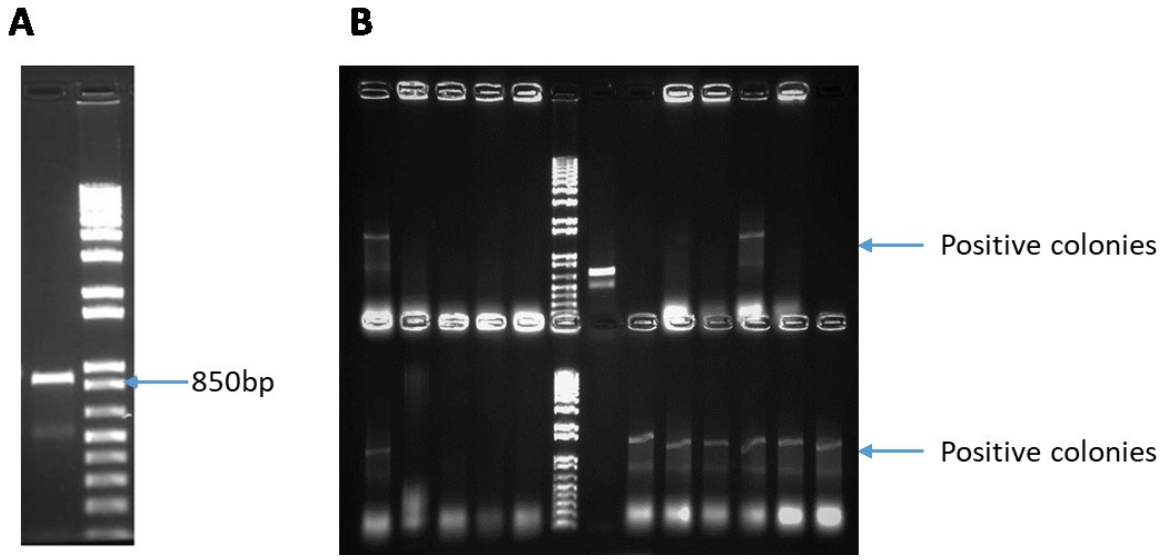


Fig5.6. Cloning of MTA from DMA A) amplified gene fragment B) colony PCR for positive colonies

5.3.5 Expression of BrilMtA(BMTA)

Protein expression was confirmed by Western blot using quiagen Penta Anti-His antibodies (Fig5.7). The expression of BMTA was obtained in both Bl21G and Bl21DTat sells at 16°C overnight induction using 0.5mM IPTG. Purification process was same as that of DMA and LDAO was used to extract protein from the membrane. The expression of BMTA caused excessive cell death after induction, leading to as little as 1gram cell pellet after overnight induction at 16°C or even 37°C for 4hours. This made it impossible to get sufficient amount of protein for further studies.

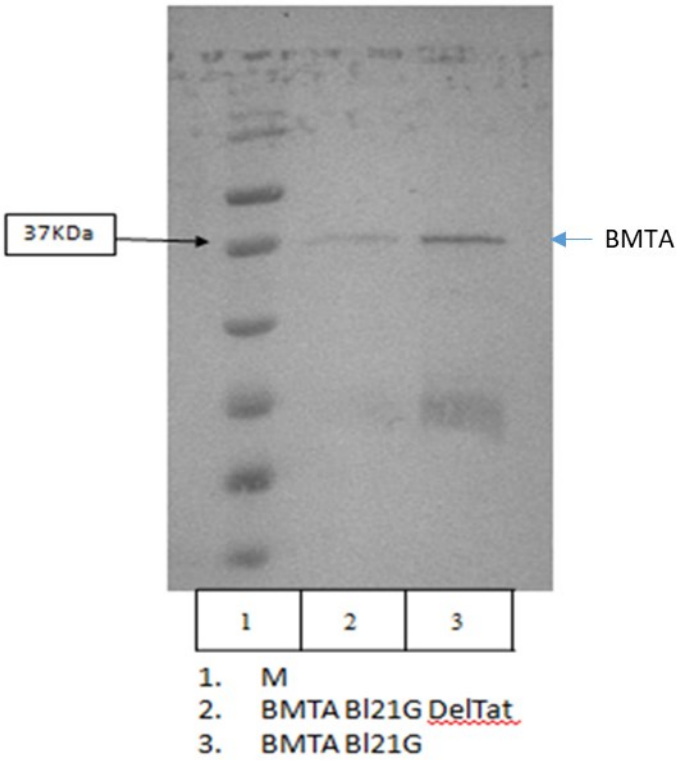
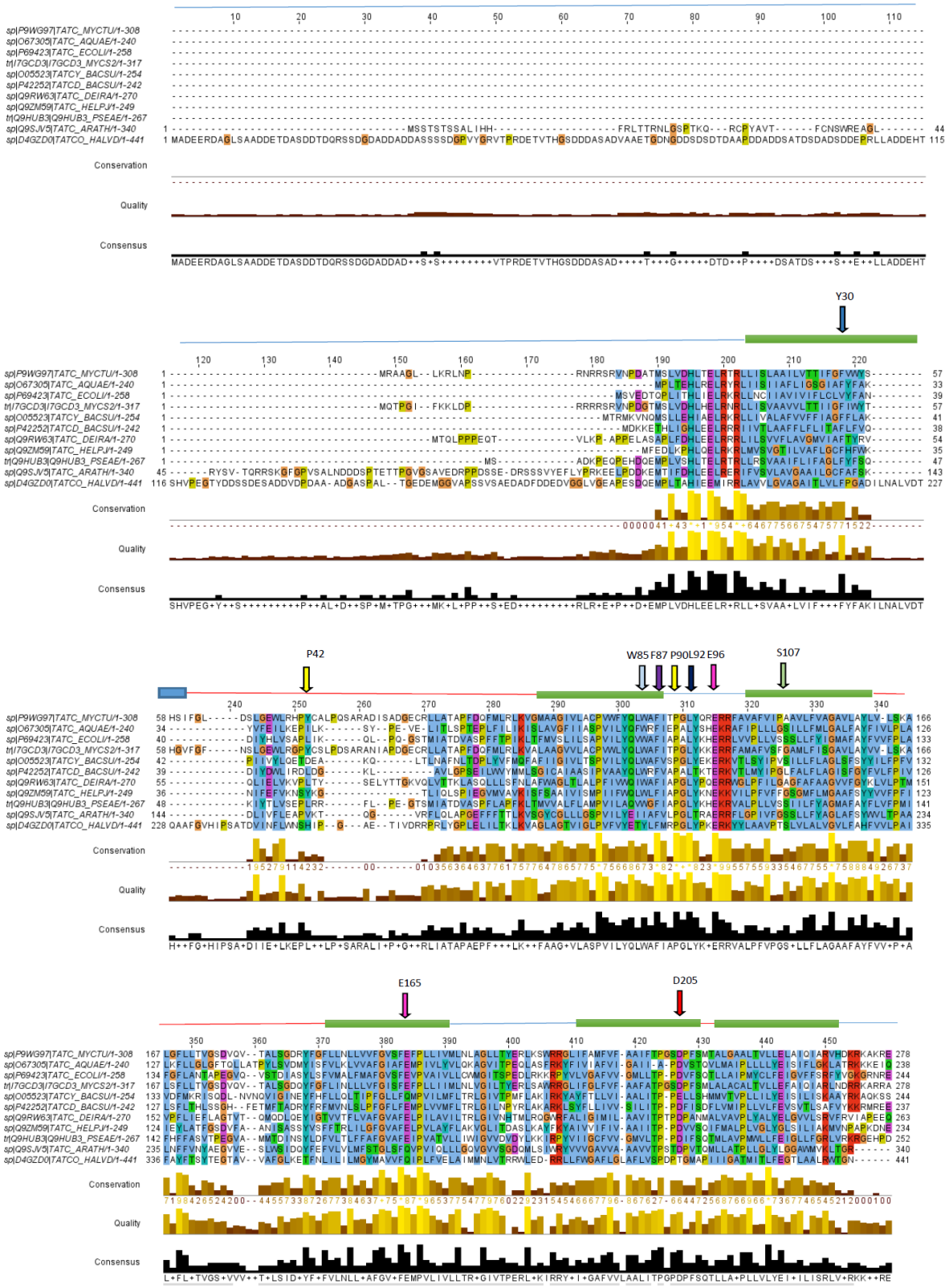


Fig5.7. Western blot confirmation of BMTA using Anti-His antibodies in DTat and Bl21Gold cells

5.3.6 Sequence analysis and Modelling of MtTatC

M. tuberculosis (strain ATCC 25618 / H37Rv) TatC sequence P9WG97 was downloaded from uniprot and aligned to other TatC using ClustalW. The predicted transmembrane(TM) regions were marked for MtTatC based on prediction from Uniprot. Important Tat residues are marked on the multiple sequence alignment (MSA) based on Aquifex TatC numbering (Fig 5.8) and tabulated (Table 5.3) [7][10].


Y30
P42
W85 F87 P90L92 E96
S107
E165
D205
Conservation
Quality
Consensus

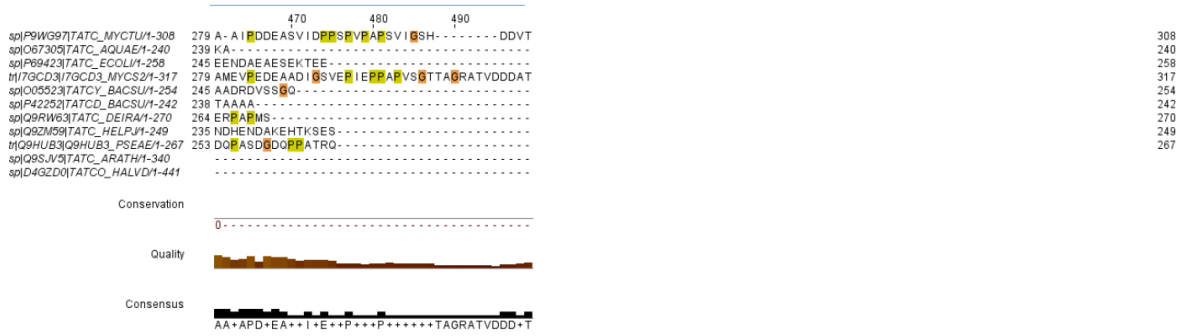


Fig5.8. Multiple Sequence alignment of the MtTatC with other TatC. The important *aquifex* residues were marked with arrows having *Aquifex* numbering.

Table5.3. MtTatC residue conservation numbered based on *Aquifex* Residue

Sr. No.	Residue (Aquifex numbering)	MtTatC
1	Glu165	+
2	Ser107	(Ala)
3	Tyr85	+
4	Glu96	+
5	Pro42	+
6	Phe87	+
7	Pro90	+
8	Leu92	+
9	D205	+
10	Y30	+

A. aeolicus Glu165, which is exposed at the center of the concave face and thus places an ionizable group in the hydrophobic interior of the bilayer [7][11], was conserved. This forms hydrogen-bond networks with Ser107 and Tyr85 [11]. Ser107 is replaced with an Ala just like in *Bacillus subtilis* TatCd which also has a hydrophobic Lys preceded by a Proline residue and Tyr85 is conserved. This might be due to the difference in membrane properties. Glu96 is another critical residue for Tat transport was conserved. The residues involved in signal binding in Aquifex Pro42, Phe87, Pro90 and Leu92 residues were conserved [7][12]. Residues

involved in TatC dimerization and interaction with other Tat pathway membrane components Asp205 was conserved and Tyr30 residue was present almost at the same location [13].

MtTatC was modelled in MODELLER ver.9.17 [6] using *A. aelocious* TatC (4HTS) as template. Due to the unavailability of template for MtTatC residues 1-20 and 282-308 the final model was generated for MtTatC 21-281. The model was refined in Galaxy web server. The final model had Ramachandran statistics[14] of Residues in most favoured regions [A,B,L] 221 94.8%, Residues in additional allowed regions [a,b,l,p] 10 4.3%, Residues in generously allowed regions [~a,~b,~l,~p] 0 0.0% and residues in disallowed regions 2 0.9% (Fig 5.9.). The RMSD of the MtTatC structure with Aa1HTS was 1.3 [15].

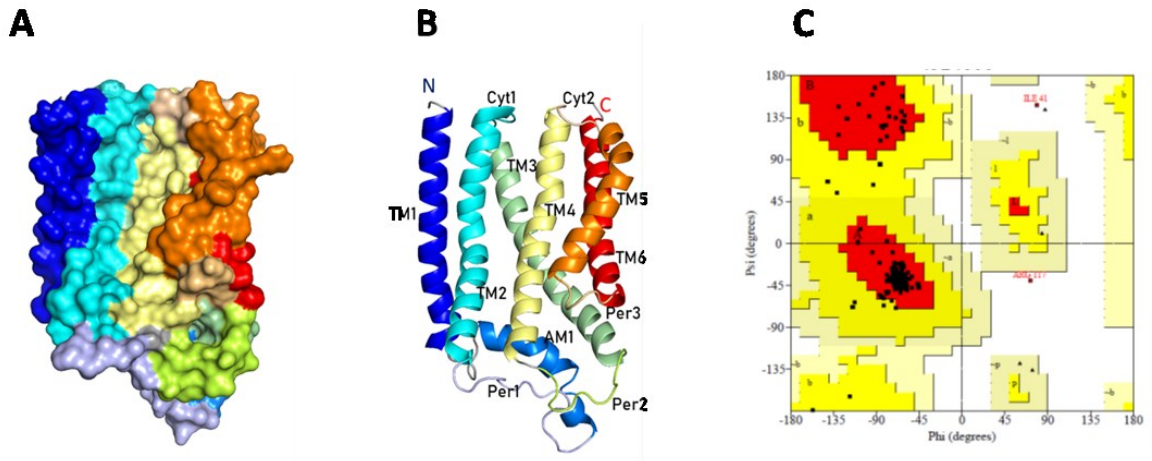


Fig.5.9. MtTatC model generated by homology modelling based on 4HTS A) Surface View B) Cartoon Representation C) Ramachandran plot

The N-terminal transmembrane starts after 30 amino acids from the cytoplasmic side with TM1 (His30-Ser59) going in perpendicular to the membrane, and continuing as an amphipathic helix -AM1 (Leu66-Asp88) arching under TM2 (Asp100-Tyr134) and constituting a large part of the periplasmic face of TatC (Figures 1A–1C). The first periplasmic loop (Per1) continues to TM2, which at the periplasmic end is angled relative to TM1 till the conserved Pro119 (*Ec*85/*Aa*78) residue generates a kink and the rest of the cytoplasmic part is parallel to TM1. The short cytoplasmic loop (Cyt1) connects TM2 to TM3. TM3 (Arg138-Thr172) is angled steeply and contacts to TM2, TM4, and TM6 from behind. Periplasmic loop

(Per2) connecting TM3 and TM4 is below TM5 and TM6. TM4 (Gly184-Ala214) is parallel to TM2, is conserved proline kink Pro205 (*Ec*172/*Aa*167) is conserved. The highly conserved glycine199 (*Ec*166/*Aa*161) of TM4 and Gly155 (*Ec*121/*Aa*114) in TM3 forms a tight interface. TM4 is connected to TM5 via a short loop (Cyt2), TM5 (Try219-Phy240) moves out from the core of the protein with a steep angle and then contacts to TM4 by kinking sharply. Periplasmic loop (Per3) post TM5 has a highly conserved Pro242 turn that lies in the hydrophobic core of the bilayer. TM6 (Pro246 – Asp271) has is slightly angled and contacts to TM3, TM4, and TM5 on the back. The C terminus of the protein extends into the cytoplasm till Tyr308

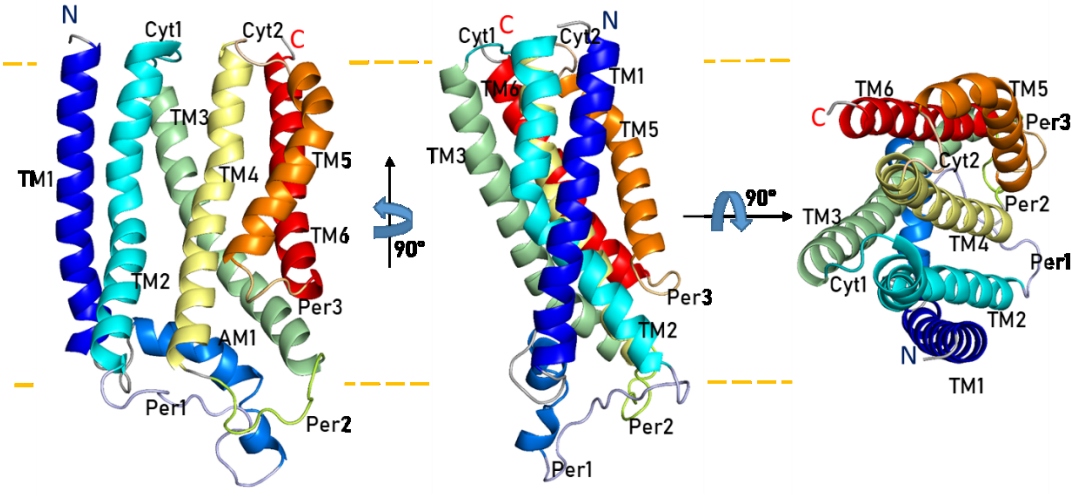


Fig.5.10. Cartoon diagram of MtTatC viewed in the plane of the membrane color-ramped from the N to the C terminus (blue to red) from the front and rotated 90°. TM1 (His30-Ser59), AM1 (Leu66-Asp88), TM2 (Asp100-Tyr134), TM3 (Arg138-Thr172), TM4 (Gly184-Ala214), TM5 (Try219-Phy240) and TM6 (Pro246 – Asp271)

Electrostatic surface potential was mapped for MtTatC (Fig 5.11) and compared to AaTatC. This analysis a number of charged patches that are buried in the membrane region of MtTatC (Fig5.11A). The charge distribution on the frontal face of MtTatC and AaTatC was similar but there is difference in charge distibution in the periplasmic cap region of the protein. There is a large positive patch on MtTatC (Fig5.11.C) whereas AaTatC has a negative patch (Fig5.11.D). The cytoplasmic cap had similar charge distribution for both the proteins.

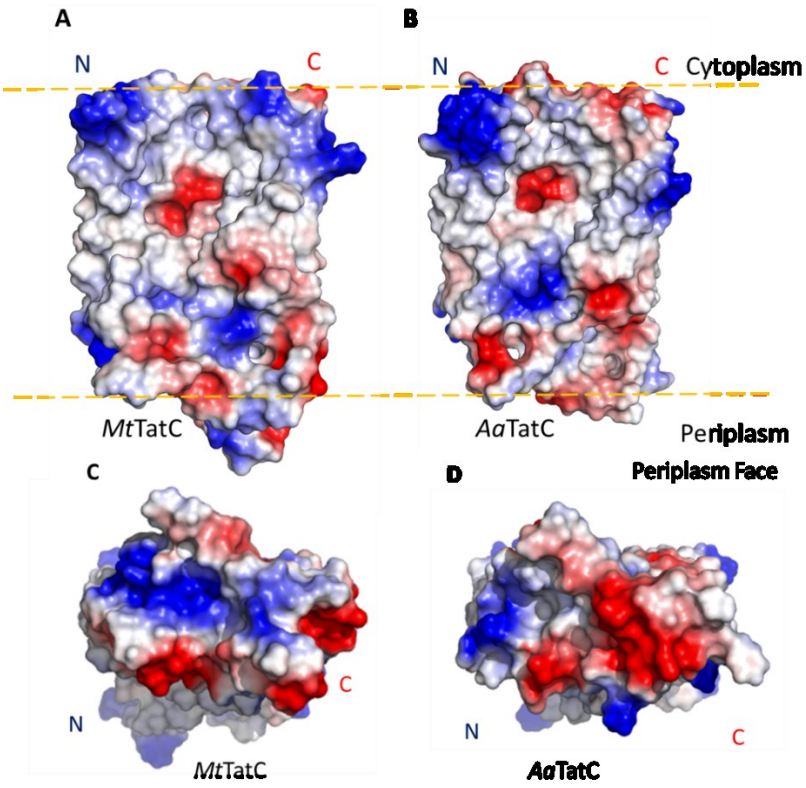


Fig.5.11. Electrostatic surface potential of MtTatC from negative (red) to positive (blue) for MtTatC and AaTatC A)MtTatC from front B) AaTatC from front C) MtTatC Periplasmic face and D) AaTatC Periplasmic face

5.3.7 Expression of E.coli Tat membrane complex

The wide variation in the sizes of the Tat Substrates causes TatA forms a pore according to the size of the substrate. The mechanism of the formation of this dynamic pore and protein translocation is not known. For structural insigts into the dynamics of this membrane complex the Cro-EM approach was also considered and the Purification of *E.Coli* Tat membrane complex was undertaken.

EcTatBC assembly in pACYC was over expressed in Bl21gold DTat cells. DmsD without 6-His tag (pACYC) and DN1 (DmsD signal sequence+GFP+6-His pET33b+) were also co-expressed in Bl21gold DTat cells. Ni-NTA purification was performed for both the complexes (Fig5.12).

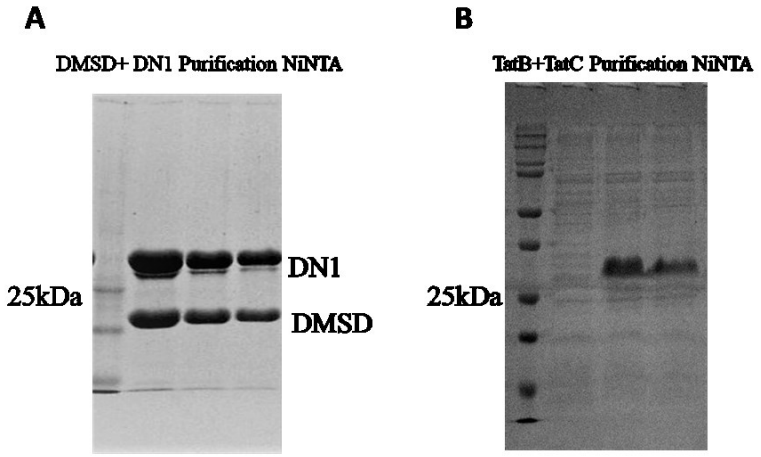


Fig5.12 Ni-NTA purification of (A) co-expressed DmsD + DN1 (DmsD signal sequence + GFP + 6-His pET33b+) and (B) EcTatBC assembly in pACYC

After Ni-NTA purification (Fig5.12) the elution fractions of EcBC were concentrated using a 100KDa cutoff filter. 250µL of protein was injected into a GE superdexS200 increase10/300 size exclusion column equilibrated with TSG100 buffer (100mM NaCl, 50mM TRIS, 10% v/v glycerol, pH7.4 and 0.03% w/v DDM). The EcTatBC complex was obtained at 10.8ml corresponding to ~347KDa (Fig15.13).

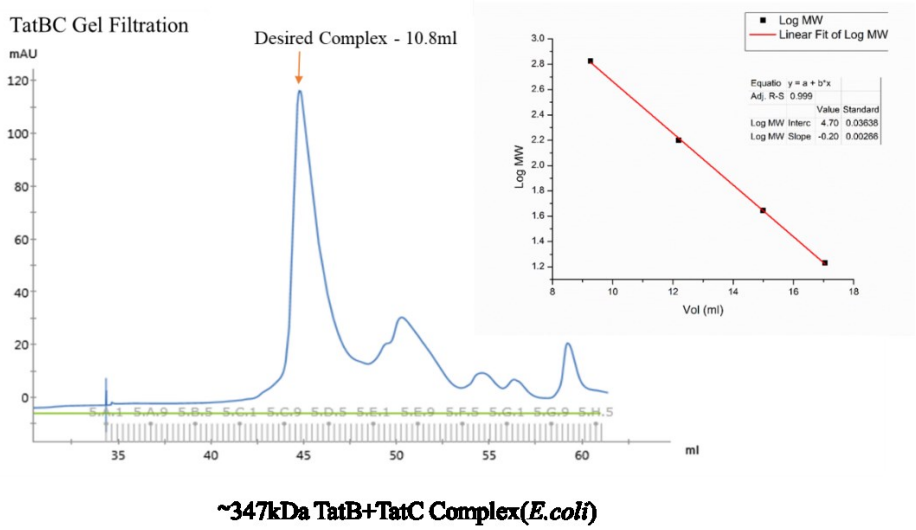


Fig5.13 EcTatBC size exclusion chromatography in Superdex S200 increase (10/300) column. EcTatBC complex was obtained at 10.8ml corresponding to ~347KDa.

Ni-NTA affinity chromatography purified DmsD + DN1 complex (Fig5.12) was concentrated using a 30KDa cutoff filter and buffer exchanged to 100mM NaCl, 50mM TRIS, 10% v/v glycerol and pH7.4. The purified *Ec*TatBC complex from Size exclusion chromatography was mixed in in 1:1 ratio with DmsD + DN1 complex for quaternary membrane pore complex formation. This was injected into a GE superdexS200 increase10/300 size exclusion column equilibrated with a buffer containing 100mM NaCl, 50mM TRIS, 10% v/v glycerol, pH7.4 and 0.03%DDM). The tetrameric protein complex peak at 13.05ml corresponded to ~124KDa complex (Fig.15.14). These EcBC and EcBC+DmsD+DN1 complex were used for negative staining.

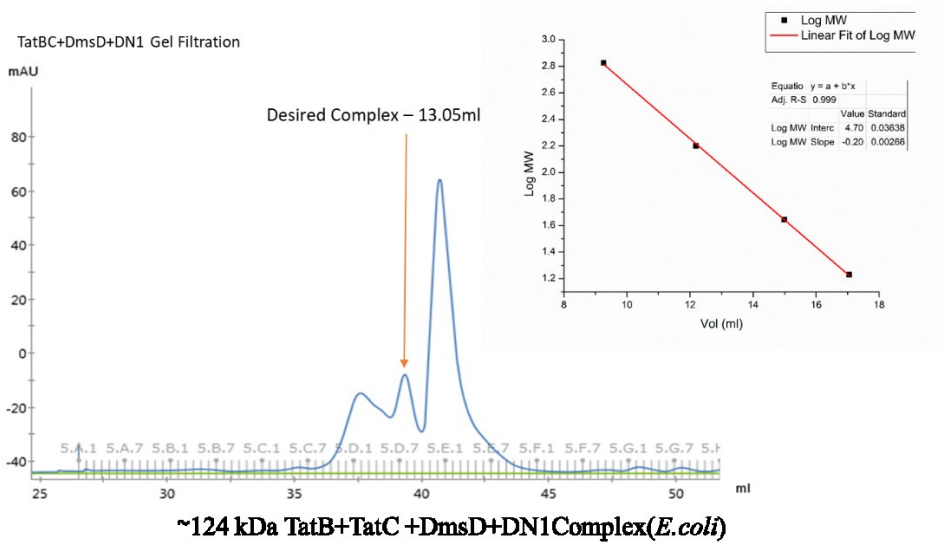


Fig5.14. EcTatBC +DmsD +DN1 size exclusion chromatography in Superdex S200 increase (10/300) column. EcTatBC +DmsD +DN1 complex was obtained at 13.05ml corresponding to ~124KDa.

5.3.8 Negative staining of *E.coli* Tat Membrane complex

The negative staining of the EcTatBC complex and EcTatBC+DmsD+DN1 complex was performed by uranyl acetate staining in an FEI Tecnai TF20 in a high resolution Transmission Electron Microscope (TEM) equipped with a FEG source. Before grid preparation the glycerol concentration of buffer was made to 0% by buffer exchange with TS100 buffer containing 100 mM NaCl, 50 mM TRIS, pH 7.4 and 0.03% w/v DDM using a

100 kDa cutoff membrane filter, this was to be as close to conditions for Cryo-EM as glycerol is a cryo-protectant and hinders proper flash freezing. Protein samples were loaded onto a Formavar-coated 300-mesh copper grids (Electron Microscopy Sciences) and stained using 2% uranyl acetate solution. The grids were visualized in at an accelerating voltage of 200 Kev. The images were recorded at 50,000x magnification at three different positions. EcTatBC complex (Fig 5.15) showed heterogeneous particles and patches of aggregates. The aggregation was probably due to the delay in making of grids due to transport delay.

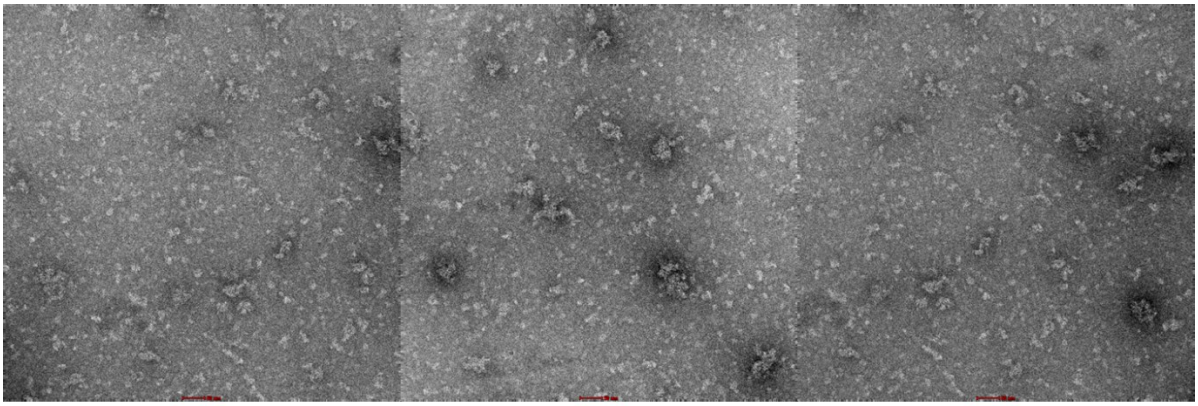


Fig5.15. Negative Stain images of EcTatBC at 50,000x magnification

Negative stain of EcTatBC+DmsD+DN1 complex (Fig 5.16) also showed heterogeneity and aggregates but had better contrast than EcTatBC. Both these complexes can be further taken ahead for structural investigation using Cryo-EM.

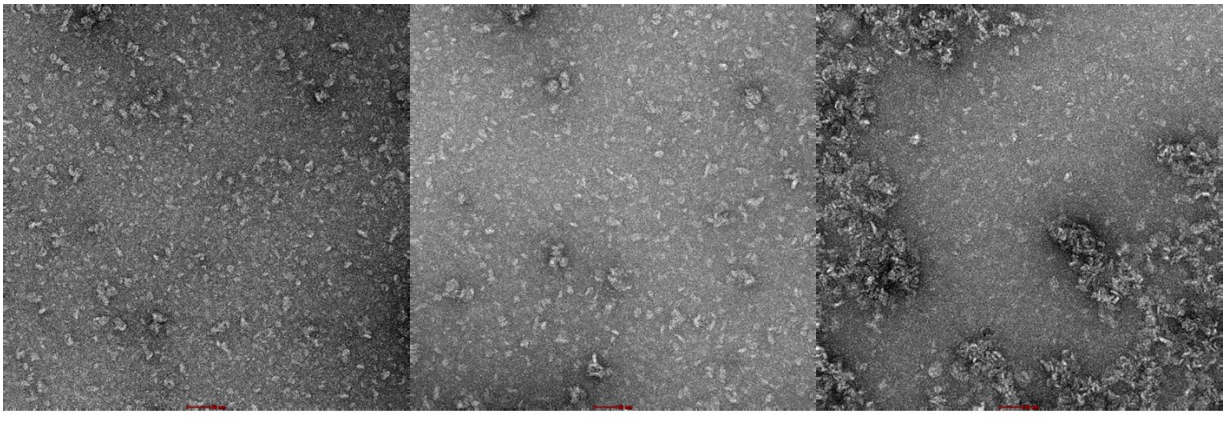


Fig5.16. Negative Stain images of EcTatBC +DmsD+DN1 complex at 50,000x magnification

To investigate the effect of glycerol removal on EcTatBC complex and its stability, Ni-NTA purified complex extracted using DDM was loaded onto SuperdexS200 5/150 size exclusion chromatography column in the presence of 10% glycerol and absence of glycerol (Fig 5.17). The protein elution pattern for the EcTatBC peak in both conditions was identical thus making it ideal for cryo-EM studies.

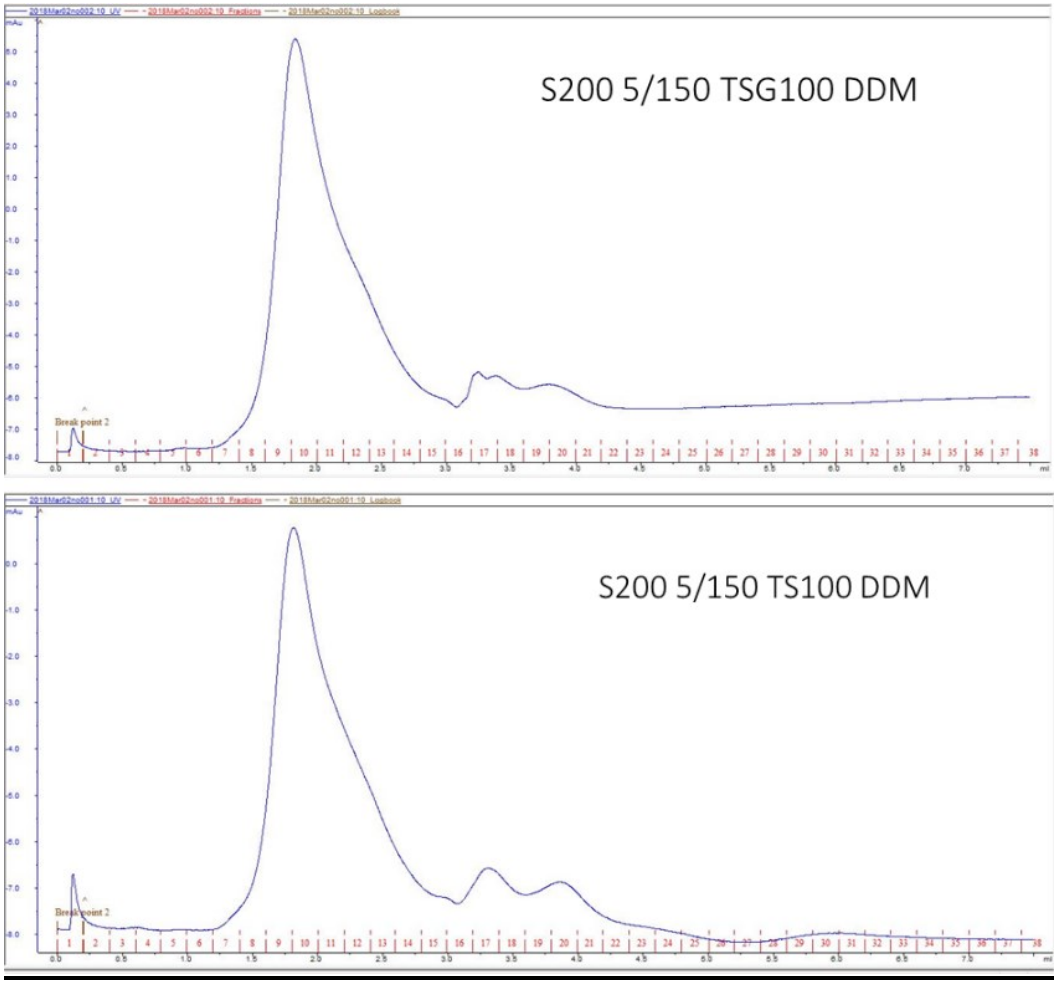


Fig5.17.EcTatBC complex SuperdexS200 5/150 size exclusion chromatography column in (A) presence of 10% glycerol and (B) absence of glycerol

5.4 Discussions

The intrinsic problem of heterologous expression of MtTatC in *E.coli* is a major bottle neck in structure determination. Swapping of the cytoplasmic Tail to a more stable Aquifex tail improved the expression but only in the case where it was expressed with a GFP which has

very stable structure. The presence of free GFP made us design a new BrilMTA construct but it cause intrinsic bacterial cell death on expression. The difference between the plasma membrane of *Mycobacterium* and *E.coli* is a major factor for the low and unstable expression of TatC which is a 6-Transmembrane protein. It is logical to shift to the *M.smegmatis* based expression system which uses the $mc^2$4517 strain and coupling it with the pYUB1602 Lac operon based plasmid. The structure of *E.coli* Tat membrane complex and receptor bound complex is a good target for Single particle cryo-EM structure determination and will provide us with great insight into the dynamics and functioning of the Tat Translocation complex having the flexible pore size. Negative Staining TEM of EcTatBC complex and EcTatBC+DmsD+DN1 complex shows the sustainability of the system to be studied further in Cryo-EM. It is necessary to decrease the time gap between Protein purification and grid preparation. Also studies emphasize the need the extraction of homogeneous Tat membrane complex and more stable complex. The use of GDN a synthetic alternate to Digitonin which can extract homogeneous EcTatABC complex right from Membrane extraction step should be considered. *E. coli* TatB mutation F13Y forms stable interaction between TatC and TatB which should also be considered. The importance of Phosphatidylethanolamine lipids in Tat activity makes it important to screen complexes with those lipids. The modelled MtTatC structure based on AaTatC structure 1HTS showed the overall structural conservation of TatC fold with the typical transmembrane proline kinks. There is conservation of all important functional residues. The key difference lies in the presence of longer N-terminal and C-terminal cytoplasmic domains and in the periplasmic cap. The MtTatC ampipathc helix is longer than that of AaTatC and surface electrostatic analysis show the presence of positively charged surface on the MtTaC when compared to AaTatC. The determination of MtTatC structure and the EcTat membrane complex will be a breakthrough for development of peptide inhibitors and also for fine tuning of the Tat pathway for therapeutic protein expression etc.

5.5 References

[1] J. E. Posey, T. M. Shinnick, and F. D. Quinn, “Characterization of the twin-arginine translocase secretion system of Mycobacterium smegmatis,” *J. Bacteriol.*, vol. 188, no. 4, pp. 1332–1340, 2006.

[2] J. A. McDonough, K. E. Hacker, A. R. Flores, M. S. Pavelka, and M. Braunstein, “The

twin-arginine translocation pathway of Mycobacterium smegmatis is functional and required for the export of mycobacterial β-lactamases," *J. Bacteriol.*, vol. 187, no. 22, pp. 7667–7679, 2005.

[3] B. C. Berks, "The Twin-Arginine Protein Translocation Pathway," *Annu. Rev. Biochem.*, vol. 84, no. 1, pp. 843–864, 2014.

[4] V. Fincher, C. Dabney-Smith, and K. Cline, "Functional assembly of thylakoid ΔpH-dependent/Tat protein transport pathway components in vitro," *Eur. J. Biochem.*, vol. 270, no. 24, pp. 4930–4941, 2003.

[5] S. S. Marshall *et al.*, "A Link between Integral Membrane Protein Expression and Simulated Integration Efficiency," *Cell Rep.*, vol. 16, no. 8, pp. 2169–2177, 2016.

[6] B. Webb and A. Sali, "Comparative Protein Structure Modeling Using MODELLER," *Curr. Protoc. Bioinforma.*, vol. 47, no. 1, pp. 5.6.1-5.6.32, Sep. 2014.

[7] S. Ramasamy, R. Abrol, C. J. M. Suloway, and W. M. Clemons, "The glove-like structure of the conserved membrane protein TatC provides insight into signal sequence recognition in twin-arginine translocation.," *Structure*, vol. 21, no. 5, pp. 777–88, May 2013.

[8] J. Ko, H. Park, L. Heo, and C. Seok, "GalaxyWEB server for protein structure prediction and refinement," *Nucleic Acids Res.*, vol. 40, no. W1, pp. 294–297, 2012.

[9] J. R. Harris, "Negative Staining and Cryoelectron Microscopy," *Micros. Today*, vol. 5, no. 1, pp. 18–18, Jan. 1997.

[10] F. Sargent, B. C. Berks, and T. Palmer, "The Tat Protein Export Pathway," *EcoSal Plus*, vol. 4, no. 1, Sep. 2010.

[11] S. E. Rollauer *et al.*, "Structure of the TatC core of the twin-arginine protein transport system.," *Nature*, vol. 492, no. 7428, pp. 210–4, Dec. 2012.

[12] S. E. Rollauer *et al.*, "Structure of the TatC core of the twin-arginine protein transport system," *Nature*, vol. 492, no. 7428, pp. 210–214, Dec. 2012.

[13] E. Holzapfel *et al.*, "The entire N-terminal half of TatC is involved in twin-arginine

precursor binding," *Biochemistry*, vol. 46, no. 10, pp. 2892–2898, 2007.

[14] G. J. Kleywegt and T. A. Jones, "Phi/Psi-chology: Ramachandran revisited," *Structure*, vol. 4, no. 12, pp. 1395–1400, 1996.

[15] L. Holm and L. M. Laakso, "Dali server update," vol. 44, no. April, pp. 351–355, 2016.

Chapter VI – Summary and conclusion

The Twin arginine translocase components have been studied in this work. This pathway transports folded proteins across the membrane and its study is of significant interest for therapeutic as well as industrial applications. The main objectives of our study were understanding the interaction between the Tat signal sequence and REMP chaperon, determining a minimal Tat signal peptide necessary for transport, development of a Novel High-throughput Twin Arginine Translocase Assay in *E. coli* for inhibitor screening, determining the atomic resolution structure of MtTatC and *E. coli* Tat membrane complex. A detailed and comprehensive analysis of the Haloarchaeal Tat pathway was also undertaken to get insights into its diversity in Haloarchaea, which almost exclusively use the tat pathway for protein secretion.

The Tat signal chaperon interaction was studied for the *E. coli* Tat substrate DmsA and DmsD. The interaction of this complex with the Tat membrane receptor TatBC was tested. This work revealed that DmsD binds to the N-terminal signal peptide of ssDmsA and forms two distinct populations. The equilibrium of these forms can be modulated by either DmsD concentration or N-terminal truncation. N- and C-terminal truncation of the ssDmsA yielded the minimum DmsA signal peptide for formation of signal chaperon complex consist of 29 amino acids. These truncations also revealed that the C-terminal part of the ssDmsA is more directly involved in DmsD binding. The kinetic analysis from SPR suggested that this peptide has the faster association rate (Ka = 10.7 X 104 Ms-1) compare to all other mutant and wild type, and the affinity is very close to the wild type signal peptide K_D = 56.5 nM. The sequence comparisons of all Tat signal, which are having binding chaperons reveals some region are more. Alanine scanning mutations of the 29 residues suggest that twin arginine motif is not essential for chaperone recognition. Further in vitro analysis for identifying the regions responsible for this interaction between TatBC with DmsD, ssDmsA and ssDmsA/DmsD complex showed that the N-terminal of the signal peptide is mainly involved in the TatBC recognition, which includes the Tat motif. The deletion of the N-terminal upstream of the Tat motif still formed complex with TatBC, but the association was weak and formed a different type of complex. The C-terminal of the signal peptide did not have affect binding. Among

monomeric and dimeric types of ssDmsA/DmsD complex, only the monomeric population interacts with TatBC and forms the tetrameric complex. The DmsD concentration depended binding towards signal peptide has been observed. Crystallization of the DDF chimeric ssDmsA-DmsD yield non diffracting crystals. This suggest that the disorder of the DmsA signal sequence is possibly a major factor of non-diffraction. The DMSD + DDF Dimer with a high melting temperature is a good candidate for further crystallization trials.

The Tat pathway is very significant for virulence of various pathogens. Moreover, the Tat pathway is specific to bacteria and transports virulence factors in pathogenic bacteria, thus making it an attractive target to screen for inhibitors to design novel drugs in human healthcare. In this context, it can be deemed necessary to develop an efficient assay that could enable the specific analysis of Tat function and mode of protein translocation. Colanic acid (CA) is an extracellular capsular polysaccharide produced by most *E. coli* strains as well as by other Enterobacteriaceae. It contains units of L-fucose, D-glucuronic acid, D-glucose, D-galactose, and pyruvate, and forms a mucoid matrix on the cell surface. Colanic acid biosynthase (WcaM) is an essential enzyme involved in the colanic acid biosynthesis pathway and is also translocated to the periplasm via the Tat pathway. The activity of the Tat pathway was measured by quantifying L-fucose one of the monomers of Colanic acid using L-Fucose. The spectrometric assay was standardized for large sample sizes and made compatible with a 96-well format microtiter plate. The use of colonic acid biosynthesis native to E. coli as a reporter of Tat export makes available a quantitative and Tat-specific method that can be used to probe Tat pathway function, and scout for new Tat inhibitors. As extraction of Colanic acid after cell growth can be performed using a very simple procedure as there is no need of heterologous expression of a reporter protein, enabling easy and swift quantification of Tat export in the different organisms where heterologous expression is not straight forward. This assay could essentially serve as a simple yet efficient indicator of Tat pathway export and function, including checking the effect of mutations in the Tat component proteins. Also to differentiate inhibitors specific for the Tat pathway against general growth inhibitors, the cell growth curve with colonic acid concentration has to be observed. In conclusion, the assay described in this study offers an easy and robust analysis system in E. coli for experiments concerning Tat pathway function and a screen for identification of new Tat substrates and inhibitors in gastrointestinal pathogens. It combines the advantages of a single-step extraction procedure

and the specificity for Tat pathway, paving the way for the development of an economical high throughput assay for Tat screening.

The evolution of the Tat pathway itself suggests that there may be a fundamental conflict between the substrates transported and the Sec export mechanism in organisms such as haloarchaea that live in extreme habitats. The diversity and multiple unique topologies of the TatC receptor in haloarchaea indicate that they may exist specifically for transport of large numbers and a wide variety of Tat substrates. The high intracellular salt concentrations cause rapid folding of proteins and the quality of protein folding is a major concern for these organisms. This is positively facilitated by the sophisticated Tat pathway components and almost the entire secretome is exported by the Tat pathway. The presence of unique TatC topologies and universal presence of two different homologs was reported with some organisms having multiple homologs of TatA. The Tat sequence analysis of TatC revealed that the 14TM TatC has two copies of TatC and present in haloarchaea along with other homologs. Conservation in the N-terminal region of TatA homologs to that of *E.coli* implies similar mechanism of transport but diversity in the C-terminal region could be useful in interaction with the wide range of substrates. Based on this the substrates of 20 Haloarchaea with complete proteome were analyzed. Unlike in other systems, many substrate groups are present here based on the class of receptor required for the translocation. This may be necessary for differential translocation dynamics and to ensure the stringent folding quality. Further *in-vivo* and *in-vitro* studies are required for dissecting the dynamics of the Tat pathway in haloarchaea. The Haloarchaeal Tat protein proofreading chaperons identified were similar to the bacterial Tat chaperons.

The *Mycobacterium tuberculosis* Tat pathway exported proteins, including virulence factors of periplasmic substrate-binding proteins of ABC transporters and Lactamases but is not essential for survival. Thus developing a MtTat inhibitor would ideally provide a drug that will make the organism non-virulent, at the same time there is no survival pressure diminishing the possibility of resistance. Heterologous expression of MtTatC was achieved in *E.coli* by swapping of the c-terminal cytoplasmic tail with an *Aquifex* TatC tail. Expression was obtained in DMA where MtTatc with AaTail was fused with GFP. This was probably due to the very

stable structure of GFP which stabilized the protein. But the presence of free GFP made us design a new BrilMTA construct but it caused intrinsic bacterial cell death of *E.coli* after induction. Absence of crystal hits due to low quality and limited protein yield is the bottleneck. MtTatC structure was modelled using the AaTatC structure for analysis. The modelled MtTatC structure based on AaTatC structure 1HTS showed the overall structural conservation of TatC fold with the typical transmembrane proline kinks. There is conservation of all important functional residues. The key difference lies in the presence of longer N-terminal and C-terminal cytoplasmic domains and in the periplasmic cap. The MtTatC ampipathc helix is longer than that of AaTatC and surface electrostatic analysis show the presence of positively charged surface on the MtTaC when compared to AaTatC. The difference between the plasma membrane of *Mycobacterium* and *E.coli* is a major factor for the low and unstable expression of TatC. It is best to shift to the *M. smegmatis* based expression system which uses the mc24517 strain and coupling it with the pYUB1602 Lac operon based plasmid.

The *E.coli* Tat membrane complex and receptor bound complex is a good target for Single particle cryo-EM structure determination and will provide us with great insight into the dynamics and functioning of the Tat Translocation complex having the flexible pore size. Negative Staining TEM of EcTatBC complex and EcTatBC+DmsD+DN1 complex shows the sustainability of the system to be studied further in Cryo-EM. The presence of aggregates due to slide preparation delay suggests the necessity to decrease the time gap between Protein purification and grid preparation. The study also emphasize the need the extraction of homogeneous Tat membrane complex and more stable complex. The use of GDN a synthetic alternate to Digitonin which can extract homogeneous EcTatABC complex right from Membrane extraction step should be considered. *E. coli* TatB mutation F13Y forms stable interaction between TatC and TatB which should also be considered. The importance of Phosphatidylethanolamine lipids in Tat activity makes it important to screen complexes with those lipids. The determination of MtTatC structure and the EcTat membrane complex will be a breakthrough for development of peptide inhibitors and also for fine tuning of the Tat pathway for therapeutic protein expression etc.

The Tat pathway is a major virulence determinant in a growing number of pathogenic bacteria including *P. aeruginosa*, *E. coli*, *Salmonella typhi*, *Bacillus anthracis* and many plant

pathogens. Since mammals lack the Tat system, it represents a major new class of antibacterial drug target, which is of interest to drug companies. In addition there is considerable potential for exploiting the rare ability of the Tat system to transport folded proteins by the biotechnology industry for heterologous protein production. Detailed knowledge of the range of substrates using the Tat system and their functional characterization would benefit both these major industries. In addition Tat pathway investigation will generate important fundamental results that will impact upon two mains areas of microbiology: traffic across the bacterial envelope and microbial pathogenesis. Advances in these areas will open up new perspectives in fundamental microbiology and may advance the development of new therapeutic agents.

List of publications

1. **Deepanjan Ghosh[#],** Shridhar Chougule[#], Vellore Sunder Avinash, Sureshkumar Ramasamy*, Evaluating a New Highthroughput Twin Arginine Translocase Assay in Bacteria for Therapeutic Applications. Current Microbiology 2017 Aug
   DOI:10.1007/s00284-017-1321-z
2. **Ghosh, D.**, Boral, D., Vankudoth, K.R. and Ramasamy, S., 2019. Analysis of haloarchaeal twin-arginine translocase pathway reveals the diversity of the machineries. *Heliyon*, *5*(5), p.e01587.
   DOI: https://doi.org/10.1016/j.heliyon.2019.e01587
3. **Ghosh D**. and Ramasamy S.*, Engineering of the tat pathway and chaperons. *Acta Crystallographica Section A: Foundations of Crystallography*, *10*, pp.483-496. (conference publication)
   DOI: 10.1107/S2053273317093007
4. Manu M.S., **Ghosh D**., Chaudhari B.P. and Ramasamy S.*, 2018. Analysis of tail-anchored protein translocation pathway in plants. *Biochemistry and biophysics reports*, *14*, pp.161-167.
   DOI: 10.1107/S2053273317093007
5. Agrawal, S.B., **Ghosh, D**. and Gaikwad, S.M., 2019. Investigation of structural and saccharide binding transitions of Bauhinia purpurea and Wisteria floribunda lectins. *Archives of biochemistry and biophysics*, *662*, pp.134-142.
   DOI: https://doi.org/10.1016/j.abb.2018.12.003
6. **Ghosh D** *et al.* Defining the DmsA Signal Sequence Characteristics Required For Binding To DmsD And TatBC. (Manuscript under preparation)
7. Pushparani D Philem[*], Yashpal Yadav[*], Avinash V Sunder, **Deepanjan Ghosh**, Asmita Prabhune, Sureshkumar Ramasamy. Expanding AHL Acylases Horizon - Insights from Structural Analysis of Choloylglycine Hydrolases from *Shewanella loihica* PV-4. (Communicated).