Structural and Functional Study of Type Three Secretion System Proteins from *Yersinia enterocolitica*

A thesis submitted for the award of degree of

Doctor of Philosophy (Science)

in Life Science and Biotechnology

by

Rajeev Kumar, (M.Sc.)

Structural Biology and Bioinformatics Division,

CSIR-Indian Institute of Chemical Biology, Kolkata, INDIA

Department of Life Science & Biotechnology

Faculty Council of Science, Jadavpur University,

Kolkata – 700032, INDIA

2022

CERTIFICATE FROM THE SUPERVISOR(S)

This is to certify that the thesis entitled "**Structural and Functional Study of Type Three Secretion System Proteins from *Yersinia enterocolitica*"** submitted by Sri / Smt. **Rajeev Kumar** who got his / her name registered on **22-11-2016** for the award of Ph. D. (Science) degree of Jadavpur University, is absolutely based upon his own work under the supervision of **Dr. Saumen Datta,** and neither this thesis nor any part of it has been submitted for either any degree/diploma or any other academic award anywhere before.

7-11-2022

Dr. Saumen Datta,
Senior principal scientist,
Structural Biology and Bioinformatics division,
CSIR- Indian Institute of Chemical Biology,
Jadavpur, Kolkata-32

डॉ. सौमेन दत्ता / Dr. Saumen Datta
प्रधान वैज्ञानिक / Principal Scientist
सीएसआईआर-भारतीय रासायनिक जीवविज्ञान संस्थान
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)
CSIR - Indian Institute of Chemical Biology
(Council of Scientific & Industrial Research)
4, राजा एस. सी. मल्लिक रोड / 4, Raja S. C. Mullick Road
कोलकाता-700032 / Kolkata-700032

<u>Dedication</u>

"Dedicated to the

hero of my life, my elder brother

***Late. Sanjay Kumar*,**

who always believed in me,

and to

my Family"

<u>Acknowledgments</u>

Table of Contents

Chapter 1

1. Introduction
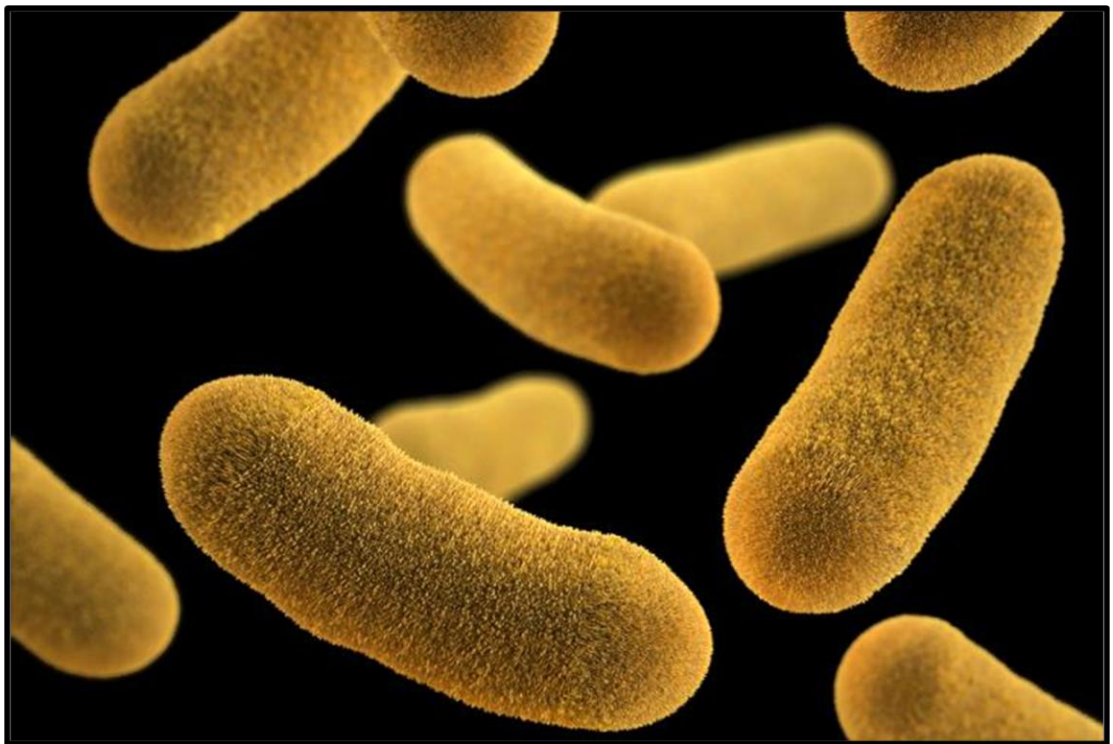
SEM Image of *Yersinia enterocolitica*. Adapted from- *Yersinia enterocolitica* infection by HEALTH JADE TEAM.

1.1. The Genus Yersinia:

The genus *Yersinia* is a gram-negative bacterium that belongs to the family of *Enterobacteriaceae*. The genus *Yersinia* was proposed by Van Loghem in 1944 to accommodate *Yersinia pestis* and *Yersinia pseudotuberculosis* which was previously put into the genus *Pasteurella* [1]. These are Gram-negative, coccobacilli bacteria, a few micrometres long and less than a micrometre in diameter, and are primarily a facultative anaerobe. Currently, this genus consists of truly 11 species [2] out of which only three are reported pathogenic to humans. The three species pathogenic to humans are *Yersinia enterocolitica, Yersinia pseudotuberculosis*, and *Yersinia pestis*. Human pathogenic *Yersinia* causes diseases ranging from gastroenteritis for *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* (who are primarily entero-pathogen), to bubonic plague caused by *Yersinia pestis* [3]. All these three species differ in their pathogenicity, *Yersinia pestis* is the deadliest among all and primarily a rodent pathogen infecting human through the bite of an infected flea, while, *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* are zoonotic food-borne pathogens having broad host range including humans, swine, dogs, rodents, birds, and wild animals [4] [5]. All the three *Yersinia* species carry a 70- kb virulence plasmid which they require during infection to the lymph tissues as well as to overcome host defence mechanism [3] [6] [7]. Based on recent development in whole-genome sequence analysis, the genus *Yersiniae* has been regarded as a model genus to study rapid evolution among bacterial pathogens [7].

1.2. Evolution of Yersinia species:

During evolution, *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* are supposed to have emerge within the last 200 million years and *Yersinia pestis* is relatively has emerged recently within the last 1500- 20000 years [7-9]. Based on 16S rDNA analysis, **Figure 1.1** represents the intrageneric phylogenetic relationship among *Yersinia* species and the position of the *Yersinia* genus within the family of *Enterobacteriaceae*. It clearly shows a coherent cluster of all *Yersinia* strains with sequence similarities ranging from 96.9%-99.8%. It also suggests the clonal diversity among strains of *Y. enterocolitica* and the close relation between *Y. pestis* and *Y. pseudotuberculosis* during evolution [10].

Yersinia enterocolitica diverged from *Yersinia pseudotuberculosis* and evolved into a genetically and biochemically heterogeneous group of organisms and has been classified into six biotypes differentiated by biochemical tests (A, 1B, 2, 3, 4 and 5) [11]. These in vitro biotypes can further can be differentiated based on their pathogenic potential: High pathogenic (biogroup 1B); nonpathogenic (biogroup 1A); and weakly pathogenic (biogroup 2, 3, 4, and5) based on the mouse infection model [11-13] (**Figure 1.2**). Biogroup 1A lacks the presence of virulence plasmid pVY and is suggested to be distantly related to other biogroups
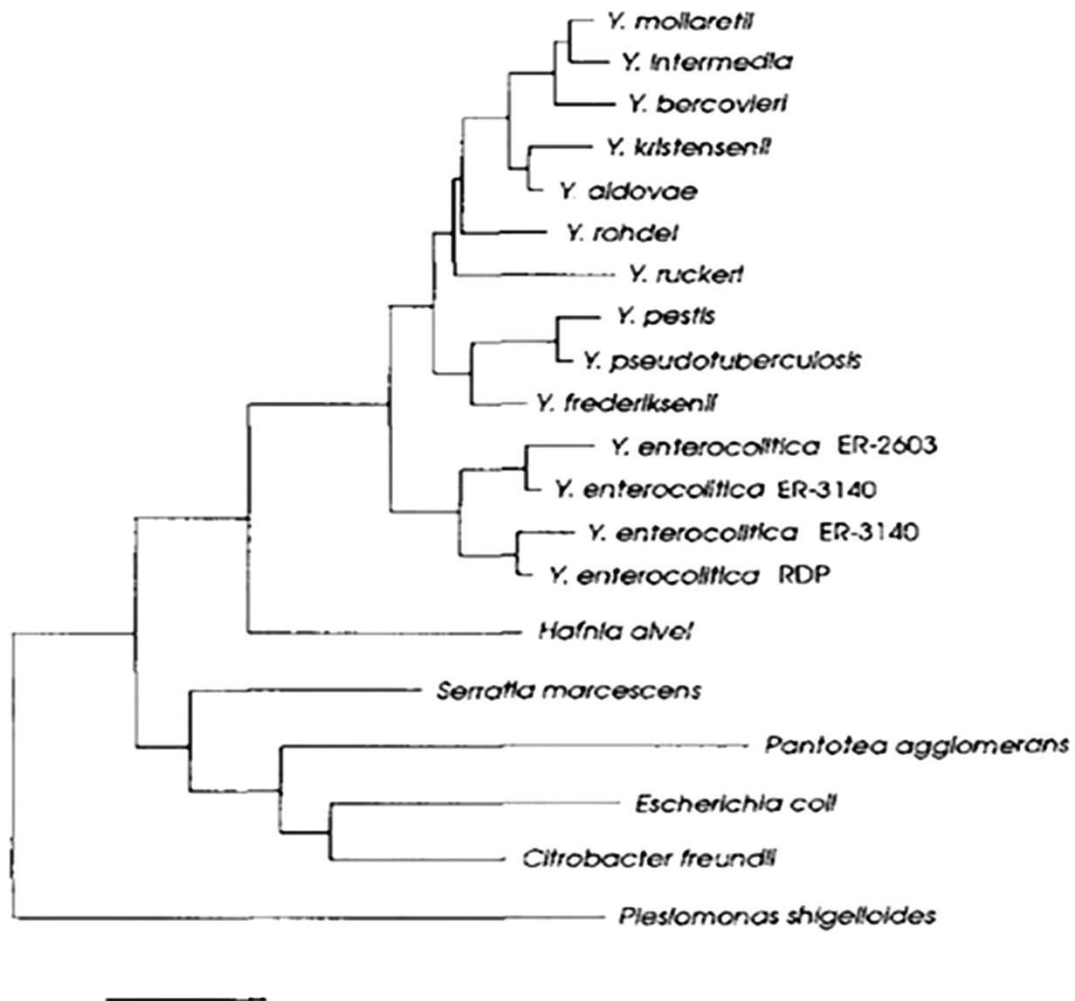
Figure 1.1: intrageneric phylogenetic relationships of *Yersinia* species and position of the genus *Yersinia* within the family *Enterobacteriaceae*, based upon 16S rDNA analysis. The bar represents 1% sequence divergence. (Adapted from Ibrahim, A., Goebel, B. M., Liesack, W., Griffiths, M., & Stackebrandt, E. (1993). The phylogeny of the genus Yersinia is based on 16S rDNA sequences. *FEMS microbiology letters*, *114*(2), 173-177.).

[10]. Biogroup 1B (highly pathogenic) frequently isolated in North America is a geographically isolated group also known as so-called '**New world**' Strains and biogroups 2- 5 (weakly pathogenic) predominantly isolated in Europe and Japan are known as 'Old world' strains. Based on variations in the O- antigen of its lipopolysaccharide (LPS) *Yersinia pseudotuberculosis* is classified into 21 different serological groups, while all *Yersinia pestis* strains are unable to express the O- antigen [7].
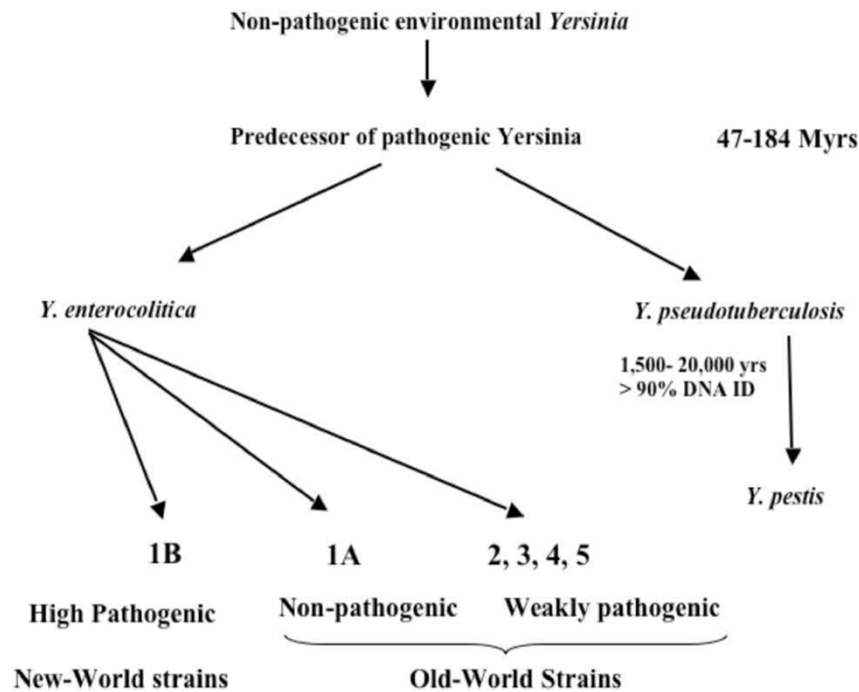
Figure 1.2: A basic model describing the evolution of the pathogenic *Yersinia* (adapted from (Wren 2003) THE GENUS YERSINIA: From Genomics to Function 2007 Springer Science Business Media, LLC).

1.3.Pathogenicity of Yersinia species:

Three important human pathogens: *Yersinia pestis*, *Yersinia enterocolitica*, and *Yersinia pseudotuberculosis* belong to the genus *Yersinia*. They cause zoonotic infections such as Yersiniosis (a Diarrheal disease) and plague (Bubonic and Pneumonic). In healthy but immunocompromised individuals, yersiniosis symptoms range from mild, self-limiting diarrhea to mesenteric lymphadenitis [14]. After ingestion of contaminated food and water, the enteropathogenic bacteria (*Yersinia enterocolitica* and *Yersinia pseudotuberculosis*) pass into the small intestine and translocate across the intestinal epithelium through Peyer's patches. Following the translocation, they migrate to the lymph node and subsequently to the liver and spleen to carry out replication externally to the host cell [15] (**Figure 1.3**). However, *Yersinia pestis* spreads across animals and occasionally to humans through infected fleas (which act as a mammalian reservoir) and reaches lymph nodes causing Bacteremia (popularly known as Bubonic plague). When this infection reaches the lungs, it results in fatal Pneumonic plague [7].

Plasmid acquisition by the Lateral gene transfer method by *Yersinia* species seems the major key resulting in virulence acquisition (see the figure below). For example, very

recently *Yersinia pestis* made an evolutionary jump by acquiring plasmids to become a flea-based systemic pathogen. A close analysis of the genomes of all *Yersinia* species suggests common virulence plasmids namely Plasmid-encoding murine toxin (pMT1), Plasminogen-activating plasmid (pPla), and, *Yersinia* virulent plasmid (pYV) present among all pathogenic *Yersinia* species which contributes to a wider host range and improves their ability to beat the host defense mechanisms [7] (**Figure 1.4**).

Among these plasmids, plasmid pYV is of particular interest as it plays a crucial role in the virulence of *Yersinia species*. pYV encodes several virulence determining factors including a specialized secretion system known as Type Three Secretion System (T3SS). This T3SS helps Yersinia to overcome the immune defense system of the host and helps *Yersinia* to establish the infection (refer to the **Table 1.1** below)**.**

Table 1.1	Plasmids important in the virulence of pathogenic yersiniae			
Plasmid name	**Other designations**	**Size (kb)**	**Virulence determinants**	**Role in disease**
Yersinia virulence plasmid, pYV	pCD	70.3	Several YOPs, type III secretion system	Avoidance of the immune system, toxicity
Plasmid-encoding murine toxin, pMT1	pFra	96.2	Murine toxin (phospholipase), F1 capsule-like antigen	Bacterial transmission by fleas
Plasminogen-activating plasmid, pPla	pPCP1, pPst	9.6	Plasminogen activator	Dissemination from intra-dermal site of infection

Table 1.1: Plasmids important in virulence of pathogenic *yersiniae*. (Adapted from- The *Yersiniae* — a model genus to study the rapid evolution of bacterial pathogens).

1.4. Secretion Systems in gram-negative bacteria:

Yersinia species harbors many different secretion systems for their growth and survival. They use these secretion systems for a different array of processes. Secretion systems are used to transport proteins of interest out of the cell into the bacterial environment or inside the host cell directly. In other cases, they use it especially for manipulating the host and a replicative niche. Different classes of bacterial secretion systems have been classified based on their mode of mechanism, whether they secrete protein substrate across a single membrane, double-membrane, or even three membranes including the host membrane. Due to the specificity of these secretion systems, few of them are present only in the pathogenic form of *Yersinia* only. Such secretion systems are a target for antimicrobials development. A brief description of the secretion systems in *Yersinia* species involved in virulence is given below.
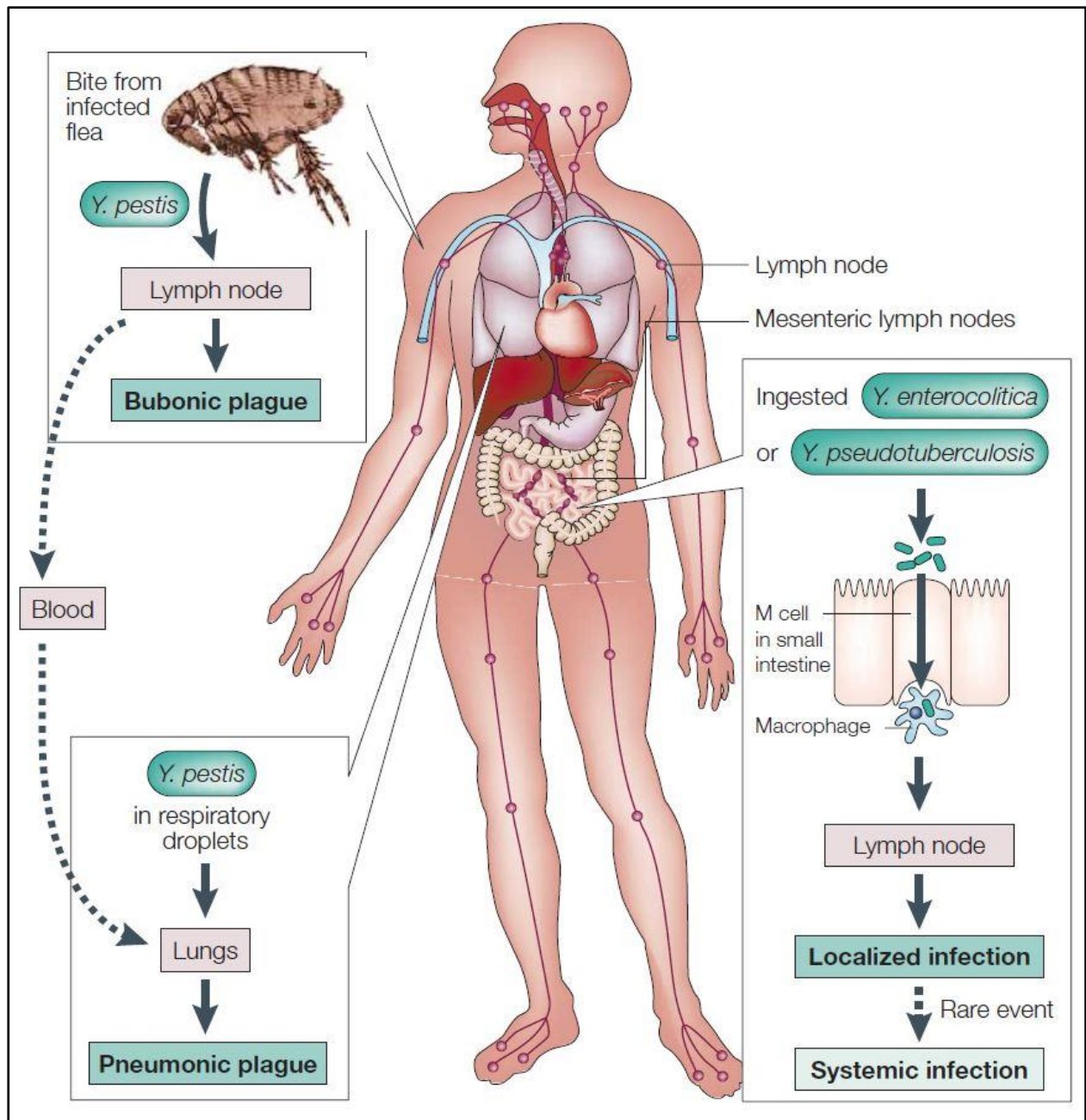
Figure 1.3: Steps in the transmission of the pathogenic *yersiniae* in humans. *Y. pestis* has a rodent reservoir. The rodent's fleas acquire *Y. pestis* from a meal of infected blood, and transmit the bacterium primarily to other rodents, or occasionally to humans — causing bubonic plague in humans. Human-to-human transmission can occur through human fleas. Pneumonic plague is transmitted from human to human through respiratory droplets, or possibly by artificially generated *Y. pestis* aerosols. *Y. enterocolitica* and *Y. pseudotuberculosis* are ingested, and in contrast to *Y. pestis*, enter the lymphatic system through the M cells of the small intestine. (Adapted from- The *Yersiniae* — a model genus to study the rapid evolution of bacterial pathogens).

Nature Reviews | Microbiology
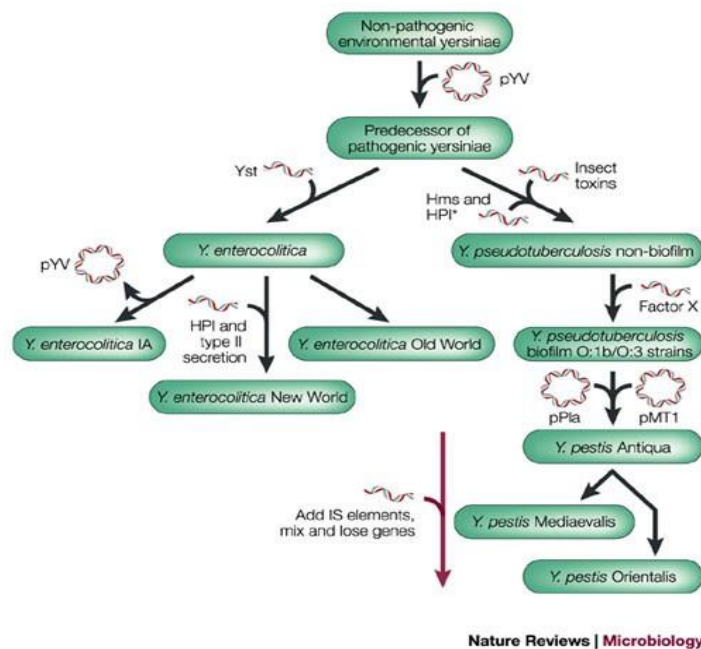
Figure 1.4 | **Simplified model of *Yersinia* species evolution based on present knowledge of genome data.** The non-pathogenic yersiniae gain the virulence plasmid pYV to form the predecessor of pathogenic Yersiniae. *Y. enterocolitica* diverges from *Y. pseudotuberculosis* and forms three lineages: 1A, Old World and New World. *Y. pseudotuberculosis* gains the ability to parasitize insects and form biofilms in hosts before evolving into *Y. pestis* through the acquisition of the plasmids pPla and pMT1, genome mixing and decay. Hms, haemin storage; HPI and HPI*, high-pathogenicity islands; IS, insertion sequence. (Adapted from- The Yersiniae — a model genus to study the rapid evolution of bacterial pathogens).

1.4.1.The Sec Secretion Pathway:

The general Secretory [9] pathway is ubiquitous and universal export machinery for most of the proteins being transported across the bacterial membranes along with the **twin-arginine translocation** [16] pathway. Both these pathways are most highly conserved among the secretion pathways across different domains of life *i.e.,* bacteria, archaea, and eukarya [17, 18]. In the **Src-dependent pathway**, the protein being secreted is synthesized as a pre-protein containing a cleavable terminal signal peptide. The Sec pathway translocate its substrate proteins primarily in their unfolded state. The Sec Secretion pathway is a system that consists of three major components: a protein targeting component, a motor protein, and a trans-membrane integrated conducting channel component (known as **SecYEG translocase**) [17]. Many proteins secreted through Sec Pathway serve different purposes, whereas many others play role in promoting the virulence of bacterial pathogens. Gram-negative bacteria like *Yersinia* use Sec dependent pathway to secrete virulence determining factors across the cytoplasmic membranes [19]. A 20 amino acid hydrophobic, N-terminal

signal sequence is required for a protein to undergo Sec dependent secretion pathway. The hydrophobic signal sequence consists of three regions: a positively charged amino terminal, a hydrophobic core, and a polar carboxy-terminal. Proteins containing Sec-B specific signal sequence are destined to be secreted outside the cell by the Sec pathway and proteins containing **signal recognition particle** (SRP) specific signal sequence will be destined to remain in the inner membrane [17]. The sec-dependent pathway operates in basically three distinct steps: Sorting and targeting, translocation, and maturation. The targeting of protein substrates operates in both co-translational and post-translational modes. The energy for this secretion pathway is provided by ATP hydrolysis by motor protein SecA and by **Proton Motive Force** (PMF). A general Sec-dependent pathway is depicted in **Figure 1.5.** For a detailed review of this system refer to [20].
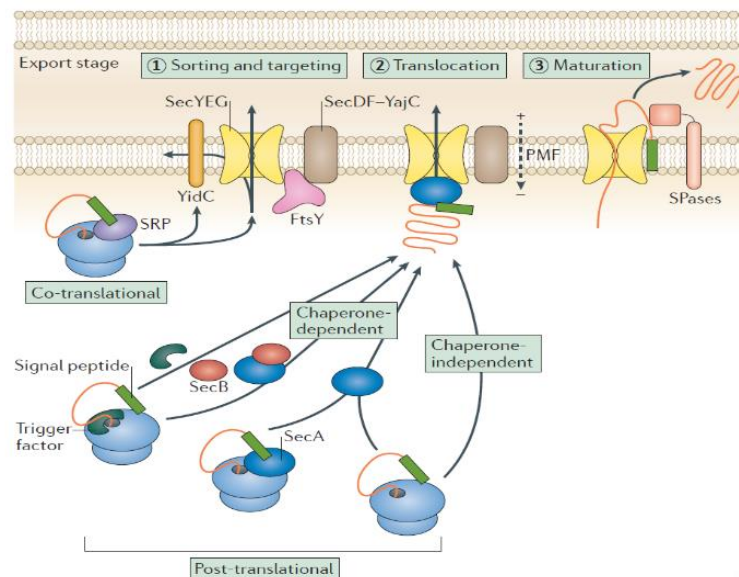
Figure 1.5: Sec-pathway-dependent export stages. Sorting and targeting (step 1): unfolded proteins (orange) that contain signal peptides and plasma membrane proteins are co-translationally sorted and targeted to the transmembrane SecYEG channel (yellow) by the signal recognition particle (SRP; purple) and its membrane receptor FtsY11 (pink), or, post-translationally, by SecA (blue)12. Chaperones aid post-translational targeting: trigger factor (dark green)22,24 and the ATPase motor SecA12,14 bind to pre-proteins on the ribosome or in the cytoplasm; the chaperone SecB binds to them in the cytoplasm2. Pre-proteins may be targeted to the SecYEG–SecA translocase in a chaperone-independent manner. Translocation (step 2): pre-proteins translocate through SecYEG to the periplasm or into the plasma membrane2, the process of which is powered by repeated cycles of ATP binding and hydrolysis by SecA16,144 and the proton motive force (PMF)16. The auxiliary components SecDF–YajC6 and YidC11 (light orange) enhance translocation efficiency. Maturation and release (step 3): signal peptidases (SPases; pale pink) cleave signal peptides and the mature domain is released into the periplasm21. Adapted from Tsirigotaki *et.al.*, (2017). Protein export through the bacterial Sec pathway. *Nature Reviews Microbiology*, *15*(1), 21-36.

1.4.2.The Tat Secretion Pathway:

Unlike the Sec secretion pathway, a functional Twin-arginine translocation [16] secretion pathway secretes proteins in the folded state across bacterial membranes [18]. The Tat pathway consists of two to three components: TatA, TatB, and TatC. In gram-negative bacteria like *E. coli*, TatB and TatC bind to the signal peptide of proteins to be secreted by Tat-Pathway and recruit TatA, present as the membrane-spanning channel. A pair of "**twin**" **arginines** in the motif S-R-R is present at the N-terminus of the folded proteins which acts as a signal sequence for Tat secretion pathway [21]. The energy for translocation of substrates by Tat Secretion is provided by Proton Motive Force (PMF) [18]. Normal physiology and survivability of pathogenic and nonpathogenic bacteria require Tat pathway. While in the case of Pseudomonas aeruginosa, Yersinia pseudotuberculosis, E. coli they require a functional Tat pathway for full virulence in animal models. [22-24]. A very good example is Phospholipase C- a Tat secreted enzyme that cleaves phospholipids immediately resulting in suppression of immune cell activity and promoting intracellular survival of the pathogenic bacteria. **Figure 1.6** provided below represent a general secretion mechanism involving Tat pathway.
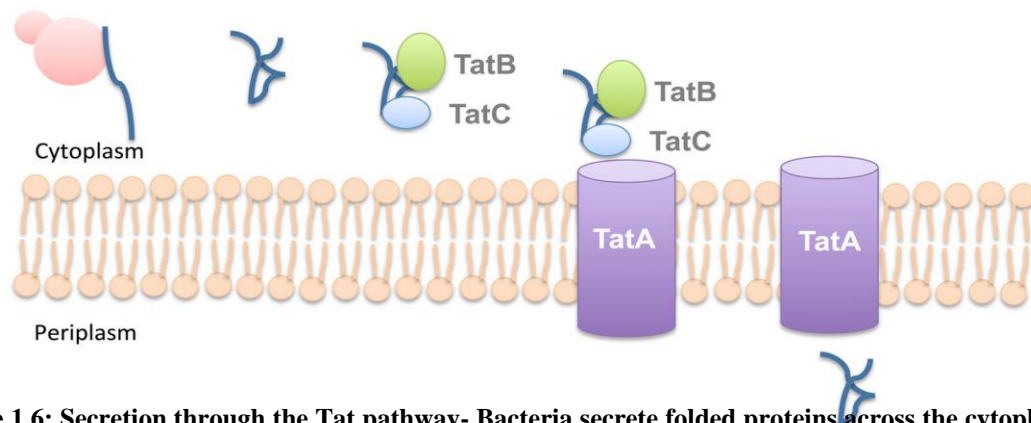
Figure 1.6: Secretion through the Tat pathway- Bacteria secrete folded proteins across the cytoplasmic membrane using the Tat secretion pathway. This pathway consists of 2–3 components (TatA, TatB, and TatC). In Gram-negative bacteria, TatB and TatC bind a specific N-terminal signal peptide containing a "twin" arginine motif on folded Tat secretion substrates. TatB and TatC then recruit TatA to the cytoplasmic membrane, where it forms a channel. Folded proteins are then translocated across the channel and into the periplasm. Adapted from (Green, E. R., & Mecsas, J. (2016). Bacterial secretion systems: an overview. *Microbiology spectrum*, *4*(1), 4-1.).

1.4.3.The Type I Secretion System:

T1SS are widely present among gram-negative bacteria and serve a variety of purposes from nutrient acquisition to bacterial virulence. This is a secretion system involving one-step translocation of proteins across both inner and outer bacterial membranes. T1SS closely resembles the ATP-binding cassette (ABC) transporter family of proteins which are involved in the export of small molecules such as antibiotics and toxins out of the cell [25]. Various substrates like digestive enzymes, such as proteases and lipases, adhesins, heme-

binding proteins, and proteins carrying repeats-in-toxins (RTX) motifs are translocated by T1SS generally in a Sec-independent manner. Such substrates usually carry a C-terminal non-cleavable signal sequence which is recognized by T1SS. The major component of T1SS is an ABC transporter protein present in the inner membrane, an outer membrane factor (OMF) in the outer membrane, and an inner membrane-embedded membrane fusion protein (MFP) associating both two components inside the periplasm [26]. In addition, the cytoplasmic side i.e., the N-terminal side of the MFP plays role in substrate selection [27, 28]. The energy for T1SS is provided by the associated ABC transporter. After substrate recognition by MFP cytoplasmic N-terminal domain, OMF generates a pore in the outer membrane through which unfolded substrate passes to the outside of the cell. The multi-purpose TolC (which is a member of the outer membrane efflux protein (OEP) family) is recruited by T1SS as their OMF component after substrate recognition by MFP and ABC transporter. Substrates secreted through T1SS contribute to virulence in many pathogenic bacteria. One widely studied example is HylA hemolysin protein of *E. Coli* (belonging to the RTX-family of proteins) which inserts into membranes of both erythrocytes and nucleated eukaryotic cells and rupture them [29]. **Figure 1.7** below represents T1SS in pathogenic bacteria.
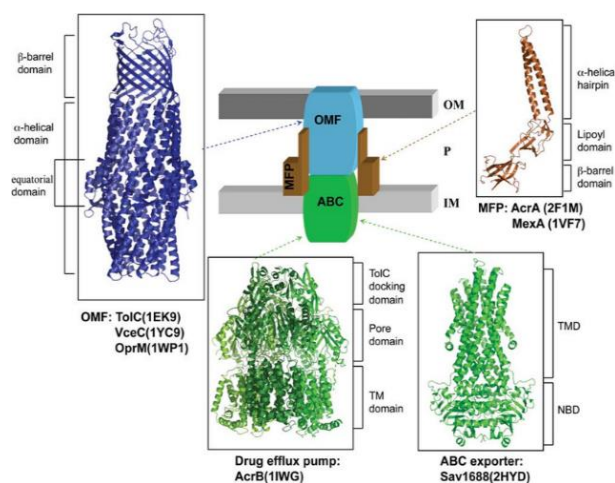
Figure 1.7: Schematic representation of the Type 1 secretion system. The tripartite system is shown schematically with the inner membrane ATPase Binding Cassette (ABC) protein shown in green, the membrane fusion protein (MFP) in brown, and the outer membrane factor (OMF) in blue. A ribbon representation of the full-length ABC exporter structure Sav1688 from *S. aureus* is shown in the inset green. Also shown is a ribbon representation of the proton antiporter AcrB from *E. coli* (PDB: 1IWG) in green. The crystal structure of AcrA from *E. coli*, an MFP from the analogous drug efflux pump system, is shown on the side inset in brown. The representative OMF crystal structure of *E. coli* TolC is also shown in blue. PDB codes are indicated. Adapted from (Durand, E., Verger, D., Rêgo, A. T., Chandran, V., Meng, G., Fronzes, R., & Waksman, G. (2009). Structural biology of bacterial secretion systems in Gram-negative pathogens-potential for new drug targets. *Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders)*, *9*(5), 518-547.).

1.4.4.The Type II Secretion System:

Gram-negative pathogenic bacteria use the Type Two Secretion System (T2SS) to secrete folded protein from the periplasm into the extracellular environment. Proteins secreted by T2SS must be present in periplasm before being transported to the extracellular matrix by T2SS present only in the outer membrane of bacteria [19]. It is the terminal step in the General Secretory pathway (Gsp) [30] which secretes proteins secreted through Sec and Tat pathway. T2SS secretes a variety of substrates including toxins required for virulence, single protein, enzymes, proteases, lipases, phosphatases, and several proteins that are required for carbohydrate metabolism in different bacterial species. The T2SS is composed of 12-15 different types of proteins encoded by a single operon. The T2SS can be further subclassified into four different sub-assemblies namely the outer membrane complex composed of secretin (which serves as a channel for translocation of folded substrates from the periplasm to the extracellular space), the inner membrane platform- embedded in the inner membrane, and extending toward the periplasm which plays a crucial role in communicating between the outer membrane secretin complex and the secretion ATPase during the secretion process, the pseudo pilus (fibrous periplasmic structure which is supposed to push the substrates outside the outer membrane complex like a piston), and a tightly associated secretion ATPase (located in the cytoplasm which acts as energy provider for the T2SS process). [19] (**Figure 1.8**).
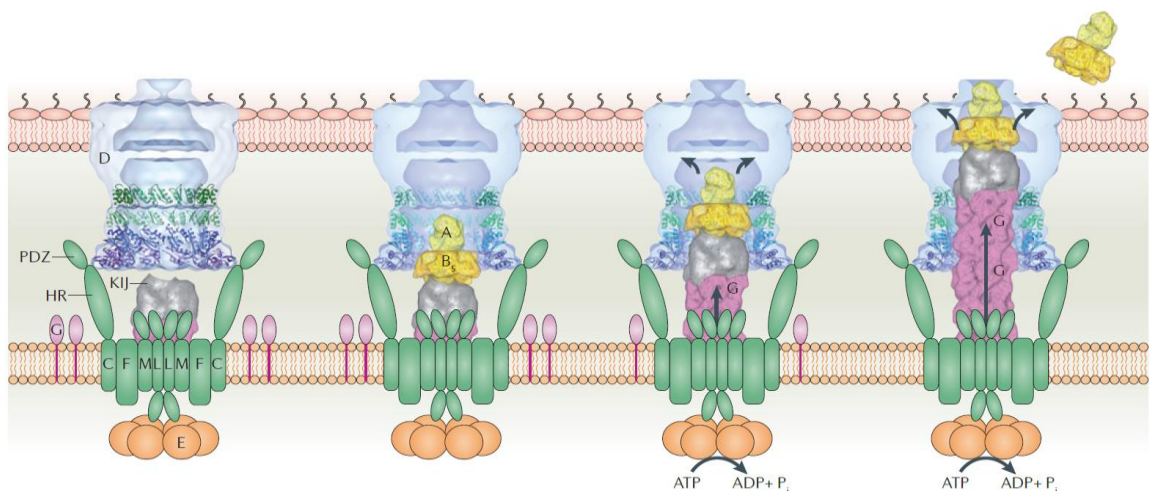
Figure 1.8: A possible mode of action of the T2SS. The hypothetical mechanism of action for the type II secretion system (T2SS) is based on numerous biochemical and structural studies. General secretory pathway (Gsp) proteins are labeled with their capital-letter identifiers (the Gsp prefix and the species-specific names are omitted). The secretion ATPase (GspE) is in orange, the inner-membrane platform proteins are in green, the trimeric tip of the pseudo pilus (GspKIJ) is in silver, the major pseudopilin (GspG) is in pink and the outer-membrane secretin (GspD) channel is in blue, cyan and purple. Also please refer to Korotkov, K. *et. al,* (2012). The type II secretion system: biogenesis, molecular architecture, and mechanism. *Nature Reviews Microbiology* for complete detail of the figure presented here. The possible architecture of the T2SS before binding an exoprotein is shown on the left. The second structure shows the T2SS with an exoprotein (in this case, the AB5 cholera toxin) in the periplasmic vestibule of the T2SS secretin. Adapted from Korotkov, K. *et.*

al, (2012). The type II secretion system: biogenesis, molecular architecture, and mechanism. *Nature Reviews Microbiology.*

1.4.5. The Type III Secretion System:

A pathogenic bacterium uses a unique multi-protein complex structure known as the Type Three Secretion System (T3SS) to manipulate its host cell physiology by secreting effector toxins directly into the host cell cytoplasm crossing three layers of membranes. These bacterial effectors interfere with the host cellular pathways to promote attachment to the host cell and help the bacteria for intracellular survival and replication or prevent host cell defense. The T3SS resembles a complex nano syringe-like structure generally referred to as Injectisome [31-34]. In all T3SS among different bacteria, it shares highly conserved nearly eight of its core components which are also evolutionary related to the flagellar system [35]. More than 25 different proteins are involved in the core architecture of T3SS complex formation however it has been classified into three different parts namely a double membrane-spanning basal body complex composed of stacked rings, a needle complex filament structure protruding away from the bacterial surface into the extracellular space and Translocon [36]. The needle complex is a ≈60nm multi-ring elongated tube structure with a base diameter of ≈26nm and a narrow channel of ≈20Å for transport of the effectors [37]. The narrow channel of the secretion needle allows only unfolded substrates to pass through it. Also associated with the basal body inside the cytoplasm a complex dynamic component structure called C-ring complex is present including hexameric ATPase ring complex which provides energy for assembly, efficient unfolding, and transport of effectors through T3SS injectisome [38]. The Tip complex present at the end of the needle complex helps in sensing contact with the host cells and regulation of secretion of effectors and also is required for insertion of the translocon to the host membrane. After the tip complex establishes contact with the host cell membrane the tip complex forms a pore in the host membrane for delivery of the effectors [39-41]. A pathogenic bacteria may secrete effectors, based secretion signal embedded within its N-terminal end, or in many cases the effectors are guided by their respective chaperones which guide them to be secreted by T3SS in an ATP-dependent manner in their unfolded state [31]. A general T3SS model is represented in **Figure 1.9**. For a more detailed review of T3SS please refer to the review of literature section.

1.4.6. The Type IV Secretion System

Type Four Secretion System (T4SS) has been reported in both Gram-positive as well as in Gram-negative bacteria. T4SS is related to the ancestral DNA conjugation process including DNA uptake and transfer. They contain a channel for translocation of protein-DNA complexes and traverse the cell envelope of the bacteria. In Gram-negative bacteria, T4SS are used to transport virulence factors into the host cell. Based on their broad function T4SS

can be divided into major three classes. The first type is T4SS used to transfer DNA from one bacterial cell to another in a contact-dependent manner. The second type includes
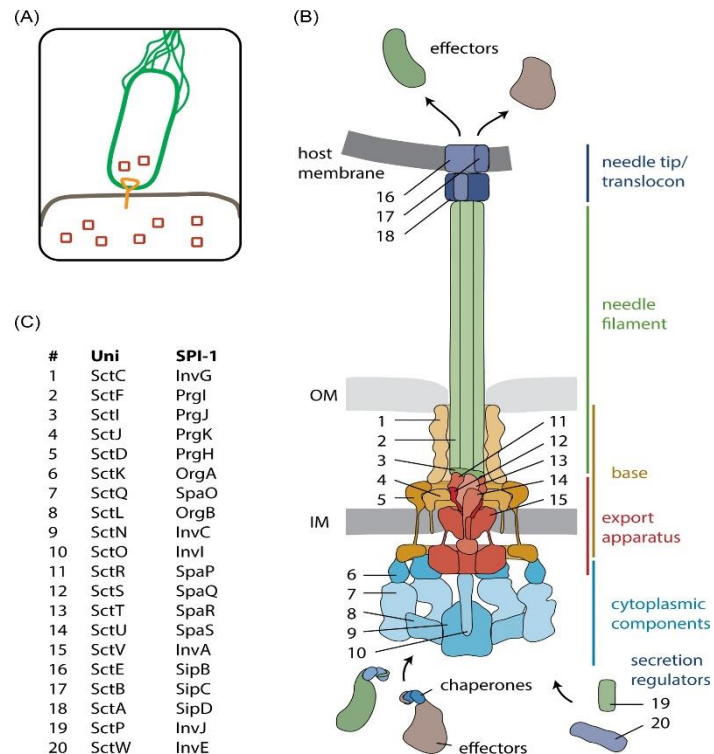
Figure 1.9: **Structure and function of the T3SS injectisome. A.** Bacteria utilize T3SS to inject bacterial effector proteins into eukaryotic host cells. **B.** Structure of the T3SS indicating the different structural units (color-coded) and the individual components. Protein names are given in C. **C.** Nomenclature of T3SS components. The unified nomenclature (Uni) according to Hueck (1998) is shown on the left, the nomenclature of the T3SS encoded by *Salmonella* pathogenicity island 1 (SPI-1) is shown on the right. The nomenclature of the most commonly studied systems in *Yersinia* and pathogenic *Escherichia coli* follows the unified nomenclature with Ysc and Esc, respectively, instead of Sct. Abbreviations: IM, inner membrane; OM, outer membrane. Adapted from Wagner, S., Grin, I., Malmsheimer, S., Singh, N., Torres-Vargas, C. E., & Westerhausen, S. (2018). Bacterial type III secretion systems: a complex device for the delivery of bacterial effector proteins into eukaryotic host cells. *FEMS microbiology letters*, *365*(19), fny201.

T4SS involved in DNA uptake and release in the extracellular environment. The third type of T4SS is involved in protein or protein complex secretion into their host cell. Most of the pathogenic gram-negative bacteria contain this type [42]. T4SS has been well characterized in the case of *Agrobacterium tumeficans* and is commonly known as the VirB/D system. *A. tumeficans* utilizes T4SS for the transfer of *Ti*-plasmid into the eukaryotic plant host cell. T4SS contains approximately 12 proteins namely VirB1- VirB11 and VirD4 [43, 44]. VirB4, VirB11, and VirD4 are ATPases and remain associated with the inner membrane and provide energy for T4SS functioning. VirD4 also acts as a coupling agent by interacting with protein

substrates before secretion. Bacteria like *Helicobacter pylori* have been reported to uptake extracellular DNA from the environment to the bacterial cytoplasm. Recently it has also been suggested that T4SS is involved in reducing competition between rival bacteria by injecting toxic effectors into them [45]. The versatility and non-specificity of T4SS function provide bacteria additional adaptability to their survival in different environmental conditions. T4SS plays a wide range of functions such as providing antibiotic resistance by DNA uptake, remodeling of the host cell architecture by secreting toxins into them, promoting environmental adaptability, and helping in virulence gene acquisition. A general model representation of T4SS is provided below in **Figure 1.10.**
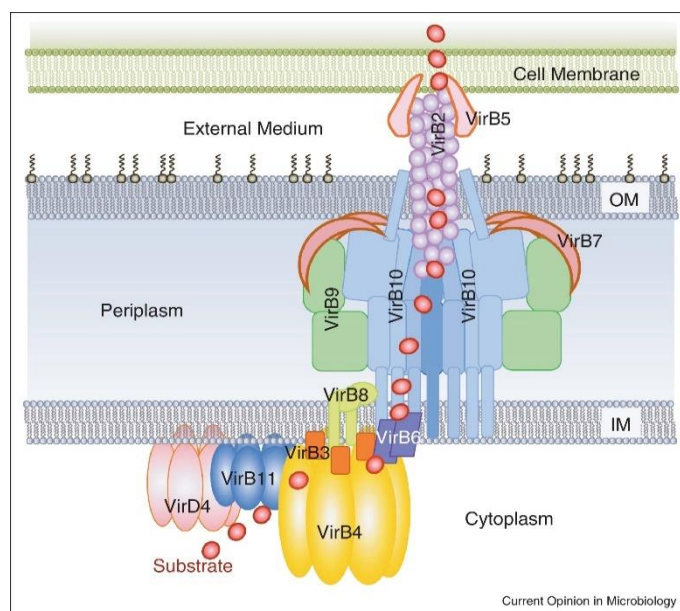
Figure 1.10. The overall organization of the T4S system. VirD4 (in pink), VirB11 (in blue), VirB4 (in gold) ATPases, polytopic VirB6 (in purple), bitopic VirB8 (in light green) and VirB3 (in orange) form the cytoplasmic IM part of the complex. VirB7 (in brown), VirB9 (in green), and VirB10 (in blue) compose the periplasmic part of the secretion system. VirB2 and VirB5 constitute the outer part of the secretion system. The red dot indicates the path of the substrate through the machinery. The stoichiometry of the various components in a native, fully assembled, T4S system is unknown. Adapted from Waksman, G., *et.al* (2014).

1.4.7. The Type V Secretion System:

Type V Secretion systems (T5SS) present in gram-negative bacteria are unique secretion processes where a substrate along with its transporter is synthesized as a single polypeptide and is secreted independently by themselves into the extracellular environment. Hence, they are also known as autotransporter systems. A membrane channel component, known as 'translocator' and its specific substrate, known as 'passenger' domain fused as a single polypeptide remain present in the periplasm of bacteria in a partially unfolded state

with help of the periplasmic chaperones (**Figure 1.11**). T5SS is functional in the outer membrane only and hence its components must be secreted to the periplasm by the Sec apparatus. For its secretion by Sec apparatus, T5SS proteins must carry an N terminal Sec signal sequence which is cleaved in the periplasm. The translocator inserts into the membrane forming a β-barrel and enabling secretion of the passenger domain through it [46-48]. Energy for this process is provided by the folding of the substrate at the exit of the pore [49]. The passenger domain secreted in many cases is auto processed and cleaved following its release into the extracellular environment, whereas in other cases the passenger domain secreted is not auto processed and remains anchored to its translocator domain and usually functions as adhesins [48].
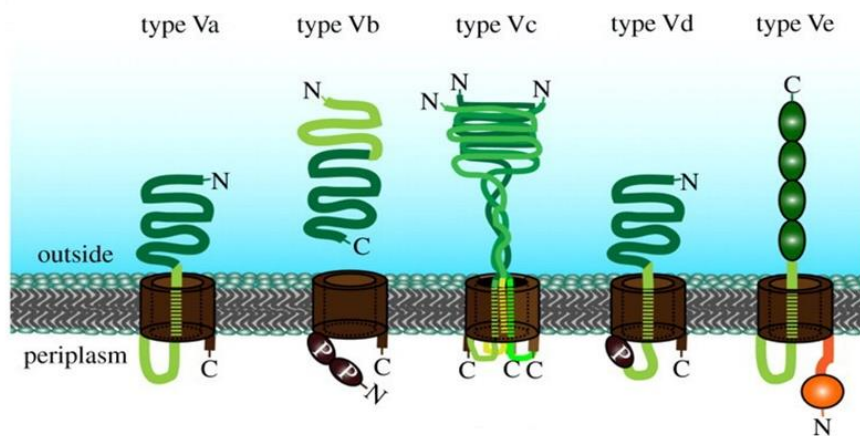
Figure 1.11. Structures and topology models. (a) Topology models of the different type V secretion systems. The translocation domain is displayed in brown, linker/Tps regions in light green, passenger domains in dark green and periplasmic domains in orange. POTRA domains are labeled (P). For clarification of the topology, N- and C-termini are indicated. Adapted from Leo, J. C., *et.al.* (2012).

1.4.8. The Type VI Secretion System

Type VI secretion systems (T6SSs) are the most recently discovered bacterial secretion systems and their structure and functions are still unknown. The transport of proteins via T6SS occurs from a donor cell to the recipient cell during interbacterial interaction in a contact-dependent mode. T6SS is widely distributed in the genomes and is composed of 13 conserved proteins along with a variable complement of accessory elements. The donor bacteria secret effectors toxic to other bacteria and to protect themselves from self-intoxication by these effectors they also encode an immunity protein adjacent to a gene encoding these effectors. These effectors are crucial for interbacterial competition for specific host niches. These effectors are single domain or multi-domain proteins and usually target bacterial cell walls (peptidoglycan) and cell membranes by phospholipases [50].

T6SS nanomachine is basically made of three major components namely the membrane complex (which is located in the inner membrane and its components are similar to the T4SS), the tail complex (which is evolutionarily related to the Phage's contractile tails), and the baseplate complex (which serves as a platform, used during assembly of the tube structure and activates the contraction of the sheath) (**Figure 1.12**). T6SS play role in interbacterial antagonism, cellular signaling between isogenic cells, and virulence [51].
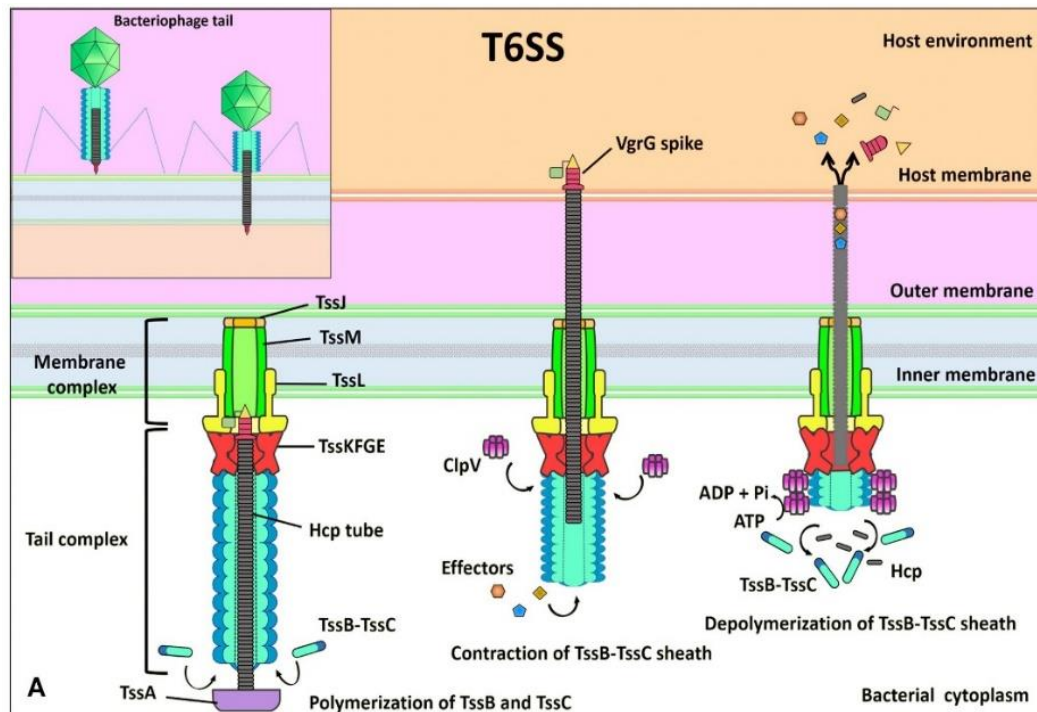
Figure 1.12: (A) The T6SS is composed of a membrane complex, a baseplate, and a tail complex. TssJ–TssL–TssM makes the membrane complex and is connected to the TssB–TssC tail sheath and the hemolysin co-regulated protein (Hcp) inner tube through the baseplate (composed of TssK, TssE, and VgrG). Effectors are recruited to the spike–tube complex through the extension domains of VgrG and/or PAAR-repeat proteins and through incorporation into the Hcp tube. An unknown extracellular signal triggers sheath contraction, which leads to the ejection of the spike–tube complex across the target membrane, thereby delivering effector proteins into the cell. The ATPase ClpV disassembles the contracted TssB–TssC sheath, which enables a new T6SS complex to be reassembled from the released subunits. Adapted from Russell, A. B., Peterson, S. B., & Mougous, J. D. (2014). Type VI secretion system effectors: poisons with a purpose. *Nature reviews microbiology*, *12*(2), 137-148.

1.5. Nomenclature of *Yersinia enterocolitica*:

Taxonomic Hierarchy

Kingdom	Bacteria
Subkingdom	Negibacteria
Phylum	Proteobacteria
Class	Gammaproteobacteria
Order	Enterobacteriales
Family	Enterobacteriaceae
Genus	Yersinia
Species	Yersinia enterocolitica
Subspecies	Yersinia enterocolitica enterocolitica

Reference: Retrieved [05/23/2022], from the Integrated Taxonomic Information System (ITIS) online database, www.itis.gov, CC0 (https://doi.org/10.5066/F7KH0KBK)

1.6.Overview of the Thesis:

The preceding introduction gives a background on different secretion systems present in *Yersinia* species and their role in virulence. The present study focuses on the regulation of the Type Three Secretion system (T3SS) in *Yersinia enterocolitica*. The T3SS system in *Y. enterocolitica* and other pathogenic bacteria is energized by both, Proton Motive Force (PMF) and by a highly conserved AAA+-ATPase present at the base of T3SS in bacterial cytosol. The highly dynamic nature of the cytosolic C-ring complex has resulted in a lack of understanding of how it regulates the T3SS. These ATPases also interact with the effector chaperone complex in the cytosol and are supposed to mediate the unfolding and translocation of the effector through the T3SS injectisome complex. In the present study, YsaN (a putative T3SS ATPase of *Y. enterocolitica*) and YsaL (a negative regulator of YsaN) were chosen and their structural, biochemical, and biophysical characterization was conducted unveiling the exact mechanism of activation and regulation of *Y. enterocolitica* T3SS by YsaN. A brief work on purification and crystallization of YopE-SycE (effector-chaperone system) was also conducted.

Chapter 1 (Review of Literature section) of this thesis critically analyses the literature present to date regarding the assembly and mechanism of function of the T3SS. It also describes different components of the T3SS injectisome and the role of each component in T3SS regulation. Also, a brief description of effector-substrate system in *Yersinia* species is given.

Chapter 2 (General methodology section) represents the general materials and methods used in this study.

Chapter 3 represents the findings based on actual purification and *in-vitro* enzymatic characterization of YsaN (a T3SS ATPase of *Y. enterocolitica*) and various deletion constructs of YsaN. Various biophysical methods like enzyme kinetics, DLS, and negative TEM were used with an emphasis on characterizing YsaN oligomerization and activation.

Chapter 4 covers the study involving the purification and characterization of YopE and SycE, a T3SS effector-chaperone protein of *Y. enterocolitica*. Also, crystallization of SycE (a cognate chaperone of effector YopE) was done and its structure was solved to 2Å resolution using the X-Ray diffraction method.

Chapter 5 the conclusion section at the end discusses the advancement in the knowledge about the activation and regulation of T3SS ATPases to date. It also discusses the difficulties faced during the conduct of this study and its shortcomings and the future perspective of this study has also been discussed at the end of this section.

Chapter 2

2. Review of Literature
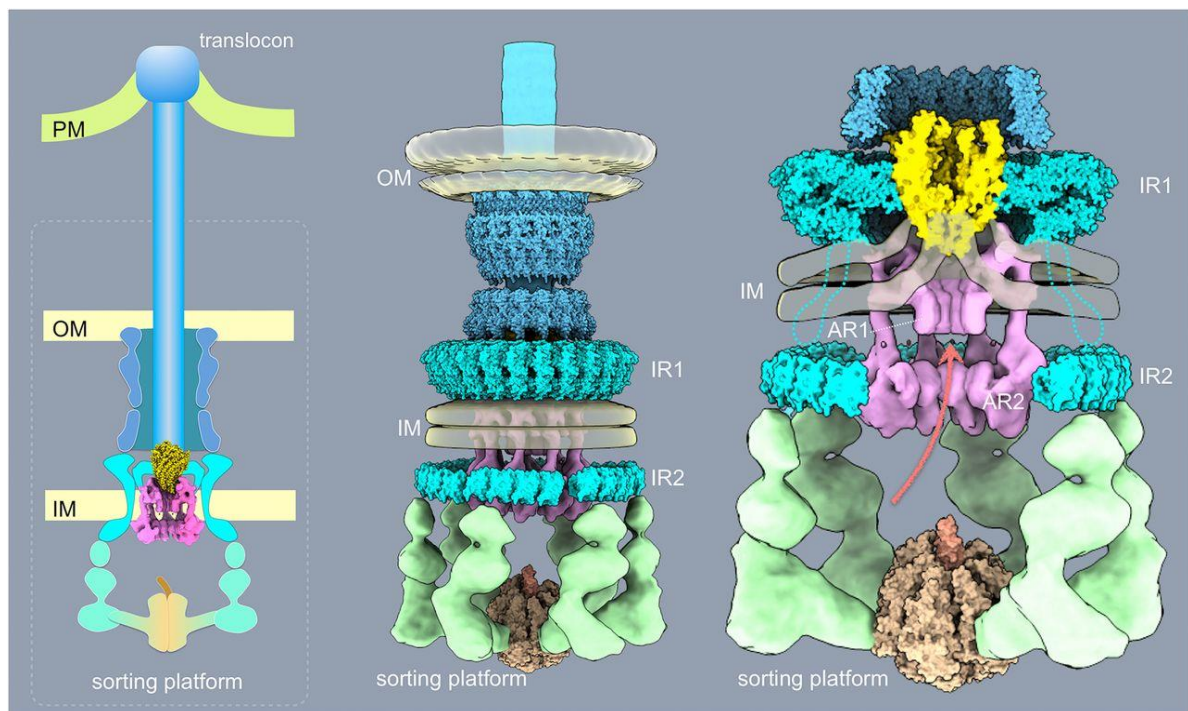
Schematic model of the export apparatus in the context of a fully assembled injectisome structure and the different bacterial envelope elements. AR1 and AR2, membrane proximal (AR1) and distal (AR2) cytoplasmic rings of InvA; IM, bacterial inner membrane; IR1 and IR2, inner rings 1 and 2 of the T3SS needle complex; OM, bacterial outer membrane; PM, target host cell plasma membrane. *Adapted from Butan, C., Lara-Tejero, M., Li, W., Liu, J., & Galán, J. E. (2019). High-resolution view of the type III secretion export apparatus in situ reveals membrane remodeling and a secretion pathway. Proceedings of the National Academy of Sciences, 116(49), 24786-24795.*

Review of Literature

A number of pathogenic Gram-negative bacteria utilize a membrane-embedded complex nanomachine popularly known as T3SS for the secretion of effectors directly into the host cell cytoplasm [32, 52]. Pathogenic bacteria like enteropathogenic *E. coli*, *Shigella*, *Salmonella,* and *Yersinia enterocolitica* have been shown to cause myriad diseases utilizing T3SS. Blocking or mutating any of T3SS components results in reduced or loss of virulence among these bacteria [52]. This membrane-localized supramolecular nanomachine of approximately 6MDa is supposed to be evolved from self-assembling flagellar T3SS [53]. Although the flagellar needle is structurally different from the T3SS needle, the cytoplasmic component of both systems is structurally and functionally similar [54].

T3SS has been discovered gradually over a period of over 30 years and was observed in the case of *Yersinia* growth defects under low calcium response, linked to virulence and secretion of effectors in the supernatant [55]. T3SS was first discovered as a novel secretion system in *Yersinia* by Michiels et al. 1990 [56]. After the discovery of a similar secretion system in other bacteria Salmond and Reeves named this system '**Type three secretion system**' in 1993. The naming of similar T3SS component proteins across different bacteria was done on an individual species basis. Now over time and currently, a unified nomenclature system has been proposed to reduce the confusion.

2.1.Introduction to the unified T3SS injectisome nomenclature system

The first attempt for unified nomenclature for T3SS injectisome proteins was suggested by Christoph Hueck in 1998 [57] who coined **'Sct'** (**Secretion and cellular translocation**) names for T3SS proteins. In the year 2016, at a meeting on T3SS in Tübingen, Germany voting was done for unifying the nomenclature of T3SS homolog proteins, and since then this nomenclature system has been used by many authors in different reviews on T3SS. **Table 2.1** represents the Sct names of different T3SS proteins and species-specific names of all the T3SS proteins [55].

2.2.T3SS biogenesis and assembly:

The T3SS is a complex structure and is made up of approximately 20 different proteins. Despite the similarity in the structural organization of T3SS injectisome among different bacterial species, the genes coding different T3SS components are organized differently in operons (**Figure 2.1**) [58]. T3SS assembly occurs in three different steps starting from the (1) Assembly of the basal body, (2) assembly of the needle complex proteins, and (3) the transport and positioning of the translocon forming proteins at the tip of the needle involved in host cell membrane integration [58]. The expression of the appropriate T3SS gene is achieved by multiple control elements present in different pathogens. Most of these elements ultimately involve a central transcriptional activator system that belongs to the AraC/XylS family of regulators. T3SS gene expression depend on the lifestyle of the pathogen and different transcriptional, post-transcriptional, and post-translational regulations.

Table 2.1: Translation table for T3SS components

Functional name	Flagellar homolog	Sct name	Ysc *Yersinia*	Inv-Mxi-Spa *Shigella*	*Salmonella* SPI-1	Ssa-Esc *Salmonella* SPI-2	*Escherichia coli*	Hrp-Hrc 1 *Pseudomonas syringae*	Hrp-Hrc 2 *Ralstonia solanacearum*
Secretin	–	SctC	YscC	MxiD	InvG	SsaC	EscC	HrcC	HrcC
Outer–inner membrane ring protein	–	SctD	YscD	MxiG	PrgH	SsaD	EscD	HrpQ	HrpW
Inner–inner membrane ring protein	FliF	SctJ	YscJ	MxiJ	PrgK	SsaJ	EscJ	HrcJ	HrcJ
Core export apparatus protein R	FliP	SctR	YscR	Spa24 (SpaP)	SpaP	SsaR	EscR	HrcR	HrcR
Core export apparatus protein S	FliQ	SctS	YscS	Spa9 (SpaQ)	SpaQ	SsaS	EscS	HrcS	HrcS
Core export apparatus protein T	FliR	SctT	YscT	Spa29 (SpaR)	SpaR	SsaT	EscT	HrcT	HrcT
Core export apparatus protein U, switch protein	FlhB	SctU	YscU	Spa40 (SpaS)	SpaS	SsaU	EscU	HrcU	HrcU
Major export apparatus protein	FlhA	SctV	YscV	MxiA	InvA	SsaV	EscV	HrcV	HrcV
Base-pod connector	–	SctK	YscK	MxiK	OrgA	SsaX	EscK	HrpD	–
Major pod protein	FliM + FliN	SctQ	YscQ	Spa33 (SpaO)	SpaO	SsaQ	EscQ	$HrcQ_{A+B}$	HrcQ
Stator	FliH	SctL	YscL	MxiN	OrgB	SsaK	EscL	HrpE	HrpF
ATPase	FliI	SctN	YscN	Spa47 (SpaL)	InvC	SsaN	EscN	HrcN	HrcN
Stalk	FliJ	SctO	YscO	Spa13 (SpaM)	InvI	SsaO	EscO	HrpO	HrpD
Needle filament protein	–	SctF	YscF	MxiH	PrgI	SsaG	EscF	HrpA	HrpY
Inner rod protein/needle adapter	FliE	SctI	YscI	MxiI	PrgJ	SsaI	EscI	HrpB	HrpJ
Needle length regulator	FliK	SctP	YscP	Spa32 (SpaN)	InvJ	SsaP	EscP	HrpP	HpaP
Hydrophilic translocator, needle tip protein	–	SctA	LcrV	IpaD	SipD	SseB	–	–	–
Hydrophobic translocator, pore protein	–	SctE	YopB	IpaB	SipB	SseC	EspD	HrpK	PopF1, PopF2
Hydrophobic translocator, pore protein	–	SctB	YopD	IpaC	SipC	SseD	EspB	–	–
Pilotin	–	SctG	YscW	MxiM	InvH	–	–	–	–
Gatekeeper	–	SctW	YopN	MxiC	InvE	SsaL	SepL	HrpJ	HpaA

(Reference: Taken from Wagner, S., et. al (2020). Bacterial type III protein secretion systems (Vol. 427). Cham, Switzerland: Springer.)

Formation of T3SS needle complex involves many factors like environmental signals which act through **Two-component system.** One important post-transcriptional factor controlling activation of T3SS gene in *Yersinia* is the temperature sensing mechanism by cellular RNA (*lcrF/virF* mRNA), also known as **RNA thermometer sensing**. The building of a T3SS injectisome is a result of expression of over 300 single copies of more than 15 different highly conserved proteins. It can be divided into three basic parts namely the **basal body complex** spanning the inner and outer bacterial membrane, **the extracellular needle component,** a hollow nano tube-like structure projecting away from the bacterial outer membrane and remains associated with the basal body complex, and the **cytoplasmic complex** associated with the base of T3SS injectisome in the bacterial cytoplasm. A brief detail of each T3SS component is given below.
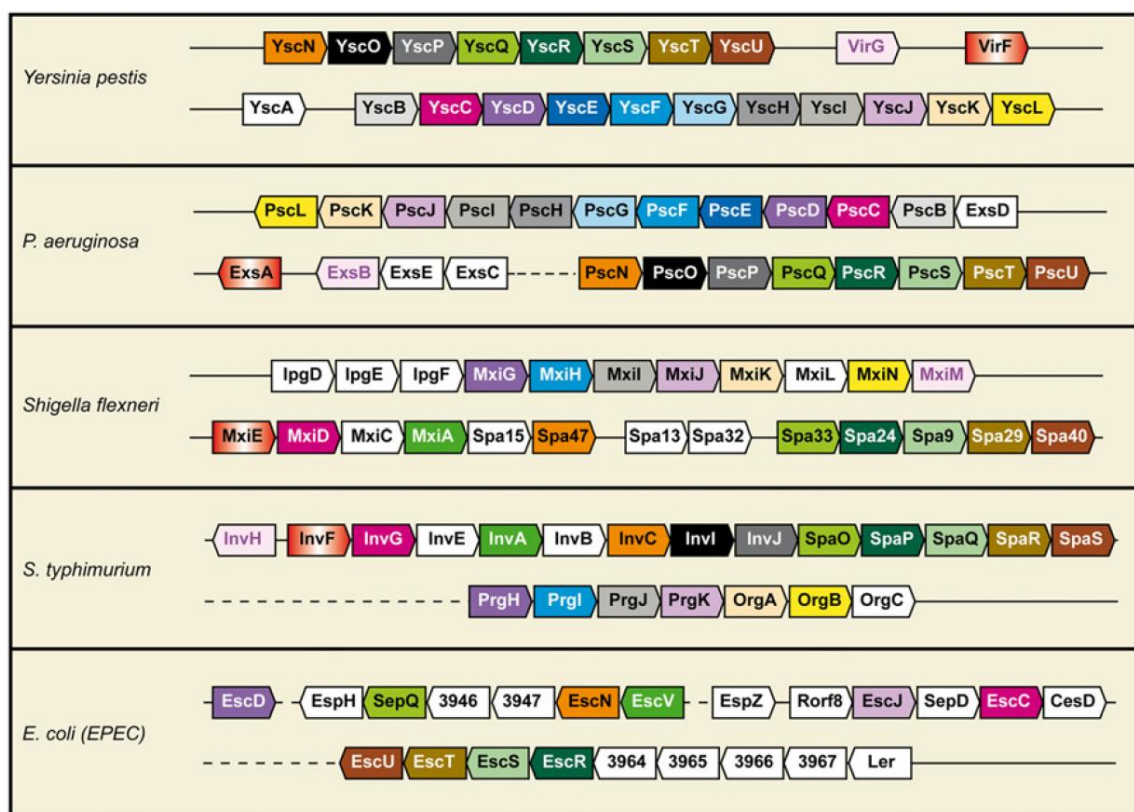
Figure 2.1: Organization of T3SS Operons in different bacterial species only operons containing transcriptional regulators, basal body, and needle formation genes are shown. Strains include *Yersinia pestis* CO29 (NC_003143), *P. aeruginosa* PAO1 (NC_002516), *S. flexneri* 301 (NC_004851), *Salmonella enterica* serovar Typhimurium LT2 (NC_003197), and E. coli (enteropathogenic) O127:H6 E2348/69 (NC_011601). Homologous genes are highlighted in the same color. Adapted from Izore, T. *et al,* (2011). Biogenesis, regulation, and targeting of the type III secretion system. *Structure*, *19*(5), 603-612.).

2.2.1. The basal body:

The basal body of the injectisome complex, consists of two co-axial homomeric protein complex rings and the core of the basal body complex is made up of a few proteins. The basal

body of T3SS is composed of multiple copies of SctC (where 12-15 SctC forms the outer membrane ring and belongs to the secretin family), SctD, and SctJ (which forms the inner membrane ring). An **export apparatus** remains associated within the center of the basal body inner ring complex and is formed by five different proteins namely SctV, SctR, SctS, SctT, and SctU [59-63]. These export complex proteins are involved in needle complex assembly. It has been also shown that SctR and SctT can assemble independently forming stable complexes [63]. Assembly of SctR, SctS, SctT, SctU and SctV are found in a 5:4:1:1:9 stoichiometry forming the needle complex core [64]. Previous TEM studies suggest that SctR forms a pentameric ring similar to a doughnut shape in association with SctT. The presence of a 15 Å wide pore suggests that this might be the entrance to the secretion channel [65]. These proteins also interact with the needle adaptor protein SctI [66].

2.2.2. The extracellular needle component:

This extracellular segment of the T3SS comprises the needle filament and a translocator pore. The needle is a hollow tube made up of multiple copies of filamentous protein SctF. The outer diameter of the needle tube is ~8 nm whereas the inner diameter of the lumen is ~1.5 nm [67]. The filament is ~50 nm long, formed by the helical assembly of needle protomers. The needle proteins are highly conserved in their primary sequence. T3SS assembly is a highly coordinated process, starting with the self-organization of the basal body complex. The whole event can be classified in early, middle, and late secretion processes. The process of needle filament assembly is considered an early substrate. The formation of the needle adaptor (SctI) inside the basal body acts as a template on which the needle extends [68]. Initiation and efficiency of the secretion process depend on the length of the needle [69]. There are two hypotheses suggesting needle length regulation. One is the **'Molecular ruler'** [70]. In Yersinia spp. a protein SctP acts as a needle length regulator by probing the degree of elongation of the needle. Once the needle grows to a suitable length SctP senses it and triggers substrate switching. Whereas, according to another study on Salmonella by Galan *et al.,*[71] suggest an alternative hypothesis where **stepwise assembly of needle protein** leads a conformational change in the needle adapter assembly at the level of the base which triggers substrate switching. SctP is required for proper assembly of needle adapter [68, 71] and is indirectly involved in needle length regulation. In some bacteria, the secretion of the T3SS needle components involve dedicated chaperones which regulate premature filament formation inside bacterial cytosol [72-75].

Synthesis of the needle of correct length induces substrate specificity switch which results in secretion of hydrophilic needle tip protein (SctA- a hydrophilic translocator protein), which eventually attaches to the distal end of the needle [76-78]. This **hydrophilic needle tip complex** is usually present in pentameric oligomer and acts as a scaffold for the pore formation in the host membrane. Tip complex formation also halts the needle elongation [79]. Host cell sensing results in the breaking of the pause after the formation of the hydrophilic tip complex. This pause in secretion involves a protein plug (SctW- gatekeeper protein) at the cytosolic face of the injectisome. Then, substrate switching takes place resulting in the secretion of another two **hydrophobic translocator proteins** (SctE and SctB) which are supposed to form a pore in the host membrane [40, 80-84]. Recently it has been suggested that host cell contact triggers the secretion of effectors by breaking the interaction of SciI (inner rod protein) and SctW (gatekeeper protein) [85].

2.2.3. The cytosolic complex:

The T3SS secretion process is a highly coordinated and regulated process and secretion occurs in a hierarchical process. The whole secretion process is usually classified into the secretion of the early substrates (composed of structural proteins required for needle complex-forming assembly), middle substrates (composed of needle tip and translocon forming proteins), and late substrates (composed of all the repertoire of effectors). The hierarchical order of secretion is maintained by a cytosolic complex structure (also known as a **sorting platform**). The cytosolic complex is a high molecular weight complex structure consisting of SctO [86] (having limited sequence similarity with the flagellar C-ring apparatus protein) in complex with the soluble cytosolic proteins SctK, SctL and SctN (ATPase) constituting a sorting platform. SctO (supposed as a positive regulator of ATPase) is present in association with the cytosolic complex. SctQ is present as 24 copies per sorting platform whereas, SctK (~6 copies), SctL (~ 12 copies), and SctN (~ 6 copies) are their stoichiometry relative to SctQ (**Figure 2.2**).

The dynamic association of the cytosolic complex with the needle complex results in the lack of its *in-situ* structural analysis. However, advancements in the Cryo-EM technique coupled with the exploitation of bacterial mini-cells have resulted in high-resolution *in-situ* visualization of the whole T3SS complex (see **Figure 2.3** below) including the associated sorting complex [87]. SctK is present in close association with the inner membrane ring (IR2) of the needle complex base. SctQ forms a major part of the C-ring complex where SctL remains associated with SctQ at one end and with the hexameric ATPase ring at the central hub. SctL acts as spokes of the wheel by interacting with the SctQ and SctN ATPase. The C-terminal side of the ATPase ring faces towards the membrane side (facing the C-terminus of the export apparatus component SctV) and which has been also predicted as a substrate docking site before its secretion [88, 89]. SctO is present between SctN and SctV suggesting its role in T3SS regulation.
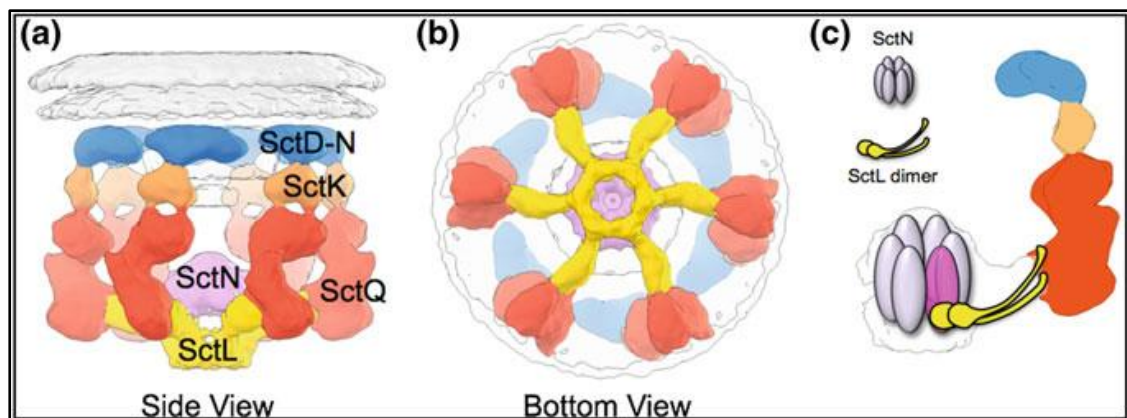
Figure 2.2: 3D representation of the cage-like structure of the type III secretion system sorting platform. A side view **(a)** and bottom view **(b)** of the sorting platform. **(c)** Schematic representation of one of the six pods that integrate the sorting platform. Adapted from Book Bacterial Type III Protein Secretion Systems 2020, Volume 427 ISBN: 978-3-030-52122-6 Andreas Diepold.

Assembly of sorting complex requires association with the needle component protein SctD [87]. However, during *in-vivo* imaging studies it has been found that a significant proportion of pre-assembled sorting complex is present in bacterial cytoplasm. This suggest that pre-assembled complex formed in cytosol is engaged with the T3SS needle complex for its

function. Also, *in-vivo* experiments in live bacteria suggest that the sorting complex remains in association and dissociation cycle with the needle complex [90, 91]. The exact mechanism of secretion and engagement of substrates by sorting complex is still debatable. The actual component within the cytosolic complex which acts as a receptor of different T3SS substrates prior to secretion is still unknown. However, many studies strongly speculate SctN ATPase is the most probable candidate as receptor for substrates. In previous studies, it has been shown that the ATPase (SctN) physically interacts with the chaperone-substrate complex and releases the substrate in the medium by unfolding them using ATP as an energy source [92]. The presence of hexameric ATPase has been observed during high-resolution *in-situ* cryo-EM studies. The C-terminal face of the ATPase is involved in chaperone-substrate docking, where the C-terminal face of the ATPase is oriented toward the membrane side [89]. The docked substrate is unfolded by the ATPase and is secreted through the channel formed by SctV and other components of the export apparatus. More experiments are required to further establish the exact mechanism of T3SS secretion.
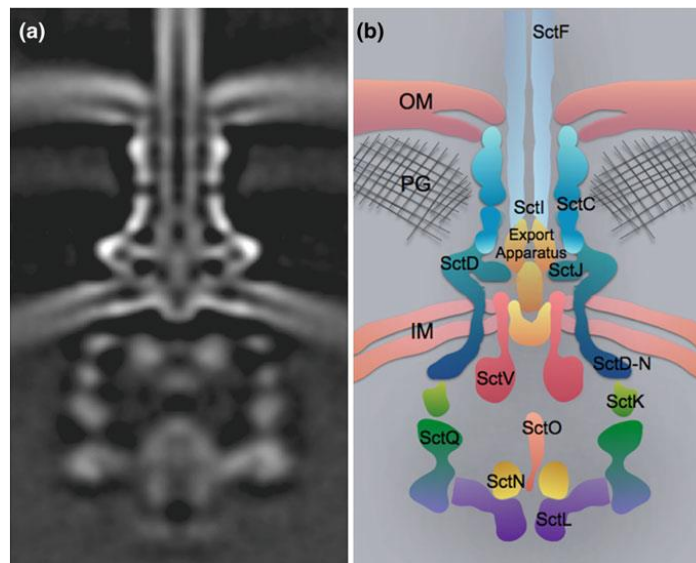
Figure 2.3: Molecular architecture of the sorting platform in the intact injectisome a. A central section of the sub tomogram average of the Salmonella injectisome as observed by cryo-ET. **b** Schematic representation of the type III secretion system with the different components labeled. Adapted from Book Bacterial Type III Protein Secretion Systems 2020, Volume 427 ISBN: 978-3-030-52122-6 Andreas Diepold.

2.3. The effector-chaperone system in *Yersinia*:

Yersinia harbors different secretion systems for effector translocation out of which majority of them are secreted by the T3SS. Pathogenic *Yersinia* uses a virulent plasmid encoded T3SS for secretion of Yops effector. Besides this, it also harbors chromosome encoded Ysa T3SS for Ysps

effector secretion (see **Figure 2.4**). The Yop secretion require specific cytosolic chaperones belonging to a family known as Syc proteins (**specific Yop chaperones**) [93, 94]. All Yops have their specific binding partner Syc chaperone. Syc proteins are encoded by a gene located beside their specific substrate with which they bind. It has also been suggested that Syc chaperones bind to their specific substrate and keep them in secretion competent state [95]. The Syc chaperones have only ~14% sequence similarity however, they are remarkably similar in their structure suggesting a common mechanism of action [96-101].

2.3.1. Yop effectors in *Yersinia*:

Yersinia spp. secretes different types of **Yersinia Outer Proteins** (Yop proteins) using a T3SS, encoded by a 70-kb virulence plasmid pYV. These Yops are encoded by the same virulent plasmid [102]. Pathogenic Yersinia secretes commonly six different types of Yop effectors (YopE, YopH, YopO, YopM YopJ/P and YopT) (**Table 2.2**). These effectors target different host immune response and help the pathogen to survive [103]. These toxic Yop effectors are capable of modifying the eukaryotic host cellular processes. For example, YopE which bears a GTPase-activating domain and downregulates the Rho activity which results in actin depolymerization and helps *yersinia* to overcome phagocytosis by macrophages [104-107]. The secretion of these Yops is supposed to be induced and triggered by host cell contact. However, low Ca^{2+} ions (known as a **low Ca^{2+} response**) in the medium also trigger Yop secretion in the culture medium. It has been also shown that the YopE protein is secreted by the Ysa T3SS apparatus under low temperature and high salt (LTHS) condition [108]. An N-terminal signal sequence is required for Yop secretion [109, 110]. In the case of YopE, the first 15 amino acids act as a signal sequence. Also, its efficient secretion requires a specific cytoplasmic chaperone SycE [93, 94].

2.3.2. Ysp effectors in *Y. enterocolitica*:

Besides Ysc T3SS *Y. enterocolitica* 8081 species utilize additional Ysa T3SS to deliver their virulence factors inside the host cell encoded by Ysa pathogenicity island (Ysa-PI) located on the chromosome. Effectors secreted by Ysa T3SS is required for gastrointestinal phase of infection. These effectors are known as Ysps (Yersinia secretory proteins). Pathogenic *Y. enterocolitica* secretes eight known Ysp effectors (YspA, YspE, YspF, YspI, YspK, YspL, YspM, and YspP) (**Table 2.2**) through Ysa T3SS. Few of the Ysa effectors (YopE, YopN and YopJ/P) are also secreted through Ysa T3SS (see Table 2.2) [108, 111]. The genes encoding Ysa T3SS effectors are widely dispersed throughout the chromosome [112]. The lack of *ysp* gene colocalization suggest that *Y. enterocolitica* biovar 1B has undergone multiple horizontal gene transfer event under selective pressure to maintain the Ysa T3SS. Ysp effectors are predicted form their conserved domains and regions having particular function. Many of Ysp effectors are predicted but their functions are still not clear [103].
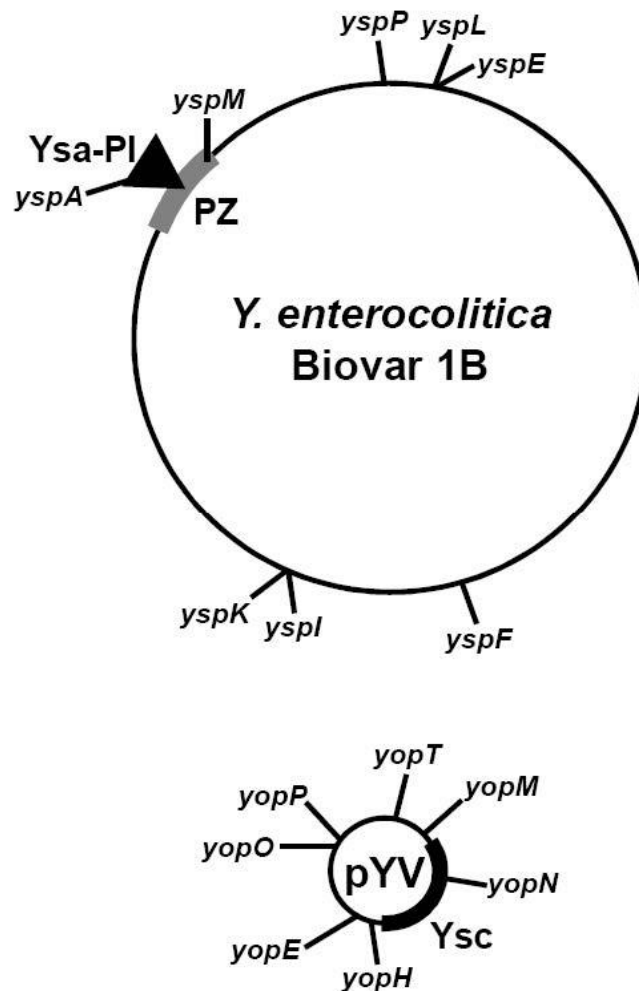
Figure 2.4: Genes of *Y. enterocolitica* Biovar 1B that encode effectors translocated into host cells by the Ysa T3SS are dispersed in the genome. Depicted is the chromosome with labels indicating the relative locations of effector-encoding ysp genes and the location of the YSA-PI encoding the Ysa T3SS. The Ysa-PI is within a region of the chromosome rich with other genes implicated in virulence called the plasticity zone (PZ). Plasmid pYV carries genes encoding the Ysc T3SS (Ysc) and the effector Yops. A similar plasmid is present in Y. pestis and Y. pseudotuberculosis (Reference: adapted from Matsumoto, H., *et al*. (2009). Translocated effectors of Yersinia. *Current opinion in microbiology*, *12*(1), 94-100).

Table 2.2: Virulence effector proteins of pathogenic *Yersinia*

Effectors	T3SS	Gene Locations	Biochemical Functions/Characteristics	Cellular Targets	Cellular Effects
Pathogenic *Y. pestis, Y. pseudotuberculosis,* and *Y. enterocolitica*					
YopE	Ysc/Ysa	Plasmid	Rho GAP mimicry	RhoA, Rac1, Cdc42	Disruption of the actin cytoskeleton, Inhibition of phagocytosis
YopH	Ysc	Plasmid	Protein tyrosine phosphatase	Focal adhesion complexes (p130Cas, FAK, paxillin, Fyb, SKAP-HOM, Crk, Pyk2) Other proteins LAT, SLP-76, Lck	Disruption of the actin cytoskeleton, Disruption of phagocytosis, Inhibition of chemoattractant, protein (MCP1) production, (P13K/Akt signaling), Suppression of adaptive immune response (cytokine IL-2)
YpkA/YopO	Ysc	Plasmid	Serine/threonine kinase RhoGDI mimicry Actin binding	Gαq, RhoA, Rac1 Actin (kinase activity activation)	Disruption of the actin cytoskeleton, Disruption of phagocytosis, Inhibition of Gαq signaling.
YopM	Ysc	Plasmid	Localization to nucleus Twelve to twenty tandem leucine-rich repeats (~20 amino-acid)	Rsk1 and Prk2?	Inhibition of cytokine production (IL-15 and IL-15Rα), Depletion of NK cells.
YopJ/P	Ysc/Ysa	Plasmid	Acetyltransferase	MAPKK and IKK family member proteins	Inhibition of MAPK and NF-κB signaling pathways. Suppression of proinflammatory cytokine and chemokine production (TNF-α, IL-8, IL-12, IL-18, etc.). Induction of apoptosis
YopT	Ysc	Plasmid	Cysteine protease	RhoA, Rac1, Cdc42	Disruption of the actin cytoskeleton Inhibition of phagocytosis
Highly Pathogenicity *Y. enterocolitica* Biovar 1B					
YspA	Ysa	Chromosome	?	?	?
YspL	Ysa	Chromosome	Two separate 15 or 17 amino-acid repeats	?	?

Table 2.2- continued…

Effectors	T3SS	Gene Locations	Biochemical Functions/Characteristics	Cellular Targets	Cellular Effects
YspP	Ysa	Chromosome	Protein tyrosine phosphatase	?	?
YspF	Ysa	Chromosome	?	?	?
YspE	Ysa	Chromosome	ADP-ribosyltransferase?	?	?
YspI	Ysa	Chromosome	Ca2+ binding protein Three tandem 12 amino-acid repeats	?	?
YspK	Ysa	Chromosome	Serine/threonine kinase	E2 ubiquitin-conjugating enzymes	Inhibition of NF-κB signaling pathway
YspM	Ysa	Chromosome	GDSL lipase motif	?	Growth inhibition of *Saccharomyces cerevisiae*

Reference: Taken from Matsumoto, H., *et. al* (2009). Translocated effectors of Yersinia. *Current opinion in microbiology, 12*(1), 94-100.

2.3.3. SycE- Specific chaperone for YopE:

The Syc (Specific Yop chaperones) chaperones are required for efficient secretion of Yops and are small acidic proteins having leucine repeats in its C-terminal moiety. SycE, is the specific chaperone for YopE and binds with the 15 to 50 residues (chaperone binding domain) of YopE. This YopE-SycE interaction is necessary to prevent degradation of YopE within the cytoplasm [113, 114]. The chaperone binding domain is not required for its secretion. Secretion of YopE also takes place in the absence of its chaperone binding suggesting that SycE is dispensable for YopE secretion [115]. However, it has also been suggested that the chaperone binding region acts as second signal sequence which is required for targeting YopE into the host cytoplasm [106, 116]. YopE (~23kDa protein), binds to homo-dimer of SycE(~14.6kDa) in an extended fashion through its chaperone binding domain (Cb domain). This interaction results in secondary structure formation in YopE Cb domain which is unstructured or absent in YopE alone. This YopE interaction with SycE dimer is also suggested to act as three-dimensional translocation signal which is recognized by the sorting complex [117].

2.4. T3SS-present model mechanism:

The T3SS is able to secrete greater diversity of proteins inside a host cell in a contact-dependent manner [118]. The secreted effector proteins serve the purpose of suppressing the host immune response to protect the bacteria by interfering with different host cell processes. They have been reported to prevent phagocytosis (as in the case of *yersinia*), the killing of the macrophages (as in the case of *Salmonella*), and acquisition of nutrients (as in the case of *Chlamydia* and *Salmonella*) [119]. Transport of effectors through T3SS is a highly regulated process. Several models have been proposed regarding how effectors are being secreted. Energy for secretion is provided by both the Proton Motive Force (PMF) and from the ATPase ring complex associated with the injectisome. The whole process of T3SS secretion can be divided in the following steps. (1) Substrate recognition by the cytoplasmic sorting complex. (2) Translocation of the substrate across the bacterial inner membrane; and, (3) Translocation of substrate further beyond the bacterial inner membrane. It has been suggested that energy for substrate translocation through inner membrane is mainly provided by the PMF [120, 121] while, the associated ATPase complex is required for substrate recruitment and unfolding prior to translocation through the injectisome pore [16, 38]. Once the contact with the host has been established, the secretion process begins. The secreted effectors contain an N-terminal secretion signal and are often present in bacterial cytosol in complex with their cognate chaperone for their targeting to the needle complex for secretion. Secretion of all the effectors is done in an unfolded state through the translocon pore of the T3SS needle in one step secretion mechanism. Energy for this process is provided by a highly conserved ATPase present at the base of the export apparatus in the cytosol and also from the PMF [122]. The ATPase is present as a hexamer ring complex and remains associated with the needle complex with the help of SctL as spokes. The C-terminal face of the ATPase complex lies towards the membrane side. The substrate docks to the C-terminal side of the ATPase ring complex prior to secretion. Unfolding of the substrate is done by the ATPase using energy by ATP hydrolysis. The presence of the hexamer ATPase has been also observed in *in-situ* cryoEM study by Hu, B., *et al* 2017 [87]. However, the exact role of ATPase in secretion process is still not understood properly.

Chapter 3

3. Materials and Methodology

3.1.General materials:

3.1.1. List of common reagents and buffer stocks

1. Acrylamide mix 30%(w/v)
- Dissolve 145 g of Acrylamide crystals (SRL 3x crystal Cat. No-61346) and 5 g of N, N-Methylene Bisacrylamide crystals (SRL 3X crystal Cat. No- 38516) in ddH$_2$O.
- Adjust the volume to 500ml

2. Ammonium sulfate 90%(w/v)
- Dissolve 90 g of Ammonium sulfate crystals (Sigma Cat. No- A4418) in 100 ml of ddH$_2$O.
- Heat to dissolve the crystals completely.
- Store at room temperature.

3. APS 20% (w/v)
- Dissolve 0.3 g of Ammonium Persulphate (SRL Cat. No- 65553) in 1.5ml of ddH$_2$O.

4. DNA gel loading dye (5X)- 10ml
- Add 25mg bromophenol blue (Sigma Cat. No- B5525) in 6.7ml of MilliQ.
- Add 25mg Xylene Cynol FF (Sigma Cat. No- X4126) and mix.
- Add to this 3.3ml of glycerol.
- Store at -20⁰C in small aliquots.

5. DTT (0.5 mM)
- Dissolve 0.7 g of 1,4-dithiothreitol crystals in 10 ml of ddH$_2$O.
- Store in small aliquot at -20⁰C.

6. EDTA pH8.0 0.5M
- Dissolve 186 g of EDTA (SRL Cat. No- 40088) crystals in 800ml of ddH$_2$O.
- Adjust the pH using Conc. HCl.
- Adjust the volume to 1liter.

7. EtBr (10mg/ml stock)
- Ethidium bromide solution was purchased from Sigma (Cat. No- E1510) as a 10mg/ml stock solution.

8. Imidazole (pH 8.0) 2.5M

- Dissolve 170 g of Imidazole (SRL Cat. No-32822) crystals in 800ml of ddH$_2$O.
- Adjust the pH using Conc. HCl.
- Adjust the volume to 1liter.

9. LB media

- Dissolve 15 g of Luria Bertani Broth, Miller powder, Himedia (Cat. No-M1245) in 600ml of ddH$_2$O.
- Autoclave to sterilize before use.

10. Tris (pH 8.0) 1M

- Dissolve 121g Trizma Base (sigma-Cat. No-93352) into 800ml of autoclaved ddH$_2$O. Adjust pH using conc. HCl.
- Adjust volume to 1 liter.
- Filter sterilize, if necessary.
- Store at 4⁰C.

11. NaCl (5M)

- Dissolve 292 g of sodium chloride (Sigma-Cat.No.- S9888) in autoclaved ddH$_2$O.
- Add ddH2O to 1 liter.

12. NaOH (10M)

- Dissolve 200 g of sodium hydroxide pellets (Merck Cat. No- 106462) in 450ml autoclaved ddH$_2$O.
- Adjust the volume to 500ml.

13. PMSF (200 mM)

- Dissolve 0.35 g phenylmethylsulphonyl fluoride crystals (Sigma Cat. No- PMSF-RO) in autoclaved ddH$_2$O.
- Make up the volume to 10 ml.

14. Protein loading dye (6X)

- Add 28ml of stacking buffer at (250mM Tris pH 6.8).
- Add 15 ml of glycerol at 40% (v/v).
- Add SDS at 10% (w/v).
- Add DTT at 100 mM.
- Add Bromophenol blue at 0.5% (w/v) maximum.
- Adjust the volume up to 50ml.

15. SDS gel running buffer (10X)

- Dissolve 144 g of Glycine crystals (SRL, Cat. No-25853) in 1 liter of ddH$_2$O.
- Add and dissolve 30 g of Tris Base (SRL Cat. No-71033).
- Add and dissolve 11 g of SDS powder (SRL Cat. No- 54468).

16. TAE (50X)

- Dissolve 242 g of Tris base (SRL Cat. No-71033) in 700 ml ddH$_2$O.
- Add 57.1 ml of Glacial Acetic acid (SRL, Cat. No-A6283).
- Adjust the volume to 1 liter.
- No need to adjust the pH of the solution.

17. SDS 10% (w/v)

- Dissolve 10 g SDS (sodium dodecyl sulfate or sodium lauryl sulfate SRL Cat. No-54468) in autoclaved ddH$_2$O to 100.
- Note that if the powder did not dissolve completely heating can be done.

18. Urea-Tris-Cl buffer (8 M)

- Dissolve 24 g of Urea crystals (SRL Cat. No- 21113) into buffer containing 50 mM Tris pH 8.0, 5% glycerol, and 100 mM NaCl.
- Adjust the volume to 50 ml using the same buffer.
- This solution should not be stored for a longer time. Make a fresh solution each time.

19. Antibiotics used

- A list of antibiotics used is given in **Table 3.1.**

S.No.	Name	Lab stock concentration	Recommended working concentration	Source
Table 3.1: List of antibiotics used.				
1	Ampicillin	200 mg/ml	100 µg/ml	Gold Biotech USA
2	Kanamycin	50 mg/ml	50 µg/ml	Gold Biotech USA
3	Chloramphenicol	34 mg/ml (dissolved in 100%EtOH)	34 µg/ml	Calbiochem

3.2.General methodology:

3.2.1.Genomic DNA isolation:

The genomic DNA isolation of *Yersinia enterocolitica* was done according to the miniprep protocol of genomic DNA isolation by Kate Wilson [123].

3.2.2.Cloning and DNA manipulation methods:

A detailed list of plasmid vectors used is provided in **Table 3.2.** Also, a list of bacterial strains used is provided in **Table 3.3.** The required plasmid vector was isolated from *E. Coli* DH5α cells by manual alkaline lysis method [124] or using Qiagen miniprep Kit following manufacturers protocol. The respective gene was PCR amplified from *Yersinia enterocolitica* genomic DNA by PCR using Phusion™ High-Fidelity DNA Polymerase (Thermo Scientific) and was inserted into the respective cloning vectors. The PCR amplified product was purified using a Qiagen PCR purification kit following the manufacturer's protocol. The PCR product was inserted into the cloning vector by Double digestion protocol using restriction enzymes from Thermo Scientific™. A list of Restriction Enzyme used is provided in **Table 3.4.** The digested products (vector and insert) were purified by Qiagen Gel Extraction protocol and ligated into the desired vector using T4 DNA ligase (Thermo Scientific). The ligated sample was transformed into *chemically competent E. coli* DH5α/ Top10 cells and spread onto an LB agar plate (with respective antibiotics). 4-to 6 colonies were selected and screened for positive clones by Double digestion and colony PCR.

Table 3.2: List of commercial vectors used.					
S.No.	Name	Promoter	Antibiotic resistance	Tag	Source
1	pET-28a (+)	T7	Kanamycin	N-terminal His-tag/thrombin/T7-Tag, an optional C-terminal His-tag sequence.	Novagen
2	pET-Duet1	T7	Ampicillin	N- Terminal His-tag (MCS 1), and C Terminal S-tag (MCS 2).	Novagen
3	pACYC-Duet1	T7	Chloramphenicol	N- Terminal His-tag (MCS 1), and C Terminal S-tag (MCS 2).	Novagen
4	pET-22b (+)	T7	Ampicillin	A *pel*B leader sequence, C-Terminal His-tag.	Novagen

Table 3.3: List of bacterial strains used.			
S.No.	Name	Details	Source
1	*Yersinia enterocolitica 8081*	A human enteropathogen.	ATCC
2	*E. coli* DH5α	They are used as cloning Host. They contain three mutations: *recA1, endA1*, and *lacZ*ΔM15.	*Invitrogen*
3	*E. coli Top10*	They are also used as Cloning Host and contain mutations: F- *mcr*Δ, Φ80*lacZ*ΔM15, Δ*lacX*74, *deo*R, *recA1, araD*139Δ, *endA1,* and *nupG.*	*Invitrogen*
4	*E. coli* BL*21*	Used as expression host. They contain mutations: *F- ompT, hsdsb,* and *(rb-mb-) gal dcm* (DE3)	Novagen

Table 3.4: List of restriction enzymes used.			
S.No.	Name	Restriction sequence	Source
1	BamH1	5' G▼GATCC 3'	Thermo Scientific
2	Bgl2	5' A▼GATCT 3'	Thermo Scientific
3	Hind3	5' A▼AGCTT 3'	Thermo Scientific
4	Nco1	5' C▼CATGG 3'	Thermo Scientific
5	Nde1	5' CA▼TATG 3'	Thermo Scientific
6	Sal1	5' G▼TCGAC 3'	Thermo Scientific
7	Xho1	5' C▼TCGAG 3'	Thermo Scientific
Note: Cut site marked by the down arrow (▼).			

3.2.3. Agarose gel electrophoresis:

To visualize DNA and plasmids agarose gel electrophoresis was used. Different gel percentage gel was used according to the size of the DNA fragment. **Table 3.5** represents the gel percentage and size of DNA separated. Approximately 0.2-0.5µg/ml of EtBr, working concentration was used and added to the gel. The gel was visualized on a Trans-UV apparatus.

Table 3.5: Agarose gel percentage Vs size of DNA table.	
Percentage of Agarose gel (w/v)	Size of DNA (kb)
0.5%	1-30 kb
0.7%	0.8-12 kb
1.0%	0.5-10 kb
1.2%	0.4- 7 kb
1.5%	0.2-3 kb
2.0%	0.05-2 kb

3.2.4. Transformation:

3.2.4.1. Competent cell preparation:
All the cells were made competent by the chemical competent cell method using Calcium chloride and Magnesium Chloride. Briefly, a 5ml primary culture of E. coli DH5α/ BL21 cell was given by picking a single colony from an agar plate and grown overnight. A 100ml of secondary culture at 37ºC was given from this primary culture till the OD (@600nm) reached 0.4. The culture was transferred to 4ºC to cool and was harvested at 6000rpm at 4ºC. The pellet was resuspended by prechilled sterile 100mM MgCl$_2$ solution and incubated at 4ºC or on ice for 15-20 minutes. Centrifugation was done pellet down the cells at 3000rpm at 4ºC. The cells were again resuspended by adding prechilled 100mM CaCl$_2$ solution and incubated at 4ºC or on ice for another 15-20 minutes. Centrifugation was done at 2500rpm for 15 minutes to pellet down the cells and the pellet was resuspended by prechilled buffer containing 85mM CaCl2 and 15% glycerol. Centrifugation was done to pellet down the cells. The cells were again resuspended in 2-4 ml of buffer containing 85mM CaCl2 and 15% glycerol. Small aliquots were made to store the competent cells and immediately stored at -80ºC.

3.2.4.2. Transformation of chemically competent cells:
Transformation of the plasmid into the chemically competent cell was done by heat shock method. Briefly approximately 1- 2µl of miniprep plasmid was added to 50µl of competent cell (thawed on ice) and incubated at ice for 10 minutes. Heat shock was given for 90 seconds at 42ºC and the sample was snap chilled again on ice for another 10 minutes. 1 ml LB (without antibiotic) was added to the sample and incubated at 37ºC for 1 hour. Then cells were pelleted down by centrifugation at 6000rpm and media was decanted. Cells were resuspended in remaining media and spread on to desired antibiotic agar plate and incubated overnight at 37ºC.

3.2.5. Protein expression and purification methods:
The positive clone plasmid was transformed into BL21 cells for recombinant expression of desired proteins. The transformed BL21 cells were tested for protein expression. BL21 cells positive for protein expression were chosen and a large-scale culture was given for large-scale protein preparation. The overnight induced (induction was done by adding 0.5-1mM IPTG working concentration) culture was harvested and stored at -80ºC or freshly purified by one of the following methods.

3.2.5.1. Induction and solubility test:
A 5ml primary culture at 37ºC was given from a single colony overnight. Another 10ml of secondary culture was given at 37ºC till OD (@600) reaches 0.4-0.6. 1ml of uninduced

culture was collected from this culture and IPTG at 1mM working concentration was added to the culture and the culture was incubated at 37°C for another 3 hours in a shaker incubator. 1ml of uninduced culture was collected and the rest was harvested at 6000rpm at 4°C. The cells were resuspended in 5ml of sonication buffer. Sonication was done using a pulse sonicator. The sonicated sample was centrifuged at 12000rpm for 30 minutes and supernatant and pellet fractions were collected. All the samples were analysed on SDS gel electrophoresis for protein expression and solubility.

3.2.5.2. Immobilized Metal Affinity Chromatography (IMAC):

Purification of proteins containing N-terminal or C-terminal His-tag was done as the first step by this method of affinity purification. The clear supernatant from a lysed-induced bacterial cell obtained after sonication and high-speed centrifugation was loaded onto pre-equilibrated Nickel beads (GE Healthcare) gravity-flow column. Incubation was done at 4°C or cold room for 30 minutes to 1 hour followed by passing the flow-through. Washing was done with wash buffer usually in 2 steps. To elute the protein elution buffer was used. A typical IMAC purification buffer table is given in **Table 3.6.**

Table 3.6: IMAC buffer chart.					
S.No.	**Buffer name**	**Tris pH 8.0**	**NaCl**	**Glycerol**	**Imidazole**
1	Equilibration buffer	50mM	100mM	5% (v/v)	5-10mM
2	Sonication buffer	50mM	100mM	5% (v/v)	5-10mM
3	Wash buffer	50mM	100mM	5% (v/v)	30-40mM
4	Elution buffer	50mM	100mM	5% (v/v)	200- 300mM

3.2.5.3. Size-exclusion Chromatography:

This step is the polishing step in the protein purification method. Proteins purified through IMAC and/or the Ion-Exchange method were subjected to this method. For large-scale protein preparation, Superdex 200 16/600 prep-garde chromatography columns and for analytical purposes Superdex 200 10/300 HR chromatography columns (GE Healthcare) were used. The column was preequilibrated with Gel-filtration buffer (20mM Tris pH 8.0, 100mM NaCl, and 2% glycerol).

3.2.6. Protein estimation:

3.2.6.1. Protein concentration estimation:

The concentration of purified proteins was estimated by the Folin Lowrey method as described in the protocol given by Waterborg, J. H. *et. al.* (1994) [125]. Also, absorbance at

280nm was done to estimate the concentration of protein samples. Briefly, 2μl of concentrated protein sample was added to 998μl of the buffer. Absorbance was read at 280nm in a spectrophotometer. Concentration was estimated by the equation given below-

$$C = \frac{A}{\epsilon l} * Df * 10^6$$

Where-

C= Concentration of protein in μM

A= Absorbance at 280nm

ε= Molar absorption coefficient of protein

l= Path length (1cm here in all cases)

Df= dilution factor (here 200 in all cases)

3.2.6.2.Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE):

To check the purity and molecular weight estimation SDS PAGE technique was used. A typical SDS gel consists of two parts, one is a stacking gel at the top and at the bottom a separating gel or resolving gel. Different gel percentage is used to separate different sizes of proteins. Various components used to prepare SDS gel are given in **Table 3.8. Table 3.9** represents the gel percentage used to separate different sizes of proteins. The SDS gel was run at a typical 16 ampere current and 100 Volt. To view the protein bands on the gel the gel is stained in staining solution and de-staining was done to remove excess stain by de-stain solution. **Table 3.10** represents a correlation chart of SDS gel percentage with the size of protein.

Table 3.8: SDS gel casting (resolving gel) buffer system.						
S.No.	**Components**	**Resolving Gel composition**				
		8%	**10%**	**12%**	**15%**	**20%**
1	**ddH₂O**	3.86 ml	3.33 ml	2.8 ml	2 ml	1.66 ml
2	**30% Polyacrylamide mix**	2.14 ml	2.67 ml	3.2 ml	4 ml	5.34 ml
3	**1.5M Tris pH 8.8 (0.4% SDS added)**	2 ml	2 ml	2 ml	2 ml	2 ml
4	**20% APS**	30μl	30μl	30μl	30μl	30μl
5	**TEMED**	18μl	18μl	18μl	18μl	18μl

Table3.9: SDS gel casting (Stacking gel) buffer system.		
S.No.	Components	Stacking Gel Composition
		5%
1	ddH$_2$O	1.112 ml
2	30% Polyacrylamide mix	200 µl
3	0.5M Tris pH 6.8 (0.4% SDS added)	187 µl
4	20% APS	6 µl
5	TEMED	6 µl

Table 3.10: SDS gel percentage vs size of protein.	
Percentage of SDS PAGE gel	Size of Protein (kDa)
20%	4-40
15%	12-45
12.5%	10-70
10%	15-100
8%	25-200

3.2.7. Glutaraldehyde Chemical Crosslinking:

A chemical crosslinking experiment was performed using glutaraldehyde (0.03- 2% (v/v) Fischer Scientific) at room temperature. Prior to glutaraldehyde treatment, all the protein samples were given buffer exchange in phosphate buffer or HEPES buffer. The reaction was terminated by adding 5X protein loading dye followed by heat denaturation by boiling the samples. All the samples were analysed by the SDS gel electrophoresis method.

Chapter 4

4. YsaN- Function and regulation in *Yersinia enterocolitica*

YsaN (T3SS ATPase) Oligomerization
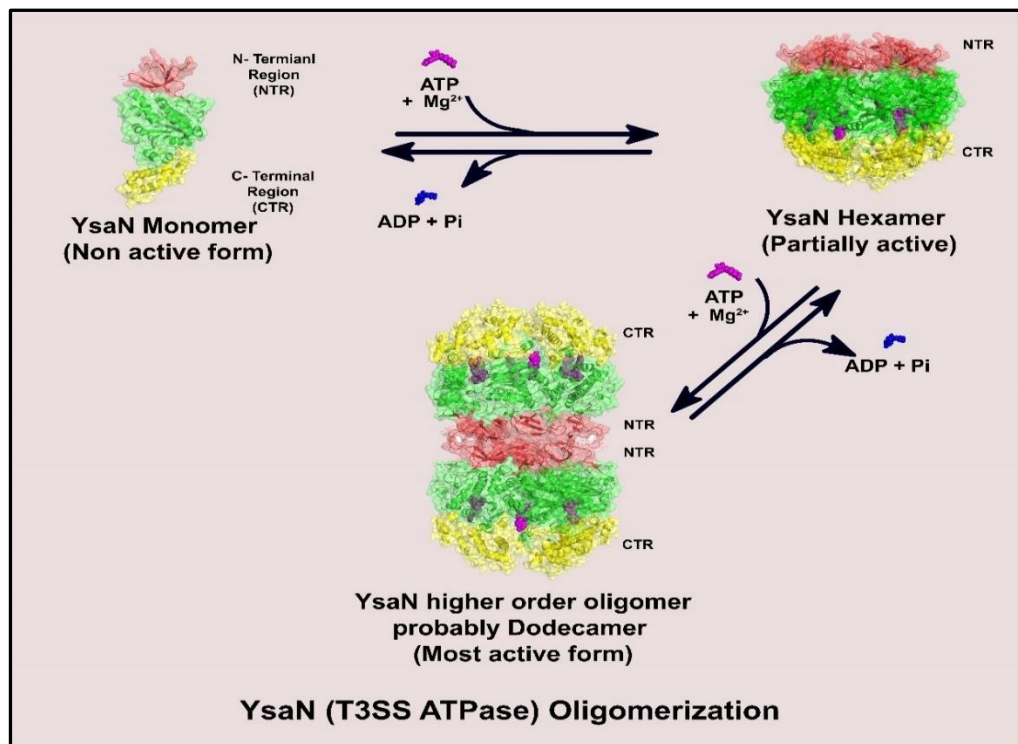
4.1. Introduction

Three species of the genus *Yersinia: Y. enterocolitica, Y. pseudotuberculosis*, and *Y. pestis* are pathogenic to rodents and humans. They primarily use *Ysc-Yop* type three secretion system (T3SS) encoded by a ~70kb virulent plasmid. Beside this pathogenic *Y. enterocolitica* strains also harbours another chromosomally encoded *Ysa-Ysp* T3SS. The *ysa* locus is absent in low virulence strains of *Y. enterocolitica* [126]. The extra virulence of *Y. enterocolitica* is due the presence of *Ysa-Ysp* T3SS and is primarily required during gastrointestinal phase of infection [127, 128]. Similar to other T3SS the *Ysa-Ysp* T3SS also requires the ATPase complex for efficient unfolding and translocation of substrates. The ATPase complex is composed of YsaN, YsaL, YsaO and YsaQ. YsaN is the highly conserved molecular ATPase and requires Mg^{2+} for its functioning. Whereas, YsaL is the negative regulator of YsaN [129]. YsaN exists in different oligomeric forms (monomer and probable dodecamer) where the higher order oligomer, probable dodecamer is the most active form[129]. Presence if higher order oligomer as the most active form and the process of oligomer formation of YsaN and homologue ATPases is poorly understood.

In the present study we focus on the formation of relevant functional YsaN oligomer and describes the mechanism of oligomer formation. We have also developed a system for soluble expression of untagged YsaN and various YsaN deletion mutants followed by their biochemical characterization. Structural visualization of the oligomer was also done using TEM. Beside that cloning and expression of untagged YsaL and co expression and purification of YsaL-YsaN-His complex was done.

4.2. Materials and methods:

4.2.1. Cloning, expression and purification:

Different YsaN construct used in the study is given in **Table 4.1** and represented in **Figure 4.1**. Classical cloning method was used for cloning all the constructs and are mentioned in material and method section. For untagged, soluble expression of YsaN and YsaN deletion constructs, pET22bΔCPD vector was used (represented as CPD vector from now onwards). YsaN was PCR amplified from *Y. enterocolitica* 8081 genomic DNA. Primers used are given in **Table 4.2.** Cloning of YsaN in CPD vector resulted in soluble expression of YsaN deletion constructs. BL21 cells were used for recombinant expression of proteins. Induction was done using IPTG (at 0.5mM) after the OD (at 600nm) reaches 0.6 for approximately 12-14hrs at 25⁰C. Cells were harvested and resuspended in sonication buffer and PMSF was added to the solution at 1mM concentration just prior to cell lysis. Lysed solution was given high speed in ultracentrifuge at 13000rpm at 4⁰C for ~1hr. The clear supernatant was loaded onto the gravity flow column containing preequilibrated Ni2+ beads. Protein purification was done according to CPD purification protocol [130] with slight modification. The sample was incubated for ~1hr at 4⁰C (or in a cold room). Unbound fraction was discarded followed by adding 2 column volume (CV)

wash buffer. Then 6-8ml of CPD buffer (50mM Tris pH 8.0, 100mM NaCl, 5% glycerol, and 100µM of Inositol-6-phosphate (Ins-6-P/ Phytic Acid-Sodium salt) was added to the column and mixed gently followed by incubation for ~1hr at 4⁰C. The Flow through containing untagged YsaN of protein of interest was collected in a fresh tube. This sample was further injected immediately into the size exclusion chromatography system (AKTA Prime) to separate leached CPD domain. The incubation time needs to be optimized for different constructs. In our case incubation time, more than ~1hr was strictly avoided to stop degradation of target protein by CPD itself. Further immediate size exclusion chromatography was necessary to separate the leached CPD domain. The respective peak containing the target protein was collected and was analysed by SDS gel. This pure sample was further used in all downstream processes. Fresh culture was done and protein was purified freshly every time for each experiment.

Table 4.1: List of clones/ vectors used in this study			
Vectors /Plasmid construct name	**Details**	**Protein size (~MW)**	**Reference(s)/ Source**
pET22bΔ50CPD*	Vibrio cholerae MARTX toxin cysteine protease domain with C terminal His Tag in pET22b.	23.1 kDa	Gifted by Matthew Bogyo, Department of Pathology, Stanford School of Medicine, Stanford, California.
YsaN	YsaN wild type full length (430 amino acids cloned in pET22bΔ50CPD vector).	47.87 kDa	This Study.
YsaNΔ83	N terminal 1- 83 amino acid deletion of YsaN (347 amino acids cloned in pET22bΔ50CPD vector)	38.6 kDa	This Study.
YsaNΔC-term	C terminal 357- 430 amino acid deletion of YsaN (349 amino acids cloned in pET22bΔ50CPD vector)	39.2 kDa	This Study.
YsaNΔN-term	N terminal 1- 140 amino acid deletion of YsaN (290 amino acids cloned in pET22bΔ50CPD vector)	32 kDa	This Study.
YsaN K166→ A	The lysine at the 166th position is mutated to alanine. It is a non-functional mutant of YsaN. (430 amino acid cloned in pET22b vector with C-terminal his tag.)	49kDa	This study.
(*) Note: Δ represents deletion in all cases.			

Figure 4.1: Representation of YsaN constructs used in this study.

Table 4.2: List of primers used in this study			
Gene/ construct name	Forward (5'- 3')	Reverse (5'- 3')	Reference(s)/ Source
YsaNCPD	GGGAATTCCATATGAATCTCTTTGATAGCTGT	CGGGATCCGAGTTTGCCAGCTC	IDT Technologies/ This study
YsaNΔ83CPD	GGGAATTCCATATGATGGGAGAACATCTG	CGGGATCCGAGTTTGCCAGCTC	GCC Biotech/ IDT Technologies / This study
YsaNΔNTerm CPD	GGAATTCCATATGGGGGGTCAGGGCAATTG	CGGGATCCGAGTTTGCCAGCTC	GCC Biotech/ This study
YsaNΔCTerm CPD	GGGAATTCCATATGAATCTCTTTGATAGCTGT	CGGGATCCGAAGTCACTTTCGAAAAC	GCC Biotech/ This study
YsaN K166→A	GCGGCCGGTAGCGGCGCAACCTCCCTGATGAGC	GCTCATCAGGGAGGTTGCGCCGCTACCGGCCGC	GCC Biotech/ This study
*Note- Restriction site highlighted in bold letters.			

4.2.2. Site-directed mutation:

A non-functional point mutation in Walker A motif was done according to [131] in YsaN pET22bΔ50CPD construct. Sequencing was done to verify positive clone. The positive clone was transformed into BL21 for its recombinant expression. Primers used is given in **Table 4.2**.

4.2.3. Malachite green ATPase assay:

ATPase assay was performed to check the ATP hydrolysis activity by Malachite green ATPase assay method with slight modification. In brief, aal the reagents were prepared sufficiently before performing the experiment according to Dey et. al: 2018. Reagents and buffers used in assay method is given in **Table 4.3**. Purified protein was incubated in the assay buffer at least ~1hr before the experiment. Other chemical reagents were filtered thorough 0.22µm filter just before use. A varying ATP concentration stock was prepared (0-2400µM). Two millilitre of reaction mixture with varying ATP

concentration was prepared by mixing 5µM of YsaN and 10µM of YsaNΔ83 protein at final working concentration in assay buffer. From this reaction mixture 100µl of sample was collected at four different time points and added to a 96 well microplate containing 50µl of malachite green-ammonium molybdate mixture and mixed by horizontal shaking. The plate was kept for incubation for at least 10 minutes at room temperature. Further, to stop colour development in the reaction mixture 100µl of 34% (v/v) citric acid was added and mixed by horizontal shaking and incubated for another ~30minutes. The absorbance was finally measured at 660nm in Hidex-Sense 96 well plate reader. The experiment was performed in triplicates separately for YsaN and YsaNΔ83 enzyme kinetics. Further a relative ATPase assay was performed to compare the activity of YsaN deletion constructs with the wild type using 5µM for YsaN, YsaN K166→A, and YsaNΔCterm whereas, 10µM for YsaNΔ83 and YsaNΔNterm respectively. For relative assay a standard 2mM working concentration of ATP was used. For this 2ml (final volume) of reaction mixture was prepared form which 100µl of reaction sample was taken at regular time interval up to 40 minutes and added to 96 well plate containing 50 µl of Malachite green-ammonium molybdate mixture. similar protocol was used as mentioned earlier for enzyme kinetics. In a similar way relative activity was also performed in presence of 2mM EDTA and 5mM NaF as a control experiment. Relative activity assay was performed in duplicates independently. A separate blank was performed for each experiment and the values were subtracted from the experimental result. Phosphate standard curve was generated using Sodium phosphate monobasic (NaH2PO4) **(Figure 4.2)**. The K_{half} and V_{max} values were calculated by plotting the kinetics data in GarphPad prism software. Other graph was plotted by fitting the experimental values in GraphPad and Origin software.

S.No.	Buffer/reagent name	Composition/Protocol
Table 4.3: Malachite green ATPase reagents and buffer.	
1	**Assay buffer**	20mM Tris-cl pH 8.0, 100mM KCl, 100mM NaCl, 10mM Mgcl2, 1mM DTT, 0.025% BSA, and 5% glycerol.
2	**Malachite green**	60ml of Conc. H_2SO_4 was slowly added to 300ml of MilliQ and cooled down to room temperature. 0.44gm of Malachite green crystals were added to this solution and dissolved. The resulting orange color solution was stable at room temperature for ~1 year.
3	**Ammonium molybdate**	7% (w/v) in 4N H_2SO_4 solution.
4	**Malachite green-ammonium molybdate mixture**	Add 2.5ml of 7.5% ammonium molybdate in 10ml of Malachite green solution. Add 200µl of 11% Tween 20. Prepare fresh for each set of experiment at least ~1hr before the start of experiment. Filter through 0.22µm, just before use.
5	**Tween 20**	11% (v/v) in MilliQ.
6	**Citric acid solution**	34% (w/v) in MilliQ.
7	**Sodium phosphate monobasic (NaH₂PO₄)**	10mM stock.
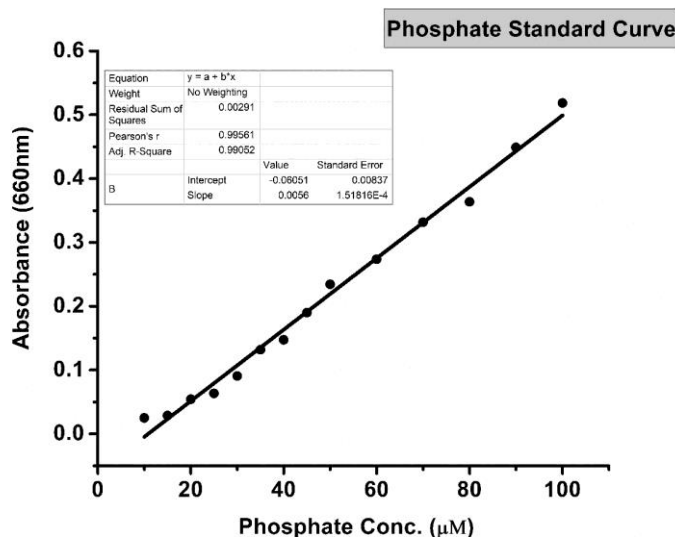
Figure 4.2: Phosphate standard curve for Malachite green assay- Inorganic phosphate standard curve was calculated from the phosphate standard stock prepared using NaH2PO4 (Sodium phosphate monobasic, Sigma) solution. Absorbance was read at 660nm.

4.2.4. Dynamic Light Scattering study:

Dynamic Light Scattering (DLS) analysis was performed to study its monodispersity and oligomerization behaviour in absence and presence of different substrates (like ATP, ADP, AMPPNP, and ADP.AlFX). DLS was performed using Zetasizer Nano ZS (Malvern instruments, UK) in DLS buffer (20mM Tris pH8.0, 100mM NaCl, 4mMMgcl2, 5% glycerol and 1mM DTT). 100µl of reaction sample was prepared in microcentrifuge tube in absence and presence of substrate and 50µl of the reaction sample was loaded in to DLS cuvette Ultra-Micro Cell ZEN2112 (Malvern, UK), and size intensity distribution was observed. Substrates were used in the concentration range of 1-3mM. All the DLS experiments were repeated more than three times in an independent experiment.

4.2.5. Analytical size-exclusion chromatography:

For the estimation of apparent molecular weight of different YsaN oligomeric complexes analytical size exclusion chromatography was performed. Oligomer formation was done using ADP.AlFX (a non-hydrolysable ATP to ADP transition state analogue) and was adapted from Rappas *et al.*, (2005) [132] with minor changes. To a 500µl of ~10 µM of purified YsaN and YsaNΔ83, 1.5mM ADP was added and incubated for 5-10 minutes at room temperature. To this sample 1.5mM of AlCl3 was added and mixed thoroughly followed by incubation for ~20 minutes at room temperature. This sample was injected into the AKTA system with Superdex 200 HR 10/30 analytical column (GE Healthcare), pre-equilibrated with equilibration buffer (50mM Tris pH 8.0, 100mM KCl, 4mM MgCl2, 5mM NaF, 1mM DTT and 5% glycerol). Sodium Fluoride (NaF) was required for maintenance of AlFX complex [132]. It was also tested that whether NaF has any effect on the activity of YsaN. It was observed that YsaN activity was not affected by

NaF **(Figure 4.12).** The flow rate was kept at 0.5ml/minutes at room temperature. All the fraction of peaks were collected, concentrated and analysed on SDS gel.

4.2.6. Negative- Transmission Electron Microscopy:

Visualization of YsaN higher order oligomer to higher resolution was done by negative Transmission Electron Microscopy (Negative-TEM). The sample was prepared by using 100µl of 0.6-0.8µM protein sample (YsaN and YsaNΔ83) was incubated with 1.5mM ADP for 10 minutes at room temperature. To this sample 1mM of AlCl3 was added and mixed thoroughly and incubated at room temperature for ~15minutes. Three microliter of this sample was loaded onto a freshly glow discharged Carbon coated copper 300 mesh grid and left for ~30 seconds. The excess sample was bloated gently using blotting paper. Precaution was taken not to dry the grid completely. Then staining was done using ~5µl of 1% Uranyl acetate solution followed by incubation of ~10 seconds. Excess of the stain was removed through bloating. Staining process was repeated three times. Finally, the grid was kept for air dry at room temperature. Visualization was done at room temperature using TALOS L 120C electron microscope (Thermo Fisher). The instrument was operated at a voltage of 120kV and a bottom mounted Flucam and Ceta 16M (35–910 kX magnification range) camera was used to capture the images.

4.3. Results:

4.3.1. Bioinformatic analysis of YsaN- Nature of N-terminal domain:

Domain prediction using P*fam* (**Figure 4.3**) suggest the presence of three domains: an N-terminal domain ~140 amino acid residues, a central ATPase domain ~140-349 amino acid residues, and a C-terminal domain ~356 to 427 amino acid residues. The N-terminal domain of such ATPase is considered as the regulatory domain and is required for its oligomerization and activation [133]. In a previous study it has been shown that YsaN is Magnesium dependent putative ATPase of *Ysa-Ysp* T3SS of *Y. enterocolitica* [129]. Further, a Multiple Sequence Alignment (MSA) using Clustal Omega [134] with few of known T3SS ATPase (see **Table 4.4** and **Figure 4.4**), suggest that the N-terminal domain is least conserved as compared with other two domains(**Figure 4.4** and (**Table 4.5 and 4.6**. Secondary structure prediction from YsaN sequence in PSIPRED 4.0 [135] suggest that the N-terminal domain consist of primarily of β-strands. Further Kyte-Doolittle's hydrophobicity in Protscale [136] result suggest that the N-terminal is primarily hydrophobic in nature (**Figure 4.5 A-B**). A YsaN homology model was also predicted using Phyre2 server [137]. Phyre2 could generate a YsaN model for only 78-430 residues only. Further to generate a full length YsaN model we used FliI-FliH crystal structure (PDB ID 5B0O) for homology modelling using MODELLER 9v.11.43 software [138]. FliI is flagellar T3SS ATPase of *E. coli* and shows 38.28% sequence identity with YsaN (see next section for details of model generation). Aligning the Phyre2 predicted YsaN model with

the YsaN model generated through MODELLER software shows an RMSD value of 1.2Å (**Figure 4.5 C**).
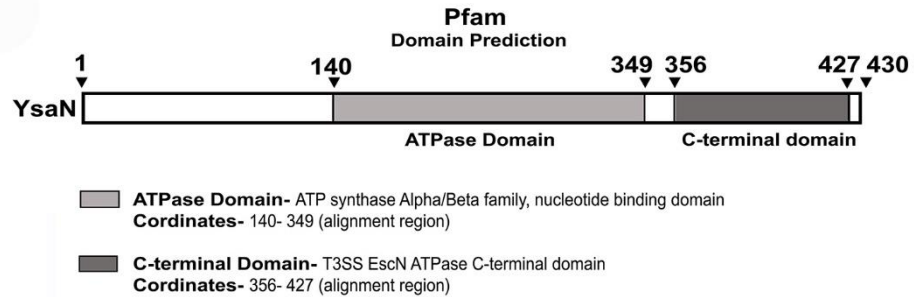
Figure 4.3: *Pfam* domain analysis of YsaN representing different predicted domains.
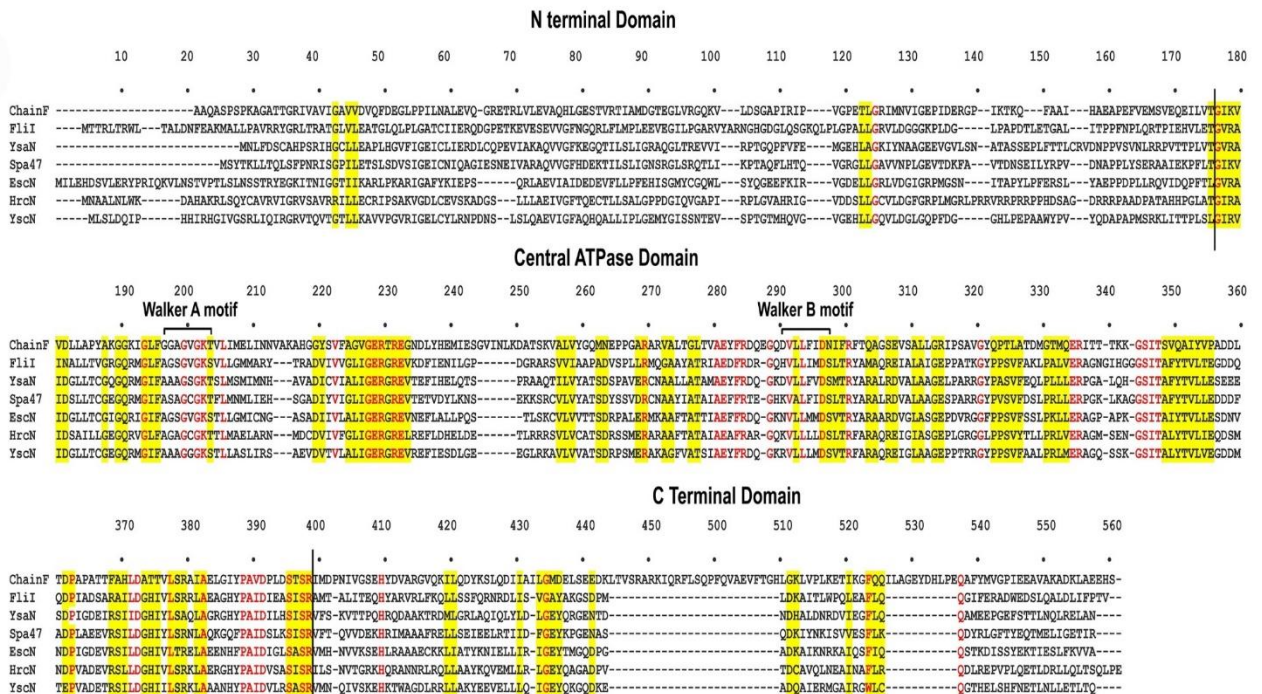
Figure 4.4: Multiple sequence alignment of YsaN and homologs (also refer to Supplementary Tables S2–S4 for details). The vertical black line represents the domain partition (i.e., N-terminal and C-terminal domains) of YsaN sequence based on *Pfam* domain analysis.

Table 4.4: List of Proteins used for Multiple Sequence alignment			
Name of protein	**Source**	**Percentage Identity with YsaN**	**PDB ID**
YsaN	*Yersinia enterocolitica* subsp. enterocolitica 8081	100 %	NA*
YscN	*Yersinia enterocolitica* subsp. enterocolitica 8081	43.60 %	NA*
Mitochondrial F1-Atpase	Bovine Mitochondrial F1-Atpase chain F	30.85 %	1BMF_F
FliI	flagellum-specific ATP synthase FliI [*Salmonella enterica* subsp. enterica serovar Typhi]	38.28 %	5B0O
Spa47	(plasmid) [*Shigella flexneri*]	42.46 %	5SWJ
EscN	[*Escherichia coli*]	45.26 %	2OBL
HrcN	[*Pseudomonas syringae* pv. *syringae*]	44.38 %	NA*
(NA*) – X-Ray structure not available.			

Table 4.5: Percent Identity For N-terminal alignment sequence							
	YsaN [*Y. enterocolitica*]	**YscN [*Y. enterocolitica*]**	**ChainF [Mitochondrial F1-Atpase]**	**FliI [*S. enterica*]**	**Spa47 [*Shigella flexneri*]**	**HrcN [*Pseudomonas syringae*]**	**EscN [*Escherichia coli*]**
YsaN [*Y. enterocolitica*]	**0**						
YscN [*Y. enterocolitica*]	21.66	**0**					
ChainF [Mitochondrial F1-Atpase]	11.54	12.93	**0**				
FliI [*S. enterica*]	16.17	24.24	15.48	**0**			
Spa47 [*Shigella flexneri*]	36.36	22.08	12.24	17.18	**0**		
HrcN [*Pseudomonas syringae*]	16.46	28.85	16.56	23.81	17.09	**0**	
EscN [*Escherichia coli*]	14.88	22.42	15.95	20.69	16.17	18.29	**0**

Table 4.6: Percent Identity For C-terminal alignment sequence						
	YsaN [*Y. enterocolitica*]	**YscN [*Y. enterocolitica*]**	**ChainF [Mitochondrial F1-Atpase]**	**FliI [*S. enterica*]**	**Spa47 [*Shigella flexneri*]**	**HrcN [*Pseudomonas syringae*]**	**EscN [*Escherichia coli*]**
YsaN [*Y. enterocolitica*]	**0**						
YscN [*Y. enterocolitica*]	32.10	**0**					
ChainF [Mitochondrial F1-Atpase]	13.01	13.82	**0**				
FliI [*S. enterica*]	21.69	24.10	12.20	**0**			
Spa47 [*Shigella flexneri*]	30.86	24.69	9.76	24.10	**0**		
HrcN [*Pseudomonas syringae*]	31.76	36.47	15.32	24.71	21.18	**0**	
EscN [*Escherichia coli*]	27.16	38.27	17.07	25.30	27.16	31.76	**0**
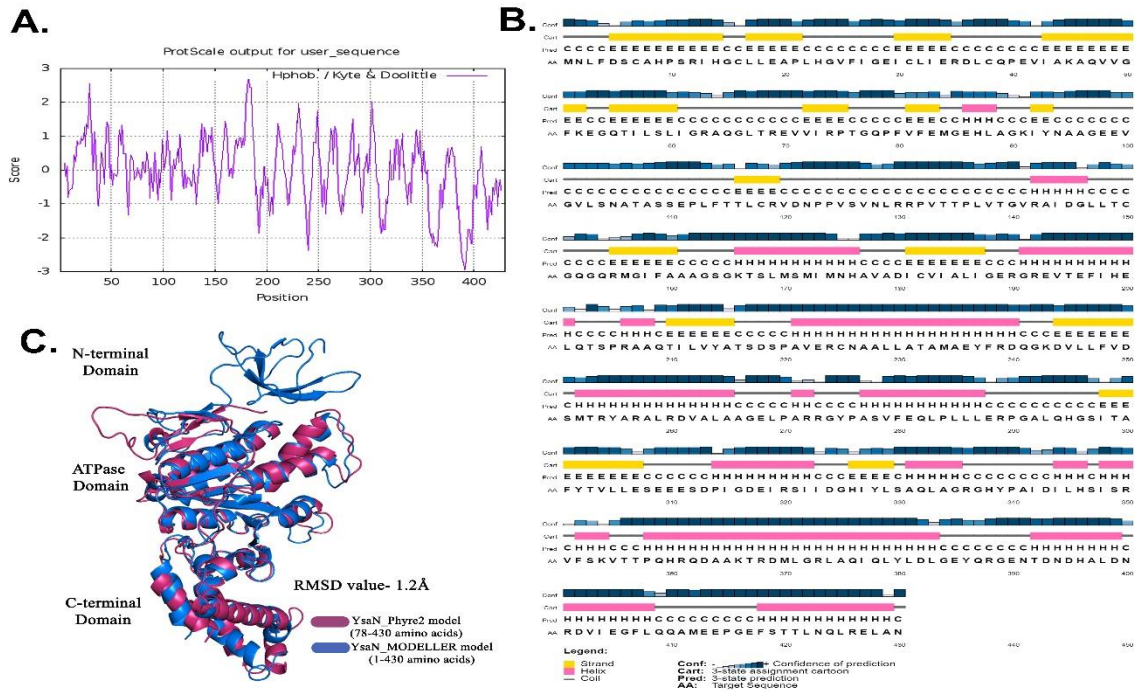
Figure 4.5: Bioinformatic analysis of YsaN- A.- Protscale Kyte & Doolittle hydrophobicity analysis suggesting the N-terminal region as primarily hydrophobic in nature. B.- Secondary structure prediction using PSIPRED shows the N-terminal region primarily as a beta-strand. C- Representation of aligned structures generated by MODELLER software and Phyre2 server. The alignment was done in PyMOL software.

4.3.2. **Functional analysis of YsaN constructs reveals role of its N-terminal domain:**

In the pursuit to obtain only the soluble and monomeric form of YsaN, many deletion constructs (a random 20 amino acid deletion) were cloned along with the deletion based of the P*fam* domain prediction (see **Figure 4.1**). **Table 4.1** and **Figure 4.1** represents various constructs used to conduct this study. All the His tagged deletion constructs of YsaN results in insoluble expression hence, we cloned all the YsaN and deletions in CPD vector. The benefit of using CPD vector was that the final product was soluble and untagged [130] and they are expressed as primarily monomeric form. A typical size exclusion chromatography curve for each construct is represented in **Figure 4.6 A-D** and **Figure 4.6 E** represents the SDS profiles of the respective peaks. Further, it was also explored that whether the presence of CPD results in observed monomeric peak of protein, we compared the size exclusion profiles of YsaN C-terminal His tag (cloned in pET22b vector) with the YsaN-CPD His tag protein. It was observed that both YsaN elutes at same elution volume (**Figure 4.7**).
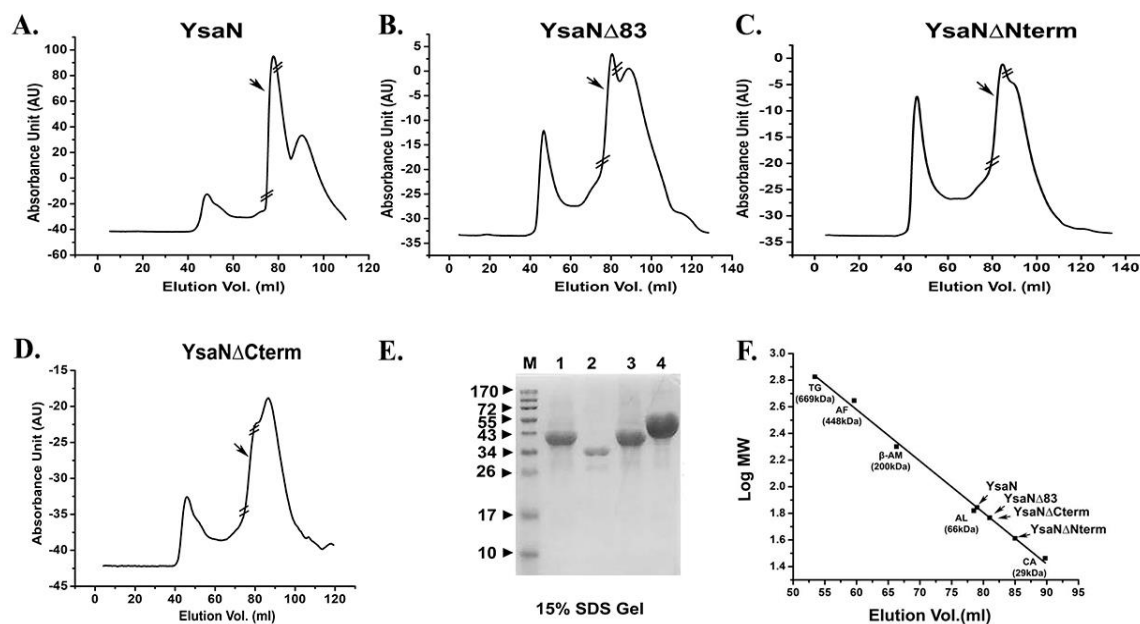
Figure 4.6: Purification and characterization of YsaN- (**A–D**)—Typical size exclusion chromatography profile of YsaN, YsaNΔ83, YsaNΔNterm, and YsaNΔCterm respectively using Superdex 16,600 columns. Respective collected peaks are delimited by crossed sections of the chromatograms. **E-** SDS gel profile of collected peaks from gel filtration (Lane M-Marker, Lane1- Peak from (D), Lane2- Peak from (C), Lane3- Peak from (B), and Lane4- Peak from (A). **F-**Log molecular weight standard curve for Superdex 16/ 60 column: Figure representing elution profiles of all YsaN constructs used in this study on the molecular weight standard curve. Molecular weight standards are represented on the curve as (TG= Thyroglobulin- 669 kDa, AF= Apoferritin 443 kDa, β-AM= β- Amylase- 200 kDa, AL= Albumin- 66 kDa, CA= Carbonic anhydrase- 29 kDa). Adapted from Kumar, R., et. al. (2022). Delineating specific regions of N-terminal domain of T3SS ATPase YsaN of Yersinia enterocolitica governing its different oligomerization states. Frontiers in Molecular Biosciences, 9.

Further, ATP hydrolysis activity was estimated through Malachite green ATPase assay. Untagged YsaN has enzyme kinetic parameters of V_{max} value of $9.07 \pm 0.72 \mu mol/min/mg$, K_{half} value of $758.7 \pm 84.35 \mu M$, and a Hill coefficient value of $h = 2.09 \pm 0.31$ (**Figure 4.8 A**) which is comparable to the previous study on YsaN (N-terminal His-tag) [129]. Whereas YsaNΔ83 has reduced activity with enzyme kinetic value of V_{max} of 3.19 ± 0.09, K_{half} value of 467.3 ± 30.15, and a hill coefficient value of $h = 1.20 \pm 0.06$ (see **Figure 4.8 B**). YsaNΔCterm has activity comparable to YsaN whereas, in the case of YsaNΔNterm, ATPase activity was barely detectable (**Figure 4.8 C-D**). These results suggest that ATPase activity of YsaN depends on the N-terminal domain while YsaNΔ83 bears cooperativity and reduced ATPase activity.
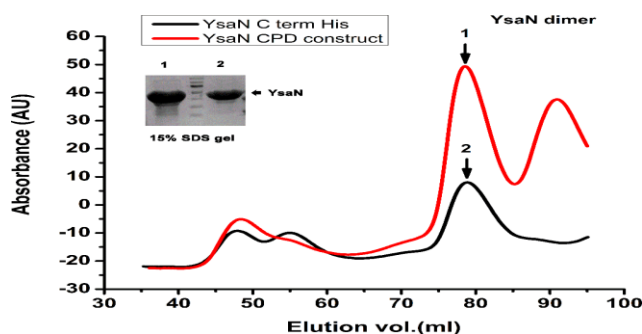
Figure 4.7: A comparison of C- terminal His tagged YsaN with the YsaN no tag obtained after CPD tag removal. C-terminal CPD fusion of YsaN does not alter the oligomeric status of YsaN in solution.
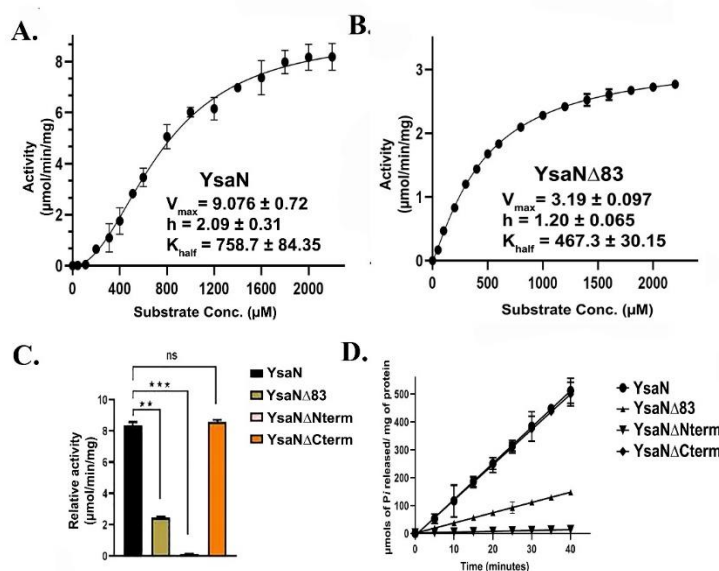
Figure 4.8: A- Standard enzyme assay curve of purified YsaN measuring μmols of phosphate released per minutes per milligrams of protein using the malachite green method for inorganic phosphate estimation, inset showing Vmax, Khalf, and Hill coefficient values. **B-** Standard enzyme curve for YsaNΔ83 in terms of μmols of phosphate released per minute per milligrams of protein, inset showing Vmax, Khalf, and Hill coefficient values. **C-** Comparison of relative ATPase activity of purified YsaN to YsaNΔ83, YsaN to YsaNΔNterm, YsaN to YsaNΔCterm. Error bars indicate standard deviation in all cases. Statistically significant differences by unpaired Student's t-test are indicated with asterisks ('***' means $p < 0.001$; '**'$p > 0.01$; "ns" means nonsignificant). **D-** Linear curve representation of relative ATPase activity.

4.3.3. DLS study and substrate dependency of YsaN:

Specificity for different substrates of YsaN was checked concerning its oligomerization behaviour and was studied through DLS. ADP (product), ATP (substrate), AMPPNP (non-hydrolysable ATP analog) and ADP.AlFX (ATP to ADP transition state analog) was tested as substrate and oligomerization behaviour was observed. Significant monodispersity change was observed as a result of oligomerization only in the case of ATP and ADP.AlFX (**Figure 4.9**). However, in the case of ADP and AMPPNP, both were incapable to promote oligomerization. ADP.AlFX has been reported to mimic the ATP to ADP transition state analogue and has been used to mimic ATP to ADP hydrolysis of γ-phosphate transition state analogue [139, 140]. Binding of ADP.AlFX to YsaN resulted in a change in monodispersity similar to that of ATP (**Figure 4.9**). Also, further to check the dependency of protein concentration on oligomerization behaviour of YsaN, we vary the protein concentration but we observe no change in the oligomeric behaviour (**Figure 4.10 D**) while varying the ATP concentration within physiological range promotes immediate oligomerization of YsaN, YsaNΔ83 (**Figure 4.10 A** and **B** respectively). DLS shows a clear shift of the peaks toward right as a result of oligomerization for YsaN and for YsaNΔ83 depending on the increasing concentration of ATP (0mM to 2mM). However, in the case of YsaNΔN-term, no shift in peak was observed under similar ATP concentration (**Figure 4.10 C**). The inability of YsaN to oligomerize in presence of ADP as compared to ATP and the difference in oligomerization behaviour of YsaN in presence of AMPPNP and ADP.AlFX (both of which are non-hydrolysable ATP analogue) suggest that YsaN oligomerization is the result of active catalysis and not just ATP binding. To evaluate this, we cloned a Walker-A lysin mutant YsaN K166→A (which can bind to ATP but is unable to hydrolyze it. Loss of

Figure 4.9: Dynamic light scattering study of YsaN oligomerization- Representation of size intensity profile of YsaN and in the presence of substrates like 2mMADP (product), 2mMATP (substrate), 2mMAMPPNP (a non-hydrolyzable ATP analog), and 1.5 mMADP.AlFX (an ATP to ADP + Pi transition state analogue).

activity was tested by ATPase assay see **Figure 4.12**), and observed its oligomerization behaviour by DLS in a similar way (**Figure 4.10 F**). Simultaneously in another experiment, YsaN was treated with 2 mM EDTA (during its purification) to quench the Mg^{2+} ion, to prevent Mg^{2+} dependent catalysis of ATP. It was observed that even in presence of 3 mM ATP, YsaN-EDTA treated sample was unable to oligomerize (**Figure 4.10 E**). The observation from both the experiment clearly indicate that YsaN oligomerization requires active catalysis and not just ATP binding. Collectively, YsaN oligomerization and its activity depend on the N-terminal domain and occurs in an ATP concentration-dependent manner, and also it requires active catalysis of ATP to ADP and Pi.
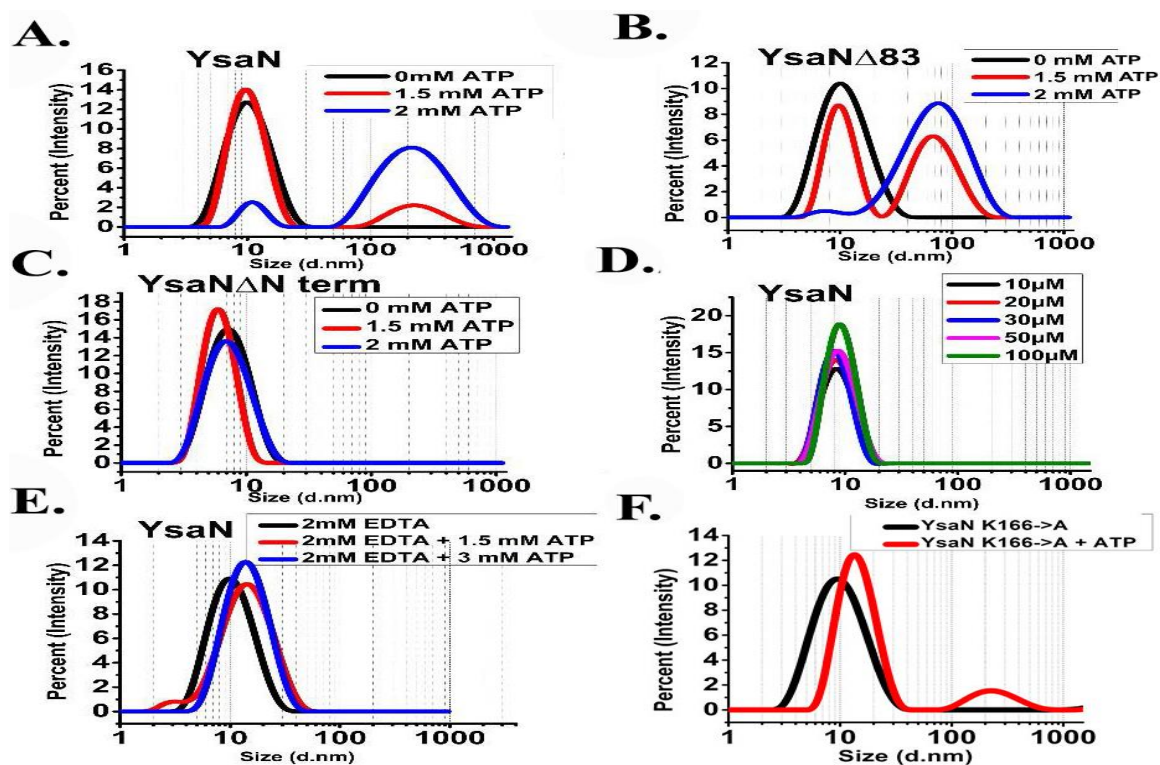
Figure 4.10: Substrate (ATP) concentration-dependent oligomerization of **A-** YsaN, **B-**YsaNΔ83, and **C-**YsaNΔNterm respectively demonstrating the role of YsaN N- terminal domains in YsaN oligomerization. **D-** Protein concentration-dependent DLS study of YsaN. **E-** Size intensity profile of YsaN in presence of 2 mMEDTA demonstrating the role of Mg2+ dependent catalysis of ATP to ADP + Pi and its role in oligomerization. **F-** Size intensity profile of YsaN Walker-A lysine mutant YsaN K166→A in the absence and presence of 2 mM ATP. Adapted from Kumar, R., et. al. (2022). Delineating specific regions of N-terminal domain of T3SS ATPase YsaN of Yersinia enterocolitica governing its different oligomerization states. Frontiers in Molecular Biosciences, 9.

4.3.4. Analytical size exclusion chromatography of YsaN oligomers:

During many studies it has been found that such AAA+ ATPases like YsaN is capable of forming different oligomeric complexes like hexamer, dodecamer, and higher-order aggregates [87, 141-143]. So far, the hexameric complex has been speculated as the most stable structure which has been also observed in many *in-vitro* and *in-situ* experiments [87,

89]. In case of YsaN we have seen that Ysan∆83 a N-terminal construct of YsaN is functionally active but with lesser strength. Also, the activity of YsaN is highly dependent on the active catalysis of ATP. So, the activity of YsaN and its oligomerization is linked to its N-terminal domain. Hence, the formation of the oligomer and its characterization was further studied through estimation of the apparent molecular weight of oligomeric complex formed through analytical size exclusion chromatography. Size exclusion chromatography was performed in absence and in the presence of ATP analogue ADP.AlFX. Apparent molecular weight estimation was done by comparing the elution profiles of the peaks with the molecular weight standard curve represented in **Figure 4.11-C**. YsaN-ADP.AlFX eluted at ~14mml and ~10ml corresponding to an equivalent molecular weight of ~288kDa (hexamer) and ~576kDa (probably dodecamer) respectively. A similar experiment with YsaN83-ADP.AlFX complex shows a complex equivalent to hexamer (molecular weight of ~234kDa) See **Figure 4.11A-B and D-E**. All the peaks collected were analysed by SDS gel electrophoresis (**Figure 4.11-F**).
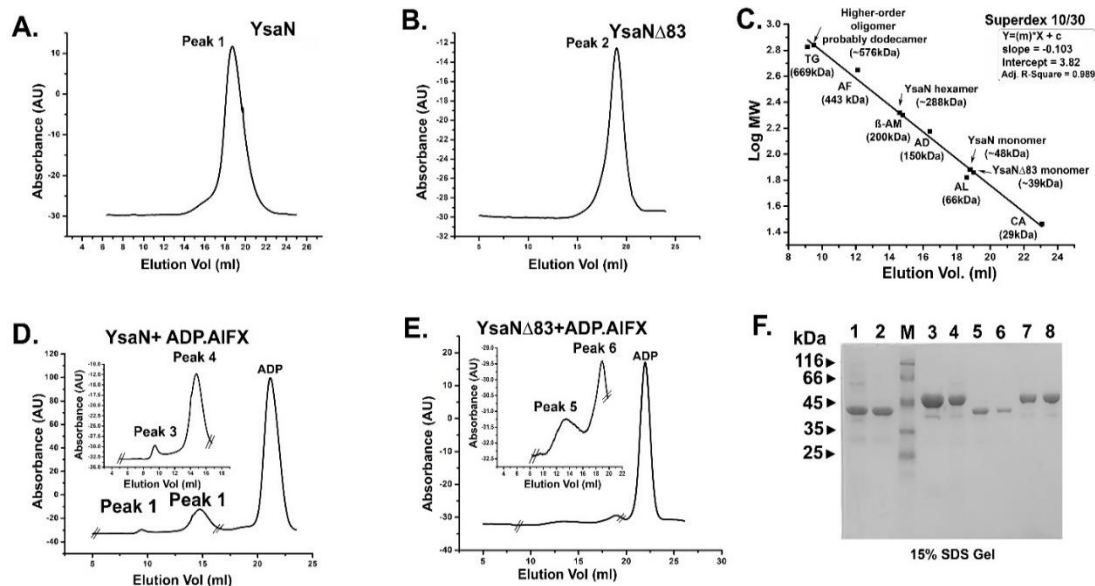
Figure 4.11: Analytical size exclusion chromatography profile of YsaN and YsaN∆83 using Superdex 10/30 column- (A) and (B)- size-exclusion profile of purified monomeric YsaN and YsaN∆83 respectively. (C)- molecular weight standard curve for Superdex 10/30 column representing elution profiles of possible oligomeric forms of YsaN and YsaN∆83. (TG = Thyroglobulin- 669 kDa, AF = Apoferritin 443 kDa, β-AM = β- Amylase- 200 kDa, AD = Alcohol dehydrogenase- 150 kDa, AL = Albumin- 66 kDa, CA = Carbonic anhydrase- 29 kDa). (D) and (E)- size-exclusion profile of YsaN and YsaN∆83 in presence of 1.5 mM ADP.AlFX in the buffer containing 5 mM NaF, inset showing peaks corresponding to probable dodecamer and hexamer respectively in the zoomed image delimited by crossed section in the full chromatogram. (F)- SDS gel profile of the purified sample and peaks collected during size exclusion chromatography steps. Lane 1 and Lane 3 represent purified YsaN∆83 and YsaN form Superdex 200 16/ 600 size exclusion chromatography column resp ctively. Lane 2 (YsaN∆83) and Lane 4 (YsaN) correspond to peak 2 of (B) and peak 1 of (A) respectively. Lane 3 represents the marker, and Lane 5 and 6 represent the peak 6 and peak 5 in (E) respectively. Lane 7 and 8 represent peak 3 and peak 4 in (D) respectively. Adapted from Kumar, R., et. al. (2022). Delineating specific regions of N-terminal domain of T3SS ATPase YsaN of Yersinia enterocolitica governing its different oligomerization states. Frontiers in Molecular Biosciences, 9.
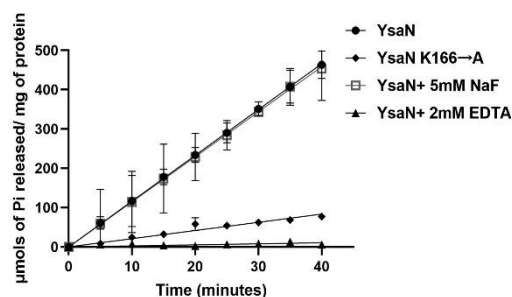
Figure 4.12: Relative activity assay: A.- Relative activity plot of all YsaN construct used in this study given as µmol of inorganic phosphate (P*i*) released per milligrams of protein. **B**.- Relative activity of YsaN, YsaN K→166A, YsaN in presence of 5mM NaF, and YsaN in presence of 2mM EDTA in the assay buffer.

4.3.5. TEM visualization of YsaN and YsaNΔ83 oligomer:

The presence of catalytic activity in YsaN as well as in YsaNΔ83 and their tendency to form oligomers prompted us to visualize the formation of oligomers to higher resolution. For this we used Negative Transmission Electroscopic (negative TEM) method. Similar experimental conditions with DLS and size exclusion chromatography were chosen here for the sample preparation. YsaN and YsaNΔ83 was viewed in absence and in presence of ATP and ADP.AlFX. While viewing YsaN and YsaNΔ83 oligomers in negative-TEM using ATP and ADP.AlFX, we were able to get better images of oligomer in the presence of ADP.AlFX as compared to ATP. In previous study it has been shown that ADP.AlFX stabilize the oligomer structure of similar ATPase and has been utilized to obtain a high-resolution structure [89]. In the case of YsaNΔ83, the presence of particle size of ≈10nm representing a hexamer formation with clear sixfold symmetry was observed (**Figure 4.13-A and C**). The particle size (10nm) of the hexamer as observed in TEM images are in line with the reported size of 10nm among the three-dimensional structures of hexameric ATPases. Whereas, in the case of YsaN the particle size of the oligomer observed was greater than 10nm and was observed as ≈ 15-20nm structure in different orientations (**Figure 4.13-B**). The presence of particle size greater than 10nm indicates the presence of a higher oligomeric form in the YsaN complex as compared to YsaNΔ83.

4.3.6. YsaN homology modelling:

Biochemical studies suggest that YsaN is the T3SS ATPase of *Y. enterocolitica*. There is no structure of YsaN available till date. Hence, to generate a full-length structure of YsaN we used homology modelling software MODELLER [138]. We chose crystal structure of FliI-FliH (PDB ID 5B0O), chain-D having sequence identity of 38.28% with YsaN. Alignment of both sequences i.e., FliI (template) and YsaN (target) was done using Clustal Omega [134]. Few of the manual improvement in the alignment files were done in

build python script followed by improvement of the PDB files using Coot software [144]. MODELLER generated five best models out of which the best model was chosen based on
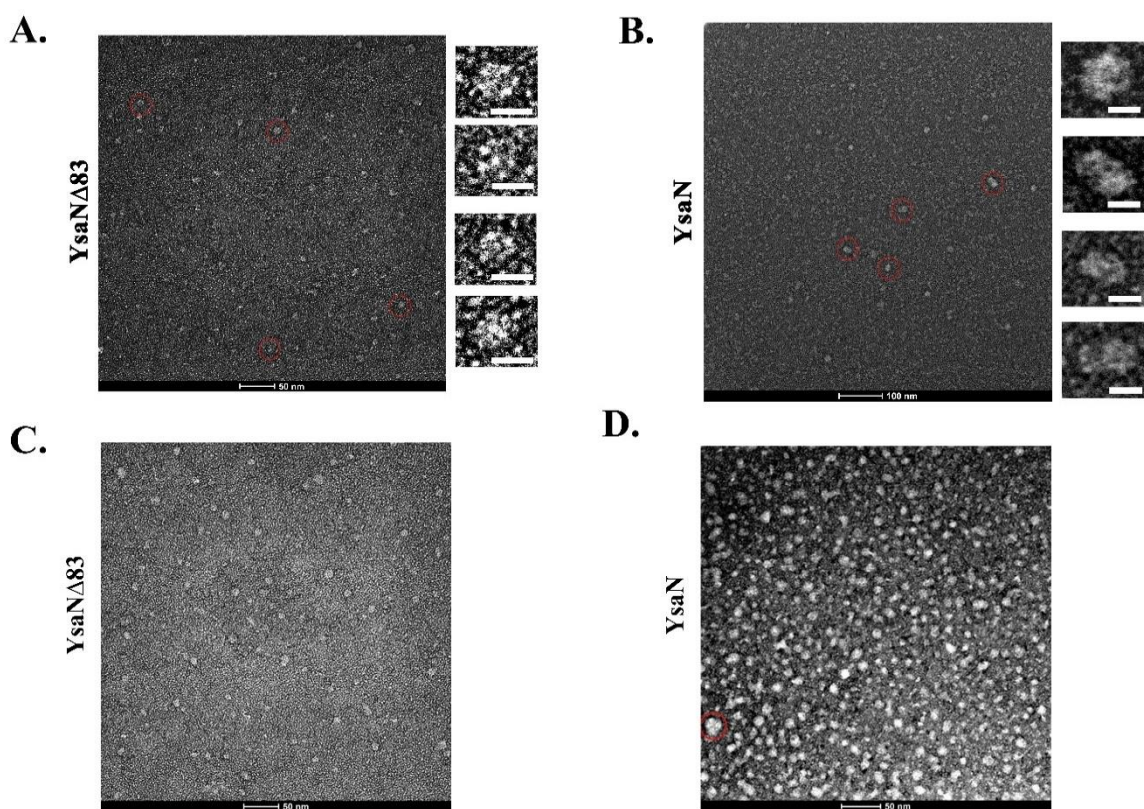
Figure 4.13: Negative TEM image of YsaN and YsaNΔ83 oligomer in presence of ADP. AlFX. (A) YsaNΔ83 oligomer. The Zoomed image represents an approximately 10 nm hexamer ring assembly of particles marked by the red circle in a different orientation. (B) The YsaN higher-order oligomer complex. The zoomed image represents particles marked by the red circle in a different orientation having a size of approximately 15–20 nm in different orientations representing probable dodecamer. The prepared sample grids were stained with 1% uranyl acetate. The white color scale bar represents 10 nm in all cases. C-YsaNΔ83 oligomers in presence of 1.5mM ADP.AlFX appears as approximately 10nm particle.D-YsaN oligomers in presence of 1.5mM ADP.AlFX. Many of the oligomers appear as approximately 15-20nm particles. The red circle represents nonspecific aggregates (**Figure 4.12-D**).

their Discrete Optimized Protein Energy (DOPE score) (see **Figure 4.14** and **Table 4.7**). Final YsaN monomer model was visualized in PyMOL [145]. Again, few manual improvements were done in the PDB text file and clashes were removed using Coot software by rotating side chains to a non-clashing rotamer position. Structure validation was done using various bioinformatic tools like RAMPAGE, ERRAT, and Verify3D. Molecular docking tool Auto Dock [146] was used for binding of small molecule ligand (ATP) to YsaN (receptor protein). The predicted domains of YsaN are coloured differently and are represented by a linear map (**Figure 4.15 A-B**). Subsequently from this monomeric YsaN model a hexameric form was generated by fitting the monomeric structure in the

hexamer of F1 ATPase (PDB ID 4XD7) in PyMOL. The hexameric PDB text file was manually edited to remove unnecessary text which hindered in visualization of the structure in VMD or PyMOL. From this hexameric text file a dodecamer model of YsaN (N-terminal 20 residue removed) was generated by using CluPro server [147, 148] in three steps: 1) rigid body docking, 2) RMSD based clustering to generate the lowest energy structures, and 3) the energy minimization step for removal of steric clashes. Docking studies with each energy parameter set generated ten models defined by centres of highly populated clusters of low energy docked structures which were further used for final model generation.
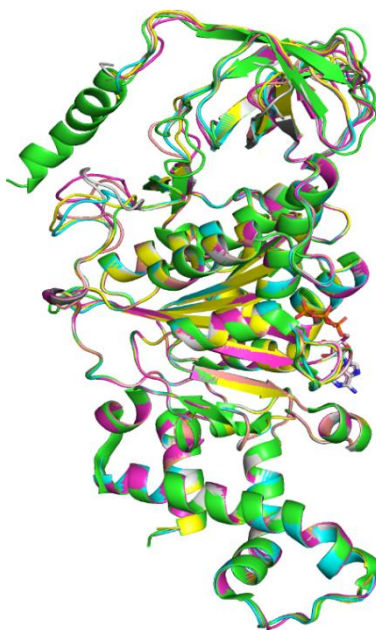
Figure 4.14: Representation of all five models aligned structures generated by MODELLER software with the template FliI. Image generated in PyMOL.

Table 4.7: Table represents the variance in the DOPE score of MODELLER generated YsaN models and the RMSD of aligned models with the template FliI.

Model Name	DOPE score	RMSD compared to Chain D of FliI-FliH structure (PDB ID-5B0O)
YsaN Model 1	-42719.90625	0.123
YsaN Model 2	-42222.11328	0.128
YsaN Model 3	-43239.66797	0.115
YsaN Model 4	-42660.85938	0.121
YsaN Model 5	-43316.37500	0.122
Note: The alignment was done in PyMOL software.		

The involvement of the N-terminal in YsaN activity, and absence of higher order oligomer in the case of YsaN83 construct guided us to generate a dodecamer model by stacking two hexamers through its N-terminal faces. The double ring dodecamer structure modelled has size of ~15nm height and ~10nm diameter along each hexameric rings. Also, a clear pore of ~1.5nm can be seen from the top of the dodecameric ring structure (see **Figure 4.16**). Following a similar protocol YsaNΔ83 homology model was also built from FliI as template using MODELLER. The best model based on DOPE score was chosen and was utilized to generate a hexamer model of YsaNΔ83 using PyMOL (see **Figure 4.17**).
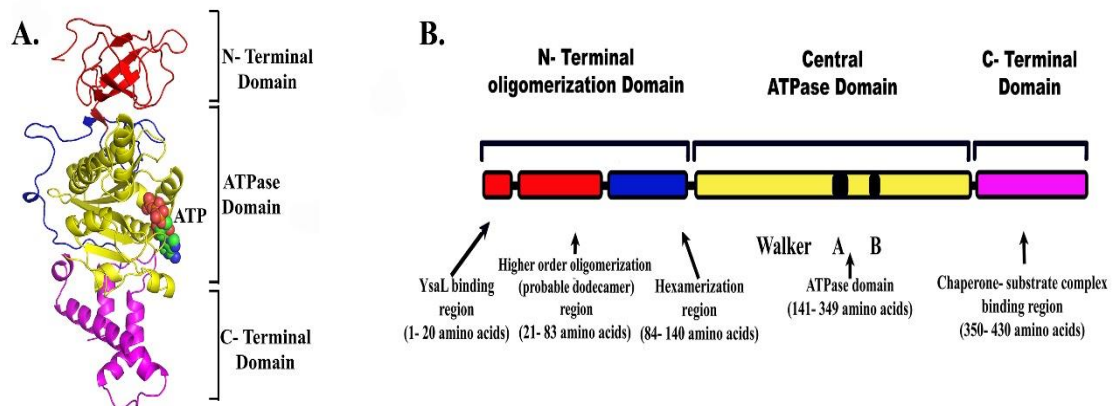
Figure 4.15: Linear representation of different predicted domains in YsaN.
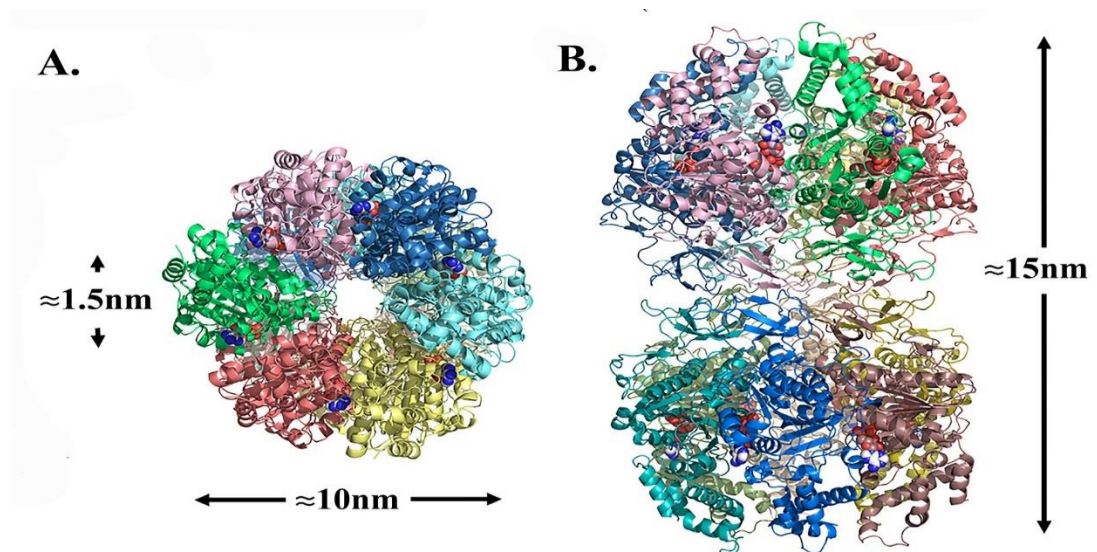
Figure 4.16: Model of probable dodecamer build by stacking two homo-hexamers through their N terminal faces following docking study. **A**- top view, **B**- side view.
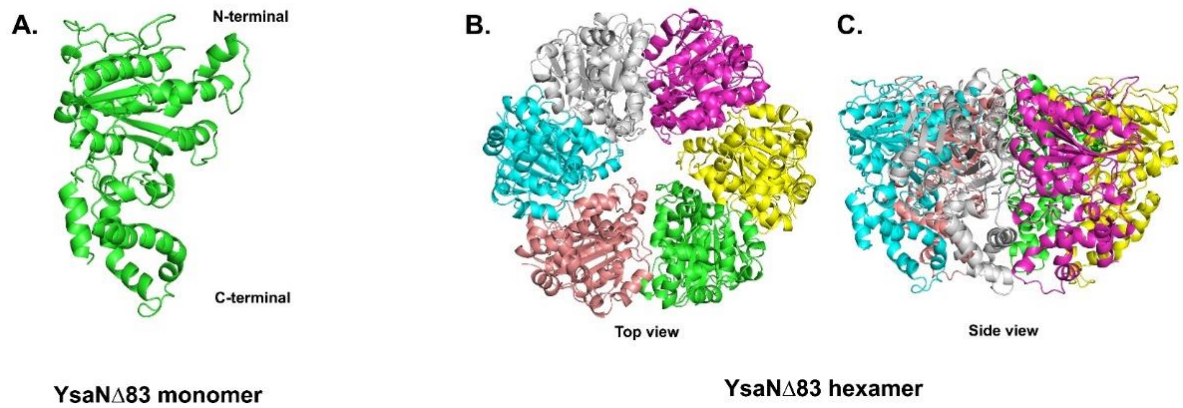
Figure 4.17: Homology model of YsaNΔ83 monomer and hexamer. A. YsaNΔ83 monomer. **B-C.**

YsaNΔ83 hexamer, Top view and Side view respectively.
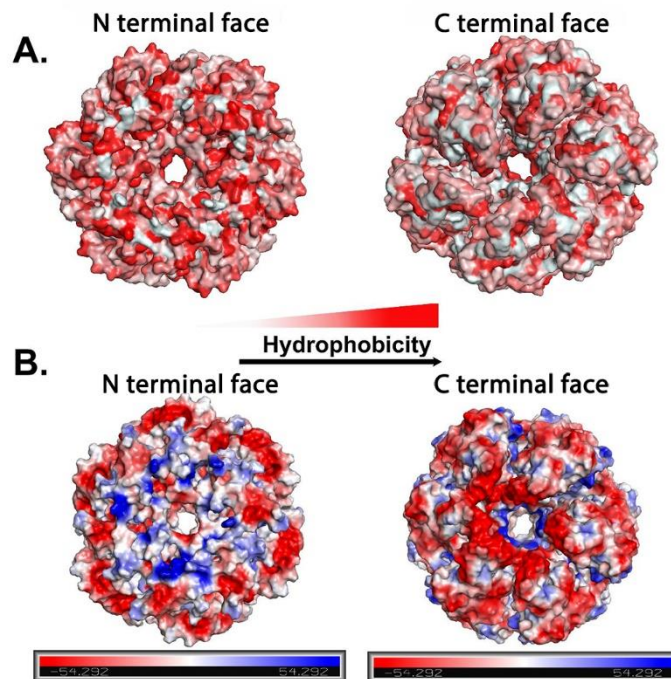
Figures 4.18: Surface map of YsaN hexameric complex- A.- Hydrophobic surface map of N terminal face
and C terminal face respectively of YsaN hexamer complex model (red indicates hydrophobic residue and
white indicates hydrophilic residues exposed). **B** - Surface map of N terminal and C terminal face of YsaN
hexamer complex based on charge (red indicate negatively charged amino acid residues and blue represents
positively charged amino acid residues.

4.4.Discussion:

Pathogenic bacteria use T3SS to deliver their effector toxins directly into host cell in single step mechanism [32, 52]. To perform this task the effectors must be unfolded first to cross the needle structure. The unfolding and translocation require a highly conserved ATPase complex which actively interacts with the effector-chaperone complex inside the bacterial cytoplasm [92]. The most functional form of such ATPase is higher order oligomer which is present at the sorting platform. Previous studies suggested that these ATPases can form hexamers and dodecamers [142, 149, 150] as the most active form. *Y. enterocolitica* harbours two T3SS where YscN (*Ysc-Yop* T3SS ATPase) is an oligomerization activated ATPase and forms hexamer [151] while YsaN (*Ysa-Ysp* T3SS ATPase) has been suggested to form probable dodecamer which is the most active form [129]. Hexamer form is the most stable and well characterized functional form of such ATPase. However, the exact nature of oligomer formation and regulation is still poorly understood. In this study we addressed the question regarding the exact nature of activation of YsaN and its oligomer formation and its correlation with the N-terminal region. We also captured the hexameric form of YsaN in a deletion construct of YsaN (YsaNΔ83), which is observed for the first time ever.

Domain analysis based on YsaN sequence and multiple sequence alignment (MSA) with another homolog ATPase suggest that YsaN is a T3SS associated ATPase of *Y. enterocolitica*. In previous study recombinant expression of N-terminal His tag YsaN resulted in YsaN mixed population and mainly higher order oligomer form [129]. Also, any terminal deletion of YsaN results in insoluble expression rendering it unusable (data not shown). To overcome the problem of solubility we have developed an CPD-vector based expression system for YsaN and its various deletion constructs. Also, the use of CPD vector resulted in purifying the target protein in untagged and monomeric form which was a crucial step in this study. Ability to purify the untagged protein is this form allows us to capture the real time oligomerization behaviour of YsaN (i.e., transition from monomer to hexamer and hexamer to higher-order oligomer as stable oligomer) based on its different N-terminal regions. Untagged YsaN shows a V_{max} value of 9.07 ± 0.72 which is comparable to the previous report [129]. Further enzyme kinetics data reveal that the ATPase activity of YsaN depends on its N-terminal region (140 residues). Whereas it is also remarkable that YsaNΔ83 (an N-terminal deletion variant of YsaN) shows ATPase activity however reduced as compared to YsaN. This finding has been observed for the first time in any T3SS ATPase and it breaches the canonical idea of loss of ATPase activity of N-terminal deletion variant (up to ~79 residues) [150]. The presence of the cooperativity in YsaNΔ83 ($h = 1.20 \pm 0.06$) also hints towards a two-step cooperative behaviour of YsaN activation. This also suggest that YsaN complete activation require both of N-terminal regions namely 1-83 and 84-140 amino acid residues in an independent manner.

In another study involving DLS YsaN shows a shift in the peak as a result of oligomerization upon substrate (ATP) binding which was absent in the case of ADP (product). It was observed that in presence of AMPPNP (non-hydrolysable ATP analog), no oligomerization was observed whereas YsaN, in presence of ADP.AlFX (non-hydrolysable ATP to ADP transition state analog) shows oligomerization similar to that in presence of ATP. This difference in behaviour of YsaN indicates that YsaN oligomerization and its activation may be a result of active catalysis and not just ATP binding. To evaluate this further we used a Walker-A lysin mutant of YsaN (YsaN166K→A- a non-functional mutant of YsaN can bind to ATP but is unable to hydrolyze it) and treated it with ATP in similar way and observed in DLS. It was observed that YsaN166K→A was unable to oligomerize. Also, in another experiment removing Mg^{2+} ions from the reaction media result in loss of ATP hydrolysis activity of YsaN. Both of these experiments confirms that YsaN activation as a result of oligomerization require active catalysis of ATP to ADP and Pi. The loss of ATPase activity in each case was studied through ATPase assay (see **Figure 4.12**). This is similar to the finding in the case of F1F0-ATPase suggesting a universal mode of mechanism among different AAA+ molecular motors across their diverse roles inside the cell [152]. YsaN in the presence of increasing concentration of ATP shows a substrate concentration dependent oligomerization property while, in previous study it was suggested that the oligomer formation in case of Spa47 (T3SS ATPase of *Shigella*) depends on protein concentration [149]. To check this, we also did DLS study in protein concentration dependent manner. Surprisingly, increasing YsaN concentration did not results in oligomerization of YsaN. Conclusively YsaN oligomerization is primarily a substrate (ATP) concentration dependent process and requires active catalysis of ATP.

Since, YsaN and YsaNΔ83 activation and oligomer formation are linked we further extended our work to estimate the size of the oligomer formed through analytical size exclusion chromatography (**Figure 4.11**). YsaN-ADP.AlFX shows formation of two oligomer complex: Peak3 (~576kDa) and Peak4 (~288kDa) corresponding to probable dodecamer and hexamer respectively. Whereas in the case of YsaNΔ83-ADP.AlFX single oligomer formation was observed: Peak5 (~234kDa) corresponding to the hexameric form. In another experiment we perform negative-TEM analysis of the oligomers of YsaN and YsaNΔ83 to visualize the oligomers to high resolution. In TEM the YsaN oligomer was seen as particles of size ranging from ~15-20nm (**Figure 4.13 B and D**) while, YsaN83 the particle size appears as ~10nm hexamer ring (**Figure 4.13 A and C**). TEM study reveals that YsaNΔ83 is capable of forming a hexamer ring of six-fold symmetry which has been observed for the first time. The presence of higher size of particles in the case of YsaN suggest that the oligomer formed here is larger than the usual observed hexamer which we suggest as probable dodecamer. In many attempts of TEM sample preparation of YsaN

oligomer complex, few particles greater than 20nm was also observed which is definitely a non-specific aggregate (**Figure 4.13 D**). It must be noted here that it seems very difficult to control YsaN oligomer formation controlling the formation of non-specific aggregates *in-vitro*. However, decreasing the protein concentration limits the formation of aggregates which also resulted in fewer particles observed in TEM images. The 10nm observed size of YsaN83 hexamer reported here is also similar to the previous studies suggesting hexamer formation of T3SS ATPase in *E. coli* [89]. Hence, YsaN higher order oligomer formation depends on the presence of 1-83 N-terminal residues while, the hexamer formation is independent to this region and 84-140 N-terminal residues are sufficient for its formation. It should be also noted that 1-20 N-terminal residues of YsaN is required for YsaL binding [129].

The formation of higher order oligomer in case of YsaN and its absence in the case of YsaNΔ83 guided us to build a dodecamer model of YsaN as probable higher order oligomer. Also, the dodecamer model was built by stacking two hexamers through its N-terminal face (**Figure 4.16**). The dodecamer ring of YsaN has a height of ~15nm and width of ~10nm (as of a hexamer ring). A clear pore through the two hexameric rings stacked, of ~1-1.5nm can be seen. The bioinformatic study show hydrophobic nature of the N-terminal residues which has been represented in surface topographic map of YsaN hexamer (**Figure 4.18**). Hence, it is suggested here that YsaN higher order oligomer formation may occur through stacking of two hexamer rings through their N-terminal faces where the hydrophobic interaction among the N-terminal surface of the hexamer rings may stabilize the dodecamer complex. Here we also suggest that the formation of higher order oligomer (probable dodecamer) occurs in two distinct steps: 1) formation of hexamer and then 2) formation of dodecamer by N to N-terminal stacking of two hexamers dependent on ATP concentration.

Significant knowledge about the structure and function of T3SS ATPase has been attained, however the exact role of ATPase in T3SS regulation and their mechanism is still poorly understood. No study has been able to answer the exact mechanism of oligomer formation and activation of ATPases. The hexamer model of ATPase is the most widely accepted and has been also observed in *in-situ* cryo-EM structure. The existence of dodecamer/higher order oligomer (greater that the hexamer) in many ATPases, where the dodecamer/higher order oligomer is the most active form *in-vitro* creates ambiguity and confusion about their role at the sorting complex. Also, how these ATPases perform the unfolding of substrates and their translocation in not completely understood. In the present study we tried to address the problem of understanding the exact mechanism of formation of different oligomers and activation of T3SS ATPase YsaN from *Y. enterocolitica.* According to one hypothesis suggested by Akeda and Galan (2005) [92], an appealing model mechanism of these ATPases similar to other AAA+ ATPases may be that the

effector chaperone complex docks to the ATPase ring complex, unfolding the effector while threading it through the ATPase ring pore. The effector docking site is present toward the C-terminal side of the ATPase ring where the C terminal side faces towards the membrane side away from the cytosol [88, 89]. These results create a great confusion of how these ATPases perform their function at the sorting platform. According to the hexamer model, when a hexamer ATPase ring complex is associated with the sorting platform the docking of effector-chaperone complex may result in possible steric hindrance for proper docking towards the buried C-terminal side of the ATPase ring complex. Instead, if the ATPase ring complex be a dodecamer (as suggested in this study) then the possibility of steric hindrance will be zero as the effector chaperone complex may dock to the free C-terminal side of the dodecamer in the cytosol. Also, the buried C-terminal side of ATPase ring will be accessible to the positive regulator SctO according to the literature [89, 153, 154].

Another aspect of the proposed stacked dodecamer ATPase ring model will also provide the possibility of passing an unfolded effector through the narrow pore of the ATPase ring complex during the secretion process like other AAA + protein translocases [92]. It should be noted that the presence of higher order oligomer (probable dodecamer) at the export gate of T3SS has not been found in any *in-situ* experiment till date. Also, it is debatable that whether unfolded effector passes through the T3SS ATPase pore or not. The absence of the probable dodecamer ATPase ring complex observed in *in-situ* Cryo-EM structures suggests the possibility that it may be required only during engagement with the effector chaperone complex representing the secretion competent phase of T3SS.

The T3SS ATPases in pathogenic bacteria are one of the most highly conserved proteins and are very crucial for their virulence hence they are a vey good drug target. Also, understanding the structure and function of these ATPase and how they regulate T3SS is necessary in this regard. The study presented here involving actual purification and characterization of oligomer formation and activation of a T3SS ATPase YsaN will provides a template for expanding our knowledge to understand T3SS regulation by these ATPases. Higher order oligomer of YsaN is the most active form and its formation can be summed up as a two-step kinetic process. However, we could not establish the exact nature of YsaN higher-order oligomer in this study. Eventually, further future experiments are also necessary to explore in depth about the true mechanism of oligomeric ATPase ring complexes in T3SS regulation in *Y. enterocolitica* and other pathogenic bacteria.

Chapter 5

5. The effector-chaperone system of *Yersinia enterocolitica*-case of YopE-SycE
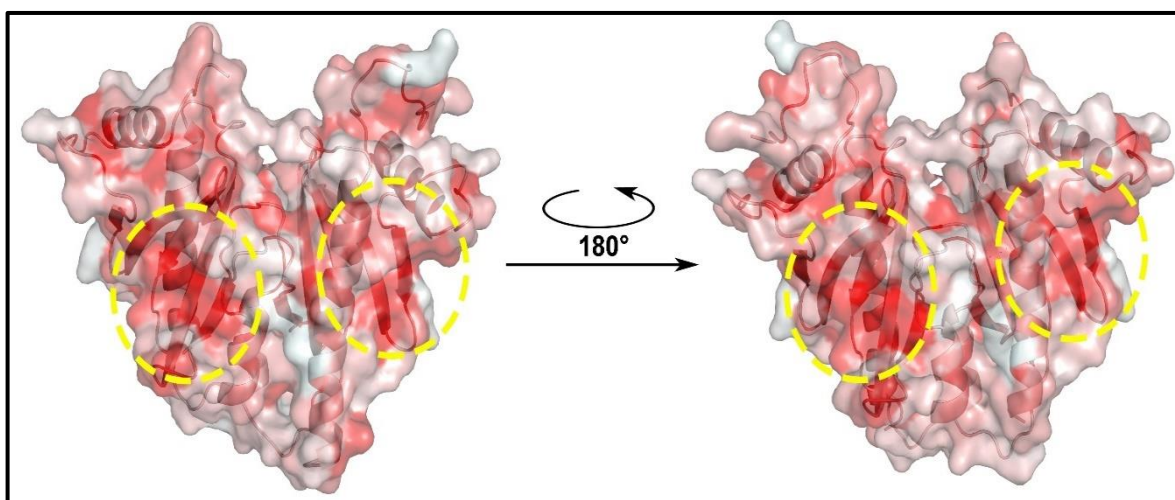
Figure of crystal structure of SycE representing its hydrophobic surface. The area highlighted by yellow dashed circle shows the hydrophobic pocket made up of beta sheet structure. This hydrophobic pocket is required by YopE to bind to SycE in extended fashion.

5.1.Introduction

Gram-negative pathogen like *Yersinia* species secrete effectors into the host macrophage cell to overcome their immune response. These effectors are presented to the secretion apparatus as protein-chaperone complexes. Few examples include YopE-SycE, YopH-SycH and YopT-SycT [57, 155]. YopE (a Rho GTPase activator-RhoGAP) protein [156] acts by disrupting the actin cytoskeleton [107, 157]. YopE has been suggested to interact and regulate via Rac1 and RhoA *in-vivo*. YopE translocation can also autoregulate its expression during infection [158]. In another study it has been suggested that GAP activity of YopE is also required during pore formation [159]. YopE plays a crucial role in virulence in association with its specific chaperone SycE [104, 156, 160]. Residue 54 to 75 of YopE comprises the **Membrane Localization Domain** (MLD). The MLD region of YopE has been suggested as a critical component and is required for proper intracellular localization within the host cell necessary for virulence of these effectors [161]. The MLD remains associated to the SycE hydrophobic region which is required for YopE interaction. The residues 96 to 219 of YopE comprise the catalytic **Rho GAP domain** while, the first 15 amino acids of YopE present at the N-terminal, acts as a non-cleavable signal sequence [162, 163]. In *Y. pseudotuberculosis,* YopE (~23kDa protein), binds to SycE(~14.6kDa) homo-dimer in an extended fashion through its **chaperone binding domain** (Cb domain: 23-78 amino acid residues). This interaction results in secondary structure formation in YopECb domain which is unstructured or absent when YopE is present alone in the solution. This YopE interaction with SycE dimer is also suggested to act as three-dimensional translocation signal recognized by the T3SS sorting complex [117].

SycE belongs to a large family of proteins whose sequence are not conserved however; they are structurally similar. Also, they have common physical property *i.e.,* they have small size and are acidic in nature [94]. T3SS effectors have their specific chaperone inside bacterial cell and they help in efficient translocation of their respective effectors. SycE exist as a global dimer structure and physically interacts with the YopE [96, 97] inside bacterial cytosol and they help secretion of YopE while they are not translocated into the host cell [115, 116]. SycE binds only to YopE containing the chaperone binding domain (Cb-domian) [106, 113, 164] which is different from the GAP domain (Ef- domain). The structure of SycE from *Y. pseudotuberculosis* [96] and *Y. pestis* [97] and *Y. enterocolitica* [100] has been already solved revealing its global dimeric structure in asymmetric unit of each crystal. Also, the crystal structure of SycE bound to the chaperone binding domain of YopE in *Y. pseudotuberculosis* (YopE-Cb) has been solved [95]. However, the physiological significance of this interaction between YopE and SycE is not yet understood clearly. How the three-dimensional signal which is structurally conserved plays its role is also not clear. Answers to such questions can be given only through visualizing its atomic structures.

The present chapter describes and characterize real time purification of YopE, SycE, and YopE-SycE complex of *Y. enterocolitica*. This chapter also present a 2Å X-Ray crystal structure of SycE as its monomer unit in the unit cell and another SycE crystal structure 2.7Å resolution representing its dimeric state in the unit cell giving its structural insight.

5.2. Materials and methods:

5.2.1. Designing of the recombinant expression vectors:

A comprehensive list of different gene construct used in this study is given in **Table 5.1**. In brief, YopE and SycE gene were PCR amplified from the plasmid isolated from *Y. enterocolitica* and cloned using conventional cloning method given in materials and method section. For co-expression of YopE and SycE complex, first we cloned a YopE C-terminal His construct. This construct was cloned in two steps: 1) cloning of YopE C-terminal in pET22b vector and, 2) PCR amplifying the YopE C-terminal His fragment from first construct and cloning it in pETDuet1 vector. SycE no tag was cloned in pACYC Duet1 vector. Then, SycE no tag PACYC Duet1 construct and YopE C-terminal His pET Duet1 constructs were co-transformed in BL21 for co-expression. Primers used are listed in T**able 5.2.**

<table>
<tr><td colspan="4" align="center">**Table 5.1: List of clones/ vectors used in this study**</td></tr>
<tr><td>**Vectors /Plasmid construct name**</td><td>**Details**</td><td>**Protein size (Appx. MW)**</td><td>**Reference(s)/ Source**</td></tr>
<tr><td>**YopE C-terminal His tag pET22b**</td><td>YopE wild type cloned in pET22b with Nde1/BamH1 restriction site containing C-terminal His Tag.</td><td>26 kDa</td><td>This Study.</td></tr>
<tr><td>**SycE N-terminal His tag pET28a**</td><td>SycE wild type cloned in pET28a with Nde1/Xho1 restriction site containing N-terminal His Tag with thrombin cut site.</td><td>17 kDa</td><td>This Study.</td></tr>
<tr><td>**YopE C-terminal His tag pETDuet1**</td><td>YopE C-terminal His tag fragment PCR amplified from YopE C-terminal His tag pET22b construct containing C-terminal His Tag cloned in pETDuet1 with Nco1/Bgl2 restriction site.</td><td>26 kDa</td><td>This Study.</td></tr>
<tr><td>**SycE no tag pACYCDuet1**</td><td>No tag SycE wild type cloned in pACYCDuet1 with Nde1/Xho1 restriction site in MCS2.</td><td>15 kDa</td><td>This Study.</td></tr>
<tr><td>**YopE no tag pACYCDuet1**</td><td>No tag YopE wild type cloned in pACYCDuet1 with Nco1/Sal1 restriction site in MCS1.</td><td>23 kDa</td><td>This Study.</td></tr>
<tr><td>**YopE Cb C-terminal His tag pETDuet1**</td><td>YopE chaperone binding domain cloned in pETDuet1 with Nde1/BamH1 restriction site containing C-terminal His Tag cloned in similar way like wild type YopE C-terminal His tag in pETDuet1.</td><td>15kDa</td><td>This study.</td></tr>
</table>

Gene/ construct name	Forward (5'- 3')	Reverse (5'- 3')	Reference(s)/ Source
YopE C-terminal His tag pET22b	GGAATTC**CATATG**AAAATATCATCATTTATTTCTACATCACTGCCC	CG**GGATCC**GACATCAATGACAGTAATTGATGCATCTGTTG	IDT Technologies/ GCC Biotech/ This study
SycE N-terminal His tag pET28a	GGAATTC**CATATG**TATTCATTTGAACAAGCTATCAC	CCG**CTCGAG**TTATCAACTAAATGACCGTG	IDT Technologies/ This study
YopE C-terminal His tag pETDuet1	TATA**CCATGG**GCATGAAAATATCATCATTTATTTCTAC	CCG**AGATCT**TTAGTGGTGGTGGTG	GCC Biotech/ This study
SycE no tag pACYCDuet1	GGAATTC**CATATG**TATTCATTTGAACAAGCTATCAC	CCG**CTCGAG**TTATCAACTAAATGACCGTG	IDT Technologies/ This study
YopE no tag pACYCDuet1	TATA**CCATGG**GCATGAAAATATCATCATTTATTTCTAC	CCC**GTCGAC**TTATCACATCAATGACAGTAATTGATGCATCTGTTG	IDT Technologies/ This study
YopE Cb C-terminal His tag pETDuet1	GGAATTC**CATATG**AAAATATCATCATTTATTTCTACATCACTGCCC	GA**GGATCC**GATTTGATAAATTCAATCGCAGAG	IDT Technologies/ This study

Table 5.1: List of primers used in this study

***Note-** Restriction site highlighted in bold letters.

5.2.2.Expression and purification of YopE, SycE and YopE-SycE complex:

Expression constructs containing YopE C-terminal His Tag and SycE N-terminal His tag was transformed into BL21 cells for recombinant expression of individual proteins. Briefly, a large-scale culture (600ml LB culture) was given from a 5ml overnight culture at 37ºC in a shaker. Induction was given when the OD of the culture reaches at ~0.6 using IPTG (0.5mM working concentration) and the culture was kept for shaking (~180 rpm) for ~12-14 hours at 25ºC. The induced cells were harvested at 4ºC in centrifuge at 6000 rpm for 10 minutes. The pellet was resuspended in Sonication buffer (see **Table 3.6**). Cells were lysed by ultrasonication method. Debris was removed by centrifugation and the clear supernatant was loaded onto pre-equilibrated gravity flow Nickel column and incubated for ~1 hour. Flow through was passed by opening the cap at bottom and the beads were washed twice using the wash buffer (see **Table 3.6**). Then elution buffer (see **Table 3.6**) was added to the beads and incubated for ~30 minutes. Eluted protein was collected in a fresh tube and the elution was subjected to size exclusion chromatography column for further purification. Peak respective to the protein was collected and concentrated using concentrating column (10kDa cutoff). Purity of the purified sample was

checked on SDS gel. For co-expression of YopE-SycE complex, YopE C-terminal His tag plasmid and SycE no tag plasmid was co-transformed into BL21 cell. Also, in other case of complex co-expression YopE no tag plasmid and SycE N-terminal His tag plasmid was co-transformed into BL21 cell for recombinant expression. Similar cloning method was applied for co-expression of YopE Cb C-terminal His-SycE. Similar purification strategy was used in all cases. **Figure 5.2** shows a typical size exclusion profile of respective proteins. **Figure 5.4** represents SDS profile of respective peaks of proteins from figure 5.1.

5.2.3. Native mass spectroscopy analysis of SycE:

MALDI-TOF mass analysis of purified SycE was done using Applied Biosystems - 4700 Proteomics Analyzer. Freshly purified SycE N-terminal His tag was subjected to gel filtration chromatography for further purification of sample. Respective peak was collected and further subjected to HiPrep 26/10 desalting column for removing excess salt from the sample. The respective peak was collected and concentrated to ~1mg/ml. **Figure 5.3** shows the peak observed during MALDI-TOF analysis. The molecular weight of the sample was calculated according to the equation given below:

$$p = m/z$$

$$p1 = (Mr + z1)/z1$$
$$p2 = [Mr + (z1 - 1)]/(z1 - 1)$$

Where:

p = a peak in the mass spectrum

m = total mass of an ion

z = total charge

Mr = average mass of the protein

5.2.4. Glutaraldehyde chemical crosslinking:

To determine the nature of complex formation between YopE and SycE and YopE Cb and SycE glutaraldehyde crosslinking was performed as mentioned in materials and method section. Briefly the purified sample was treated with 0.3- 0.5% (v/v) glutaraldehyde and incubated at room temperature for ten minutes. Crosslinking reaction was stopped by adding 1X SDS gel loading dye. Samples were run on SDS gel (**Figure 5.4**)

5.3. Results

5.3.1. In-silico study and domain analysis of YopE:

Outer membrane virulence protein of *Y. enterocolitica* containing 218 amino acids has two predicted domains: first is its N-terminal domain (1-126 residues) and, second is the GTPase

activator domain (GAP domain). The first 19 residues of the N-terminal contain a signal sequence while, residues from 22-78 is suggested as chaperone binding domain and residues 127-195 contains its catalytic GAP domain. Also, residues 1-58 is disordered region [165] (see **Figure 5.1**).
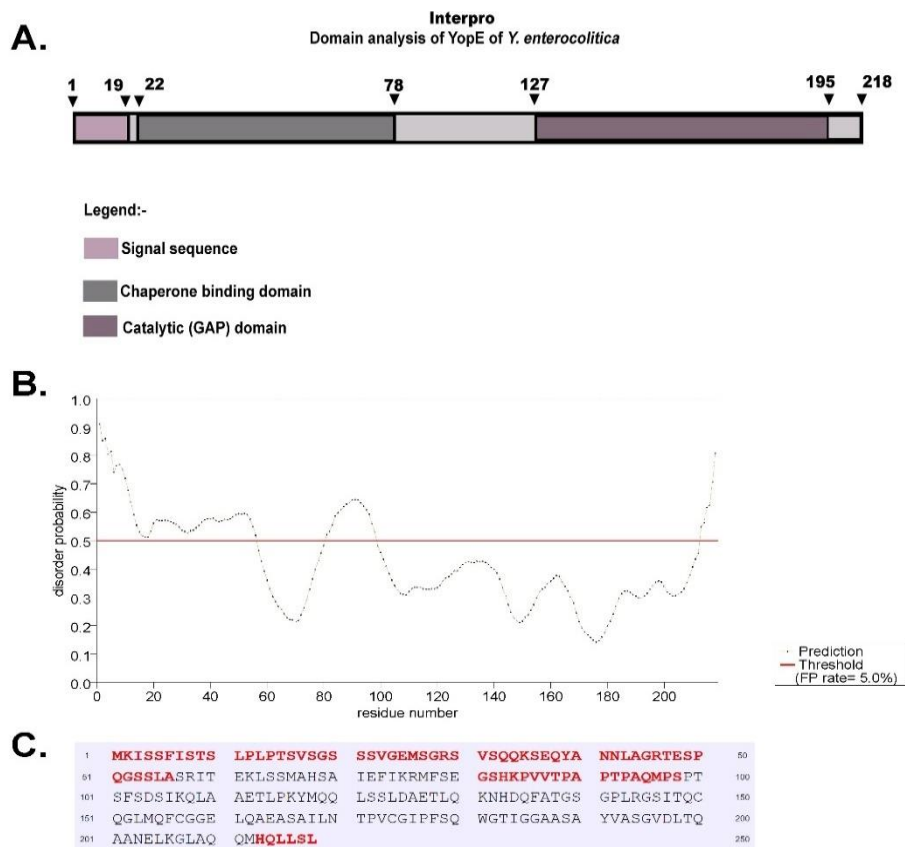
Figure 5.1: A.- Linear representation of domain prediction of YopE using interpro domain analyser. **B.-C.-** prediction of disorder regions in YopE using PrDOS.

5.3.2.YopE-SycE interaction and complex formation:

YopE is known to form a heterotrimeric complex with SycE chaperone in *Y. pseudotuberculosis* and *Y. pestis*. Expression of YopE alone resulted in formation of soluble aggregate (**Figure 5.2A**) while, SycE was highly soluble and remains as dimer in solution (see **Figure 5.2B and Figure 5.3**.). Hence, we tried co-expression of the YopE-SycE complex for their purification. Co-expression resulted in formation of heterotrimeric complex of YopE:SycE :: 1:2. Both, YopE C-terminal His and YopE no tag binds to SycE resulting in formation of YopE-SycE heterotrimeric complex (**Figure 5.2 C and D**). Like this YopE Cb C-terminal His (YopE containing only the chaperone binding domain) also interacts with SycE resulting in formation of the heterotrimeric complex (**Figure 5.2 E**). The peaks collected respective of each
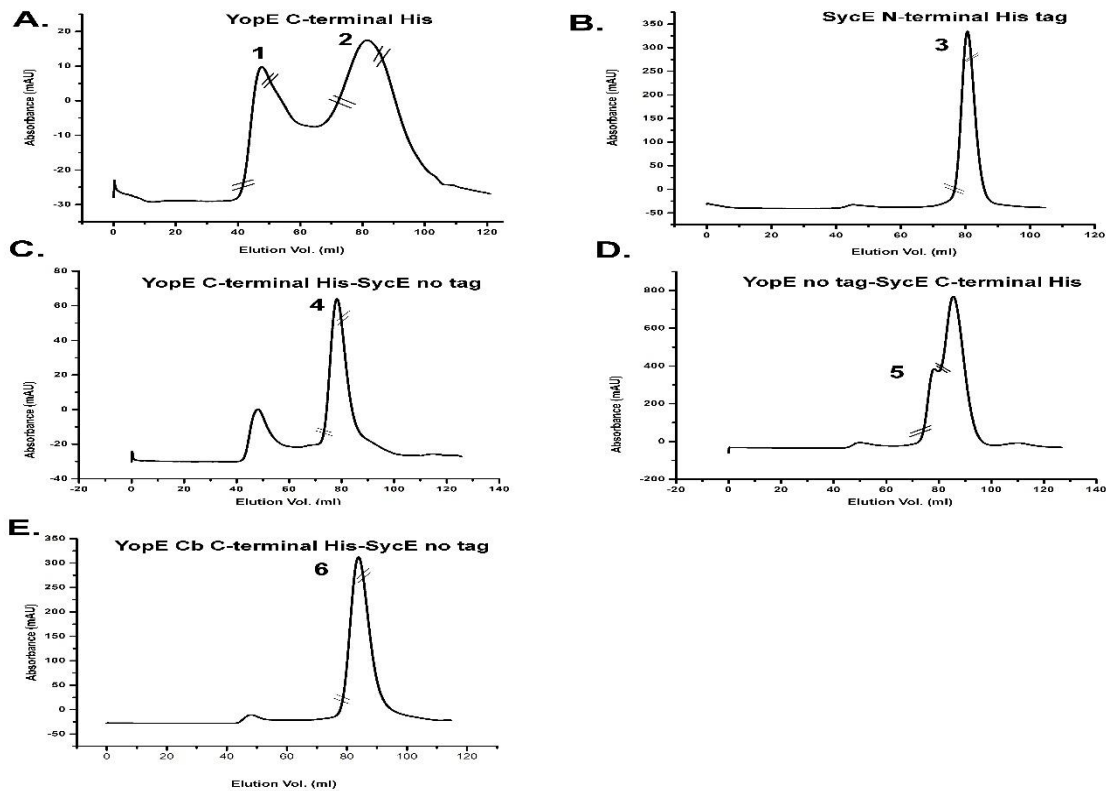
Figure 5.2: Size Exclusion profile of YopE, SycE and YopE-SycE complexes. **A.** Size exclusion curve of YopE C-terminal His. **B.** Curve representing profile of SycE. **C. & D.** size exclusion profile of YopE C-terminal His-SycE no tag and YopE no tag- SycE N-terminal His tag complex respectively. **E.** Curve representing size exclusion profile of YopE Cb C-terminal His- SycE complex. Size exclusion was performed using Superdex 200 16-60 column (GE Healthcare) at room temperature. Flow rate was kept at 1ml/minutes. All the respective peaks are delimited by cross sections of the chromatograms
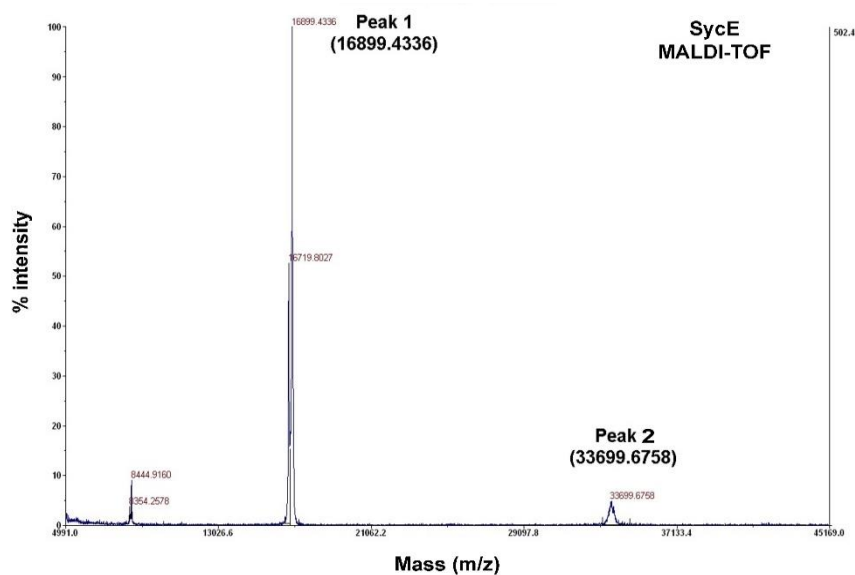
Figure 5.3: MALDI-TOF analysis profile of SycE. Different peaks representing differentially charged ions. The estimated molecular weight calculated from this peak is **33796.87** Dalton representing dimer state of SycE.

complex was analysed on SDS gel (Figure 5.4). This purified sample was further used for downstream experiments.
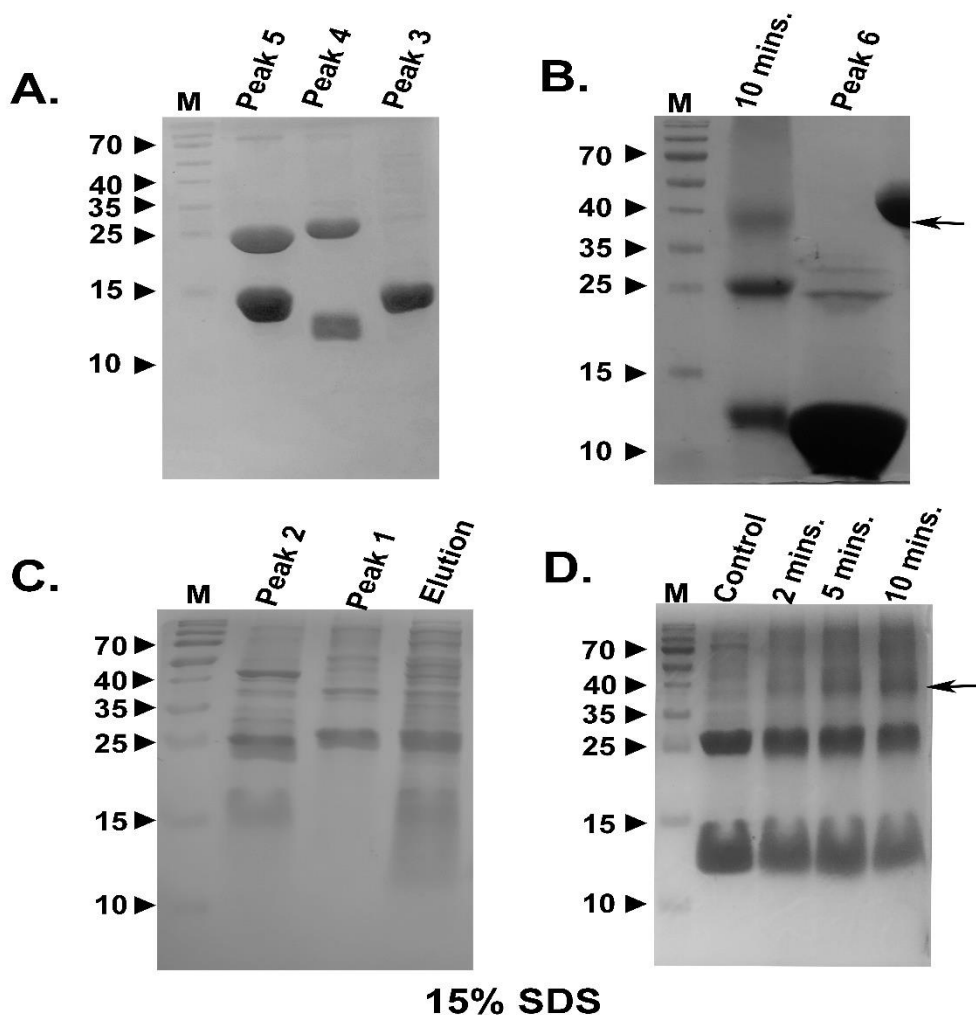
Figure 5.4: SDS profile of YopE, SycE proteins peaks collected from size exclusion chromatography column. **A.** Lane 2, 3, and 4 represents Peak 5, Peak 4, and Peak 3 respectively from figure 5.2. **B.** Glutaraldehyde crosslinking profile of YopE Cb-SycE his complex peak collected from figure 5.2 E. **C.** SDS profile of peaks from figure 5.2 A. **D.** Glutaraldehyde crosslinking profile of YopE C-terminal His-SycE no tag complex peak collected from figure 5.2 C and D. The heterotrimeric complex is indicated by arrow.

5.3.3. Crystallization trials of YopE-SycE complex and structure solving of SycE:

Crystallization trials and screening was conducted for YopE-SycE complex, YopE Cb-SycE complex, and SycE in vapour diffusion sitting drop method. Crystallization of YopE-SycE complex was unsuccessful because of degradation of YopE, while in few cases setting crystallization tray with commercial screen resulted in SycE crystals after some time. Similar observation was observed in the case of crystallization trials of YopE Cb-SycE complex. Crystallization of SycE (20mg/ml) resulted in diffraction quality crystals in more than one condition having similar morphology with commercial screen Rigaku Wizard 1, 2. We optimized a condition for SycE crystallization. SycE was purified in buffer containing 100mM Tris pH 8.0, 100mM NaCl, 2% glycerol. SycE (30mg/ml) was finally crystallized in crystallization buffer- 100mM MES pH 6.0- 6.5, 50-

80mM Na-Citrate, 5-10% PEG 3350, 100mM NaCl at 20ºC using vapour diffusion sitting method. Crystals appear in drop ratio was kept 1:1 with the reservoir solution within 2 weeks. In another crystallization trials of SycE we also tried additive screening method using screen from Hampton Research (USA). Larger size crystals of different morphology were observed within two weeks in two conditions. We optimized one condition containing 1M Potassium Sodium tartrate tetrahydrate. The final crystallization condition was 100mM MES pH 6.0- 6.5, 50-80mM Na-Citrate, 5-10% PEG 3350, 100mM NaCl and 1M Potassium Sodium tartrate tetrahydrate.

Preliminary screening of unit cell was done by mounting single crystal soaked in paratone-N (Hampton Research). The crystal diffracted well at ~2Å resolution. Full data collection was done at home X-ray source Bruker D8 venture using PHOTON III CCD detector. Data set collected were processed using PROTEUM3 data processing software suit. **Table 5.3** and **Table 5.4** represents the details of the data collection and processing. SycE N-terminal his was crystallized in the first condition (see **Figure 5.5 C) PDB ID:7EVA** contains a monomer unit in the asymmetric unit while in the other second condition containing 1M Sodium Potassium tartrate SycE crystallized as a dimer in the asymmetric unit cell (see **Figure 5.5 F).** In both the case the structure was determined using molecular replacement method using SycE crystal structure **(PDB ID 1n5b)** as template. Phasing was done using in built programme PHASER in PHENIX followed by refinement using PHENIX refine [166]. Model building and density adjustment was done using Coot software. The SycE structure containing monomer unit in the asymmetric unit has been submitted to PDB (**PDB ID:7EVA**) while the other SycE dimer structure is still unpublished.
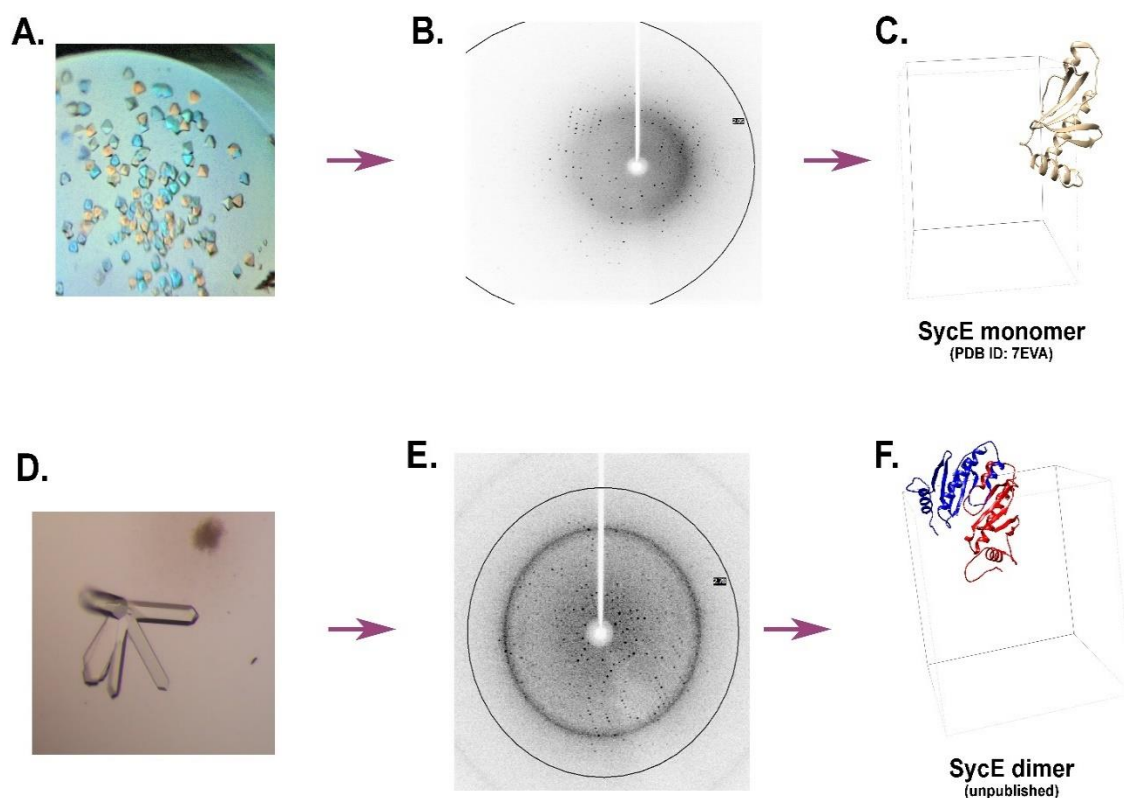
Figure 5.5: Crystallization of SycE. A. Crystals of SycE N-terminal his in buffer containing 100mM MES pH 6.0-6.5, 50-80mM Na-Citrate, 5-10% PEG 3350, and 100mM NaCl at 20ºC. **B.** Diffraction spot of crystal shown in A. **C.** Asymmetric unit cell containing SycE monomer. **D.** Crystals of SycE N-terminal his in buffer containing 100mM MES pH 6.0- 6.5, 50-80mM Na-Citrate, 5-10% PEG 3350, 100mM NaCl and 1M sodium potassium tartrate at 20ºC. **E.** Diffraction spot of crystal shown in D. **F.** Asymmetric unit cell containing SycE dimer.
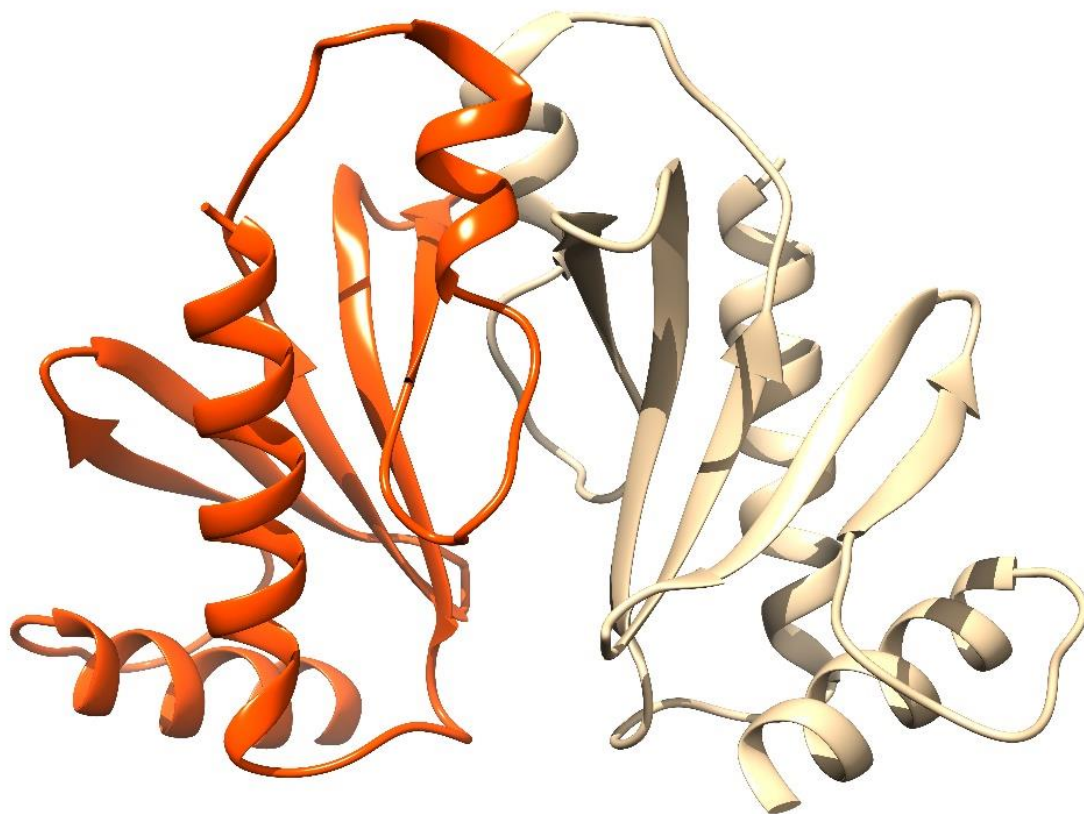
SycE
PDB ID: 7EVA
(Biological assembly)

Figure 5.6: Crystals of SycE N-terminal his (**PDB ID: 7EVA**), obtained in buffer containing 100mM MES pH 6.0-6.5, 50-80mM Na-Citrate, 5-10% PEG 3350, and 100mM NaCl at 20⁰C. The crystal contains SycE monomer in the asymmetric unit. Figure here represents the biological assembly of SycE dimer. Structure was visualized in Chimera software.

Table 5.3. Data collection and refinement statistics for SycE **(PDB ID: 7EVA)**

X-ray Wavelength	1.54 A
Detector Type	CCD
Space Group	P 43 21 2
Resolution Range	25.94 - 2.08
Unit cell - a, b, c,	58.00Å, 58.00Å, 77.00Å
α, β, γ	90.00, 90.00, 90.00
Number of unique reflections	8304
Completeness (%)	99.8 (25.94-2.08)
Mean I/sigma(I)	1.51
R$_{work}$	0.1995
R$_{free}$	0.2345
R$_{merge}$	0.20
Clashscore (%)	5.39
Ramachandran Favored (%)	94.78
Ramachandran Outliers (%)	0.87
Refinement program used	PHENIX 1.15.2-3472-000
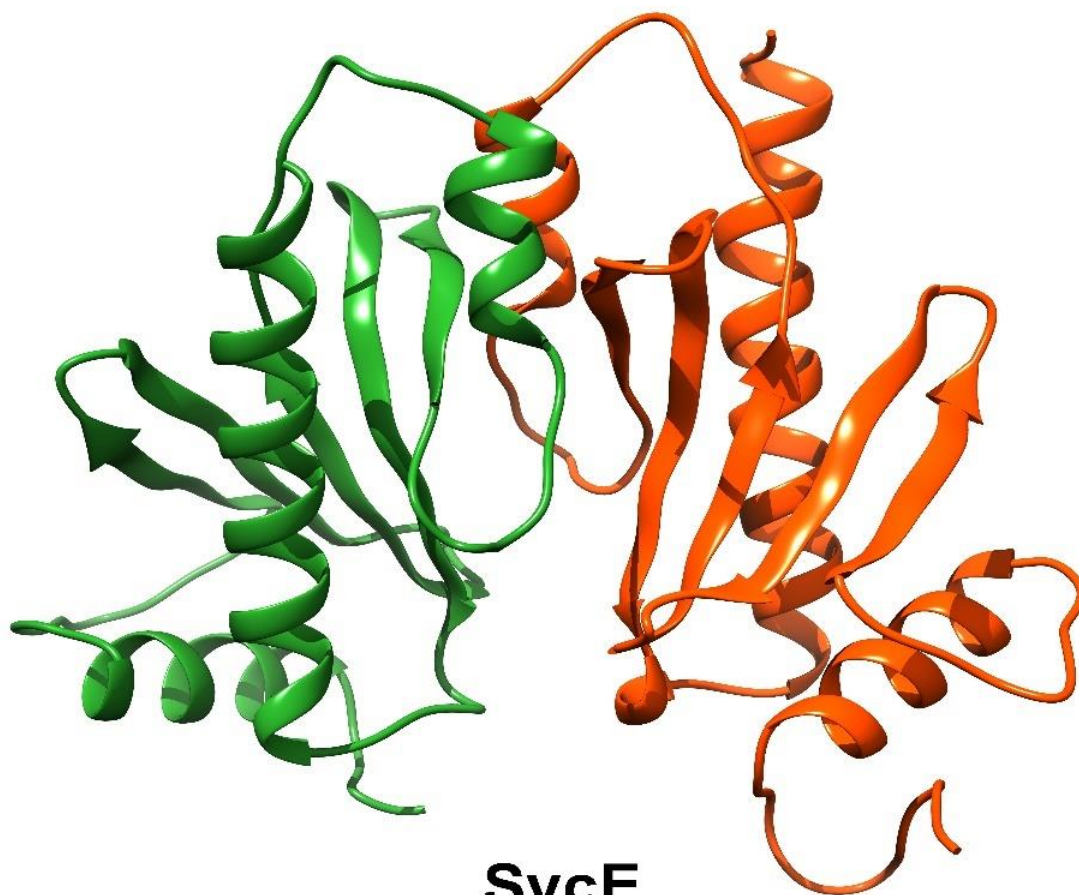
SycE
Unpublished
(Biological assembly)

Figure 5.7: Crystals of SycE N-terminal his (**unpublished**), obtained in buffer containing 100mM MES pH 6.0- 6.5, 50-80mM Na-Citrate, 5-10% PEG 3350, 100mM NaCl and 1M sodium potassium tartrate at 20°C. The crystal contains SycE dimer in the asymmetric unit. Figure here represents the biological assembly of SycE dimer. Structure visualization was done in Chimera software.

Table 5.4. Data collection and refinement statistics for SycE **(unpublished).**

X-ray Wavelength	1.54 A
Detector Type	CCD
Space Group	P 43 21 2
Resolution Range	26.613 - 2.789
Unit cell - a, b, c,	87.00Å, 87.00Å, 105.00Å
α, β, γ	90.00, 90.00, 90.00
Number of unique reflections	10505
Completeness (%)	99.82 (26.6 - 2.7)
Mean I/sigma(I)	4.34
R_{work}	0.2069
R_{free}	0.2668
Clashscore (%)	6.40
Ramachandran Favored (%)	96.08
Ramachandran Outliers (%)	0.34
Refinement program used	PHENIX 1.15.2-3472-000

5.3. Discussion:

YopE in *Y. pestis* and *Y. pseudotuberculosis* is a GTPase activator protein and regulates the host cellular pathway after being secreted inside the host cell. YopE in association with its cognate chaperone SycE is required for virulence [104, 156, 160]. Different distinct regions on YopE are required for different function. In the case of *Y. enterocolitica* the first 19 residues of its N-terminal domain contain the signal sequence. Residues 22 to 78 is the chaperone binding domain of YopE and is required for SycE interaction. YopE Cb C-terminal his can bind to SycE forming a heterotrimeric complex (see Figure 5.2 and Figure 5.4). YopE when expressed in the absence of SycE form soluble aggregate and its major fraction elutes in void volume during size exclusion chromatography. YopE in presence of SycE during co-expression results in formation of heterotrimeric complex. This suggest that SycE is indispensable for YopE stabilization. Similar to this has been suggested for YopE in other *Yersinia* species. Whereas, SycE when expressed alone is highly soluble and form dimer in solution.

Crystallization set up trial for complex crystallization was extensively performed using commercially available crystal screens such as Rigaku wizard screen and Hampton crystal screen. The degrading nature of YopE resulted in failure of complex crystallization trials. Similar observation was seen for YopE Cb his-SycE crystallization. After some time, crystals of SycE appears in the tray set up with the YopE- SycE complex. Whereas SycE shows a high tendency for crystallization. Crystals of SycE appears within 1-2 weeks in more than one condition. SycE being a homodimer in solution however in low salt condition, SycE form crystals with monomer in asymmetric unit (see SycE crystal **Figure 5.5**, PDB ID:7EVA). Whereas in one condition with high salt (i.e., 1M sodium potassium tartrate) SycE form crystal with different morphology but similar space group (P 43 21 1) containing its homodimer in the asymmetric unit. When both the SycE structures were aligned with each other they bear minor differences (RMSD value 0.344; see **Figure 5.8**). This suggest that SycE has highly stable structure.

In previous studies it has been suggested that SycE plays role in YopE binding and the keeping the YopE in soluble and secretion competent state [92]. This interaction is necessary for formation of three-dimensional secretion signal which required for docking to the export apparatus. However, how the YopE-SycE interaction plays role in secretion is still not understood properly. Indeed, future work is required to reveal necessary requirement for YopE-SycE complex formation and regulation of secretion through its structural motif.
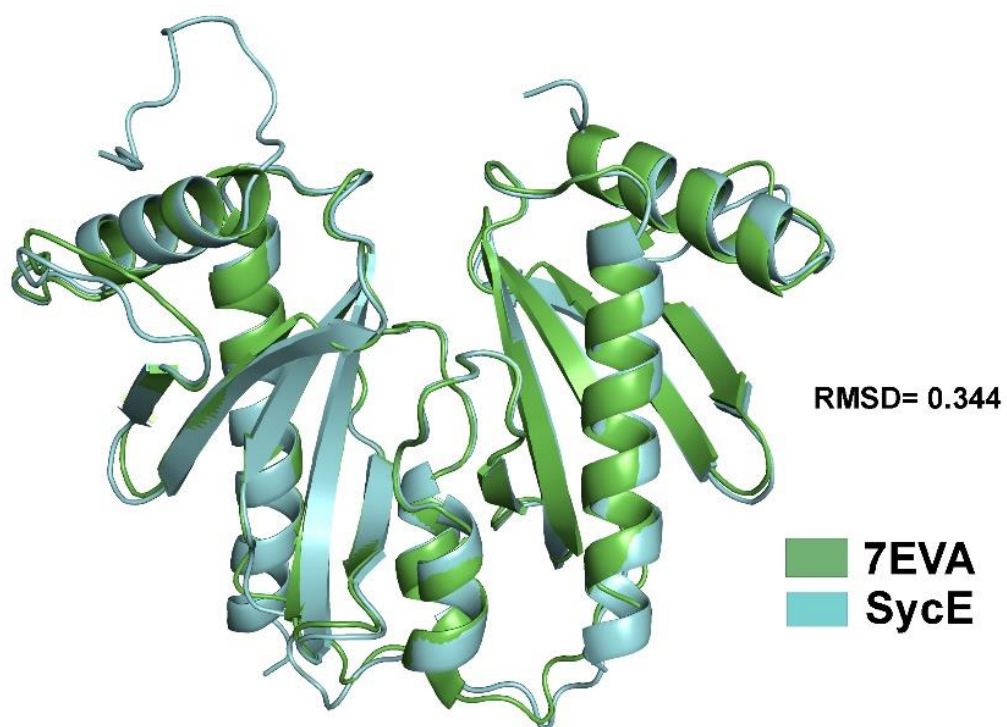
Figure 5.8: Structure alignment between 7EVA dimer and SycE dimer (unpublished). The alignment and visualization were done in PyMOL.

Chapter 6

6. Concluding remarks

Concluding remarks:

T3SS is required by many gram-negative pathogens and has been established as a major component, playing a crucial role in their virulence. In almost all the T3SS among different pathogenic bacteria the major structural component remains conserved, primarily the basal body components. All T3SS required a highly conserved AAA+ ATPase family protein for their function. Also, the same ATPase complex is present at the export gate plays a crucial role in effector unfolding and translocation by physically interacting with the effector-chaperone complex. Another factor adding to the virulence of these pathogens are the large repertoire of the effectors and their specific chaperones. These chaperones play a role in efficient secretion of these effectors by keeping them in secretion competent state. The work presented in this thesis primarily focuses on understanding the regulation of T3SS in *Yersinia enterocolitica.* As *Y. enterocolitica* harbours two distinctly regulated T3SS: one is chromosomal based Ysa-Ysp T3SS and the second is Ysc-Yop T3SS encoded by a ~70kb virulent plasmid.

YsaN is the conserved ATPase of Ysa-Ysp T3SS, however the mechanism of its activation remains elusive. Chapter 4 of this thesis represents the study involving actual purification and characterization of YsaN, and its different deletion and mutation construct revealing the insight into the mechanism of its activation and oligomerization. The role of its N-terminal region of YsaN in its activity was also revealed. The challenges during conduct of this study include insolubility of YsaN deletion which was overcome using a specialized vector (pET22bΔ50CPD vector) system mentioned in detail in chapter 4. Using this vector also results in mostly monomeric population of target proteins which was a crucial step in this study. During TEM, visualization of the higher-order oligomer of YsaN was one of the challenging tasks as YsaN has a great tendency to form non-specific aggregates *in-vitro*. This was partially overcome reducing the protein concentration in micromolar to nanomolar range, which results in fewer oligomer particles observed in TEM images. Presence of hexamer of ATPase in a deletion variant of YsaN (YsaNΔ83) suggest that YsaN can form hexamer in the absence of its 83 N-terminal residues which is unique to such ATPase and observed for the first time. Reduced ATP hydrolysis activity and formation of hexamer in the case of YsaNΔ83 suggest that the hexamer may not be the fully functional form of ATPase ring present at the export gate of T3SS during secretion phase. The ATP concentration dependence of YsaN oligomerization suggest that such ATPases can form its fully functional oligomeric state (probable dodecamer state in case of YsaN) at the export gate only when a sufficient ATP threshold has been achieved, suggesting that secretion of effectors is a high energy consuming process. Similar to other AAA+ ATPases, YsaN also has a tendency to oligomerize only after the active catalysis of phosphodiester bond of ATP. The study presented here overall suggest that such T3SS ATPase function like other AAA+- ATPases hence, suggest a universal mechanism among such family of proteins.

Prior to secretion of effectors through the effector-chaperone complex interacts with the ATPase ring complex. *Y. enterocolitica* bears two distinctly operating T3SS. Few of the known effectors of Ysc-Yop T3SS has been identified to be secreted through Ysa-Ysp T3SS such as YopE. Before getting secreted the YopE-SycE complex must interact with YsaN ATPase for unfolding and secretion through the T3SS needle complex. It has been suggested that a three-dimensional secretion signal formed only when YopE remain associated with SycE through its chaperone binding domain. YopE when expressed alone results in soluble aggregate. SycE interaction with the YopE is required to maintain YopE in secretion competent state. Hence, we tried co-expression of YopE and SycE together. YopE co-expression results in soluble heterotrimeric complex (YopE:SycE :: 1:2). However, this complex was not sufficient to protect YopE degradation after purification in our case. This stability issue hindered crystallization of YopE-SycE complex. Even the chaperone binding region of YopE forming a heterotrimeric complex (YopECb-SycE) was also not stable. Setting crystal tray with YopECb-SycE purified protein resulted in SycE crystal after some time due to degradation of YopECb. However, we were able to crystalize SycE in two different conditions. SycE exist as a dimer in solution however in one crystallization condition the unit cell contains SycE monomer. Generating the symmetry mates shows the dimeric nature of SycE. In other crystallization condition presence of high salt resulted in crystallization of SycE with dimer in unit cell. In both the case of SycE crystallization the space group remains same.

Significant understanding of T3SS regulation by the soluble cytosolic components has been known. The dynamic nature of the association of cytosolic components to the injectisome restricts *in-situ* studies. However, recent cryo-EM studies at high resolution have provided detail information of the cytoplasmic complex associated still the exact role of ATPase in T3SS regulation is not understood. This work will provide a template to explore the possibility of ATPase higher order complexes in regulating type three secretion. It should be noted here that presence of dodecamer complex of ATPase has not been observed in any *in-situ* studies till date whereas presence of hexamer has been observed in recent *in-situ* Cryo-EM study. Future studies are certainly required to get a better picture of how these ATPase complexes regulate T3SS and secretion of effectors.

Chapter 7

7. References
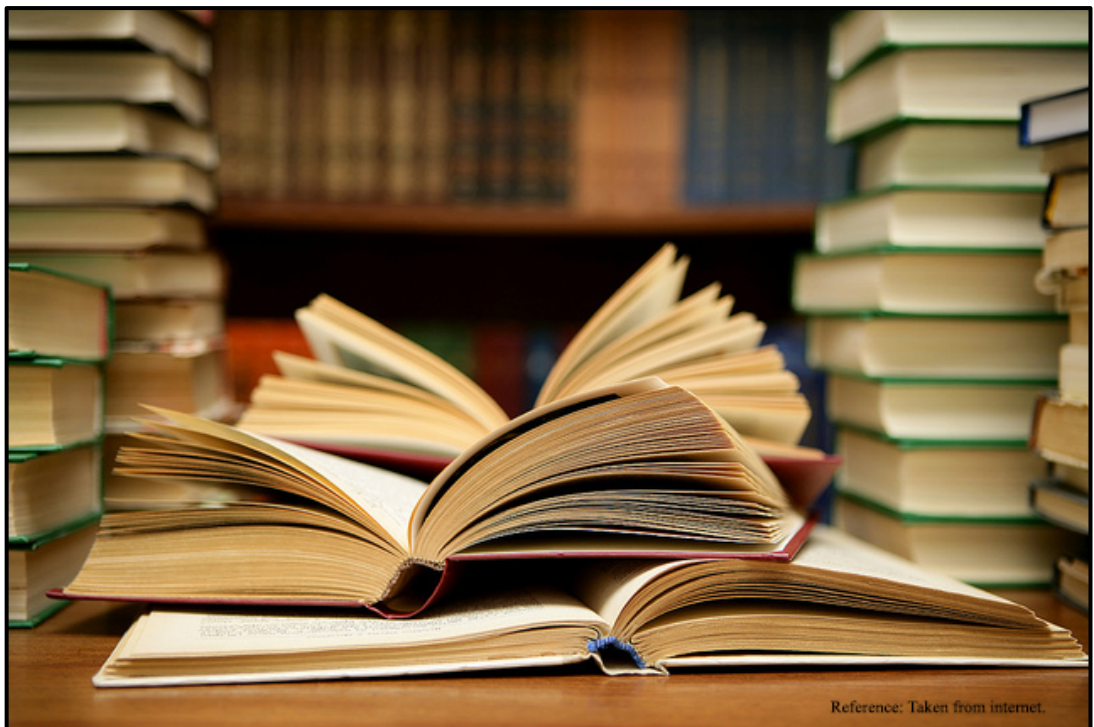
Reference: Taken from internet.

References:

1.	Van Loghem, J.J., *The classification of the plague-bacillus.* Antonie Van Leeuwenhoek, 1944. **10**(1-2): p. 15.

2.	Ewing, W.H., et al., *Yersinia ruckeri sp. nov., the Redmouth (RM) Bacterium.* International Journal of Systematic Bacteriology, 1978. **28**(1): p. 37-44.

3.	Perry, R.D. and J.D. Fetherston, *Yersinia pestis--etiologic agent of plague.* Clinical Microbiology Reviews, 1997. **10**(1): p. 35-66.

4.	Thoerner, P., et al., *PCR detection of virulence genes in Yersinia enterocolitica and Yersinia pseudotuberculosis and investigation of virulence gene distribution.* Appl Environ Microbiol, 2003. **69**(3): p. 1810-6.

5.	Wang, X., et al., *Distribution of pathogenic Yersinia enterocolitica in China.* Eur J Clin Microbiol Infect Dis, 2009. **28**(10): p. 1237-44.

6.	Naktin, J. and K.G. Beavis, *Yersinia enterocolitica and Yersinia pseudotuberculosis.* Clin Lab Med, 1999. **19**(3): p. 523-36, vi.

7.	Wren, B.W., *The yersiniae--a model genus to study the rapid evolution of bacterial pathogens.* Nat Rev Microbiol, 2003. **1**(1): p. 55-64.

8.	Achtman, M., et al., *Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 14043-8.

9.	Achtman, M., et al., *Microevolution and history of the plague bacillus, Yersinia pestis.* Proc Natl Acad Sci U S A, 2004. **101**(51): p. 17837-42.

10.	Ibrahim, A., et al., *The phylogeny of the genus Yersinia based on 16S rDNA sequences.* FEMS Microbiol Lett, 1993. **114**(2): p. 173-7.

11.	Wauters, G., K. Kandolo, and M. Janssens, *Revised biogrouping scheme of Yersinia enterocolitica.* Contrib Microbiol Immunol, 1987. **9**: p. 14-21.

12.	McNally, A., et al., *Comparison of the biotypes of Yersinia enterocolitica isolated from pigs, cattle and sheep at slaughter and from humans with yersiniosis in Great Britain during 1999-2000.* Lett Appl Microbiol, 2004. **39**(1): p. 103-8.

13.	Van Noyen, R., et al., *Yersinia enterocolitica: its isolation by cold enrichment from patients and healthy subjects.* J Clin Pathol, 1981. **34**(9): p. 1052-6.

14.	Bottone, E.J., *Yersinia enterocolitica: the charisma continues.* Clin Microbiol Rev, 1997. **10**(2): p. 257-76.

15.	Lian, C.J., et al., *Invasiveness of Yersinia enterocolitica lacking the virulence plasmid: an in-vivo study.* J Med Microbiol, 1987. **24**(3): p. 219-26.

16. Terashima, H., et al., *In Vitro Reconstitution of Functional Type III Protein Export and Insights into Flagellar Assembly.* mBio, 2018. **9**(3).

17. Papanikou, E., S. Karamanou, and A. Economou, *Bacterial protein secretion through the translocase nanomachine.* Nat Rev Microbiol, 2007. **5**(11): p. 839-51.

18. Robinson, C. and A. Bolhuis, *Tat-dependent protein targeting in prokaryotes and chloroplasts.* Biochim Biophys Acta, 2004. **1694**(1-3): p. 135-47.

19. Korotkov, K.V., M. Sandkvist, and W.G. Hol, *The type II secretion system: biogenesis, molecular architecture and mechanism.* Nat Rev Microbiol, 2012. **10**(5): p. 336-51.

20. Tsirigotaki, A., et al., *Protein export through the bacterial Sec pathway.* Nat Rev Microbiol, 2017. **15**(1): p. 21-36.

21. Muller, M., *Twin-arginine-specific protein export in Escherichia coli.* Res Microbiol, 2005. **156**(2): p. 131-6.

22. Ochsner, U.A., et al., *Effects of the twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis.* Proc Natl Acad Sci U S A, 2002. **99**(12): p. 8312-7.

23. Lavander, M., et al., *The twin arginine translocation system is essential for virulence of Yersinia pseudotuberculosis.* Infect Immun, 2006. **74**(3): p. 1768-76.

24. Pradel, N., et al., *Contribution of the twin arginine translocation system to the virulence of enterohemorrhagic Escherichia coli O157:H7.* Infect Immun, 2003. **71**(9): p. 4908-16.

25. Symmons, M.F., et al., *The assembled structure of a complete tripartite bacterial multidrug efflux pump.* Proc Natl Acad Sci U S A, 2009. **106**(17): p. 7173-8.

26. Thomas, S., I.B. Holland, and L. Schmitt, *The Type 1 secretion pathway - the hemolysin system and beyond.* Biochim Biophys Acta, 2014. **1843**(8): p. 1629-41.

27. Pimenta, A.L., et al., *Antibody analysis of the localisation, expression and stability of HlyD, the MFP component of the E. coli haemolysin translocator.* Mol Gen Genet, 1999. **261**(1): p. 122-32.

28. Lee, M., et al., *Membrane fusion proteins of type I secretion system and tripartite efflux pumps share a binding motif for TolC in gram-negative bacteria.* PLoS One, 2012. **7**(7): p. e40460.

29. Welch, R.A., et al., *Haemolysin contributes to virulence of extra-intestinal E. coli infections.* Nature, 1981. **294**(5842): p. 665-7.

30. Pugsley, A.P., *The complete general secretory pathway in gram-negative bacteria.* Microbiol Rev, 1993. **57**(1): p. 50-108.

31. Green, E.R. and J. Mecsas, *Bacterial Secretion Systems: An Overview.* Microbiol Spectr, 2016. **4**(1).

32. Buttner, D., *Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria.* Microbiol Mol Biol Rev, 2012. **76**(2): p. 262-310.

33. Cornelis, G.R., *The type III secretion injectisome.* Nat Rev Microbiol, 2006. **4**(11): p. 811-25.

34. Gaytan, M.O., et al., *Type Three Secretion System in Attaching and Effacing Pathogens.* Front Cell Infect Microbiol, 2016. **6**: p. 129.

35. Troisfontaines, P. and G.R. Cornelis, *Type III secretion: more systems than you think.* Physiology (Bethesda), 2005. **20**: p. 326-39.

36. Abrusci, P., et al., *Building a secreting nanomachine: a structural overview of the T3SS.* Curr Opin Struct Biol, 2014. **25**: p. 111-7.

37. Radics, J., L. Konigsmaier, and T.C. Marlovits, *Structure of a pathogenic type 3 secretion system in action.* Nat Struct Mol Biol, 2014. **21**(1): p. 82-7.

38. Lee, P.C. and A. Rietsch, *Fueling type III secretion.* Trends Microbiol, 2015. **23**(5): p. 296-300.

39. Price, S.B., et al., *The Yersinia pestis V antigen is a regulatory protein necessary for Ca2(+)-dependent growth and maximal expression of low-Ca2+ response virulence genes.* J Bacteriol, 1991. **173**(8): p. 2649-57.

40. Picking, W.L., et al., *IpaD of Shigella flexneri is independently required for regulation of Ipa protein secretion and efficient insertion of IpaB and IpaC into host membranes.* Infect Immun, 2005. **73**(3): p. 1432-40.

41. Holmstrom, A., et al., *LcrV is a channel size-determining component of the Yop effector translocon of Yersinia.* Mol Microbiol, 2001. **39**(3): p. 620-32.

42. Wallden, K., A. Rivera-Calzada, and G. Waksman, *Type IV secretion systems: versatility and diversity in function.* Cell Microbiol, 2010. **12**(9): p. 1203-12.

43. Bundock, P., et al., *Trans-kingdom T-DNA transfer from Agrobacterium tumefaciens to Saccharomyces cerevisiae.* EMBO J, 1995. **14**(13): p. 3206-14.

44. Fronzes, R., P.J. Christie, and G. Waksman, *The structural biology of type IV secretion systems.* Nat Rev Microbiol, 2009. **7**(10): p. 703-14.

45. Bayer-Santos, E., et al., *The opportunistic pathogen Stenotrophomonas maltophilia utilizes a type IV secretion system for interbacterial killing.* PLoS Pathog, 2019. **15**(9): p. e1007651.

46. Costa, T.R., et al., *Secretion systems in Gram-negative bacteria: structural and mechanistic insights.* Nat Rev Microbiol, 2015. **13**(6): p. 343-59.

47. Tseng, T.T., B.M. Tyler, and J.C. Setubal, *Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology.* BMC Microbiol, 2009. **9 Suppl 1**: p. S2.

48. Fan, E., et al., *Type V Secretion Systems in Bacteria.* Microbiol Spectr, 2016. **4**(1).

49. Junker, M., R.N. Besingi, and P.L. Clark, *Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion.* Mol Microbiol, 2009. **71**(5): p. 1323-32.

50.   Russell, A.B., S.B. Peterson, and J.D. Mougous, *Type VI secretion system effectors: poisons with a purpose.* Nat Rev Microbiol, 2014. **12**(2): p. 137-48.

51.   Navarro-Garcia, F., et al., *Type VI Secretion System in Pathogenic Escherichia coli: Structure, Role in Virulence, and Acquisition.* Front Microbiol, 2019. **10**: p. 1965.

52.   Coburn, B., I. Sekirov, and B.B. Finlay, *Type III secretion systems and disease.* Clin Microbiol Rev, 2007. **20**(4): p. 535-49.

53.   Abby, S.S. and E.P. Rocha, *The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems.* PLoS Genet, 2012. **8**(9): p. e1002983.

54.   Galan, J.E., et al., *Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells.* Annu Rev Microbiol, 2014. **68**: p. 415-38.

55.   Wagner, S. and J.E. Galán, *Bacterial type III protein secretion systems*. Vol. 427. 2020: Springer.

56.   Michiels, T., et al., *Secretion of Yop proteins by Yersiniae.* Infect Immun, 1990. **58**(9): p. 2840-9.

57.   Hueck, C.J., *Type III protein secretion systems in bacterial pathogens of animals and plants.* Microbiol Mol Biol Rev, 1998. **62**(2): p. 379-433.

58.   Izore, T., V. Job, and A. Dessen, *Biogenesis, regulation, and targeting of the type III secretion system.* Structure, 2011. **19**(5): p. 603-12.

59.   Galan, J.E., C. Ginocchio, and P. Costeas, *Molecular and functional characterization of the Salmonella invasion gene invA: homology of InvA to members of a new protein family.* J Bacteriol, 1992. **174**(13): p. 4338-49.

60.   Groisman, E.A. and H. Ochman, *Cognate gene clusters govern invasion of host epithelial cells by Salmonella typhimurium and Shigella flexneri.* EMBO J, 1993. **12**(10): p. 3779-87.

61.   Fields, K.A., G.V. Plano, and S.C. Straley, *A low-Ca2+ response (LCR) secretion (ysc) locus lies within the lcrB region of the LCR plasmid in Yersinia pestis.* J Bacteriol, 1994. **176**(3): p. 569-79.

62.   Allaoui, A., et al., *YscU, a Yersinia enterocolitica inner membrane protein involved in Yop secretion.* J Bacteriol, 1994. **176**(15): p. 4534-42.

63.   Wagner, S., et al., *Organization and coordinated assembly of the type III secretion export apparatus.* Proc Natl Acad Sci U S A, 2010. **107**(41): p. 17745-50.

64.   Zilkenat, S., et al., *Determination of the Stoichiometry of the Complete Bacterial Type III Secretion Needle Complex Using a Combined Quantitative Proteomic Approach.* Mol Cell Proteomics, 2016. **15**(5): p. 1598-609.

65.   Dietsche, T., et al., *Structural and Functional Characterization of the Bacterial Type III Secretion Export Apparatus.* PLoS Pathog, 2016. **12**(12): p. e1006071.

66. Torres-Vargas, C.E., et al., *The inner rod of virulence-associated type III secretion systems constitutes a needle adapter of one helical turn that is deeply integrated into the system's export apparatus.* Mol Microbiol, 2019. **112**(3): p. 918-931.

67. Hu, J., et al., *Cryo-EM analysis of the T3S injectisome reveals the structure of the needle and open secretin.* Nat Commun, 2018. **9**(1): p. 3840.

68. Lefebre, M.D. and J.E. Galan, *The inner rod protein controls substrate switching and needle length in a Salmonella type III secretion system.* Proc Natl Acad Sci U S A, 2014. **111**(2): p. 817-22.

69. Mota, L.J. and G.R. Cornelis, *The bacterial injection kit: type III secretion systems.* Ann Med, 2005. **37**(4): p. 234-49.

70. Journet, L., et al., *The needle length of bacterial injectisomes is determined by a molecular ruler.* Science, 2003. **302**(5651): p. 1757-60.

71. Marlovits, T.C., et al., *Assembly of the inner rod determines needle length in the type III secretion injectisome.* Nature, 2006. **441**(7093): p. 637-40.

72. Quinaud, M., et al., *The PscE-PscF-PscG complex controls type III secretion needle biogenesis in Pseudomonas aeruginosa.* J Biol Chem, 2005. **280**(43): p. 36293-300.

73. Quinaud, M., et al., *Structure of the heterotrimeric complex that regulates type III secretion needle formation.* Proc Natl Acad Sci U S A, 2007. **104**(19): p. 7803-8.

74. Chatterjee, S., et al., *Structure and biophysics of type III secretion in bacteria.* Biochemistry, 2013. **52**(15): p. 2508-17.

75. Sal-Man, N., et al., *EscE and EscG are cochaperones for the type III needle protein EscF of enteropathogenic Escherichia coli.* J Bacteriol, 2013. **195**(11): p. 2481-9.

76. Mueller, C.A., et al., *The V-antigen of Yersinia forms a distinct structure at the tip of injectisome needles.* Science, 2005. **310**(5748): p. 674-6.

77. Johnson, S., et al., *Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD.* J Biol Chem, 2007. **282**(6): p. 4035-44.

78. Broz, P., et al., *Function and molecular architecture of the Yersinia injectisome tip complex.* Mol Microbiol, 2007. **65**(5): p. 1311-20.

79. Poyraz, O., et al., *Protein refolding is required for assembly of the type three secretion needle.* Nat Struct Mol Biol, 2010. **17**(7): p. 788-92.

80. Hakansson, S., et al., *The YopB protein of Yersinia pseudotuberculosis is essential for the translocation of Yop effector proteins across the target cell plasma membrane and displays a contact-dependent membrane disrupting activity.* EMBO J, 1996. **15**(21): p. 5812-23.

81. Blocker, A., et al., *The tripartite type III secreton of Shigella flexneri inserts IpaB and IpaC into host membranes.* J Cell Biol, 1999. **147**(3): p. 683-93.

82. Neyt, C. and G.R. Cornelis, *Insertion of a Yop translocation pore into the macrophage plasma membrane by Yersinia enterocolitica: requirement for translocators YopB and YopD, but not LcrG.* Mol Microbiol, 1999. **33**(5): p. 971-81.

83. Tardy, F., et al., *Yersinia enterocolitica type III secretion-translocation system: channel formation by secreted Yops.* EMBO J, 1999. **18**(23): p. 6793-9.

84. Marenne, M.N., et al., *Genetic analysis of the formation of the Ysc-Yop translocation pore in macrophages by Yersinia enterocolitica: role of LcrV, YscF and YopN.* Microb Pathog, 2003. **35**(6): p. 243-58.

85. Cherradi, Y., et al., *Interplay between predicted inner-rod and gatekeeper in controlling substrate specificity of the type III secretion system.* Mol Microbiol, 2013. **87**(6): p. 1183-99.

86. Lara-Tejero, M., et al., *A sorting platform determines the order of protein secretion in bacterial type III systems.* Science, 2011. **331**(6021): p. 1188-91.

87. Hu, B., et al., *In Situ Molecular Architecture of the Salmonella Type III Secretion Machine.* Cell, 2017. **168**(6): p. 1065-1074 e10.

88. Imada, K., et al., *Insight into the flagella type III export revealed by the complex structure of the type III ATPase and its regulator.* Proc Natl Acad Sci U S A, 2016. **113**(13): p. 3633-8.

89. Majewski, D.D., et al., *Cryo-EM structure of the homohexameric T3SS ATPase-central stalk complex reveals rotary ATPase-like asymmetry.* Nat Commun, 2019. **10**(1): p. 626.

90. Zhang, Y., et al., *Visualization and characterization of individual type III protein secretion machines in live bacteria.* Proc Natl Acad Sci U S A, 2017. **114**(23): p. 6098-6103.

91. Diepold, A., et al., *A dynamic and adaptive network of cytosolic interactions governs protein export by the T3SS injectisome.* Nat Commun, 2017. **8**: p. 15940.

92. Akeda, Y. and J.E. Galan, *Chaperone release and unfolding of substrates in type III secretion.* Nature, 2005. **437**(7060): p. 911-5.

93. Wattiau, P. and G.R. Cornelis, *SycE, a chaperone-like protein of Yersinia enterocolitica involved in Ohe secretion of YopE.* Mol Microbiol, 1993. **8**(1): p. 123-31.

94. Wattiau, P., et al., *Individual chaperones required for Yop secretion by Yersinia.* Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10493-7.

95. Birtalan, S.C., R.M. Phillips, and P. Ghosh, *Three-dimensional secretion signals in chaperone-effector complexes of bacterial pathogens.* Mol Cell, 2002. **9**(5): p. 971-80.

96. Birtalan, S. and P. Ghosh, *Structure of the Yersinia type III secretory system chaperone SycE.* Nat Struct Biol, 2001. **8**(11): p. 974-8.

97. Evdokimov, A.G., et al., *Three-dimensional structure of the type III secretion chaperone SycE from Yersinia pestis.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 3): p. 398-406.

98.   Luo, Y., et al., *Structural and biochemical characterization of the type III secretion chaperones CesT and SigE.* Nat Struct Biol, 2001. **8**(12): p. 1031-6.

99.   Stebbins, C.E. and J.E. Galan, *Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion.* Nature, 2001. **414**(6859): p. 77-81.

100.   Trame, C.B. and D.B. McKay, *Structure of the Yersinia enterocolitica molecular-chaperone protein SycE.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 2): p. 389-92.

101.   van Eerde, A., et al., *Structure of Spa15, a type III secretion chaperone from Shigella flexneri with broad specificity.* EMBO Rep, 2004. **5**(5): p. 477-83.

102.   Cornelis, G.R., et al., *The virulence plasmid of Yersinia, an antihost genome.* Microbiol Mol Biol Rev, 1998. **62**(4): p. 1315-52.

103.   Matsumoto, H. and G.M. Young, *Translocated effectors of Yersinia.* Curr Opin Microbiol, 2009. **12**(1): p. 94-100.

104.   Rosqvist, R., et al., *The cytotoxic protein YopE of Yersinia obstructs the primary host defence.* Mol Microbiol, 1990. **4**(4): p. 657-67.

105.   Rosqvist, R., A. Forsberg, and H. Wolf-Watz, *Intracellular targeting of the Yersinia YopE cytotoxin in mammalian cells induces actin microfilament disruption.* Infect Immun, 1991. **59**(12): p. 4562-9.

106.   Sory, M.P., et al., *Identification of the YopE and YopH domains required for secretion and internalization into the cytosol of macrophages, using the cyaA gene fusion approach.* Proc Natl Acad Sci U S A, 1995. **92**(26): p. 11998-2002.

107.   Von Pawel-Rammingen, U., et al., *GAP activity of the Yersinia YopE cytotoxin specifically targets the Rho pathway: a mechanism for disruption of actin microfilament structure.* Mol Microbiol, 2000. **36**(3): p. 737-48.

108.   Foultier, B., et al., *Identification of substrates and chaperone from the Yersinia enterocolitica 1B Ysa type III secretion system.* Infect Immun, 2003. **71**(1): p. 242-53.

109.   Michiels, T. and G.R. Cornelis, *Secretion of hybrid proteins by the Yersinia Yop export system.* J Bacteriol, 1991. **173**(5): p. 1677-85.

110.   Sory, M.P. and G.R. Cornelis, *Translocation of a hybrid YopE-adenylate cyclase from Yersinia enterocolitica into HeLa cells.* Mol Microbiol, 1994. **14**(3): p. 583-94.

111.   Young, B.M. and G.M. Young, *Evidence for targeting of Yop effectors by the chromosomally encoded Ysa type III secretion system of Yersinia enterocolitica.* J Bacteriol, 2002. **184**(20): p. 5563-71.

112.   Matsumoto, H. and G.M. Young, *Proteomic and functional analysis of the suite of Ysp proteins exported by the Ysa type III secretion system of Yersinia enterocolitica Biovar 1B.* Mol Microbiol, 2006. **59**(2): p. 689-706.

113. Cheng, L.W., D.M. Anderson, and O. Schneewind, *Two independent type III secretion mechanisms for YopE in Yersinia enterocolitica.* Mol Microbiol, 1997. **24**(4): p. 757-65.

114. Frithz-Lindsten, E., et al., *The chaperone-like protein YerA of Yersinia pseudotuberculosis stabilizes YopE in the cytoplasm but is dispensible for targeting to the secretion loci.* Mol Microbiol, 1995. **16**(4): p. 635-47.

115. Woestyn, S., et al., *The cytosolic SycE and SycH chaperones of Yersinia protect the region of YopE and YopH involved in translocation across eukaryotic cell membranes.* Mol Microbiol, 1996. **20**(6): p. 1261-71.

116. Schesser, K., E. Frithz-Lindsten, and H. Wolf-Watz, *Delineation and mutational analysis of the Yersinia pseudotuberculosis YopE domains which mediate translocation across bacterial and eukaryotic cellular membranes.* J Bacteriol, 1996. **178**(24): p. 7227-33.

117. Rodgers, L., et al., *The type III secretion chaperone SycE promotes a localized disorder-to-order transition in the natively unfolded effector YopE.* J Biol Chem, 2008. **283**(30): p. 20857-63.

118. Galan, J.E., *SnapShot: effector proteins of type III secretion systems.* Cell, 2007. **130**(1): p. 192.

119. Galan, J.E., *Common themes in the design and function of bacterial effectors.* Cell Host Microbe, 2009. **5**(6): p. 571-9.

120. Paul, K., et al., *Energy source of flagellar type III secretion.* Nature, 2008. **451**(7177): p. 489-92.

121. Minamino, T. and K. Namba, *Distinct roles of the FliI ATPase and proton motive force in bacterial flagellar protein export.* Nature, 2008. **451**(7177): p. 485-8.

122. Wilharm, G., et al., *Yersinia enterocolitica type III secretion depends on the proton motive force but not on the flagellar motor components MotA and MotB.* Infect Immun, 2004. **72**(7): p. 4004-9.

123. Wilson, K., *Preparation of genomic DNA from bacteria.* Curr Protoc Mol Biol, 2001. **Chapter 2**: p. Unit 2 4.

124. Ehrt, S. and D. Schnappinger, *Isolation of plasmids from E. coli by alkaline lysis.* Methods Mol Biol, 2003. **235**: p. 75-8.

125. Waterborg, J.H. and H.R. Matthews, *The Lowry method for protein quantitation.* Methods Mol Biol, 1994. **32**: p. 1-4.

126. Foultier, B., et al., *Characterization of the ysa pathogenicity locus in the chromosome of Yersinia enterocolitica and phylogeny analysis of type III secretion systems.* J Mol Evol, 2002. **55**(1): p. 37-51.

127. Bent, Z.W., et al., *Transcriptomic Analysis of Yersinia enterocolitica Biovar 1B Infecting Murine Macrophages Reveals New Mechanisms of Extracellular and Intracellular Survival.* Infect Immun, 2015. **83**(7): p. 2672-85.

128. Haller, J.C., et al., *A chromosomally encoded type III secretion pathway in Yersinia enterocolitica is important in virulence.* Mol Microbiol, 2000. **36**(6): p. 1436-46.

129. Chatterjee, R., P.K. Halder, and S. Datta, *Identification and molecular characterization of YsaL (Ye3555): a novel negative regulator of YsaN ATPase in type three secretion system of enteropathogenic bacteria Yersinia enterocolitica.* PLoS One, 2013. **8**(10): p. e75028.

130. Shen, A., et al., *Simplified, enhanced protein purification using an inducible, autoprocessing enzyme tag.* PLoS One, 2009. **4**(12): p. e8119.

131. Ho, S.N., et al., *Site-directed mutagenesis by overlap extension using the polymerase chain reaction.* Gene, 1989. **77**(1): p. 51-9.

132. Rappas, M., et al., *Structural insights into the activity of enhancer-binding proteins.* Science, 2005. **307**(5717): p. 1972-5.

133. Minamino, T., et al., *Oligomerization of the bacterial flagellar ATPase FliI is controlled by its extreme N-terminal region.* J Mol Biol, 2006. **360**(2): p. 510-9.

134. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Mol Syst Biol, 2011. **7**: p. 539.

135. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices.* J Mol Biol, 1999. **292**(2): p. 195-202.

136. Wilkins, M.R., et al., *Protein identification and analysis tools in the ExPASy server.* Methods Mol Biol, 1999. **112**: p. 531-52.

137. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nat Protoc, 2015. **10**(6): p. 845-58.

138. Webb, B. and A. Sali, *Comparative Protein Structure Modeling Using MODELLER.* Curr Protoc Bioinformatics, 2014. **47**: p. 5 6 1-32.

139. Chen, B., et al., *ATP ground- and transition states of bacterial enhancer binding AAA+ ATPases support complex formation with their target protein, sigma54.* Structure, 2007. **15**(4): p. 429-40.

140. Schumacher, J., et al., *ATP-dependent transcriptional activation by bacterial PspF AAA+protein.* J Mol Biol, 2004. **338**(5): p. 863-75.

141. Halder, P.K., C. Roy, and S. Datta, *Structural and functional characterization of type three secretion system ATPase PscN and its regulator PscL from Pseudomonas aeruginosa.* Proteins, 2019. **87**(4): p. 276-288.

142. Pozidis, C., et al., *Type III protein translocase: HrcN is a peripheral ATPase that is activated by oligomerization.* J Biol Chem, 2003. **278**(28): p. 25816-24.

143. Muller, S.A., et al., *Double hexameric ring assembly of the type III protein translocase ATPase HrcN.* Mol Microbiol, 2006. **61**(1): p. 119-25.

144. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

145. DeLano, W.L., *Pymol: An open-source molecular graphics tool.* CCP4 Newsletter on protein crystallography. **40**(1): p. 82-92.

146. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* J Comput Chem, 2010. **31**(2): p. 455-61.

147. Kozakov, D., et al., *How good is automated protein docking?* Proteins, 2013. **81**(12): p. 2159-66.

148. Kozakov, D., et al., *The ClusPro web server for protein-protein docking.* Nat Protoc, 2017. **12**(2): p. 255-278.

149. Andrade, A., et al., *Enzymatic characterization of the enteropathogenic Escherichia coli type III secretion ATPase EscN.* Arch Biochem Biophys, 2007. **468**(1): p. 121-7.

150. Burgess, J.L., et al., *Spa47 is an oligomerization-activated type three secretion system (T3SS) ATPase from Shigella flexneri.* Protein Sci, 2016. **25**(5): p. 1037-48.

151. Woestyn, S., et al., *YscN, the putative energizer of the Yersinia Yop secretion machinery.* J Bacteriol, 1994. **176**(6): p. 1561-9.

152. Martin, J.L., et al., *Anatomy of F1-ATPase powered rotation.* Proc Natl Acad Sci U S A, 2014. **111**(10): p. 3715-20.

153. Zarivach, R., et al., *Structural analysis of a prototypical ATPase from the type III secretion system.* Nat Struct Mol Biol, 2007. **14**(2): p. 131-7.

154. Kato, J., M. Lefebre, and J.E. Galan, *Structural Features Reminiscent of ATP-Driven Protein Translocases Are Essential for the Function of a Type III Secretion-Associated ATPase.* J Bacteriol, 2015. **197**(18): p. 3007-14.

155. Iriarte, M. and G.R. Cornelis, *YopT, a new Yersinia Yop effector protein, affects the cytoskeleton of host cells.* Mol Microbiol, 1998. **29**(3): p. 915-29.

156. Forsberg, A. and H. Wolf-Watz, *Genetic analysis of the yopE region of Yersinia spp.: identification of a novel conserved locus, yerA, regulating yopE expression.* J Bacteriol, 1990. **172**(3): p. 1547-55.

157. Black, D.S. and J.B. Bliska, *The RhoGAP activity of the Yersinia pseudotuberculosis cytotoxin YopE is required for antiphagocytic function and virulence.* Mol Microbiol, 2000. **37**(3): p. 515-27.

158. Aili, M., et al., *Functional analysis of the YopE GTPase-activating protein (GAP) activity of Yersinia pseudotuberculosis.* Cell Microbiol, 2006. **8**(6): p. 1020-33.

159. Viboud, G.I. and J.B. Bliska, *A bacterial type III secretion system inhibits actin polymerization to prevent pore formation in host cell membranes.* EMBO J, 2001. **20**(19): p. 5373-82.

160. Forsberg, A. and H. Wolf-Watz, *The virulence protein Yop5 of Yersinia pseudotuberculosis is regulated at transcriptional level by plasmid-plB1 -encoded trans-acting elements controlled by temperature and calcium.* Mol Microbiol, 1988. **2**(1): p. 121-133.

161. Isaksson, E.L., et al., *The membrane localization domain is required for intracellular localization and autoregulation of YopE in Yersinia pseudotuberculosis.* Infect Immun, 2009. **77**(11): p. 4740-9.

162. Anderson, D.M. and O. Schneewind, *A mRNA signal for the type III secretion of Yop proteins by Yersinia enterocolitica.* Science, 1997. **278**(5340): p. 1140-3.

163. Lloyd, S.A., et al., *Yersinia YopE is targeted for type III secretion by N-terminal, not mRNA, signals.* Mol Microbiol, 2001. **39**(2): p. 520-31.

164. Boyd, A.P., et al., *Yersinia enterocolitica can deliver Yop proteins into a wide range of cell types: development of a delivery system for heterologous proteins.* Eur J Cell Biol, 2000. **79**(10): p. 659-71.

165. Ishida, T. and K. Kinoshita, *PrDOS: prediction of disordered protein regions from amino acid sequence.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W460-4.

166. Liebschner, D., et al., *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix.* Acta Crystallogr D Struct Biol, 2019. **75**(Pt 10): p. 861-877.

Chapter 8

8. Appendix-
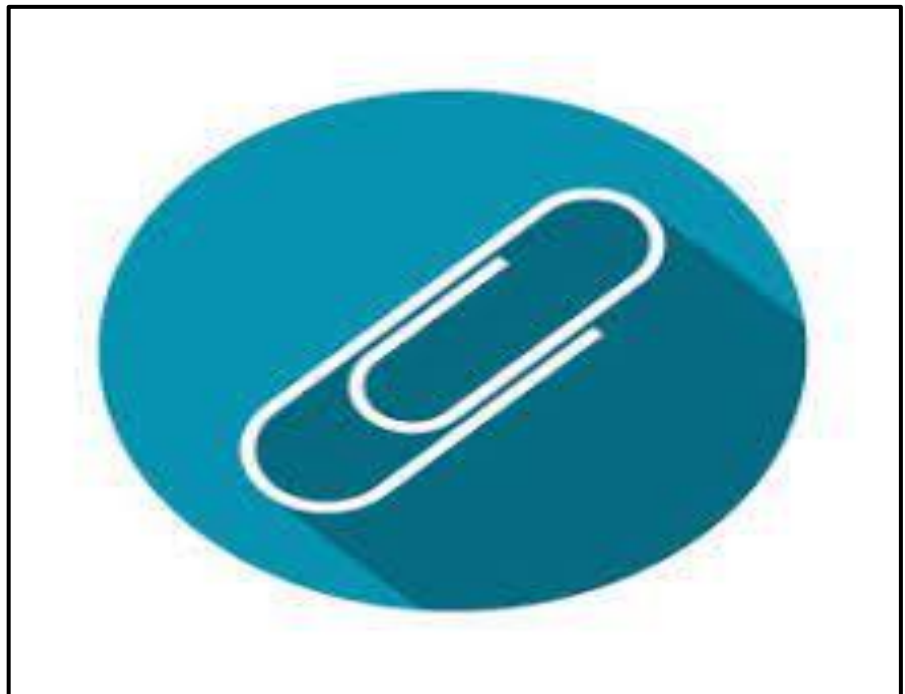
8.3. Abbreviations:

μg:	Microgram
μl:	Microliter
Å:	Angstrom
AAA+ ATPase:	ATPases Associated with diverse cellular Activities
ABC:	ATP binding cassette
ADP:	Adenosine di phosphate
ADP.AlFX:	Adenosine di phosphate Aluminum fluoride complex
AMPPNP:	Adenylyl-imidodiphosphate
ATCC:	American Type Culture Collection
ATP:	Adenosine tri phosphate
APS:	Ammonium per sulfate
BSA:	Bovine serum albumin
Cryo-EM:	Cryo electron microscopy
DLS:	Dynamic light scattering
DNA:	Deoxyribonucleic acid
DOPE	Discrete Optimized Protein Energy
DTT:	Dithiothreitol
EtBr:	Ethidium Bromide
EDTA:	Ethylenediamine tetra acetic acid
GAP:	GTPase activating protein
IM:	Inner membrane
IR1:	Inner membrane ring 1

IR2:	Inner membrane ring 2
IMAC:	Immobilized metal affinity chromatography
kb:	Kilo base
LB:	Luria-Bertani
LPS:	Lipopolysaccharide
M:	Molar
MCS:	Multiple cloning site
mM:	Milli molar
NaCl:	Sodium chloride
NaOH:	Sodium hydroxide
Ni–NTA:	Nickel–nitrilotriacetic acid
nm:	Nanometer
OM:	Outer membrane
PCR:	Polymerase chain reaction
PDB:	Protein Data Bank
PMF:	Proton motive force
PMSF:	phenylmethylsulphonyl fluoride
RNA:	Ribonucleic acid
rpm:	Rotation per minute
S:	Svedberg Unit
SDS:	Sodium Dodecyl Sulfate
Sct:	Secretion and cellular translocation
SRP:	Signal recognition Particle

T3SS: Type three secretion system

TAE: Tris-acetate-EDTA

TEM: Transmission electron microscopy

TEMED: Tetramethyl ethylenediamine

UV: Ultra violet

8.4. List of Figures:

Chapter 4

Chapter 5

8.5. List of Tables:

8.6. Publications

Reference: Taken from internet.

Delineating specific regions of N-terminal domain of T3SS ATPase YsaN of Yersinia enterocolitica governing its different oligomerization states

Rajeev Kumar, Chittran Roy and Saumen Datta*

Structural Biology and Bioinformatics Division, Council of Scientific and Industrial Research—Indian Institute of Chemical Biology, Kolkata, West Bengal, India

Oligomerization of YsaN, a putative T3SS-ATPase is a necessary and crucial event for T3SS functioning in *Y. enterocolitica*. Different oligomeric states have been proposed for similar ATPases, yet, the true nature of its activation and formation of different oligomers is still poorly understood. *In-vitro* studies of YsaN reveal that its activation and oligomerization depend on its N-terminal region and occur as a result of active catalysis of ATP in an ATP concentration-dependent manner following two-step cooperative kinetics. Also, the N-terminal 83 amino acid residues of YsaN are crucial for higher-order oligomer formation while YsaNΔ83 is capable of hexamer formation upon oligomerization. Enzyme kinetics study shows reduced ATPase activity of YsaNΔ83 (3.19 ± 0.09 μmol/min/mg) in comparison to YsaN (9.076 ± 0.72 μmol/min/mg). Negative-TEM study of YsaN and YsaNΔ83 oligomer suggests that the formation of higher-order oligomer (probably dodecamer) occurs by stacking of two hexamers through their N-terminal faces involving N-terminal 83 amino acid residues which have been further supported by the docking of two hexamers during the *in-silico* study. These results suggest that YsaN is an oligomerization-activated T3SS ATPase, where distinct regions of its N-terminal domain regulate its different oligomeric nature and is essential for its activation.

Abbreviations: AAA + ATPases, ATPase associated with various cellular activities; ADP- AlFX, ADP aluminum fluoride complex; AMPPNP, Adenylyl- imidodiphosphate, CPD, cysteine protease domain; DTT, Dithiothreitol; EDTA, Ethylenediaminetetraacetic acid; IPTG, Isopropyl β- d-1, thiogalactopyranoside; NaF, Sodium fluoride; PDB, Protein data bank; PMF, Proton Motive Force; RMSD, Root mean square deviation; RPM, rotation per minute; SDS, Sodium dodecyl sulphate.

1 Introduction

The type three secretion system (T3SS) has been regarded as major virulence determining factor in numerous plants and animal pathogenic bacteria (Coburn et al., 2007; Buttner, 2012). A broad spectrum of diseases is caused by pathogenic bacteria such as enteric infections caused by enteropathogenic *E. coli*, *Shigella*, *Salmonella,* and *Y. enterocolitica* containing T3SS. These pathogens use T3SS to inject effector toxins directly inside the host cell to manipulate host cellular processes. Mutations, deletions, or blocking of the T3SS apparatus components result in reduced virulence and disease manifestations in mouse model experiments (Coburn et al., 2007). T3SS has been predicted to be evolved from flagellar T3SS (fT3SS) through horizontal gene transfer and both of them share a common ancestor. Both T3SS and fT3SS share many structurally similar components at their core (Gophna et al., 2003; Diepold and Armitage, 2015). Bacterial T3SS is a complex structure composed of approximately 20 different proteins and broadly can be divided into three basic parts namely the basal body, needle complex, and a large cytosolic component known as the sorting platform or the C- ring complex. The basal body consists of two co-axial homomeric protein complex rings across the inner and outer plasma membrane including the peptidoglycan layer. The needle is present in association with the outer ring projecting away from the bacterial membrane providing approximately 1.5–2.5 nm conduit for translocation of unfolded effectors. Tip complex protein associated with the needle helps the bacteria to integrate with the host plasma membrane. Once the association with the host membrane is established, the export apparatus helps in the translocation of effectors through the needle to the host cell cytoplasm (Kubori et al., 1998; Ghosh, 2004; Diepold and Wagner, 2014). Previous structural studies involving Cryo-Electron microscopy have revealed the global structure of T3SS Injectisome. However, the precise details of the cytoplasmic components are largely undetermined because of their dynamic nature of association with the injectisome.

Energy for unfolding and translocation of effectors is provided by both, Proton Motive Force (PMF) (Wilharm et al., 2004) and ATP hydrolysis by a highly conserved ATPase ring complex which remains in close association with the active needle complex at the sorting platform (Lee and Rietsch, 2015). These ATPases also have structural and functional similarities with the F_0- F_1 ATPase—an AAA + ATPase (ATPase associated with various cellular activities) (Akeda and Galan, 2005; Yoshida et al., 2014). Structural and functional similarity between T3SS ATPases and the AAA + enzymes suggest the possibility of a universal model of mechanism among such molecular motors.

Based on the results from Electron microscopy, high-resolution Cryo-electron tomography, and biochemical studies, it has already been suggested that T3SS ATPases function as homo-oligomers namely hexamer (Hu et al., 2017; Halder et al.,

2019) and dodecamer (Pozidis et al., 2003; Muller et al., 2006) which are located at the central pore of the T3SS apparatus on the cytoplasmic side. Besides, previous studies have also shown that during effector unfolding and translocation through the needle complex, the effector chaperone complex docks to the C-terminal side of the ATPase ring complex at the export gate. Also, the C-terminal side of this ATPase ring complex faces the pore of the needle complex (Akeda and Galan, 2005; Allison et al., 2014; Majewski et al., 2019).

Y. enterocolitica uses contact-dependent T3SS (Abrusci et al., 2013) for the delivery of anti-host effector proteins directly into the eukaryotic host cell (Cornelis, 2002; Edgren et al., 2012; Bent et al., 2013). In addition to fT3SS, *Y. enterocolitica* maintains two distinct, independently regulated T3SSs. The first one is *Ysa- Ysp* T3SS encoded by *Ysa* Pathogenicity Islands located on the chromosome and another one is *Ysc- Yop* T3SS which is encoded by ≈70 Kb virulent plasmid pYV/pCD1 (Cornelis et al., 1998; Foultier et al., 2002). The *Ysa- Ysp* T3SS is required for the gastrointestinal phase of infection and intracellular survival of *Y. enterocolitica* within macrophages (Bent et al., 2015) and is important in virulence (Haller et al., 2000). In our previous study, it was shown that YsaN, a T3SS ATPase in *Y. enterocolitica* encoded by *Ysa- Ysp* T3SS is a magnesium-dependent ATPase and the probable dodecamer state is the most active form of YsaN. Moreover, it was also shown that the N- terminal 1–20 amino acid residues of YsaN is crucial for YsaL binding (a negative regulator of YsaN ATPase) (Chatterjee et al., 2013).

In the present study, we have performed the *in-vitro* characterization of untagged YsaN and various deletion constructs. Another focus of this study is the characterization of the formation of different functionally relevant higher-order oligomers based on its N- terminal region i.e., partially active hexamer and the most active higher-order oligomer forms of YsaN. The present study indicates the involvement of distinct regions on the N- terminal domain of YsaN in the formation of hexamer and its higher-order oligomer. It has also been shown that the transition to a higher oligomer state of YsaN is an ATP concentration-dependent event following two-step kinetics. The existence of the hexamer as well as the higher-order oligomer complex has also been observed in negative TEM which suggest that the formation of higher-order oligomer occurs by stacking of two homo- hexamers through their N- terminal faces which are also supported by perfect docking between two hexamer rings during *in-silico* studies.

2 Materials and methods

2.1 Cloning expression and purification

Plasmid vectors and constructs used in this study are given in Table 1. Genomic DNA was isolated from *Y. enterocolitica*

8081 culture according to (Wilson, 2001). *YsaN* wild-type gene was PCR amplified from this purified genomic DNA using *Pfu* DNA polymerase Thermo Scientific™ and cloned into *pET*22bΔ50CPD vector containing C- terminal His- tagged cysteine protease domain (hereafter CPD vector) with Nde1 and BamH1 restriction sites. Similarly, other *YsaN* deletions (based on P*fam* domain analysis, refer to Figure 1B) were cloned in fusion with the CPD domain in the *pET*22bΔ50CPD vector. Primers used in this study are listed in Supplementary Table S1. Cloning in CPD vector resulted in soluble expression of all YsaN deletions which were usually not soluble when expressed alone (data not shown) (Shen et al., 2009). DH5α was used as a cloning host in all cases. For recombinant expression of the protein, the plasmid constructs were transformed into BL21 DE3 chemically competent cells. IPTG (at 0.5 mM working concentration) was added after OD 600 reached 0.6 and kept for constant shaking for approximately 12–14 h at 298 K. Induced cells were harvested at 5,000 g at 277 K for 10 min. Cells were lysed in sonication buffer (50 mM Tris pH 8.0, 100 mM NaCl, 5% glycerol, and 5 mM imidazole) by ultrasonication method using a sonicator (Q-Sonica 125). Protease inhibitor PMSF at 1 mM working concentration was added to the resuspended pellet just before sonication. Centrifugation was done to remove the cellular debris at 18,000 g at 277 K for approximately 60 min. Protein was purified according to the CPD purification protocol (Shen et al., 2009) with few modifications. All the purification steps were carried out at 277 K in a cold room. Briefly, the clear supernatant was loaded onto a pre-equilibrated gravity-flow purification column, containing Nickel beads (Nickel Sepharose™ Fast Flow, GE Healthcare) followed by incubation of approximately 60 min. The beads were washed with two column volumes (CV) of wash buffer (50 mM Tris pH 8.0, 100 mM NaCl, 5% glycerol, and 35 mM imidazole). Then, the beads were mixed and incubated with CPD buffer (50 mM Tris pH 8.0, 100 mM NaCl, 5% glycerol) added with 100 μM (working) Inositol-6- phosphate (Phytic acid sodium salt hydrate, sigma) for approximately 60 min for CPD-His-tag removal. The incubated sample (containing untagged YsaN and other deletion proteins) was collected in a fresh collection tube and immediately injected into the gel filtration system AKTA prime plus (GE Healthcare). CPD buffer incubation time for more than 60 min was strictly avoided in our case to reduce the degradation of YsaN and other YsaN deletion proteins by CPD itself. Gel filtration was done to further purify the protein using the Hi-load Superdex 200 16/60 gel filtration column (GE Healthcare). Immediate gel filtration was necessary in this case to avoid further degradation of target proteins due to leached CPD from the nickel beads which was separated only after gel filtration. 250 mM Imidazole in elution buffer was used to elute the residual proteins bound to the beads. The corresponding peak in size exclusion respective to YsaN and its deletions were collected and analyzed by SDS gel

electrophoresis. This purified protein sample was used for all downstream experiments. Freshly purified samples were used in all experiments each time.

2.2 Site-directed mutation

To create a non-functional Walker-A lysine to alanine mutant of YsaN i.e., YsaN K166→A, site-directed mutation was performed according to (Ho et al., 1989) in YsaN pET22bΔ50CPD construct. Primers used for mutation studies are mentioned in Supplementary Table S1. The positive clones were verified by sequencing. The positive clone plasmid was further transformed into BL21 for further purification. ATPase assay was performed to verify its loss of activity (Supplementary Figure S5B).

2.3 ATPase assay

To check the functional activity of all the constructs, malachite- green ATPase assay was performed according to (Lanzetta et al., 1979; Baykov et al., 1988) with slight modification. Briefly, protein in assay buffer (20 mM Tris pH8.0, 100 mM NaCl, 10 mM $Mgcl_2$, 1 mM DTT, 0.025% BSA, and 5% glycerol) was prepared sufficiently before performing the experiment. Malachite green- ammonium molybdate reagent mixture was prepared according to (Dey et al., 2018). All the reagents and buffers were filtered using a 0.22 μm syringe filter every time just before use. Varying ATP concentration stock was prepared (0–2,400 μM). Two milliliter of the reaction mixture was prepared by mixing protein at final working concentration of 5 μM for YsaN, and 10 μM for YsaNΔ83 in the assay buffer with varying ATP concentrations. 100 μl of reaction sample was taken at four different time points and added to 96 well plate containing 50 μl of the malachite green reaction mixture and incubated for 10–20 min. To stop further change in color development, 100 μl of 34% citric acid (Merck) solution was added to the reaction and incubated for another 30 min at 300 K. Finally, the absorbance was read at 660 nm in Hidex- sense 96 well plate reader. Separate independent experiments were performed in triplicates for YsaN and YsaNΔ83 enzyme kinetics. To further compare the activity of different YsaN constructs relative ATPase activity graph was obtained by using a final working concentration of 5 μM for YsaN, YsaN K166→A, and YsaNΔCterm whereas 10 μM for YsaNΔ83 and YsaNΔNterm respectively (Figure 2H, Supplementary Figure S5). Briefly, in all cases, a standard 2 mM ATP (working) was incubated for up to 40 min in a total of 2 ml reaction volume and 100 μl of reaction sample was taken at eight different time intervals (from 5 to 40 min) and added to 96 well plate containing 50 μl of malachite green- ammonium molybdate reagent mixture as mentioned

TABLE 1 List of clones/vectors used in this study.

Vectors/Plasmid construct name	Details	Protein size (approximate molecular weight)	Reference(s)/source
pET22bΔ50CPD[a]	*Vibrio cholerae* MARTX toxin cysteine protease domain with C terminal His Tag in pET22b	23.1 kDa	Gifted by Matthew Bogyo, Department of Pathology, Stanford School of Medicine, Stanford, California
YsaN	YsaN wild type full length (430 amino acids cloned in pET22bΔ50CPD vector)	47.87 kDa	This Study
YsaNΔ83	N terminal 1–83 amino acid deletion of YsaN (347 amino acids cloned in pET22bΔ50CPD vector)	38.6 kDa	This Study
YsaNΔC-term	C terminal 357–430 amino acid deletion of YsaN (349 amino acids cloned in pET22bΔ50CPD vector)	39.2 kDa	This Study
YsaNΔN-term	N terminal 1–140 amino acid deletion of YsaN (290 amino acids cloned in pET22bΔ50CPD vector)	32 kDa	This Study
YsaN K166→ A	The lysine at the 166th position is mutated to alanine. It is a non-functional mutant of YsaN. (430 amino acid cloned in pET22b vector with C-terminal his tag.)	49 kDa	This study

[a]Δ represents deletion in all cases.

FIGURE 1

Bioinformatic analysis of YsaN- **(A)**, Linear representation of the YsaN constructs used in this study (based on domain analysis of YsaN using Pfam, also refer to Table 1 for complete details) Different color shading represents respective domains of YsaN, Red- represents N terminal domain, Gray- represents a central domain, and green- represents C-terminal domain. The red star shows site-directed mutation site on the central domain. **(B)**, Pfam Domain analysis of YsaN sequence. **(C)**, Multiple sequence alignment of YsaN and homologs (also refer to Supplementary Tables S2−S4 for details). The vertical black line represents the domain partition (i.e., N-terminal and C-terminal domains) of YsaN sequence based on Pfam domain analysis.

FIGURE 2

Purification and characterization of YsaN- **(A–D)**—Typical size exclusion chromatography profile of YsaN, YsaNΔ83, YsaNΔNterm, and YsaNΔCterm respectively using Superdex 16,600 columns. Respective collected peaks are delimited by crossed sections of the chromatograms. **(E)** SDS gel profile of collected peaks from gel filtration (Lane M-Marker, Lane1- Peak from **(D)**, Lane2- Peak from **(C)**, Lane3- Peak from **(B)**, and Lane4- Peak from **(A)**. **(F)**- Standard enzyme assay curve of purified YsaN measuring μmols of phosphate released per minutes per milligrams of protein using the malachite green method for inorganic phosphate estimation, inset showing V_{max}, K_{half}, and Hill coefficient values. **(G)**- Standard enzyme curve for YsaNΔ83 in terms of μmols of phosphate released per minute per milligrams of protein, inset showing V_{max}, K_{half}, and Hill coefficient values. **(H)**- Comparison of relative ATPase activity of purified YsaN to YsaNΔ83, YsaN to YsaNΔNterm, YsaN to YsaNΔCterm (also please see Supplementary Figure S5A,B for further details of enzyme kinetics of YsaNΔ83 and YsaN K166→A and relative assay). Error bars indicate standard deviation in all cases. Statistically significant differences by unpaired Student's t-test are indicated with asterisks ("★★★" means $p < 0.001$; "★★"$p > 0.01$; "ns" means non-significant).

above. Following a similar assay protocol relative assay was also performed in an assay buffer containing 2 mM EDTA and with the assay buffer containing 5 mM NaF as a control experiment to test the effect of NaF on YsaN activity (Supplementary Figure S5B). Relative ATPase activity experiments were performed in duplicate in an independent experiment. A separate blank for each experiment was generated using assay buffer and ATP (respective concentration) with no protein and the values were subtracted from the experimental results. Inorganic phosphate released was calculated from the phosphate standard curve generated (Supplementary Figure S4) using NaH_2PO_4 (Sodium phosphate monobasic, Sigma). The K_{half}, V_{max}, and Hill coefficients were observed by plotting the non-linear regression plot in GraphPad prism8 software. The graph was plotted using Origin8 and GraphPad prism 8 software.

2.4 Dynamic light scattering experiment

Dynamic light scattering (DLS) was performed using Zetasizer Nano ZS (Malvern Instruments, United Kingdom)

with the samples to check mono-dispersity of samples in the presence and absence of various concentrations of ATP in DLS buffer (20 mM Tris pH8.0, 100 mM NaCl, 4mM $Mgcl_2$, 5% glycerol and 1 mM DTT). 100 μl of 10–100 μM protein samples with varying ATP concentrations were mixed and incubated for 2–5 min at 298 K. 50 μl of this sample was added in Ultra-Micro Cell ZEN2112 (Malvern Panalytical, United Kingdom), and size intensity distribution was observed. Substrates like ADP, AMP-PNP, and ADP. AlFX (1–3.0 mM range) was also used in the DLS experiment. All the DLS experiments were performed more than three times in all cases with freshly purified samples in independent experiments.

2.5 Size exclusion chromatography

Further, the analytical size exclusion chromatography (or gel filtration) method was performed to characterize the oligomeric behavior of YsaN. We used ADP.AlFX (a non-hydrolyzable ATP to ADP transition state analog, refer to Figure 4), to mimic ATP. YsaN-ADP.AlFX and YsaNΔ83-ADP.AlFX complex formation

was done according to Rappas et al. (2005) with slight modification. Briefly, 500 μl of approximately 10 μM purified YsaN and YsaNΔ83 protein was incubated with 1.5 mM of ADP for 5–10 min at room temperature. To this pre-incubated sample, 1.5 mM AlCl₃ (working concentration) was added and mixed thoroughly followed by incubation for another 20 min at room temperature. This sample was injected into gel filtration system with pre-equilibrated Superdex 200 H R 10/30 analytical column (GE Healthcare) with 50 mM Tris pH 8.0, 100 mM KCl, 4 mM MgCl₂, 5 mM NaF, 1 mM DTT and 5% glycerol. NaF in gel filtration buffer was required for maintenance of the AlFX complex (Rappas et al., 2005). The flow rate was kept at 0.5 ml/min at room temperature. All the respective peaks were collected, concentrated, and analyzed by SDS gel electrophoresis (Figure 4F).

2.6 Transmission electron microscopy

Negative TEM was performed to visualize YsaN and YsaNΔ83 oligomers to a higher resolution. Briefly, approximately 0.6–0.8 μM of purified YsaN and YsaNΔ83 was incubated with 1.5 mM ADP for 10 min at room temperature. To this sample 1 mM (working concentration) of AlCl₃ was added and incubated for another 15 min at room temperature. 3 μl of this sample was loaded onto a freshly glow discharged carbon-coated copper 300 mesh grid and left for 30 s. Excess of the sample was blotted gently using blotting paper. The precaution was taken not to dry the grid completely. Then 5 μl of 1% uranyl acetate solution (Cornelis et al., 1998) was loaded onto the grid and left for 10 s. Excess stain was blotted gently using blotting paper. The staining process was repeated three times. The stained grids were left to air dry at room temperature and stored for viewing purposes. The negatively stained samples were visualized at room temperature using a TALOS L 120C electron microscope (Thermo Fisher). The instrument was operated at 120 kV and the images were captured using a bottom-mounted Flucam and Ceta 16M Camera (35–910 kX magnification range).

3 Results

3.1 In- silico study of YsaN

T3SS ATPase family proteins contain three predicted domains: an N-terminal domain ≈with 100 residues involved in oligomerization and stabilization of the ATPase ring complex, a central ATPase domain, and a C-terminal domain for effector-chaperone complex interaction. In our previous study, we have shown that YsaN is the putative ATPase of Ysa T3SS in *Y. enterocolitica* 8081 (26). To investigate it further we did various *in-silico* studies mentioned here to analyze the YsaN sequence and the nature of its N-terminal domain. Multiple sequence

alignment (MSA) of YsaN was conducted using Clustal Omega (Sievers et al., 2011) with the homolog proteins mentioned in Supplementary Table S2. The MSA (refer to Figure 1C; Supplementary Tables S3, S4) shows that the N-terminal domain is less conserved unlike the central domain and the C-terminal domain among different homologs. Further, domain analysis of YsaN using P*fam* (Bateman et al., 2000) suggests the probable presence of ATP synthase alpha and beta subunits signatures at 139–349 amino acids with a bit score of 246.8 and E-value score equals 1.9e-76 (Figure 1B). Secondary structure prediction using PSIPRED 4.0 with YsaN sequence (Jones, 1999) (Supplementary Figure S1B) suggests that the N terminal region primarily consists of β strands. Also, Kyte & Doolittle's hydrophobicity result in ProtScale (Gasteiger et al., 2005) suggests that the N- terminal region (up to 80 amino acids) is primarily hydrophobic (Supplementary Figure S1A). Further a YsaN model was generated using Phyre2 server (Kelley et al., 2015). Phyre2 could generate a YsaN model for 78–430 amino acids only. To generate the full length YsaN homology structure we used MODELLER 9v.11.43 software and for template we used chain D of FliI-FliH crystal structure (PDB ID 5B0O). FliI is fT3SS ATPase in *E. coli* and has 38.28% sequence identity with YsaN (details of model generation process is provided in Section 3.5). Also, using PyMOL both the structures were aligned. The aligned structure has a RMSD value of 1.2Å (Supplementary Figure S10).

3.2 Purification and characterization of untagged YsaN

In pursuit of obtaining monomeric soluble YsaN, we tried different YsaN deletions (a random 20 amino acid N-terminal deletion constructs were cloned along with the deletions based on P*fam* domain prediction; Figure 1B), out of which only the constructs mentioned in Table 1 and represented in Figure 1A were used in this study. Since all the YsaN His tagged deletion constructs were insoluble during recombinant expression (data not shown) hence, we cloned YsaN and various deletions in the CPD vector (Table 1) for their soluble expression. One benefit of this CPD fusion was getting untagged protein in soluble form. A typical size-exclusion profile of all the constructs used in this study is shown in (Figures 2A–D) followed by their molecular weight analysis by SDS gel (Figure 2E). Also, to check whether CPD fusion resulted in any change or had any effect on oligomeric behaviour of YsaN in solution we compared the size exclusion profiles of C- terminal His tagged YsaN (cloned in *pET*22b vector) with untagged YsaN obtained after removal of C- terminal CPD-His tag. Both YsaN- His and YsaN (untagged) elute at the same elution volume (Supplementary Figure S3). To estimate the functional efficiency of different constructs concerning YsaN enzyme kinetics their ATPase activity was studied by the Malachite green ATPase assay method. YsaN

(untagged) shows kinetic parameters–V_{max} value of 9.07 ± 0.72 μmol/min/mg, K_{half} value of 758.7 ± 84.35 μM, and Hill coefficient value of $h = 2.09 ± 0.31$ (Figure 2F), comparable to our previous study on YsaN (N-terminal His-tag) (Chatterjee et al., 2013). YsaNΔ83 shows reduced ATPase activity whereas, in the case of YsaNΔNterm, the activity was barely detectable and YsaNΔCterm (a C terminal deletion variant of YsaN), shows ATPase activity comparable to YsaN (refer to Figure 2H, Supplementary Figure S5A). These observations suggest that the ATPase activity of YsaN is controlled by its N terminal region. Further to investigate the cooperative nature of YsaNΔ83 we perform enzyme kinetics for YsaNΔ83 (at 10 μM working concentration, also refer to Figure 2G). YsaNΔ83 has a reduced V_{max} of 3.19 ± 0.09, K_{half} value of 467.3 ± 30.15, and a hill coefficient value of $h = 1.20 ± 0.06$. The hill coefficient value obtained for YsaNΔ83 suggest that YsaNΔ83 bears cooperative nature.

3.3 Oligomerization behaviour of YsaN and YsaNΔ83

The cooperative nature of YsaN and YsaNΔ83 prompted us to check the substrate-dependent oligomerization nature of YsaN and YsaNΔ83. ATP (substrate), ADP (product), AMP-PNP (non-hydrolysable ATP analog), and ADP.AlFX (ATP to ADP transition state analog) were tested for their role in the oligomerization behaviour of YsaN and was studied by DLS, represented in Figure 3. Figure 3A shows the oligomerization behaviour of YsaN where a single monodisperse peak of is visible for YsaN only, YsaN in presence of 2 mM ADP, and YsaN in presence of 2 mM AMP-PNP. However, the peak got shifted toward right indicating formation of higher order oligomer in the case of YsaN incubated with 2 mM ATP and 1.5 mM ADP.AlFX. Further DLS was also performed in an ATP concentration-dependent and protein concentration-dependent manner to evaluate the oligomeric behaviour of YsaN. Figure 3E shows that increasing protein concentration does not have any effect on YsaN oligomerization whereas increasing ATP concentration has a significant role in its oligomerization (Figure 3B). Similar substrate-dependent behaviour was also observed in the case of YaNΔ83 (a partially active construct of YsaN) (Figure 3C) while the oligomerization behaviour was absent in the case of YsaNΔNterm (a non-functional YsaN construct) (Figure 3D). From Figure 3A it was also observed that AMP-PNP (an ATP analog) was unable to promote oligomerization of YsaN while ADP.AlFX [also an ATP analog, which mimics ATP to ADP transition state (Schumacher et al., 2004; Chen et al., 2007)] was capable of inducing oligomerization. To investigate this behaviour of YsaN we ask the question that whether ATP binding is the key to induce YsaN oligomerization or YsaN oligomerization is a result of the active catalysis of ATP. To evaluate this, we created a Walker-A lysin mutant YsaN

K166→A which is a non-functional mutant (Supplementary Figure S5B) of YsaN, and observed its oligomerization behaviour by DLS in a similar way (Figure 3G). In another experiment, we treated YsaN with 2 mM EDTA (during its purification) to quench the Mg^{2+} ion, to prevent Mg^{2+} dependent ATP hydrolysis. YsaN EDTA treated sample was unable to oligomerize even in presence of 3 mM ATP (Figure 3F). In both experiments it was observed that YsaN oligomerization requires active catalysis. Collectively, YsaN oligomerization and its activity depend on the N terminal domain and is an ATP concentration-dependent event and requires active catalysis of ATP to ADP and Pi.

3.4 Formation of YsaN higher-order oligomer by YsaN and YsaNΔ83

In the previous section, we observed and identified YsaNΔ83 a YsaN construct that is functionally active but, with lesser strength. It is also observed that the formation of YsaN oligomers is very much dependent on the active catalysis of ATP. So, active catalysis and oligomer formation are very much controlled by a specific region of the N-terminal part of YsaN. From the literature, we know that AAA + ATPase like YsaN is capable of forming various higher-order oligomers like hexamer, dodecamer, and higher-order aggregates. So far, the hexamer is the most stable variant of such ATPase complex which has been observed in various *in-situ* studies.

The role of the N-terminal region in YsaN oligomerization and its ATPase activity was further studied through estimation of the apparent molecular weight by analytical size exclusion chromatography (Superdex 200 H R 10/30 column) of the oligomeric complexes of YsaN and YsaNΔ83 in the absence (Figures 4A,B respectively) and in presence of ADP.AlFX (Figures 4D,E respectively). Apparent molecular weight estimation was done by comparing the elution profiles with the molecular weight standard curve represented in Figure 4C. YsaN-ADP.AlFX, complex eluted at ≈14 ml and ≈10 ml which corresponds to the molecular weight of ≈288 kDa (hexamer), and ≈576 kDa (probably dodecamer) respectively. A similar experiment was performed with YsaNΔ83-ADP.AlFX a peak corresponding to hexamer (molecular weight of ≈234 kDa) along with an un-oligomerized peak (Figures 4B,E respectively) was observed. The respective peaks were collected and analyzed by SDS gel electrophoresis (Figure 4F).

Following the previous sections, YsaN and YsaNΔ83 have catalytic activity and they can form functional oligomeric structures in solution, we extended our work to visualize these oligomers in greater detail. We used negative-TEM imaging techniques to visualize these oligomers. Similar experimental conditions were chosen for these imaging studies which have been used in DLS and gel filtration studies. YsaN and YsaNΔ83 was viewed in absence and in presence of ATP and
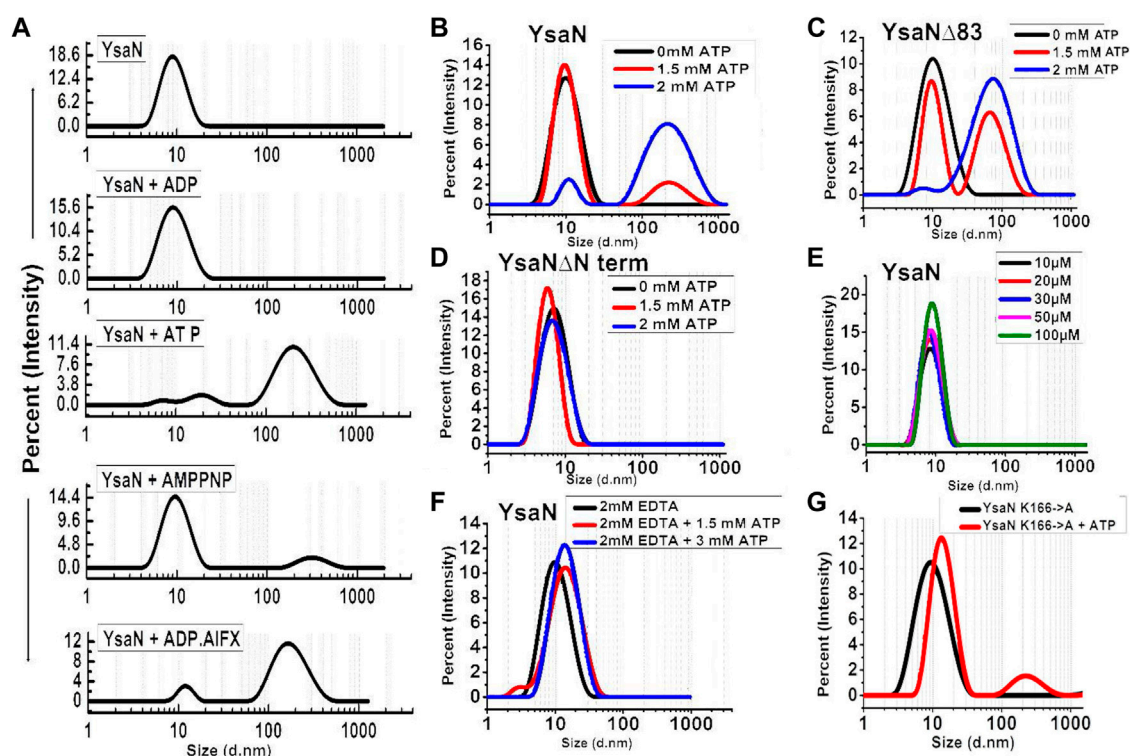
FIGURE 3
Dynamic light scattering study of YsaN oligomerization—**(A)**- Representation of size intensity profile of YsaN and in the presence of substrates like 2 mM ADP (product), 2 mM ATP (substrate), 2 mM AMPPNP (a non-hydrolyzable ATP analog), and 1.5 mM ADP.AlFX (an ATP to ADP + Pi transition state analogue). **(B–D)-** Substrate (ATP) concentration-dependent oligomerization of YsaN, YsaNΔ83, and YsaNΔNterm respectively demonstrating the role of YsaN N- terminal domains in YsaN oligomerization. **(E)-** Protein concentration-dependent DLS study of YsaN. **(F)-** Size intensity profile of YsaN in presence of 2 mM EDTA demonstrating the role of Mg²⁺ dependent catalysis of ATP to ADP + Pi and its role in oligomerization. **(G)-** Size intensity profile of YsaN Walker-A lysine mutant YsaN K166→A in the absence and presence of 2 mM ATP.

ADP.AlFX. While viewing YsaN and YsaNΔ83 oligomers in negative-TEM using ATP and ADP.AlFX, we were able to get better images of oligomer in the presence of ADP.AlFX as compared to ATP. In the case of YsaNΔ83, we were able to visualize the hexameric ring structure of YsaNΔ83. It should be noted here that ADP.AlFX has been reported to stabilize the oligomeric structure of similar ATPase in earlier studies and has been used extensively to obtain high-resolution TEM and Cryo-EM structures (Majewski et al., 2019). In the case of YsaNΔ83-ADP.AlFX, the observed particle size of ≈10 nm representing a hexamer formation with clear sixfold symmetry was observed (Figure 5A). The particle size (10 nm) of the hexamer as observed in TEM images are similar to the previously reported size of 10 nm among the three-dimensional structures of hexameric ATPases. Whereas, in the case of YsaN-ADP.AlFX the particle size of the oligomer observed was greater than 10 nm and was observed as ≈ a 15–20 nm structure in different orientations (Figure 5B). The presence of particle size greater than 10 nm indicates the presence of a higher oligomeric form in the YsaN complex as compared to YsaNΔ83.

The presence of particles having more than 10 nm in size indicates the presence of a higher oligomeric form in the YsaN complex, as compared to YsaNΔ83. We assume this higher-order oligomeric form to be probable YsaN dodecamer. It is also notable that the presence of a few other intermediate oligomers in the YsaNΔ83-ADP.AlFX in TEM image suggests that the formation of the hexamer complex was weak which may be a result of the deletion of the N-terminal 83 residues which may be required for stable hexamer ring formation post oligomerization. Collectively YsaNΔ83 is capable of hexamer formation independent of 83 N-terminal residues.

3.5 Homology model structure of YsaN oligomer

Since the only full-length crystal structure of FliI (fT3SS ATPase) is present to date, hence, to generate the monomeric homology model, Chain D was chosen (having sequence identity of 38.28% with YsaN) from the crystallographic structure of

FIGURE 4
Analytical size exclusion chromatography profile of YsaN and YsaNΔ83 using Superdex 10/30 column- **(A)** and **(B)**- size-exclusion profile of purified monomeric YsaN and YsaNΔ83 respectively. **(C)**- molecular weight standard curve for Superdex 10/30 column representing elution profiles of possible oligomeric forms of YsaN and YsaNΔ83. (TG = Thyroglobulin- 669 kDa, AF = Apoferritin 443 kDa, β-AM = β- Amylase- 200 kDa, AD = Alcohol dehydrogenase- 150 kDa, AL = Albumin- 66 kDa, CA = Carbonic anhydrase- 29 kDa). **(D)** and **(E)**- size-exclusion profile of YsaN and YsaNΔ83 in presence of 1.5 mM ADP.AlFX in the buffer containing 5 mM NaF, inset showing peaks corresponding to probable dodecamer and hexamer respectively in the zoomed image delimited by crossed section in the full chromatogram. **(F)**- SDS gel profile of the purified sample and peaks collected during size exclusion chromatography steps. Lane 1 and Lane 3 represent purified YsaNΔ83 and YsaN form Superdex 200 16/600 size exclusion chromatography column respectively. Lane 2 (YsaNΔ83) and Lane 4 (YsaN) correspond to peak 2 of **(B)** and peak 1 of **(A)** respectively. Lane 3 represents the marker, and Lane 5 and 6 represent the peak 6 and peak 5 in **(E)** respectively. Lane 7 and 8 represent peak 3 and peak 4 in **(D)** respectively.

FliI–FliH complex (PDB Id: 5B0O). The 3D-structure prediction of YsaN was done using the homology modeling software package MODELLER 9v.11.43 (Webb and Sali, 2014). Alignment between template and target was performed by Clustal Omega (https://www.ebi.ac.uk) (Sievers et al., 2011). A few manual improvements were done in alignment files and an in-build python script. These manual improvements of PDB files was done using Coot software (Emsley et al., 2010) to remove the clashes. Finally, at least five models were taken from each modeling procedure. The best model was then selected based on the Discrete Optimized Protein Energy (DOPE) score (Supplementary Figure S9; Supplementary Table S5). The different predicted domains of YsaN were represented by a linear map in (Figure 6A). The final model generated was viewed using PyMOL software (DeLano, 1874) (Figure 6B) and subsequently, the hexameric form was built by fitting the monomeric structure into the hexamer of the F1 ATPase (PDB Id: 4XD7) by the use of PyMOL software. Again, some manual improvement of a hexameric text file was done. After generating the model some unnecessary text appears in the PDB text file, which breaks chain continuity and cause a problem during

visualization in softwares like vmd, PyMOL etc. To solve this problem, we have to manually delete such kinds of unnecessary text from PDB text file. Using this hexameric text file the dodecameric form of the YsaN (N terminal 20 amino acids deleted) structure was generated by the ClusPro server (Kozakov et al., 2013; Kozakov et al., 2017). Also, Coot was used manually to remove the clashes caused by close contacts of amino acid side chain. These clashes were removed by rotating side chains to a non-clashing rotamer. Structural evaluation and stereochemical analysis of modeled protein were checked by various bioinformatics tools and software packages such as RAMPAGE, ERRAT, and Verify3D. Molecular docking was performed by an automated docking tool, Auto Dock (Trott and Olson, 2010). The software was used for the binding of small molecules (ATP) with a structural model of its known three-dimensional receptor protein (YsaN). It uses the genetic algorithm for conformational search and it is regarded as a popular method to study docking. The technique combines simulated annealing for conformation searching with a rapid grid-based method of energy evaluation. During the docking simulations, the

FIGURE 5
Negative TEM image of YsaN and YsaNΔ83 oligomer in presence of ADP. AlFX. **(A)** YsaNΔ83 oligomer. The Zoomed image represents an approximately 10 nm hexamer ring assembly of particles marked by the red circle in a different orientation. **(B)** The YsaN higher-order oligomer complex. The zoomed image represents particles marked by the red circle in a different orientation having a size of approximately 15–20 nm in different orientations representing probable dodecamer. The prepared sample grids were stained with 1% uranyl acetate. The white color scale bar represents 10 nm in all cases.

inhibitors were regarded as flexible and subjected to energy minimization and scoring system based on force-field energy scoring function. By the use of PyMOL, the hexameric structure of YsaN was generated by fitting into F1 ATPase (PDB Id: 4XD7). Some manual improvements were done in the hexameric form of YsaN to generate a dodecameric structure. The dodecameric structure was built using the ClusPro server (Kozakov et al., 2017) *via* three computational steps: 1) rigid-body docking, 2) RMSD based clustering of the lowest energy structures, and 3) the removal of steric clashes by energy minimization. Docking with each energy parameter set results in ten models defined by centers of highly populated clusters of low energy docked structures which were picked up for final model

generation. Based on the above study, the YsaN hexamer- hexamer docked model was generated to represent the pre assumed dodecame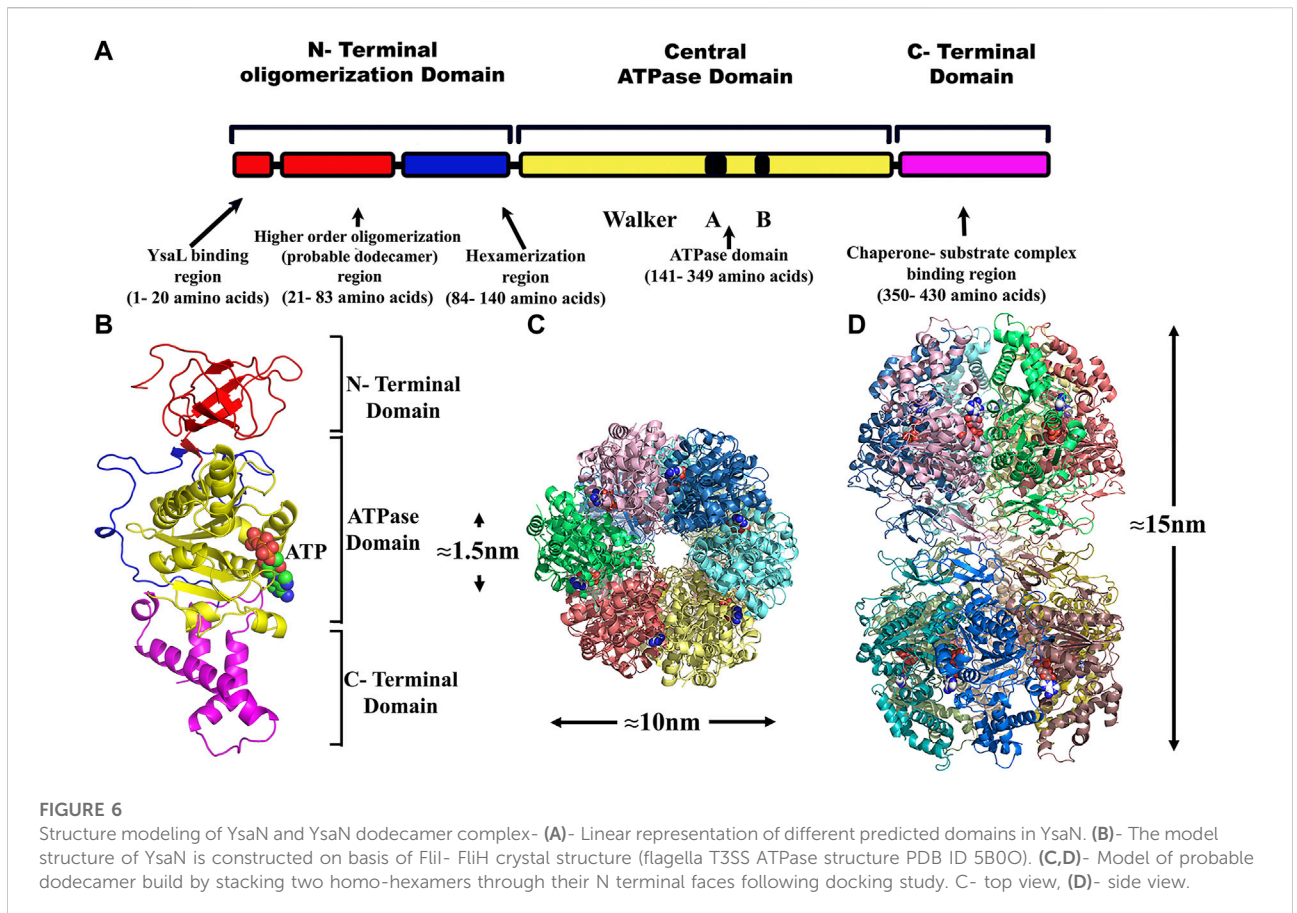ric double-ring structure. The double ring structure was modeled by docking the N- Terminal residue faces of two hexamers (Figures 6C,D). The dodecamer model of YsaN has approximately 15 nm length along with the stacked interface and approximately 10 nm diameter along each hexameric ring. A clear approximately 1.5 nm pore is visible from the top of the dodecamer structure. Following a similar protocol, YsaNΔ83 homology model was built from FliI template using MODELLER software. Based on their DOPE score, five models were chosen. The best model was utilized to construct a YsaN83 hexameric structure. YsaN83 hexamer was constructed using PyMOL software by fitting YsaN83 into the F1 -ATPase (PDB ID 4XD7) hexamer structure (Supplementary Figure S11).

4 Discussion

T3SSs are complex nano-syringe-like structures used by many pathogenic bacteria to inject effector toxins directly into the host cell cytoplasm (Coburn et al., 2007; Buttner, 2012). Efficient unfolding and translocation of an effector require active interaction of the effector-chaperone complex with the highly conserved T3SS ATPases inside (Akeda and Galan, 2005) bacterial cytoplasm. The oligomeric form of these ATPases is the most active form present at the export gate. In previous studies, hexamer and dodecamer have been reported as functional oligomeric forms of these ATPases (Pozidis et al., 2003; Andrade et al., 2007; Burgess et al., 2016). In *Y. enterocolitica*, YscN (a putative energizer of *ysc-yop* T3SS for Yop secretion) has been suggested to be activated through oligomerization and work as a hexamer (Woestyn et al., 1994). However, the true nature of the activation and formation of the different oligomers is still poorly understood. Here, we report the existence of a higher-order oligomer (probable dodecamer) as well as a hexamer, as the functional oligomeric form of ATPase YsaN from *Y. enterocolitica ysa-ysp* T3SS. In our previous *in-vitro* study, the possible dodecamer has been identified as the most active form of YsaN over the weakly active monomeric form (Chatterjee et al., 2013). Based on different biochemical and biophysical methods, the present study analyzes the mechanism of formation of different functional oligomers of YsaN and its correlation with the N-terminal region.

P*fam* domain analysis of YsaN protein sequence and its sequence homology with the identified and well characterized T3SS ATPase from *Shigella flexineri* (Spa47), *E. coli* (EscN), *Pseudomonas syringae* (HrcN), flagellar ATPase FliI and with the β-subunit of F_1-F_0 ATPase suggest that YsaN is also an T3SS associated ATPase of *Y. enterocolitica*. Recombinant expression of YsaN N-terminal His resulted in mixed YsaN population predominantly higher oligomeric form. Also, any deletion of YsaN render it insoluble and difficult to purify (Chatterjee et al., 2013). We have developed an expression system which allowed

FIGURE 6
Structure modeling of YsaN and YsaN dodecamer complex- **(A)**- Linear representation of different predicted domains in YsaN. **(B)**- The model structure of YsaN is constructed on basis of FliI- FliH crystal structure (flagella T3SS ATPase structure PDB ID 5B0O). **(C,D)**- Model of probable dodecamer build by stacking two homo-hexamers through their N terminal faces following docking study. C- top view, **(D)**- side view.

expression of untagged YsaN and soluble expression of any YsaN deletion construct in its native monomeric form. The ability to purify all the YsaN deletion construct in soluble monomeric form resulting from CPD fusion using CPD-vector has been crucial for this study (Shen et al., 2009). This allows us to capture the real-time oligomerization behaviour of YsaN (i.e., transition from monomer to hexamer and hexamer to higher-order oligomer as stable oligomeric form) and the involvement of its N-terminal region. Enzyme kinetics study of YsaN (untagged) obtained here are comparable to previous studies and other homologs (Andrade et al., 2007; Chatterjee et al., 2013; Case and Dickenson, 2018). Also, similar ATPase activity observed for YsaN and YsaNΔCterm suggests that YsaN ATPase activity is independent of its C-terminal domain. Whereas, the absence of ATPase activity in YsaNΔNterm and the presence of activity in YsaNΔ83 (however reduced as compared to YsaN) suggest that YsaN activity depends on its N-terminal domain (1–140 residues). Also, the presence of activity in YsaNΔ83 (an N-terminal deletion variant of YsaN) breaches the canonical thinking of loss of activity of N-terminal deletions variants of homolog ATPases (Burgess et al., 2016). The presence of the cooperative nature of YsaNΔ83 (h = 1.20 ± 0.06) also reveals the two-step cooperative behaviour of

YsaN activation. This indicates that the ATPase activity of YsaN requires both N-terminal regions: 1–83 and 84–140 residues in an independent manner. Hence, we propose that the complete activation of YsaN requires both of its N terminal regions including 1–83 and 84–140 amino acid residues.

DLS study suggests that YsaN changes its form upon ATP binding as a result of its oligomerization. YsaN in presence of AMPPNP and ADP.AlFX (both of which are ATP analog), shows different oligomerization behaviour in the DLS study. This difference in oligomerization behaviour may be a result of the requirement of active catalysis of ATP for YsaN oligomerization and not just ATP binding. This hypothesis was supported by YsaN 166K→A site-directed mutation study. In another experiment, removing Mg^{2+} ions to stop magnesium-dependent hydrolysis of ATP by using EDTA also resulted in the loss of oligomerization and activity of YsaN. Hence, we propose that the binding of ATP to YsaN is not the prime factor responsible for its oligomerization. Instead, the active catalysis of ATP to ADP + P_i may be the prime factor for YsaN oligomerization. This is similar to the observation in the case of F_1F_o- ATPase suggesting the idea of a universal mechanism in these molecular AAA + ATPases across their diverse roles inside the cell (Martin et al., 2014).

From Figures 3B,C, it was observed that both YsaN and YsaNΔ83 have the oligomerization tendency dependent on increasing ATP concentration. Also, the estimated size of the oligomer in the case of YsaN is larger than the oligomer formed in the case of YsaNΔ83. Oligomerization of T3SS ATPase of *Shigella*, Spa47 has been shown to be dependent on protein concentration (Andrade et al., 2007), whereas YsaN oligomerization occurs in an ATP (substrate) concentration-dependent manner. To investigate this, we have performed DLS of YsaN in a concentration-dependent manner. It was observed that by increasing the concentration of YsaN, the formation of higher-order oligomers was absent (Figure 3E). Conclusively, YsaN oligomerization is a substrate (ATP) concentration-dependent event and requires active catalysis of ATP to ADP + P*i*. It is also proposed that YsaNΔ83 bears the oligomerization behavior upon ATP binding and hydrolysis unlike, other homolog ATPases N-terminal deletion variants, which was previously considered non-functional. Further, to estimate the apparent molecular weight of the oligomeric complexes we performed analytical size exclusion chromatography (Figure 4). YsaN treated ADP.AlFX shows the presence of two complex forms corresponding to YsaN oligomer equivalent to probable dodecamer and hexamer corresponding to a molecular weight of approximately 576 and 288 kDa respectively. Whereas, ADP.AlFX treated YsaNΔ83 shows only a single complex of approximately 234 kDa corresponding to hexameric form along with an un-oligomerized peak. Also, we observed the oligomers in negative-TEM. In TEM study the observed size of YsaN and YsaNΔ83 oligomer was seen as ≈15–20 nm (Figure 6B) and ≈10 nm (Figure 6A) respectively. The observed sixfold symmetry of the YsaNΔ83 hexamer in TEM suggests that YsaN is capable of forming hexamer independent of N terminal 1–83 amino acid residues when bound to ADP.AlFX. This property of hexamer formation in T3SS ATPase with N-terminal 83 residues deleted has been observed for the first time. While in the case of YsaN, the higher observed size of the particles suggests the presence of a higher-order oligomeric form greater than hexamer which we propose as the YsaN probable dodecamer. In many attempts, during TEM sample preparation of YsaN oligomer, few of the particles greater than 20 nm were also observed which is certainly non-specific aggregates resulting from binding of ADP.AlFX (Supplementary Figure S6) which was absent in the case of YsaNΔ83 (Supplementary Figure S7). It seems difficult to control the *in-vitro* oligomerization, avoiding the formation of non-specific aggregates for YsaN. It is to mention that decreasing protein concentration limits the formation of aggregates resulting in fewer particles observed in TEM images. The 10 nm size of the YsaNΔ83 hexamer reported here is similar to the observed size of 10 nm for the homo hexameric T3SS ATPase in *E. coli* (Majewski et al., 2019). Conclusively, the N- terminal 1–83 residue of YsaN is crucial for higher-order oligomer formation while the formation of hexamer is independent of this region and requires 84–140 residues. The N-terminal

1–20 residues of YsaN are required for YsaL interaction (Chatterjee et al., 2013). The formation of higher-order oligomer was absent in the case of YsaNΔ83. Also, it has been observed that the YsaNΔNterm is incapable to form oligomer. The above observed results guided us to build a dodecamer model by stacking two hexamers through their N terminal faces (Figures 6C,D). A clear pore of about ≈1–1.5 nm can be easily seen confirming the possibility of N-to-N terminal stacking between two hexamers. Also, bioinformatic sequence analysis of YsaN shows the hydrophobic nature of the predicted beta-strand of the N-terminal domain primarily from 21 to 83 amino acid residues (Supplementary Figure S1). The hydrophobic nature of the predicted beta-strand of the N terminal domain of YsaN might play a crucial role in higher-order oligomer formation as shown in our model of YsaN hexameric complex surface topology map (Supplementary Figure S8). Hence, there exists the possibility that the formation of the higher-order oligomer (possible dodecamer) may take place by stacking two homo-hexamers by their N-terminal faces. Also, the dodecamer complex might be stabilized primarily by the hydrophobic interaction between N-terminal faces of two independent hexamers. We also propose that the formation of the most active form of YsaN oligomer proceeds in two distinct steps i.e., beginning with the formation of hexamer followed by the formation of higher-order oligomer depending on the ATP concentration.

Significant advancement in the knowledge has been achieved to date about the structure and mechanism of functionally least active monomeric and most active oligomeric form (hexamer or dodecamer) of T3SS ATPases. Yet, no study has been able to answer exactly which active oligomeric structure is required for substrate unfolding and translocation through the T3SS needle complex. The presence of hexamer ATPase ring complex at the T3SS export gate is widely accepted, and also has been observed in *in-situ* cryo-EM structures available to date. Whereas existence of hexamer and dodecamer ATPase complex during *in-vivo* and *in-vitro* studies (where the dodecameric form shows maximum activity) creates ambiguity and confusion about their role. Also, there is lacking of study addressing the exact mechanism of oligomerization among these highly conserved proteins. In the present study, we demonstrate the mechanism of formation of oligomers giving insight into the oligomeric nature of such T3SS ATPase, YsaN as an example. However, the exact mechanism of how these ATPases perform the unfolding and translocation of effectors is still not known. According to Akeda and Galan (2005), an appealing model mechanism among these T3SS ATPases similar to other AAA + ATPases is that the effector chaperone complex docks to such ATPase ring complex while unfolding the effector by threading them through the pore of the ATPase ring complex. The effector docking C- terminal side of the ATPase ring complex faces away from the cytosol towards the needle complex (Imada et al., 2016; Majewski et al., 2019). The hexamer model of ATPase at the export gate may result in possible steric hindrance for proper and free docking of the

chaperone effector complex to the C-terminal end of the ATPase ring complex. Instead, the proposed probable dodecameric ATPase ring complex, if present at the active T3SS removes the possibility of steric hindrance and allows free access for docking of chaperone effector complex to the ATPase ring complex towards the cytosolic C- terminal side. The other C-terminal side would be freely accessible for interacting with the positive regulator (SctO) linking it to the membrane component SctV towards the membrane side as has been reported in earlier studies (Zarivach et al., 2007; Kato et al., 2015; Majewski et al., 2019). Another aspect of the proposed dodecamer ATPase model also allows the possibility that unfolded effectors can pass through the narrow pore of the ATPase ring complex during the secretion process similar to the other AAA + protein translocases (Akeda and Galan, 2005). It is worth mentioning that the presence of higher order oligomer (probable dodecamer) at the export gate of T3SS has not been found in any *in-situ* experiment till date. While the presence of hexamer ATPase ring complex has been observed in recent *in-situ* Cryo-EM structure of *Salmonella* T3SS by Hu et al. (2017). The absence of the probable dodecamer ATPase ring complex observed in *in-situ* Cryo-EM structures suggests the possibility that it may be required only during engagement with the effector chaperone complex representing the secretion competent phase of T3SS. It may be noted that whether unfolded effector passes through the ATPase pore or not is still a matter of debate.

These highly conserved T3SS ATPases plays a crucial role in effector secretion and may be a significant target for drug development. To reveal the exact functioning of T3SS it is necessary to understand the role of such ATPases in T3SS regulation. The work reported in this study involving characterization of nature of YsaN oligomerization will provides a foundation for expanding our knowledge to understand T3SS regulation by these ATPases. Overall, the present study suggest that YsaN is an oligomerization activated T3SS ATPase in *Y. enterocolitica*. Also, formation of YsaN oligomer depend on active catalysis of ATP and the whole process of YsaN oligomerization and activation can be summed up as a two-step cooperative kinetic process involving formation of YsaN hexamer followed by formation of higher order oligomer. However, we could not establish the exact oligomeric nature of the most active form of YsaN higher order oligomer in the present study. Eventually, further experiments are necessary to explore the true mechanism of oligomeric ATPase ring complexes in T3SS regulation in *Yersinia enterocolitica* and other pathogenic bacteria.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

RK: Conceptualization, methodology and experimentation, writing original draft preparation, reviewing and editing. CR: Homology modelling, software validation. SD: Conceptualization, Writing- Reviewing and Editing of manuscript, and supervision and project administration

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.967974/full#supplementary-material

References

Abrusci, P., Vergara-Irigaray, M., Johnson, S., Beeby, M. D., Hendrixson, D. R., Roversi, P., et al. (2013). Architecture of the major component of the type III secretion system export apparatus. *Nat. Struct. Mol. Biol.* 20 (1), 99–104. doi:10.1038/nsmb.2452

Akeda, Y., and Galan, J. E. (2005). Chaperone release and unfolding of substrates in type III secretion. *Nature* 437 (7060), 911–915. doi:10.1038/nature03992

Allison, S. E., Tuinema, B. R., Everson, E. S., Sugiman-Marangos, S., Zhang, K., Junop, M. S., et al. (2014). Identification of the docking site between a type III secretion system ATPase and a chaperone for effector cargo. *J. Biol. Chem.* 289 (34), 23734–23744. doi:10.1074/jbc.M114.578476

Andrade, A., Pardo, J. P., Espinosa, N., Perez-Hernandez, G., and Gonzalez-Pedrajo, B. (2007). Enzymatic characterization of the enteropathogenic *Escherichia coli* type III secretion ATPase EscN. *Arch. Biochem. Biophys.* 468 (1), 121–127. doi:10.1016/j.abb.2007.09.020

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L. (2000). The Pfam protein families database. *Nucleic Acids Res.* 28 (1), 263–266. doi:10.1093/nar/28.1.263

Baykov, A. A., Evtushenko, O. A., and Avaeva, S. M. (1988). A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Anal. Biochem.* 171 (2), 266–270. doi:10.1016/0003-2697(88)90484-8

Bent, Z. W., Branda, S. S., and Young, G. M. (2013). The Yersinia enterocolitica Ysa type III secretion system is expressed during infections both *in vitro* and *in vivo*. *Microbiologyopen* 2 (6), 962–975. doi:10.1002/mbo3.136

Bent, Z. W., Poorey, K., Brazel, D. M., LaBauve, A. E., Sinha, A., Curtis, D. J., et al. (2015). Transcriptomic analysis of Yersinia enterocolitica biovar 1B infecting murine macrophages reveals new mechanisms of extracellular and intracellular survival. *Infect. Immun.* 83 (7), 2672–2685. doi:10.1128/IAI.02922-14

Burgess, J. L., Jones, H. B., Kumar, P., Toth, R. T. t., Middaugh, C. R., Antony, E., et al. (2016). Spa47 is an oligomerization-activated type three secretion system (T3SS) ATPase from Shigella flexneri. *Protein Sci.* 25 (5), 1037–1048. doi:10.1002/pro.2917

Buttner, D. (2012). Protein export according to schedule: Architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiol. Mol. Biol. Rev.* 76 (2), 262–310. doi:10.1128/MMBR.05017-11

Case, H. B., and Dickenson, N. E. (2018). MxiN differentially regulates monomeric and oligomeric species of the Shigella type three secretion system ATPase Spa47. *Biochemistry* 57 (15), 2266–2277. doi:10.1021/acs.biochem.8b00070

Chatterjee, R., Halder, P. K., and Datta, S. (2013). Identification and molecular characterization of YsaL (Ye3555): A novel negative regulator of YsaN ATPase in type three secretion system of enteropathogenic bacteria Yersinia enterocolitica. *PLoS One* 8 (10), e75028. doi:10.1371/journal.pone.0075028

Chen, B., Doucleff, M., Wemmer, D. E., De Carlo, S., Huang, H. H., Nogales, E., et al. (2007). ATP ground- and transition states of bacterial enhancer binding AAA+ ATPases support complex formation with their target protein, sigma54. *Structure* 15 (4), 429–440. doi:10.1016/j.str.2007.02.007

Coburn, B., Sekirov, I., and Finlay, B. B. (2007). Type III secretion systems and disease. *Clin. Microbiol. Rev.* 20 (4), 535–549. doi:10.1128/CMR.00013-07

Cornelis, G. R., Boland, A., Boyd, A. P., Geuijen, C., Iriarte, M., Neyt, C., et al. (1998). The virulence plasmid of Yersinia, an antihost genome. *Microbiol. Mol. Biol. Rev.* 62 (4), 1315–1352. doi:10.1128/MMBR.62.4.1315-1352.1998

Cornelis, G. R. (2002). Yersinia type III secretion: Send in the effectors. *J. Cell. Biol.* 158 (3), 401–408. doi:10.1083/jcb.200205077

DeLano, W. L. (1874). Pymol: An open-source molecular graphics tool. *CCP4 Newsl. protein Crystallogr.* 40 (1), 82

Dey, S., Biswas, C., and Sengupta, J. (2018). The universally conserved GTPase HflX is an RNA helicase that restores heat-damaged *Escherichia coli* ribosomes. *J. Cell. Biol.* 217 (7), 2519–2529. doi:10.1083/jcb.201711131

Diepold, A., and Armitage, J. P. (2015). Type III secretion systems: The bacterial flagellum and the injectisome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20150020. doi:10.1098/rstb.2015.0020

Diepold, A., and Wagner, S. (2014). Assembly of the bacterial type III secretion machinery. *FEMS Microbiol. Rev.* 38 (4), 802–822. doi:10.1111/1574-6976.12061

Edgren, T., Forsberg, A., Rosqvist, R., and Wolf-Watz, H. (2012). Type III secretion in Yersinia: Injectisome or not? *PLoS Pathog.* 8 (5), e1002669. doi:10.1371/journal.ppat.1002669

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* 66 , 486–501. doi:10.1107/S0907444910007493

Foultier, B., Troisfontaines, P., Muller, S., Opperdoes, F. R., and Cornelis, G. R. (2002). Characterization of the ysa pathogenicity locus in the chromosome of Yersinia enterocolitica and phylogeny analysis of type III secretion systems. *J. Mol. Evol.* 55 (1), 37–51. doi:10.1007/s00239-001-0089-7

Gasteiger, H. C. E., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., Bairoch, A., et al. (2005). "Protein identification and analysis tools on the ExPASy server," in *The proteomics protocols handbook*. Editor John M. Walker (Humana PressHumana Press), 571

Ghosh, P. (2004). Process of protein transport by the type III secretion system. *Microbiol. Mol. Biol. Rev.* 68 (4), 771–795. doi:10.1128/MMBR.68.4.771-795.2004

Gophna, U., Ron, E. Z., and Graur, D. (2003). Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene* 312, 151–163. doi:10.1016/s0378-1119(03)00612-7

Halder, P. K., Roy, C., and Datta, S. (2019). Structural and functional characterization of type three secretion system ATPase PscN and its regulator PscL from *Pseudomonas aeruginosa*. *Proteins* 87 (4), 276–288. doi:10.1002/prot.25648

Haller, J. C., Carlson, S., Pederson, K. J., and Pierson, D. E. (2000). A chromosomally encoded type III secretion pathway in Yersinia enterocolitica is important in virulence. *Mol. Microbiol.* 36 (6), 1436–1446. doi:10.1046/j.1365-2958.2000.01964.x

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77 (1), 51–59. doi:10.1016/0378-1119(89)90358-2

Hu, B., Lara-Tejero, M., Kong, Q., Galan, J. E., and Liu, J. (2017). *In situ* molecular architecture of the Salmonella type III secretion machine. *Cell* 168 (6), 1065–1074. doi:10.1016/j.cell.2017.02.022

Imada, K., Minamino, T., Uchida, Y., Kinoshita, M., and Namba, K. (2016). Insight into the flagella type III export revealed by the complex structure of the type III ATPase and its regulator. *Proc. Natl. Acad. Sci. U. S. A.* 113 (13), 3633–3638. doi:10.1073/pnas.1524025113

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292 (2), 195–202. doi:10.1006/jmbi.1999.3091

Kato, J., Lefebre, M., and Galan, J. E. (2015). Structural features reminiscent of ATP-driven protein translocases are essential for the function of a type III secretion-associated ATPase. *J. Bacteriol.* 197 (18), 3007–3014. doi:10.1128/JB.00434-15

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10 (6), 845–858. doi:10.1038/nprot.2015.053

Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., et al. (2013). How good is automated protein docking? *Proteins* 81 (12), 2159–2166. doi:10.1002/prot.24403

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., et al. (2017). The ClusPro web server for protein-protein docking. *Nat. Protoc.* 12 (2), 255–278. doi:10.1038/nprot.2016.169

Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., et al. (1998). Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* 280 (5363), 602–605. doi:10.1126/science.280.5363.602

Lanzetta, P. A., Alvarez, L. J., Reinach, P. S., and Candia, O. A. (1979). An improved assay for nanomole amounts of inorganic phosphate. *Anal. Biochem.* 100 (1), 95–97. doi:10.1016/0003-2697(79)90115-5

Lee, P. C., and Rietsch, A. (2015). Fueling type III secretion. *Trends Microbiol.* 23 (5), 296–300. doi:10.1016/j.tim.2015.01.012

Majewski, D. D., Worrall, L. J., Hong, C., Atkinson, C. E., Vuckovic, M., Watanabe, N., et al. (2019). Cryo-EM structure of the homohexameric T3SS ATPase-central stalk complex reveals rotary ATPase-like asymmetry. *Nat. Commun.* 10 (1), 626. doi:10.1038/s41467-019-08477-7

Martin, J. L., Ishmukhametov, R., Hornung, T., Ahmad, Z., and Frasch, W. D. (2014). Anatomy of F1-ATPase powered rotation. *Proc. Natl. Acad. Sci. U. S. A.* 111 (10), 3715–3720. doi:10.1073/pnas.1317784111

Muller, S. A., Pozidis, C., Stone, R., Meesters, C., Chami, M., Engel, A., et al. (2006). Double hexameric ring assembly of the type III protein translocase ATPase HrcN. *Mol. Microbiol.* 61 (1), 119–125. doi:10.1111/j.1365-2958.2006.05219.x

Pozidis, C., Chalkiadaki, A., Gomez-Serrano, A., Stahlberg, H., Brown, I., Tampakaki, A. P., et al. (2003). Type III protein translocase: HrcN is a peripheral ATPase that is activated by oligomerization. *J. Biol. Chem.* 278 (28), 25816–25824. doi:10.1074/jbc.M301903200

Rappas, M., Schumacher, J., Beuron, F., Niwa, H., Bordes, P., Wigneshweraraj, S., et al. (2005). Structural insights into the activity of enhancer-binding proteins. *Science* 307 (5717), 1972–1975. doi:10.1126/science.1105932

Schumacher, J., Zhang, X., Jones, S., Bordes, P., and Buck, M. (2004). ATP-dependent transcriptional activation by bacterial PspF AAA+protein. *J. Mol. Biol.* 338 (5), 863–875. doi:10.1016/j.jmb.2004.02.071

Shen, A., Lupardus, P. J., Morell, M., Ponder, E. L., Sadaghiani, A. M., Garcia, K. C., et al. (2009). Simplified, enhanced protein purification using an inducible, autoprocessing enzyme tag. *PLoS One* 4 (12), e8119. doi:10.1371/journal.pone.0008119

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi:10.1038/msb.2011.75

Trott, O., and Olson, A. J. (2010). AutoDock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31 (2), 455–461. doi:10.1002/jcc.21334

Webb, B., and Sali, A. (2014). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.* 47, 5 6 1–32. doi:10.1002/0471250953.bi0506s47

Wilharm, G., Lehmann, V., Krauss, K., Lehnert, B., Richter, S., Ruckdeschel, K., et al. (2004). Yersinia enterocolitica type III secretion depends on the proton motive force but not on the flagellar motor components MotA and MotB. *Infect. Immun.* 72 (7), 4004–4009. doi:10.1128/IAI.72.7.4004-4009.2004

Wilson, K. (2001). Preparation of genomic DNA from bacteria. *Curr. Protoc. Mol. Biol.* 2, 2.4. doi:10.1002/0471142727.mb0204s56

Woestyn, S., Allaoui, A., Wattiau, P., and Cornelis, G. R. (1994). YscN, the putative energizer of the Yersinia Yop secretion machinery. *J. Bacteriol.* 176 (6), 1561–1569. doi:10.1128/jb.176.6.1561-1569.1994

Yoshida, Y., Miki, T., Ono, S., Haneda, T., Ito, M., and Okada, N. (2014). Functional characterization of the type III secretion ATPase SsaN encoded by Salmonella pathogenicity island 2. *PLoS One* 9 (4), e94347. doi:10.1371/journal.pone.0094347

Zarivach, R., Vuckovic, M., Deng, W., Finlay, B. B., and Strynadka, N. C. (2007). Structural analysis of a prototypical ATPase from the type III secretion system. *Nat. Struct. Mol. Biol.* 14 (2), 131–137. doi:10.1038/nsmb1196

Biophysical and Computational Approaches to Unravel pH-Dependent Conformational Change of PspA Assist PspA-PspF Complex Formation in *Yersinia enterocolitica*

Chittran Roy[1] · Rajeev Kumar[1] · Md Maruf Hossain[1] · Arkaprava Das[1] · Saumen Datta[1]

Abstract

In enteropathogen, *Yersinia enterocolitica*, the genes encoding phage shock proteins are organized in an operon (*pspA-E*), which is activated at the various types of cellular stress (i.e., extracytoplasmic or envelop stress) whereas, PspA negatively regulates PspF, a transcriptional activator of *pspA-E* and *pspG*, and is also involved in other cellular machinery maintenance processes. The exact mechanism of association and dissociation of PspA and PspF during the stress response is not entirely clear. In this concern, we address conformational change of PspA in different pH conditions using various in-silico and biophysical methods. At the near-neutral pH, CD and FTIR measurements reveal a ß-like conformational change of PspA; however, AFM measurement indicates the lower oligomeric form at the above-mentioned pH. Additionally, the results of the MD simulation also support the conformational changes which indicate salt-bridge strength takes an intermediate position compared to other pHs. Furthermore, the bio-layer interferometry study confirms the stable complex formation that takes place between PspA and PspF at the near-neutral pH. It, thus, appears that PspA conformational change in adverse pH conditions abandons PspF from having a stable complex with it, and thus, the latter can act as a trans-activator. Taken together, it seems that PspA alone can transduce adverse signals by changing its conformation.

Keywords Phage shock protein A · pH · Intermediate helix (3/10-helix) · Salt-bridge

Abbreviations
PspA Phage shock protein A
PspF Phage shock protein F
CD   Circular dichrorism
FTIR Fourier transform infrared spectroscopy
BLI   Bio-layer interferometry
SB   Salt-bridge

1 Introduction

The phage shock protein (*psp*), operon (*pspA-E*), system protects the cell from a variety of physical and chemical shocks across the cytoplasmic membrane. The adaptor protein, PspA of the PspA/IM30 family, is a ubiquitous and conserved member of this system. PspA regulates *pspA-E* and *pspG* at their transcription level with the help of PspF (an AAA + ATPase), a partner protein in the cytoplasm [1–4]. PapA also interacts with membrane-bound PspBC proteins in a cellular stress condition [1, 5]. Based on extracellular environment (normal or stressed) and interaction with partner proteins, it has been observed that PspA interacts with different cellular components in various locations of the cell such as at the inner or peripheral membrane and in the cytosol [1, 6, 7]. PspA typically consists of three to four coiled-coil domains. The N-terminal coiled-coil domain of PspA (144 residues) is involved in interaction with PspF [1, 4]. The membrane interaction and oligomerization domains follow this PspF interaction domain [1]. The lower oligomeric form of PspA interacts with PspF while, PspA forms a higher

Chittran Roy and Rajeev Kumar have contributed equally to this work.

✉ Saumen Datta
  saumen_datta@iicb.res.in

[1] Structural Biology and Bioinformatics Division, Council of Scientific and Industrial Research – Indian Institute of Chemical Biology, 4, Raja S. C. Mullick Road, Jadavpur, Kolkata, West Bengal 700032, India

oligomeric structure during membrane interaction [4, 6, 8]. Although this PSP system has been extensively studied in *E. coli*, it works similarly in *Y. enterocolitica* and *Salmonella enterica* [9–11]. The genes for the PSP system in *Y. enterocolitica* are located at two unlinked loci arranged as pspF-pspABCDycjXF and pspG [9]. The conserved pspFABC gene regulates the physiological function of the cell. It has been reported earlier in *Y. enterocolitica* that the association of PspA with membrane via PspBC depends on the monomer or pore-forming multimer state of secretin5 [12–14]. The location of PspA inside the cell depends on its interaction with partner proteins [15]. It has been shown in previous studies that PspA deletion is associated with different kinds of physiological defects such as—a defect in prolonging stationary phase at highly alkaline pH or reduced PMF due to proton leakage across the cytoplasmic membrane and increased stored curvature elastic (SCE) stress [11, 16, 17]. However, what regulates the interaction of PspA with different partner proteins in different subcellular locations in response to various physiological conditions of the cell is still not clear. It is also a matter of concern regarding the effect of change in the cell pH value as a result of proton leakage on the structure of PspA. Although PspA of *Y. enterocolitica* is similar to *E. coli* (identity ~ 78%), its structural information is still a matter of future research and examination. PspA has at least three coiled-coil sequence stretches present within 1–181 residues which has a heptad repeat. Leu and Ala are highly abundant in this repeat along with the polar and charged residues which are almost equally predominant [18]. Again, Leu, Ala (hydrophobic), charged residues (except Asp), and Gln (polar) have high helix-propensity that further supports the helical structure of the coiled-coil [19]. It is worth mentioning that the C-terminal part (beyond 181 residues) of the PspA sequence needs further investigation. Since there is no Tyr and only a Trp, PspA is an ideal system for intrinsic fluorescence-based characterization [20]. Since PspA is rich in titratable amino acid side chains, it is relevant to know how a change in pH value affects the conformation of the protein.

The present study focuses on the research and investigation of the issues mentioned above. In this study, a detailed characterization of the acidic, neutral and alkaline states of PspA by different biophysical methods has been described. Besides, the focus is also given on the effects of different pH states of PspA governing its interaction with PspF. Further, a description of the differential structural features including specific electrostatic strength of the end-states from the pH-dependent molecular dynamic simulation is also presented. We believe that the insights we have received from this study to transduce pH-dependent signals will shed light on the characteristic function of PspA.

2 Materials and Methods

2.1 Sequence Retrieval, Generation of PspA and PspF Model

The amino acid sequence of PspA and PspF from *Y. enterocolitica* were retrieved from the UniProt public protein database (http://www.uniprot.org) and using the FASTA file, the secondary structure was predicted using the SOPMA web tool [21]. The model of PspA and PspF were generated by the I-Tasser server [22]. In this methodology, both the protein sequences were loaded in FASTA format to I-Tasser as the input file. Additional restrain or templates were not assigned by the user. The threading programs of I-Tasser assigned the top 10 threading templates. From the output files, the best model was represented in the VMD molecular graphic system [23]. The PDB file of the best model of both the proteins generated by I-Tasser was again run by the NAMD simulation package [24] in presence of an explicit water box. The minimization was performed and the best frame with the lowest potential energy was selected. The water box and added hydrogen atoms were removed and the coordinate was normalized and finally saved in PDB format. Structural evaluation and stereochemical analysis of both modeled proteins were analyzed by various bioinformatics tools and software packages such as RAMPAGE, and ERRAT (Table S1).

2.2 Cloning, Expression, and Purification of PspA and PspF

pspA (untagged) gene was PCR amplified from *Y. enterocolitica* genomic DNA using a set of primers for *pspA* gene: Forward:(GATCAGATCTCATGGGTATTTTTTCTCGTTTTG) and Reverse: (TCGACTCGAGTTATTGCGCTGAATTCATTTTTG). The 666 bp PCR amplified fragment was then cloned in pET Duet1 vector (Novagen) with Bgl2 and Xho1 restriction sites in MCS 2. Cloning of *pspF* was done using forward primer- GGGAATTCCATATGAGTGAGCAATTAGAAAATC and reverse primer- CGGGATCCGATTGTGGATCATTCTCAC in vector pET22b (+) for C terminal His-tagged expression. Expression of both no-tagged PspA and His-tagged PspF was carried out in *E. coli* BL21 (DE3). Cells were grown at 37 °C at 180 rpm until it has reached an optical density of 0.6 to 0.8. Protein expression was induced using IPTG with a concentration of 0.5 mM and the culture was further grown for 4 h at 37 °C for PspA and 10 h at 200C for PspF. After that cell were harvested by centrifugation at 6000 rpm for 10 min at 4⁰C. Cell pellets were lysed using ultrasonication with the addition of 1 mM phenyl

methyl sulfonyl fluoride (PMSF). Cell debris was pellet down by centrifugation at 10,000 rpm for 45 min at 40C. PspA (untagged) was purified further by the Ion-Exchange method. The clear supernatant was loaded into a pre-equilibrated anion exchanger column (Hi-Trap QFF) and a flow rate of 0.5 ml/min was maintained throughout the chromatography. To increase the purity further, gel filtration was performed using Superdex 200 16/60 pg (GE Healthcare) column. Gel filtration was done in gel filtration buffer containing 10 mM MES (pH 6.3), 40 mM NaCl, and 1% glycerol v/v. Most of the PspA was eluted in void showing its higher-order oligomer and a small amount of protein was also eluted in the monomeric form (Fig. 1b). Whereas, refolding was performed for PspF due to the insoluble nature of PspF (N-terminal His tagged) during expression. The lysed cell pellet obtained after centrifugation containing the inclusion was treated with buffer containing chaotropic agents (such as 8 M Urea). After that, dialysis with buffer (10 mM Tris pH8.0, 60 mM NaCl, 1% glycerol) was used to remove the chaotropic agents from the protein solution. To achieve better purity of the protein

solution, Gel filtration in Superdex 200 10/30 pg column was performed using a buffer containing 10 mM Tris pH 8.0, 60 mM NaCl, and 1% Glycerol buffer (Fig. 1d). To check the purity of both the proteins, eluted proteins were analyzed on 15% w/v SDS-PAGE (Fig. 1c, e).

2.3 CD Spectra

Circular dichroism [25, 26] measurements of PspA and PspF were performed using a Jasco J-815 spectrophotometer at 250C for far-UV CD experiments with a scanning speed of 50 nm/min, and spectra were recorded over the wavelength range 190–240 nm (Fig. 3). Far UV CD spectra were performed at a protein concentration of 20 μM (working) using a 0.2 cm path-length cuvette. Far-UV CD spectra were recorded with a step size of 1 nm and bandwidth of 1 nm. Buffer contribution was subtracted from each of the protein spectra to eliminate the background effect. The experiments were performed in triplicates and the final spectra were generated by averaging three scans. The mean residual ellipticity
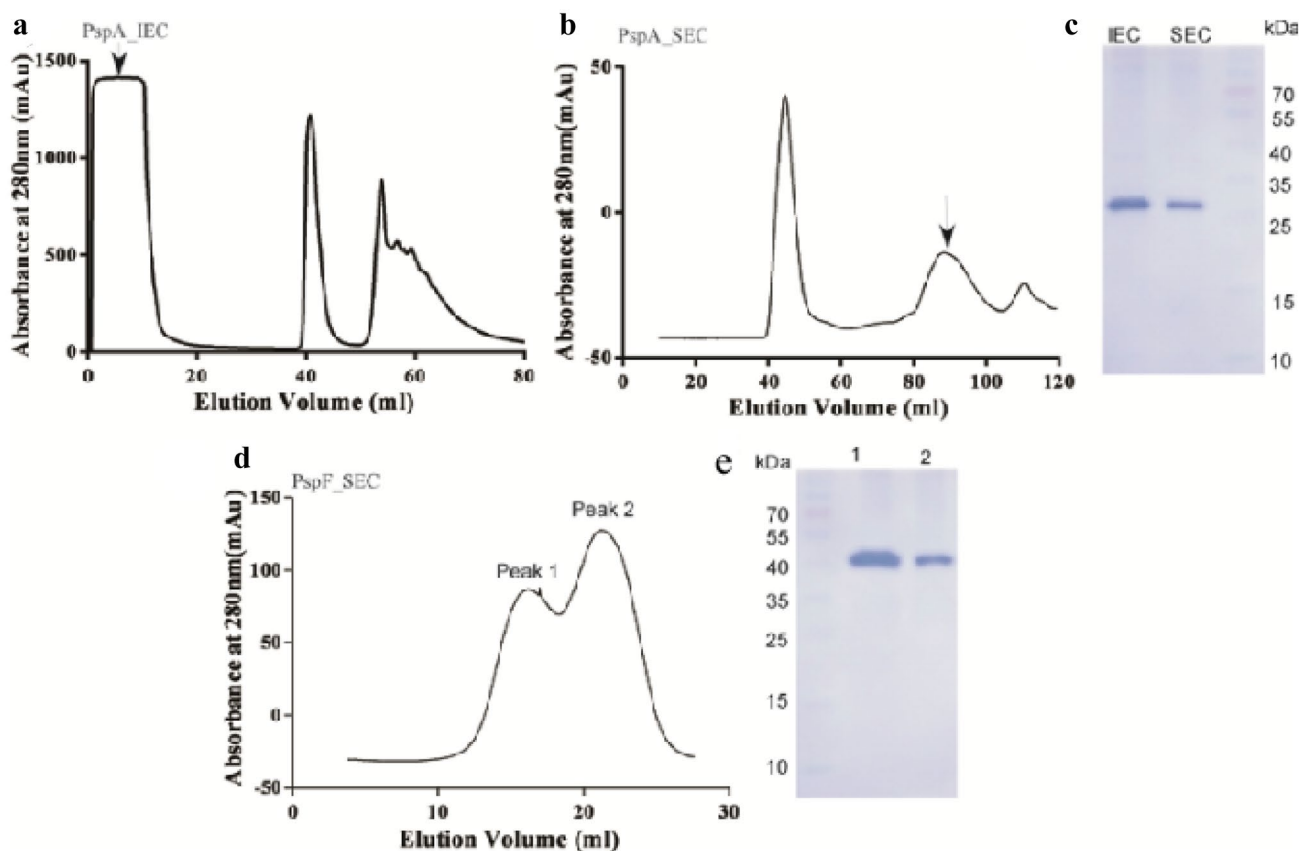
Fig. 1 Assessment of the oligomeric and monomeric states of PspA and PspF. **a** Ion-exchange chromatography of PspA shows, maximum elution of PspA came out in flow-through (arrow mark). **b** Size exclusion chromatography separates oligomeric (peak 1) and monomeric (peak 2) states of PspA. **c** Purity checked by SDS-PAGE. **d** Size exclusion chromatography separates oligomeric (peak 1) and monomeric (peak 2) state of PspF. **e** Purity checked by SDS-PAGE

(MRE) was calculated using a mean residual weight (MRW) of 114 for the PspA as per the following expression:

$$[\theta]_{mrw.\lambda} = MRW * \left(\frac{\theta_\lambda}{10 * d * c} \right)$$

here θ_λ is the observed ellipticity at wavelength λ, d is the path length (cm), and c is the concentration (g/mL) [27].

Experimental data were analyzed in the K2D2 server [28] to calculate the percentage of different secondary structure content of PspA protein(Table S2).

Near-UV CD was done at 25 °C with a scanning speed of 50 nm/min for tertiary structure identification, and spectra were acquired over the wavelength range of 250–350 nm using a Jasco J-815 spectrophotometer. PspA, PspF, and the PspA-PspF complex had working protein concentrations of 20 μM, 20 μM, and 20 μM, respectively. Figure S1 shows the near-UV CD results.

2.4 FTIR Spectra

Conformational changes of PspA at different pH values were recorded using Bruker Tensor 27 FTIR spectrometer. Before the experiment, the protein was incubated at various pH conditions overnight. 20 μM protein concentrations were used for the FTIR experiment. Samples were ten drop-cast onto FTIR substrate and allowed for reading. Buffer baseline subtractions were done before taking each spectrum. The deconvolution of raw spectra in the amide I region (1600 cm^{-1} to 1700 cm^{-1}) was performed by curve fitting shape by Prism software. The assignment of peck value was taken using the criteria proposed by Chirgadze & Nevskaya [29] and Krimm & Bandekar [30].

2.5 AFM Experiment

AFM in dry mode using a mica chip was performed using a 5 μM purified protein sample which was incubated for 1 min and diluted at least 100-fold in 0.22 m filtered milliQ, and 2 μl of this sample was loaded onto the center of MUSCO-VITE MICA, 20 mm diameter chip (Electron Microscopy Sciences, Hatfield, PA). The sample was left to air dry at room temperature. Acoustic alternative current (AAC) mode was performed on a Pico plus 5500 AFM instrument (Agilent Technologies, USA). Microfabricated silicon cantilevers of 225 μm in length with a nominal spring force constant of 21–98 N/m were used using Nanosensors (PPP-NCL). The cantilever scanned the surface with a speed of 0.8- 1.2 μm/s and a resonance frequency of 168.4 kHz. Micrographs showing the surface topology of the protein samples were captured after scanning each sample several times and images were processed using Pico View 1.10.1 software (Agilent Technologies, USA) and Pico image advanced software for 3D rendering process and height profile analysis. All AFM experiments were repeated at least three times with freshly purified samples each time.

2.6 Weak Force Estimation of PspA

Weak forces such as hydrogen bond, electrostatic (π-cation, π-anion, ion-pair, salt-bridge, etc.) interaction, hydrophobic (π–π, π–σ, π–amide, π–alkyl, alkyl–alkyl) interaction were extracted using BIOVIA Discovery Studio 2020 with the default parameters for distance and angle.

2.7 Molecular Dynamics and Salt-Bridge Energy Calculation of PspA at Different pH

Protonated states were induced in PspA at different pH values using a freely available PropKa server (https://server.poissonboltzmann.org) to calculate the pK value. The generated protonation states for PspA were used for MD simulation at the pH of 5.3, 7.3, and 9.3 (simulation protocols are described in Supplementary materials). In most of the pH conditions, the two-alpha helix of PspA remains in the coil-coil conformation whereas, at pH 7.3 shows a significant change. The isolated pair method has been used earlier for the extraction of isolated-pair salt-bridge energy terms using the PDB2PQR [31] and APBS [32] programs [33–35]. The energy terms for NU salt-bridge were extracted using the NUM method, since, unlike IP, NU is made up of more than two salt-bridge partners [33, 36]. The net energy of a salt bridge is the sum of the component energy terms. Average accessibility was determined using the NACCESS program [33, 36, 37].

2.8 Biolayer Interferometry

Biolayer interferometry (BLI) experiments were done using the OctetRED96e with Ni–NTA coated biosensor tips. Analyte (PspA) of 6 different concentrations (2.5, 5, 10, 30, 50, and 70 μM) were overnight incubated in three different pH (5.3, 7.3, and 9.3) conditions. During the experiment, a sample volume of 200 μL was used in 96 well plates and kept agitated. The Ni–NTA biosensor tips were hydrated in a buffer containing 60 mM Tris pH 8.0, 60 mM NaCl, and 1% glycerol throughout all the experiments. Baseline adjustments were done with the biosensor tips dipped in buffer solution without any analyte. After the baseline adjustment with a buffer solution containing tris-buffer (pH 8), the biosensor tips were incubated in the well containing ligand (PspF) with a concentration of 40 μM for 300 s for binding of the ligands to the biosensor tips. Ligand bound biosensor tips were then dipped into the buffer solution for 60 s to remove any excess ligand bound to the biosensor tips followed by incubating the biosensor tips for 120 s in the

wells containing different concentrations of the untagged analyte (PspA) in an increasing order ranging from 2.5 to 70 μM. In the last step, the tips were again dipped in the buffer solution for 180 s to analyze the dissociation step. The Biolayer interferometry experiments used His-tagged PspF as the analyte, binding 40 uM of the protein to Ni–NTA biosensor tips. Then the tips were transferred to solutions with PspA at different concentrations at the studied pH values. Though the binding affinity of His-tags to Ni–NTA is pH-dependent yet, there is no significant dissociation pattern observed (Fig. S5).

3 Results

3.1 Physicochemical Analysis and Structure Prediction of PspA and PspF

In the protein sequence, the code of the 3D structure and interaction are intertwined [38]. Thus, to understand the distinguishing features of the PspA and PspF sequence, we have compared their physicochemical properties. The average hydrophobicity and hydrophilicity of PspA are much lower and higher than PspF respectively (Fig. 2a, b). Again, in PspA, although, hydrophobic residues (HB) are somewhat

compromised due to the predominance of charged residues (CR) compared to PspF (Fig. 2c). The 90–110 region of the sequence carries a strong hydrophobic character (Fig. S3A). On the other hand, a large region (~ 100 residues) in the middle of the sequence of PspF has strong hydrophobic characters (Fig. S4B). Overall, the PspA sequence has more (18.6%) strong disorder forming residues (S, E, R, and P) than PspF (13%), which is due to its C-terminal end (Fig. 2d). The results of the secondary structure prediction of these two proteins are different. The major part of the PspA secondary structure is made up of helix (~ 95%), and in the case of PspF, it is much more consistent (Fig. 2e).

Since, PspA and PspF of *Y. enterocolitica* have no structure in the protein structure database, we have determined their structure using the *ab-initio* method [22]. Both PspA and PspF structures are based on their entire sequence. It is to be mentioned that these model structures also qualified by authentic validation procedures (Table S1). Interestingly, the PspA structure is largely composed of the helix as was found in secondary structure prediction where the N-terminal helices (from residue 24 to residue 144) produce two prominent characteristic coiled-coil domain structures. It is worth noting here that the same structure can be seen in case of *E Coli*. Significantly, the sequence from the residue, 144 to the end of the C-terminal region shows fragments

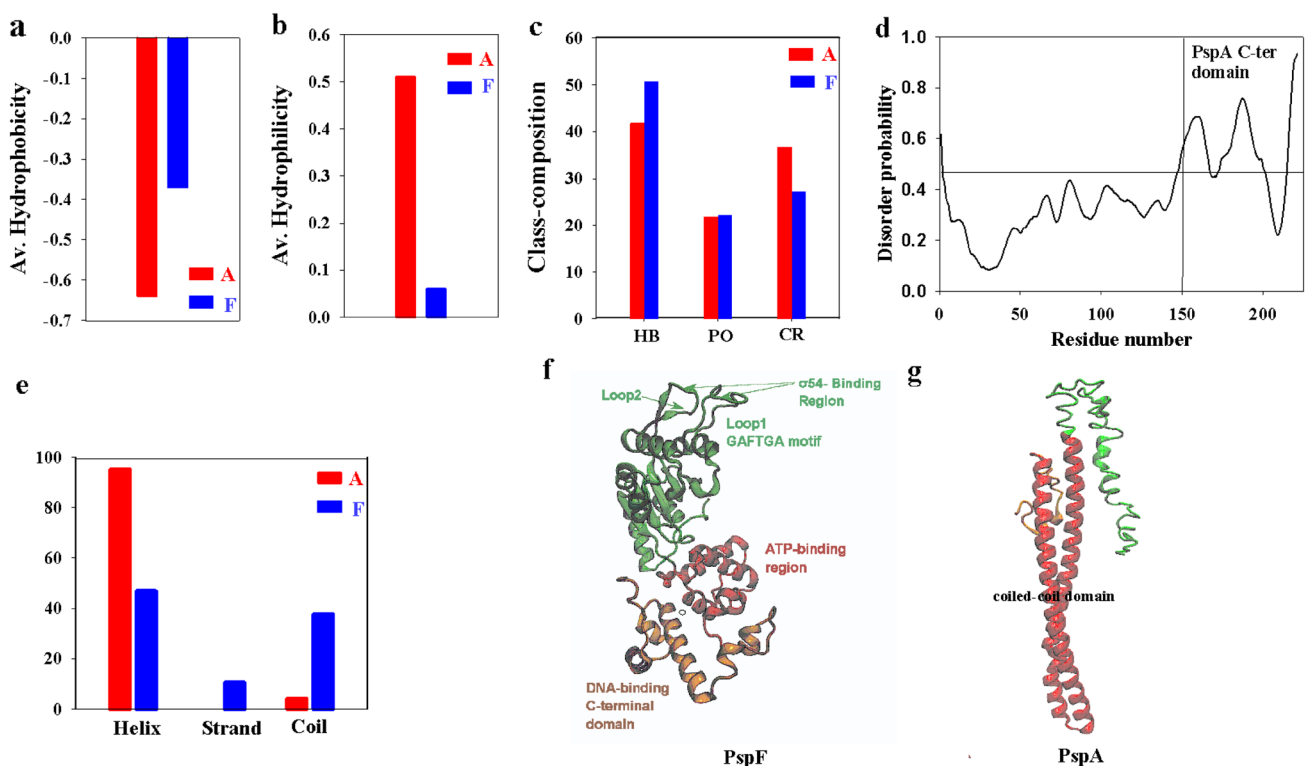**Fig. 2** Physicochemical properties and model structure of PspA and PspF. Comparison of PspA and PspF average hydrophobicity (**a**), average hydrophilicity (**b**), and amino acid class composition (**c**). Dis-order probability of PspA (**d**). Comparison between different types of the secondary structure of PspA and PspF (**e**). PspF (**f**) and PspA (**g**) predicted model structure

of short secondary structures, but this region is in disarray (Fig. 2g). PspF, which is an enhancer-binding protein (EBP), on the other hand, distinctly owns at least two domains, the C-terminal domain with helix and turns and the N-terminal domain including the central domain with helix and beta structure (i.e., α/β domain). It was discovered that this protein interacts with DNA via the C-terminal domain (helical-domain). The N-terminal and central domains are responsible for interaction with ATP for its hydrolysis and activation of the N-terminal bound RNAP-σ54 holoenzyme. The cleft of the ATP interaction region is also highlighted (Fig. 2f). Here, the two highly conserved loops namely; loop-1 and loop-2 of the N-terminal domain are noteworthy. Sequence-motif of the tip of loop-1, GAFTGA, and motif-sequence of loop-2, VGGSNP interact with RNAP-σ54 complex. Conclusively, it can be said that in our two model structures of PspA and PspF of *Y. enterocolitica*, the characteristic features found in the case of other species are well revealed. Below we show the influence of pH on the structure and structural elements.

3.2 The Conformation of PspA Secondary Structure Depends on the Environmental pH

The secondary structure can be considered as the determinant of protein topology and conformation, which can be probed accurately using circular dichroism and FTIR spectroscopy. We asked a question about whether a change in the environment pH affects the PspA confirmation or not. To check this, we recorded CD spectra at various pH values (Fig. 3aI). The far-UV CD spectra of PspA at pH 5.3, 6.3, 8.3, and 9.3 show minima at 208 nm and 224 nm which characteristic of α-helix structure. But, in the case of pH 7.3, it is much different (Fig. 3a). The difference between this pH form and other pHs is much clear in the ellipticity ratio (CD222:CD208) (Fig. 3b). Notably, at pH7.3, a minima on at around 214 nm, and a prominent maxima at 200 nm are observed (Fig. 3a), implying helix to coil transition. Apart from the above observation, there are also slightly negative minima present between 210–220 nm across all pH values, which may indicate a β-like conformer. Such conformation is most prominently visible at pH 7.3 (Fig. 3c). Furthermore, at the pH of 6.3, a single minima was also observed at the position near 200 nm (Fig. 3aI), suggesting its disorderliness [39]. The CD spectra collected at several pH values were further analyzed using the K2D2 online web server (Supplementary materials Table S2).

The secondary structure of PspA at various pH values was further explored by FTIR spectroscopy, represented in Table S3. The FTIR outcome substantiates the CD results. The FTIR spectra of PspA at different pH values within the range of 1600–1700 cm-1 probe the vibrational frequency of different amide bonds of the main protein chain
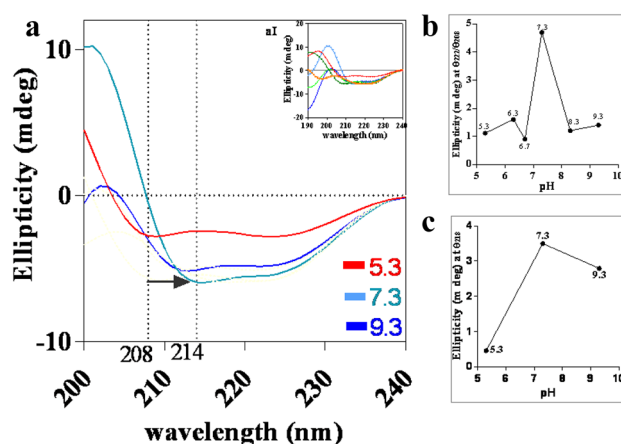
Fig. 3 Change of secondary structure of PspA in response to pH. **a** Comparison of far-uv CD spectra of PspA at pH 5.3 (red), 7.3 (cyan), and 9.3 (blue). **b** Change of θ_{222} to θ_{208} ratio of PspA at different pH. **c** Relative value of β structure with pH (218 nm)
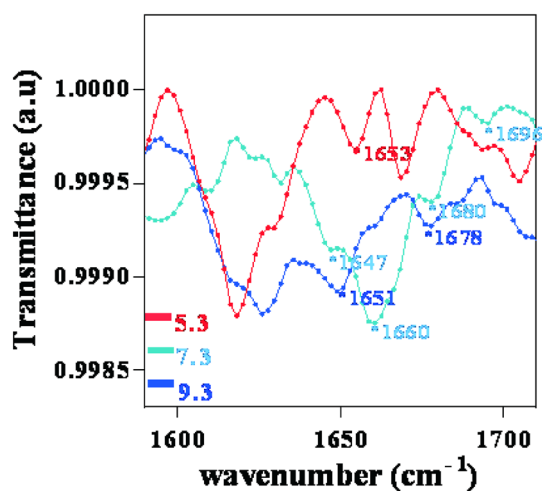
Fig. 4 Comparison of FTIR spectral characteristics of PspA at different pH. Spectra showing conformation of secondary structure where, α-helix, intermediate helix (3/10-helix), β-turn, and β-sheet conformer denoted as an asterisk (*). Spectra of PspA at pH 7.3 (cyan) shows β-turn β-sheet and intermediate of α-helical conformation (3/10-helix where, at the pH of 5.3 (red) and 9.3 (blue) shows abundance of α-helix

related to the secondary structure are shown in Fig. 4. At pH 5.3 and 9.3, the secondary structure of PspA has likely a-helical conformation (both having a wavelength of 1653 cm-1 and 1651 cm-1 respectively). The intermediate helix (3/10-helix) at 1647 cm-1 is also faintly visible at pH 7.3, but noticeable changes have been observed at pH 7.3 (Fig. 4), where, the protein has a little tendency for combining extended β-strand (1660 cm-1, 1696 cm-1) and β-turn (1680 cm-1) like conformation. We wanted to know how PspA changes its oligomeric state, which helps it respond to external stress during pH changes. To

visualise the oligomers, we performed a pH-dependent AFM experiment.

3.3 PspA Oligomerization may be Aided by Conformational Changes at Certain pH Levels

PspA oligomeric state depends on its cellular location, interaction, and function [40]. We know that lower oligomeric form interacts with PspF in the cytosol and higher oligomeric form interacts with the PspB-PspC complex at the membrane [41]. What has pH got to do with pspA oligomeric association and dissociation equilibria? To check this out, we've performed the results of the pH-dependent AFM experiment. It is found that the lower oligo-form prevails more at neutral pH than in acidic and alkaline pH conditions (Fig. 5). Furthermore, as a chain of adhering beads at acidic and basic pH values. That means the structure of PspA at neutral pH is most likely the ideal form for interaction with PspF. Following the above study, prompted us to understand the week-force interaction of a modelled PspA structure at various pH levels using an in-silico technique. Is there any variation of weak force stability with pH changes? To understand this, we investigate specific electrostatic and other weak interactions.

3.4 Salt-Bridge Helps to Maintain the Overall Tertiary Structure of Various Structural Conformers at Different pH Conditions of PspA

To find out the effect of pH on PspA confirmation, we have done pH-dependent MDS. The result of which is presented in Fig. 6. Several points are noteworthy. First, the N-terminal coiled-coil interacts with PspF. This coiled-coil structure of 24 to 144 residues shows pH-dependent bending, extending from residue 58 to 110 at pH 7.3. Although, this characteristic bending is present at all pH values, the residue expansion of the bend region and bend-curvature show a wide variation. Judging by these criteria, it can be said that this bending in the case of pH 7.3 proves to be more flexible in comparison to other pH values (Fig. 6aII). However, 61%, 65%, and 68% of the helix are present at pH 5.3, pH 7.3, and pH 9.3 respectively. These analyses were performed using the Stride web server [42]. At Neutral pH values, the presence of an anti-parallel β-like structure is another contrasting characteristic as compared to other pH values (Fig. 6bI).

Second, we see the change of conformation with the change in pH values. How does this change in conformation affect the salt-bridge and its energetics? Salt-bridges are generally of two types namely IP and NU types. With the change of pH, there is a wide variation in the frequency of IP and NU types of salt-bridge shown in Table 1. Notable, the specificity of the salt-bridge is altered due to the change in the pH-dependent conformation. The overall frequency of salt-bridge at pH values of 5.3, 7.3, and 9.3 is 19, 16, and 18, respectively. In other words, the changes in salt-bridge frequency with the change of pH supports the alteration of conformation of PspA. With the change in specificity, the change in the location (core or surface) of the salt-bridge is also significant. The pH 5.3 form has some NU types of salt-bridge in the core, whereas the pH 7.3 and pH 9.3 have all the IP and NU type salt-bridges at the surface (Table S5-S10). Although at pH 7.3 salt-bridge frequency is minimal, per-salt-bridge stability point is highest (Table 1). Similarly, the analysis of other weak interactions shows some variations where the hydrogen bonds are the most followed by hydrophobic and electrostatic interactions (Table S11). Third, to find out the conformational difference of PspA in alkaline and acidic with compared to neutral pH, we have superposed the structures of pH 5.3 and pH 9.3 with pH 7.3. The RMSD of structures at pH 7.3 vs. pH 5.3 and pH 7.3 vs. pH 9.3 are 9.1 Å and 8.7 Å, respectively. This proves that the conformation of the structures of acidic and alkaline conditions deviates significantly from the structure at neutral pH. Now, the question arises, which of these structures
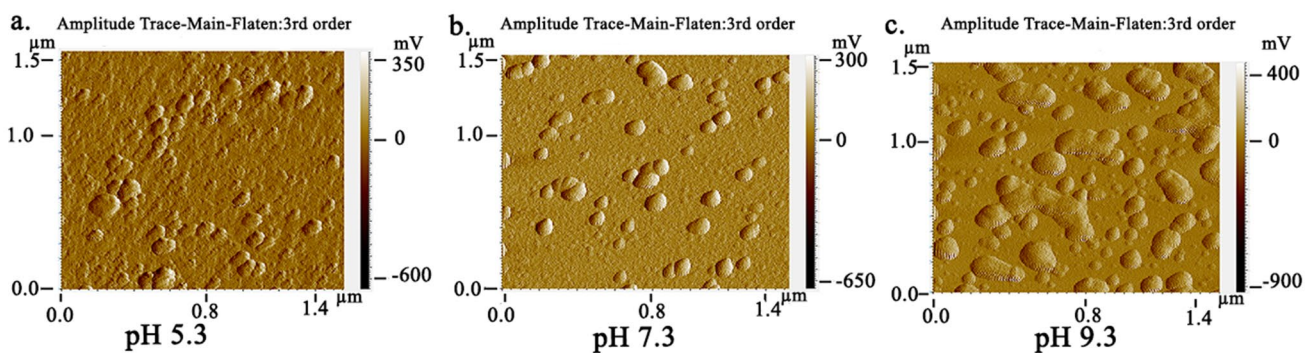

Fig. 5 Levels of oligomerization of PspA at different pH. Compared to pH 7.3 (**b**), the level of oligomerization in pH 5.3 (**a**) and 9.3 (**c**) is visible. In pH 5.3 and 9.3 PspA look higher order chain like oligomers whereas less/scattered oligomeric nature observed in pH 7.3
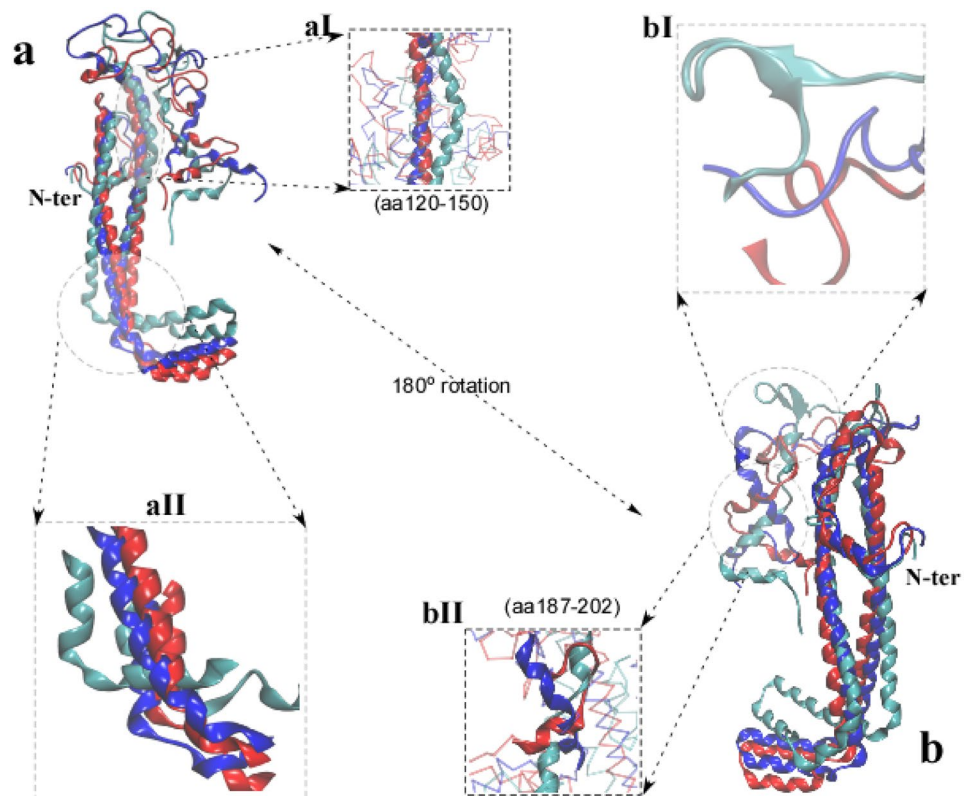
Table 1 pH-dependent IP and NU salt-bridge frequency, overall and per-salt-bridge energetic

Protein	IP pair	NU pair	Total pair	IP energy	NU energy	Total energy	Energy/SB
ph5.3	6	13	19	− 24.9	− 57.91	− 82.81	− 4.35842
ph7.3	10	6	16	− 53.79	− 32.21	− 86	− 5.375
ph9.3	13	5	18	− 68.42	− 15.79	− 84.21	− 4.67833

IP isolated pair, *NU* network unit

binds most strongly with PspF. To understand this, we have performed the binding experiment (see below).

3.5 The Binding Stability of PspA with PspF is Regulated by pH

Interaction between PspA and PspF was monitored at different pH conditions using the optical analytical technique, bio-layer interferometry (Fig. 7). The dissociation constant of PspA and PspF in pH 5.3 is very high i.e., fast dissociation. In this respect, if we move the pH towards neutral i.e., 7.3, the average dissociation constant is the lowest, meaning that the interaction affinity of PspA and PspF is the highest (1.55 μM). Interestingly, at the alkaline pH (pH 9.3), the interaction efficiency of these two proteins reduces. Taken together it can be said that the affinity of the interaction between PspA and PspF is highest at near-neutral pH and has maximum and moderate deviation in acidic and alkaline conditions, respectively. The best fit model of PspA and

pspF interaction data shows that their binding stoichiometry is 2: 1 that is, one mole of PspF interacts with two moles of PspA (Table S4).

4 Discussion

The structure, function, and metabolic state of the cell are normal if the components of the environment in which the cell is functional are in order. When stress emerges in the environment, the cell may change its normal metabolic state to protect itself from it. In gram-negative bacteria, a multimeric secretin complex is needed to trade the protein and other cellular materials across the cell membrane. In *Y. enterocolitica* secretin mislocalization results in activation of the PSP system [14]. PspA acts as a major player by mitigating the stress of the cell by freeing its negative control partner, PspF which also enhances the impact of the innate response system [6]. In the absence of stress, PspA remains

Fig. 7 Overview of interaction between PspF and PspA at different pH. Binding affinity of PspF with PspA at various pH, determined by binding kinetics assay of bio-layer interferometry. The affinity of the interaction between PspA and PspF is highest at pH 7.3 (cyan), which shows the maximum and moderate deviation in acidic (pH 5.3) and alkaline (pH 9.3) conditions, respectively. The best fit m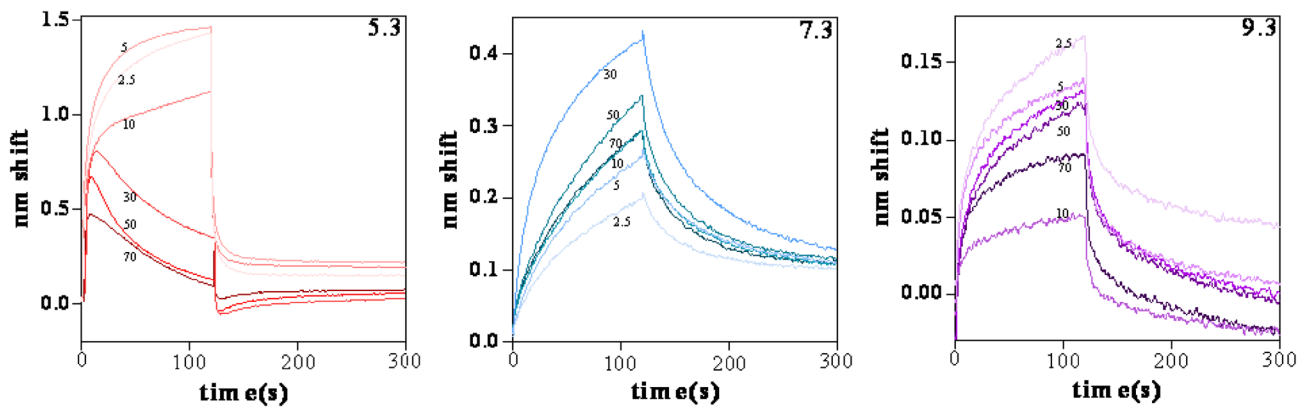odel representing PspA-PspF complex formation shows that their binding stoichiometry is 2:1 that is, one mole of PspF interacts with two moles of PspA

in association with its cytosolic partner PspF while during stress conditions PspA is released from PspF and localizes to the membrane. Since *Y. enterocolitica* is a pathogen that causes many infections and deaths every year. So, the understanding of the PSP system seems to be useful for basic and application purposes.

Hydrophilicity and hydrophobicity in the PspA sequence are largely region-specific, which can probably help maintain its interaction and specificity with multiple partners. Again, it is natural that the conformation of PspA is related to a change in the environmental pH, as it has a higher proportion of charged residues. Extracellular stress is accompanied by protrusion of cell protons, which can be particularly sensitive to the structure of PspA and its conformation. For all these reasons, we undertook the study of pH-dependent conformational change of PspA and its binding strength with PspF.

The codes of the structure of a protein are embedded in the protein sequence. Due to the lack of available structure of PspA and PspF in the structure database, an *ab-initio* procedure was used to generate the structures. The N-terminal of PspA is composed of a coiled-coil domain while its C-terminal domain is often disordered. Although all the basic weak forces are important in this structure, their normalized frequency shows some differences in different pH with the highest count of hydrogen bonds. Such interaction of the inter-helix interface is also possible with other incoming helices in the unoccupied coiled-coil interface. Thus, it appears that a coiled-coil motif is helpful in protein–protein interaction and oligomerization to form a much higher order coiled-coil structure. Again, the intrinsically disordered region of the C-terminal part of PspA can act as a determinant of specificity in this interaction. It has been shown earlier that disorder proteins achieve their appropriate

geometry and orientation only after interaction with their partner proteins [43].

Apart from specific-electrostatic (salt-bridge or ion-pairs) interaction, other interaction like pi-cation, pi-anion, pi-alkyl, etc., although having low or negligible frequency, appear to act as determinants of specificity. It has been shown that pi–pi interaction, in general, has abundance at the core of a typical protein. While in the case of PspA, the absence of pi–pi interaction suggests the importance of the weak forces described above. It is especially important in the structure of PspA at pH 7.3 that has the highest stability from a small number of pairs. These interactions help to maintain more stability at pH 7.3 as compared to at pH 5.3 and pH 9.3.

In addition, the bending and the nature of flexibility at bend-junction of PspA coiled-coil at pH 7.3 are also different from other pH. Such bending is suitable for docking of a partner protein. From this, it can be inferred that the interactivity of this pH form is higher than others. In fact, the Kd value of PspA interaction with PspF, was found to be the support the above-mentioned contention. The interaction between PspA and PspF, which prefers a low oligomeric state [41], is favorable at pH 7.3. This means, that at pH (pH 7.3), PspA has a role to play in the equilibrium states. Needless to say, the sub-cellular location of interaction partner-protein, and oligomerization states of PspA are different [16]. While the interaction of PspA and PspF favors the lower oligomeric form of PspA. The interaction of PspA and PspBC favors the higher oligomeric form. The change in conformation of PspA with environment pH has also been observed in molecular dynamic data. Circular dichroism and FTIR spectroscopy seem to be closely related to the orientation, and geometry of the backbone. The observed binding ratio during PspA-PspF

complex formation in case of *E. coli* is 6:6 [4], however, in PspA-PspF complex formation in *Y. enterocloitica* suggest a 2:1 binding ratio. Similar observation has been found earlier by Heidrich et al.[44]. The exact cause of this slight decline is unknown, but the solution state of PspA points to the variation of its interactivity. The fact that the interaction of PspA with PspF and stabilization of PspA-PspF complex structure at near-neutral pH is greater than the other pH. This interaction does not depend solely on the helix of the N-terminal part of PspA because the protein at the near neutral pH has an anti-parallel β like structure in the C-terminal region, which is missing from the structure at other pH. Also, the C-terminal region is more disordered so the appearance of this structure at neutral pH seems to be a result of the coil to β transition.

Finally, PspA is said to be a protective protein against cell stress that can transduce various types of stress into its conformation change. Broadly, this conformation change works like a binary switch, which controls PspF mediated transactivation of RNAP-σ54. Now if this PspA binds to PspF in cytosol it means the normal metabolic state of the cell. In other words, PspA negatively regulates PspF. In this case, PspA has a low oligomeric state. If extra-cellular stress occurs in the cell membrane, the oligomeric state of the PspA changes. As a result, this oligomer of PspA interacts with PspBC in the membrane, while PspF is freed from PspA and begins the work of the above-mentioned transactivation. In this context, our study demonstrates that acidic and alkaline pH environment may act as a stress signal for PSP system activation via PspA. Since the predominance of a charged residue is inherently higher in the sequence of PspA, the variation of pH affects the titration state of the side chain of these residues. In addition, the salt-bridges that are in a stable form in near-neutral pH (pH 7.3) are also affected by this pH variation. As a result, the overall stability of a protein varies with the alteration of pH, which may affect the proteins global conformation. We believe that this PspA in neutral pH is a negative regulator of PspF and at acidic or alkaline pH, it adopts very useful higher oligomeric form that interact with PspBC on the membrane. Further, it may be that in the in vivo situation, the state of PspA that exists in the cytosol or membrane associated can be mimicked by the change of pH in the in vitro condition. We think that it will be possible to shed light on the mechanism of stress signal transduction by PspA via its pH-dependent conformational change.

Declarations

Competing interests The authors declare no competing interests.

References

1. Osadnik H, Schopfel M, Heidrich E, Mehner D, Lilie H, Parthier C, Risselada HJ, Grubmuller H, Stubbs MT, Bruser T (2015) PspF-binding domain PspA1–144 and the PspA.F complex: New insights into the coiled-coil-dependent regulation of AAA+ proteins. Mol Microbiol 98:743–759

2. Brissette JL, Weiner L, Ripmaster TL, Model P (1991) Characterization and sequence of the *Escherichia coli* stress-induced psp operon. J Mol Biol 220:35–48

3. Dworkin J, Jovanovic G, Model P (2000) The PspA protein of *Escherichia coli* is a negative regulator of sigma(54)-dependent transcription. J Bacteriol 182:311–319

4. Joly N, Burrows PC, Engl C, Jovanovic G, Buck M (2009) A lower-order oligomer form of phage shock protein A (PspA) stably associates with the hexameric AAA(+) transcription activator protein PspF for negative regulation. J Mol Biol 394:764–775

5. Srivastava D, Moumene A, Flores-Kim J, Darwin AJ (2017) Psp stress response proteins form a complex with mislocalized secretins in the *Yersinia enterocolitica* cytoplasmic membrane. mBio 8:e01088-17

6. Adams H, Teertstra W, Demmers J, Boesten R, Tommassen J (2003) Interactions between phage-shock proteins in *Escherichia coli*. J Bacteriol 185:1174–1180

7. Engl C, Jovanovic G, Lloyd LJ, Murray H, Spitaler M, Ying L, Errington J, Buck M (2009) In vivo localizations of membrane stress controllers PspA and PspG in *Escherichia coli*. Mol Microbiol 73:382–396

8. Hankamer BD, Elderkin SL, Buck M, Nield J (2004) Organization of the AAA(+) adaptor protein PspA is an oligomeric ring. J Biol Chem 279:8862–8866

9. Darwin AJ, Miller VL (2001) The psp locus of Yersinia enterocolitica is required for virulence and for growth in vitro when the Ysc type III secretion system is produced. Mol Microbiol 39:429–444

10. Eriksson S, Lucchini S, Thompson A, Rhen M, Hinton JC (2003) Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. Mol Microbiol 47:103–118

11. Beloin C, Valle J, Latour-Lambert P, Faure P, Kzreminski M, Balestrino D, Haagensen JA, Molin S, Prensier G, Arbeille B, Ghigo JM (2004) Global impact of mature biofilm lifestyle on *Escherichia coli* K-12 gene expression. Mol Microbiol 51:659–674

12. Flores-Kim J, Darwin AJ (2016) The phage shock protein response. Annu Rev Microbiol 70:83–101

13. Joly N, Engl C, Jovanovic G, Huvet M, Toni T, Sheng X, Stumpf MP, Buck M (2010) Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology. FEMS Microbiol Rev 34:797–827

14. Yamaguchi S, Gueguen E, Horstman NK, Darwin AJ (2010) Membrane association of PspA depends on activation of the phage-shock-protein response in Yersinia enterocolitica. Mol Microbiol 78:429–443

15. Yamaguchi S, Reid DA, Rothenberg E, Darwin AJ (2013) Changes in Psp protein binding partners, localization and behaviour upon activation of the Yersinia enterocolitica phage shock protein response. Mol Microbiol 87:656–671

16. Southern SJ, Male A, Milne T, Sarkar-Tyson M, Tavassoli A, Oyston PC (2015) Evaluating the role of phage-shock protein A in Burkholderia pseudomallei. Microbiology (Reading) 161:2192–2203

17. Weiner L, Model P (1994) Role of an *Escherichia coli* stress-response operon in stationary-phase survival. Proc Natl Acad Sci USA 91:2191–2195

18. Chambers TJ, Hahn CS, Galler R, Rice CM (1990) Flavivirus genome organization, expression, and replication. Annu Rev Microbiol 44:649–688

19. Williams RW, Chang A, Juretic D, Loughran S (1987) Secondary structure predictions and medium range interactions. Biochim Biophys Acta 916:200–204

20. Hennig R, West A, Debus M, Saur M, Markl J, Sachs JN, Schneider D (1858) The IM30/Vipp1 C-terminus associates with the lipid bilayer and modulates membrane fusion. Biochim Biophys Acta Bioenerg 2017:126–136

21. Geourjon C, Deleage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11:681–684

22. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738

23. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(33–38):27–38

24. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802

25. Greenfield N, Fasman GD (1969) Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry 8:4108–4116

26. B.J.A.C. Franck, Optical Circular Dichroism. Principles, Measurements, and Applications. Von L. Velluz, M. Legrand und M. Grosjean, übers. von J. MacCordick. Verlag Chemie GmbH., Weinheim/Bergstr., und Academic Press, New York-London, 1965. XII, 247 S., 149 Abb., 10 Tab., geb. DM 40.–, 77 (1965) 875–875.

27. Joshi V, Shivach T, Yadav N, Rathore AS (2014) Circular dichroism spectroscopy as a tool for monitoring aggregation in monoclonal antibody therapeutics. Anal Chem 86:11606–11613

28. Louis-Jeune C, Andrade-Navarro MA, Perez-Iratxeta C (2012) Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. Proteins 80:374–381

29. Nevskaya NA, Chirgadze YN (1976) Infrared spectra and resonance interactions of amide-I and II vibration of alpha-helix. Biopolymers 15:637–648

30. Krimm S, Bandekar J (1986) Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. Adv Protein Chem 38:181–364

31. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB-2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res 32:W665-667

32. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 98:10037–10041

33. Banerjee S, Gupta PSS, Islam RNU, Bandyopadhyay AK (2021) Intrinsic basis of thermostability of prolyl oligopeptidase from *Pyrococcus furiosus*. Sci Rep 11:11553

34. Nayek A, Sen Gupta PS, Banerjee S, Mondal B, Bandyopadhyay AK (2014) Salt-bridge energetics in halophilic proteins. PLoS One 9:e93862

35. Roy C, Kumar R, Datta S (2020) Comparative studies on ion-pair energetic, distribution among three domains of life: Archaea, eubacteria, and eukarya. Proteins 88:865–873

36. Bandyopadhyay AK, Islam RNU, Mitra D, Banerjee S, Goswami A (2019) Stability of buried and networked salt-bridges (BNSB) in thermophilic proteins. Bioinformation 15:61–67

37. Chen J, Sawyer N, Regan L (2013) Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. Protein Sci 22:510–515

38. Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230

39. Lopes JL, Miles AJ, Whitmore L, Wallace BA (2014) Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: applications in secondary structure analyses. Protein Sci 23:1765–1772

40. Junglas B, Huber ST, Heidler T, Schlosser L, Mann D, Hennig R, Clarke M, Hellmann N, Schneider D, Sachse C (2021) PspA adopts an ESCRT-III-like fold and remodels bacterial membranes. Cell 184:3674–3688

41. Mehta P, Jovanovic G, Lenn T, Bruckbauer A, Engl C, Ying L, Buck M (2013) Dynamics and stoichiometry of a regulated enhancer-binding protein in live *Escherichia coli* cells. Nat Commun 4:1997

42. Heinig M, Frishman D (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res 32:W500-502

43. Widłak W (2013) Protein structure and function. In: Widłak W (ed) Molecular biology: not only for bioinformaticians. Springer, Berlin, pp 15–29

44. Heidrich ES, Bruser T (2018) Evidence for a second regulatory binding site on PspF that is occupied by the C-terminal domain of PspA. PLoS ONE 13:e0198564

RESEARCH ARTICLE

PROTEINS WILEY

Comparative studies on ion-pair energetic, distribution among three domains of life: Archaea, eubacteria, and eukarya

Chittran Roy | Rajeev Kumar | Saumen Datta

Structural Biology and Bioinformatics Division, Council of Scientific and Industrial Research—Indian Institute of Chemical Biology, Kolkata, West Bengal, India

Correspondence
Saumen Datta, Structural Biology and Bioinformatics Division, Council of Scientific and Industrial Research—Indian Institute of Chemical Biology, 4, Raja S. C. Mullick Road, Jadavpur, Kolkata 700032, West Bengal, India.
Email: saumen_datta@iicb.res.in

Abstract

Salt-bridges play a unique role in the structural and functional stability of proteins, especially under harsh environments. How these salt-bridges contribute to the overall thermodynamic stability of protein structure and function across different domains of life is elusive still date. To address the issue, statistical analyses on the energies of salt-bridges, involved in proteins' structure and function, are performed across three domains of life, that is, archaea, eubacteria, and eukarya. Results show that although the majority of salt-bridges are stable and conserved, yet the stability of archaeal proteins ($\Delta\Delta G_{net}$ = −5.06 ± 3.8) is much more than that of eubacteria ($\Delta\Delta G_{net}$ = −3.7 ± 2.9) and eukarya ($\Delta\Delta G_{net}$ = −3.54 ± 3.1). Unlike earlier study with archaea, in eukarya and eubacteria, not all buried salt-bridge in our dataset are stable. Buried salt-bridges play surprising role in protein stability, whose variations are clearly observed among these domains. Greater desolvation penalty of buried salt-bridges is compensated by stable network of salt-bridges apart from equal contribution of bridge and background energy terms. On the basis proteins' secondary structure, topology, and evolution, our observation shows that salt-bridges when present closer to each other in sequence tend to form a greater number. Overall, our comparative study provides insight into the role of specific electrostatic interactions in proteins from different domains of life, which we hope, would be useful for protein engineering and bioinformatics study.

KEYWORDS
accessible surface area, ion-pair, PDB, PDB2PQR, salt-bridge

1 | INTRODUCTION

Proteins maintain an intricate balance between rigidity and flexibility for function. The majority of charged residues of protein that form salt-bridges are present on proteins' surface, reducing excess protein hydration and making it less hydrophobic and more flexible. It also helps to overcome the deleterious effect of salt by promoting non-specific electrostatic interaction with salts in solution.[1,2] So, ion-pair or salt-bridge is one of the major contributors to the stability of protein structure and function.[3-5] It is much more visible when the proteins are adapted in the extremes of physical and chemical environments such as high salt and high temperature.[6-9] It has been demonstrated that by inserting a single salt-bridge by mutation in the

protein surface increases protein stabilization in the tertiary structure.[10] So, engineered salt-bridge on the protein surface may contribute to the excess of stability.

The energy of salt-bridge can be subdivided into three components, that is, columbic attraction between opposite charge, bridge interactions ($\Delta\Delta G_{brd}$), their desolvation ($\Delta\Delta G_{dslv}$), and background ($\Delta\Delta G_{prt}$) interactions. Columbic attraction is always contributing in tertiary structure of proteins and other two terms could be either contributing or costly. The favorable charge-charge attraction within a salt-bridge is often opposed by unfavorable desolvation of charges and is further modulated by charge-dipole interaction as well as by the ionization behavior of nearest charged group.[11] Thus, net ion-pair energy distribution could either be stabilizing[12-15] or destabilizing[11,16,17] or insignificant.[18]

Computational study based on Poisson-Boltzmann-Equation (PBE), a theoretical analysis, was carried out on the crystal structure of ferredoxin and malate dehydrogenase from *Haloarcula marismortui* and glutamate dehydrogenase from thermophlilic *Pyrococcus furiosus* and mesophilic from *Clostridium symbiosum* for understanding contribution of salt and pH on classical electrostatic stability.[19,20] Calculation of net energy of salt-bridge is easily possible by the use of PBE. It is an ideal electrostatic descriptor for bimolecular system and Delphi[9,11,14,15,21,22] and APBS[23] are most popular solver.

Here we analyze the distribution, energy terms, and contribution of salt-bridges to the overall stability of proteins using a dataset of 24 high resolution (≤1.5 Å) crystal structure from three domains of life (eukarya, eubacteria, and archaea). Comparative analysis and understanding of the differences in salt-bridge and its energetics in archaea relative to others are the main highlights of this study. This comparative study with salt-bridge seems to have potential applications in the understanding of protein adaptation in extreme environments.

2 | MATERIALS AND METHODS

2.1 | Salt-bridge dataset

We obtain salt-bridge from 24 high resolution with Debye-Waller B-factor or temperature factor (atomic mean square displacement from a solved crystal structure), whose Enzyme Commission (EC) number is 3 and are available in protein data bank (PDB).[24] The three-dimensional (3D) structure of this protein has a resolution better than or equal to 1.5 Å with R factor of not greater than 0.25 and sequence identity of ~30%. The PDB code of these proteins are: 1l7m, 2bjd, 3g91, 3wr0, 4e19, 5dhd, 5fot, 5j8n, 1is9, 1kg2, 1nz0, 2c71, 2pof, 2pqx, 2vng, 4b1m, 4ql3, 5a67, 5b8d, 5jig, 5uq6, 5wlf, 6avx, and 6dg4 (Supporting Information Table S1).

2.2 | Extraction and categorization of salt-bridge

In 3D protein structure, ion-pair forms due to interaction of side chains of oppositely charged residues (Asp or Glu pair with Lys or Arg or His) and an ion-pair is defined as a salt-bridge if it fulfills the criteria: (a) the centroid of the side chain charged groups in two oppositely charged residues lying within 4.0 Å[25] and (b) at least one pair carboxyl oxygen atom of Asp or Glu side chain and nitrogen atom of Lys, Arg, or His side chain are within 4.0 Å distance.[15,21] Of the observed salt-bridges, we calculated the percentage of individual acidic and basic residue involves in salt-bridge. We also categorize the salt-bridge based on hydrogen bonded or nonhydrogen bonded, buried or solvent accessible and network or isolated.

2.3 | Sequence enrichment

Within secondary structure of protein, the side chain-side chain contacts were included: between $i \rightarrow i + 4$, $i \rightarrow i + 3$, $i + 2$, and $i + 1$

(Supporting Information Figure S1) for inter- or intra-helical/strand/coil or inter-helical/strand/coil. Analysis of all possible salt-bridge interaction study was done by SBION2.[26]

2.4 | Protein-energy minimization

The 3D protein structure was subjected for refinement via conjugate gradient energy minimization by NAMD simulation package[27] in presence of explicit water box. The minimization was performed for 5000 steps with a collection of frames at the interval of 200 steps and thus a total of 25 frames were collected. Each frame potential energy was judged by NAMD analyses plug-in in VEGAZZ protein visualizing interface.[28] The best frame with lowest potential energy was selected. The water box and added hydrogen atoms were removed and coordinate was normalized and finally saved in PDB format.

2.5 | Salt-bridge electrostatic energy contribution by computational approaches

Using an APBS methodology, we can analyze the electrostatic interaction of protein by computational approaches. APBS calculates free energy based on the distribution of electrostatic potential within the atomic charge, according to the equation:

$$G_{elec} = \frac{1}{2}\sum q_i \varphi_i,$$

where, q_i and φ_i are the charges and electrostatic potential at each grid point, respectively. PBE is a successful description of electrostatics in 3D structure of a protein. In this method, dielectric constant inside the protein is relatively low. So, the charges of ionizable groups and the partial charges of permanent dipole are assigned to the corresponding atoms according to the 3D structure of proteins. The solvent is represented as higher dielectric medium, which correlates with a higher ability of the solvent to dissolve salts. APBS method[23] and Delphi software package[9,11,14,15,21,22] of PBE calculation are popularly used for finding salt-bridge energy. Salt-bridge energy calculation by computational approaches was observed by experimental method.[17,29] We have followed the former procedure as constructed by Hendsch and Tidor (1994) along with ADSBET2.[30] The total electrostatic free energy ($\Delta\Delta G_{net}$) can be decomposed into three energy terms: (a) $\Delta\Delta G_{dslv}$ is the energy difference caused by desolvation of charged residues forming salt-bridge in the folded state of protein. It is an unfavorable term. (b) $\Delta\Delta G_{prt}$ is the energy difference due to background interaction of charge with permanent dipole of the peptide backbone of helices or nonionizable polar side chain. (c) $\Delta\Delta G_{brd}$ is the favorable bridge-energy term that represents the interaction among various charge residue pairs constituting the network. So, total electrostatic energy is the sum of $\Delta\Delta G_{dslv}$, $\Delta\Delta G_{prt}$, and $\Delta\Delta G_{brd}$.

$$\Delta\Delta G_{net} = \Delta\Delta G_{dslv} + \Delta\Delta G_{prt} + \Delta\Delta G_{brd}.$$

Based on the APBS method, PBE is solved on a cubic grid box with finer grid spacing (0.5 Å per grid step) using iterative finite-difference method.[25,31] Box size is determined based on the size of protein. To define molecular surface, 1.4 Å radius is used as solvent probe. The internal protein dielectric constant and external solvent dielectric constant are 4 and 76, respectively. The 0.2 M NaCl is used for ionic strength at neutral pH.

3 | RESULTS

3.1 | General characterization of salt-bridge forming residues among three domains of life

Salt-bridge is the specific electrostatic interaction between the side chains of positive and negative charge residues that contribute to the overall stability of proteins. It can be either isolated or networked. Salt-bridge can also be formed either in core or at the

TABLE 1 Absolute and normalize frequency of total and salt-bridge forming acidic and basic residues of archaeal proteins (total residues: 1596)

Residues	f_t^d	f_t^d in%	f_{sb}^d	f_{sb}^d in%	f_{sb}^R in%
Glu	131	8.20	68	4.26	**51.90**
Asp	94	5.88	34	2.13	36.17
Lys	113	7.08	34	2.13	30.08
Arg	95	5.95	59	3.69	**62.10**
His	25	1.56	9	0.56	36.00
Total	458	28.69	204	12.78	44.54

Note: f_t^d: amino acid absolute frequency in total protein chain. f_t^d in%: percent frequency w.r.t. total residues. f_{sb}^d: residues salt-bridge absolute frequency. f_{sb}^d in%: percent salt-bridge frequency w.r.t. total frequency. f_{sb}^R in%: percent salt-bridge frequency w.r.t corresponding salt-bridge forming residues. Total values are bold.

surface of proteins. To check the preferences for the formation of salt-bridges, we have computed normalized frequency of acidic and basic residues from eight unique chains of eight crystal structures from each domain and presented in Tables 1–3. Although the normalized frequency of basic and acidic amino acid residues is almost equal, percentile salt-bridge frequency with respect to corresponding salt-bridge forming residues of basic amino acids are higher than acidic amino acid residues. In archaea, the percentile basic residues (f_t^d in% of Lys, Arg, and His are 113, 95, and 25, respectively) are higher than as compared to other two domains. In particular, Glu and Arg have higher salt-bridge forming tendency in archaea (f_{sb}^d in% of Glu and Arg are 4.26 and 3.69, respectively) (Table 1) whereas Asp, Arg, and His frequency are higher in case of eubacteria and eukarya, respectively.

3.2 | Frequency of salt-bridge against intervening residues and possible pair

To test the location of salt-bridging partners in the sequence of proteins, we have created 10 intermediate group frequency distribution plots and the same is depicted in Figure 1. In this plot, we have distributed first 100 residues of 8 proteins from each domain, with an interval of 10 residues each. It is seen in the plot that the frequency of salt-bridge formation decreases with increase in intervening non-salt-bridge forming amino acid residues. In case of archaea, there are tendencies of salt-bridge formation with higher (≈40-50 and ≈50-60) intervening residues (Figure 1). Based on all six possible combinations viz Arg-Glu, Lys-Glu, Arg-Asp, His-Glu, Lys-Asp, and His-Asp, involved in forming salt-bridge that are present in all three domains have been shown in Figure 2. Here questions may arise about the preference for any chosen specificity in these pairs within different domains of life. Notably, a distinguishable higher specificity is observed for Arg-Glu and Lys-Glu pairs in archaea, Arg-Asp and Lys-Asp pairs in eubacteria and His-Asp pair in eukarya (Figure 2).

TABLE 2 Absolute and normalize frequency of total and salt-bridge forming acidic and basic residues of eubacterial proteins (total residues: 1704)

Residues	f_t^d	f_t^d in%	f_{sb}^d	f_{sb}^d in%	f_{sb}^R in%
Glu	86	5.04	41	2.40	47.67
Asp	105	6.16	59	3.46	**56.19**
Lys	108	6.33	34	1.99	31.48
Arg	94	5.51	50	2.93	**53.19**
His	31	1.81	16	0.93	**51.61**
Total	424	24.88	200	11.73	47.16

Note: f_t^d: amino acid absolute frequency in total protein chain. f_t^d in%: percent frequency w.r.t. total residues. f_{sb}^d: residues salt-bridge absolute frequency. f_{sb}^d in%: percent salt-bridge frequency w.r.t. total frequency. f_{sb}^R in%: percent salt-bridge frequency w.r.t. corresponding salt-bridge forming residues. Total values are bold.

TABLE 3 Absolute and normalize frequency of total and salt-bridge forming acidic and basic residues of eukaryotic proteins (total residues: 1439)

Residues	f_t^d	f_t^d in%	f_{sb}^d	f_{sb}^d in%	f_{sb}^R in%
Glu	95	6.60	32	2.22	33.68
Asp	79	5.48	37	2.57	**46.83**
Lys	84	5.83	25	1.73	29.76
Arg	85	5.90	32	2.22	37.64
His	47	3.26	12	0.83	25.53
Total	390	27.10	138	9.58	35.38

Note: f_t^d: amino acid absolute frequency in total protein chain. f_t^d in%: percent frequency w.r.t. total residues. f_{sb}^d: residues salt-bridge absolute frequency. f_{sb}^d in%: percent salt-bridge frequency w.r.t. total frequency. f_{sb}^R in%: percent salt-bridge frequency w.r.t. corresponding salt-bridge forming residues. Total values are bold.
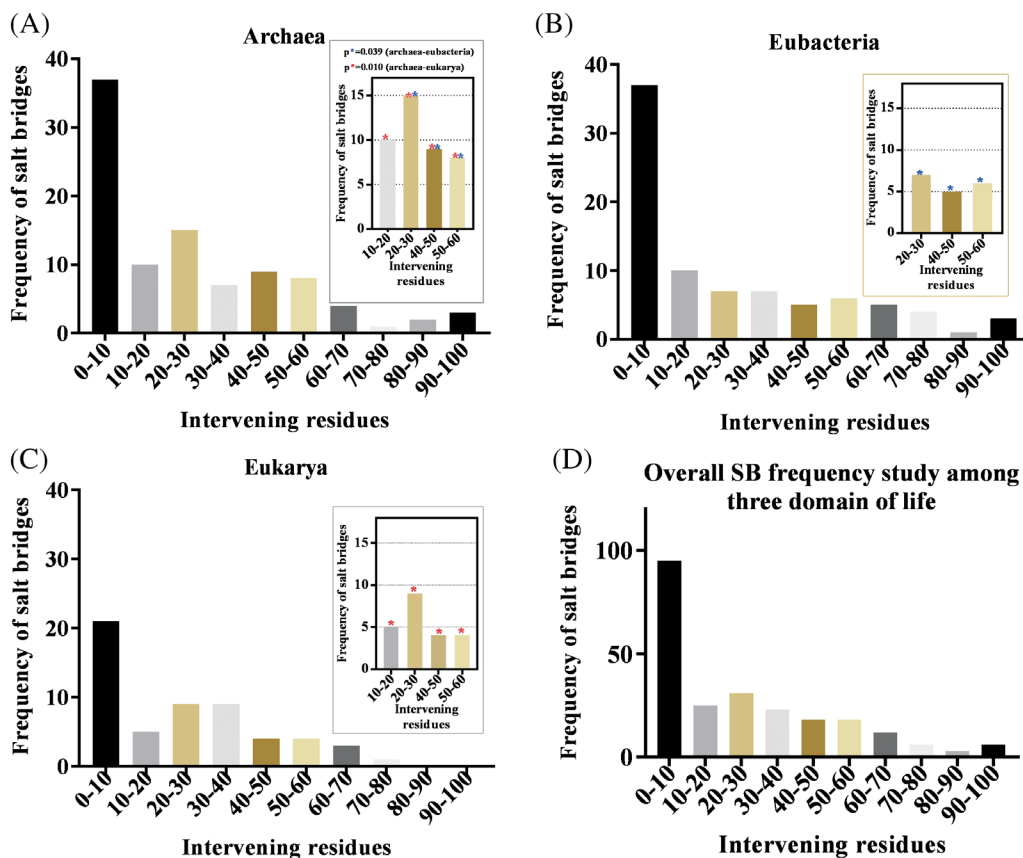
FIGURE 1 Histogram showing the frequency of salt-bridge against intervening residues of 100 amino acids from N-terminal. Tendency of short range salt-bridge formation is higher in all domain but in archaea, the salt-bridge formation also be seen at long range (40-50 and 50-60)
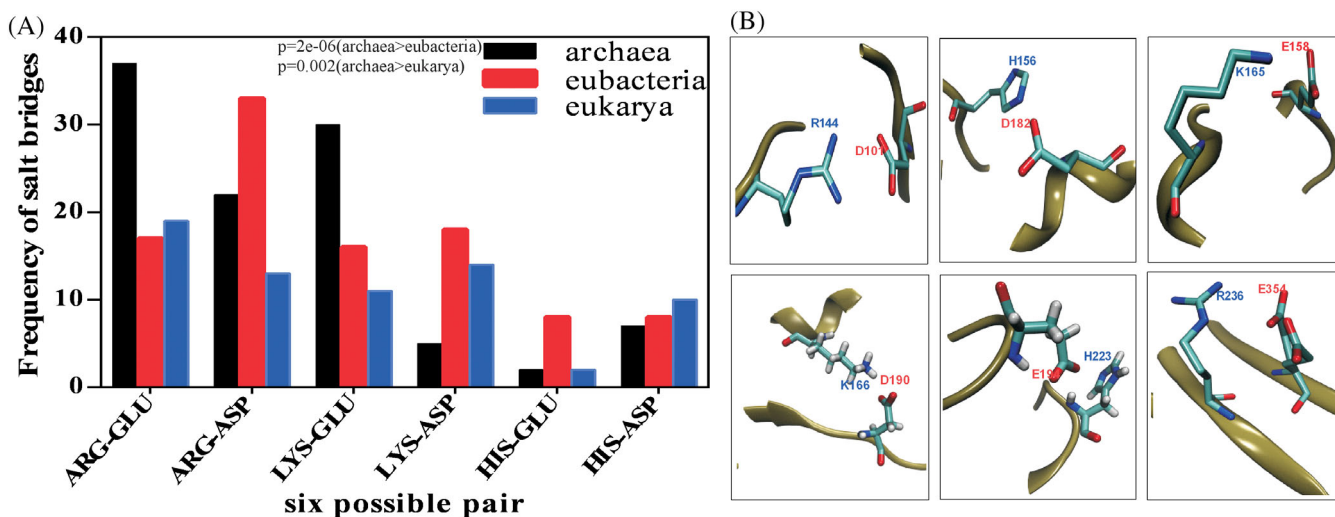
FIGURE 2 A, Histogram showing the frequency of salt-bridges formed by each of six possible pairs among three domains of life. B, Schematic representation of six possible ion-pairs [Color figure can be viewed at wileyonlinelibrary.com]

3.3 | Buried and exposed salt-bridges among three domains

The location of a salt-bridge within protein structure depends on the average accessibility of candidate salt-bridge partner using a probe radius of 1.4 Å.[32] If the average accessible surface area (ASA) is less than or equal to 20% it indicates salt-bridge is present in protein core otherwise exposed to higher ASA cutoff values. Here, we have divided the salt-bridge into buried and exposed categories based on computational method (Table 4). By computational study, we can see

that in eukarya 27% of salt-bridge are located in the core, but in archaea and eubacteria normally it stands at 26% and 23%, respectively (Table 4).

However, unlike earlier study, in eukarya and eubacteria, not all buried salt-bridge in our dataset are stable. In archaea, all buried-networked salt-bridges are stable as earlier,[6] whereas only 36% and 72% of these salt-bridges are stable in eubacteria and eukarya, respectively (Supporting Information Table S4). In Figure 3, we see that the desolvation energy of salt-bridge in protein core is much higher than other types of salt-bridge. However, the sum of bridge term and background energy term may compensate the effect of desolvation energy term and thus, giving the overall protein stability. In general, bridge energy term is always favorable ($\Delta\Delta G_{brd} < 0$) but background or microenvironment could be either favorable or unfavorable. Our study shows that background energy play an important role in proteins' core. The role of microenvironment in proteins' core within archaea is much more prominent than that in the eubacteria and eukarya (Supporting Information Table S2). On the other hand, microenvironment in proteins' exposed area plays more or less similar role in the domains of archaea eubacteria and eukarya (Supporting Information Table S3).

TABLE 4 Percentile calculation of various salt-bridges among three domains

	Archaea	Eubacteria	Eukarya
Stable	89 (88%)	82 (78%)	50 (67%)
Unstable	12 (12%)	23 (22%)	24 (33%)
Network	57 (56%)	35 (33%)	31 (41%)
Isolated	44 (44%)	70 (67%)	43 (59%)
Buried/core	26 (26%)	24 (23%)	20 (27%)
Exposed	75 (74%)	81 (77%)	54 (73%)

3.4 | Comparison of the electrostatic contribution of salt-bridge

Like other noncovalent interaction, salt-bridge plays a significant role in the stability of protein structure and function. In addition to amino acid-amino acid interactions that provide stability to proteins' core and side chains within protein structure, ionic bonds within these subunits can help in stabilizing proteins, which may lead to a higher melting temperature.[21] Here, we used APBS method[23] along with ADSBET2[30] for calculation of all three associated energy term ($\Delta\Delta G_{dslv}$: desolvation; $\Delta\Delta G_{prt}$: background, and $\Delta\Delta G_{brd}$: bridge) for finding net salt-bridge energy ($\Delta\Delta G_{net}$: sum of above three-component terms). Both, $\Delta\Delta G_{dslv}$ and $\Delta\Delta G_{prt}$ are indirect and pH-independent energy terms,[30] where $\Delta\Delta G_{dslv}$ is unfavorable term due to desolvation of charges during folding and $\Delta\Delta G_{prt}$ could be favorable or unfavorable that largely depends on the composition of microenvironment. $\Delta\Delta G_{brd}$ is always favorable due to interaction of opposite charges in folded state conformation of protein structure. With a view that statistical analysis might affect, due to heterogeneity of the database,[31] we chose homogeneous datasets from three domains. Our observations suggest that archaea utilize more of specific electrostatic interactions than that of eubacteria and eukarya. Figure 3 shows that although the total ($\Delta\Delta G_{net}$) average network energy is much higher than the isolated types in all three domains, total ($\Delta\Delta G_{net}$) salt-bridge energy is much more (2-3 times) in archaea than eubacteria and eukarya. Overall energy for buried but stable type salt-bridges also shows similar tendency. Figure 3 also tells us the overall stability of salt-bridge is much higher in archaea ($\Delta\Delta G_{net} = -5.06 \pm 3.8$) than eubacteria ($\Delta\Delta G_{net} = -3.7 \pm 2.9$) and eukarya ($\Delta\Delta G_{net} = -3.54 \pm 3.1$). More detailed distribution of salt-bridge energy from all three domains is shown in Figure 4 and Supporting Information Figures S2 and S3. Although, symmetric
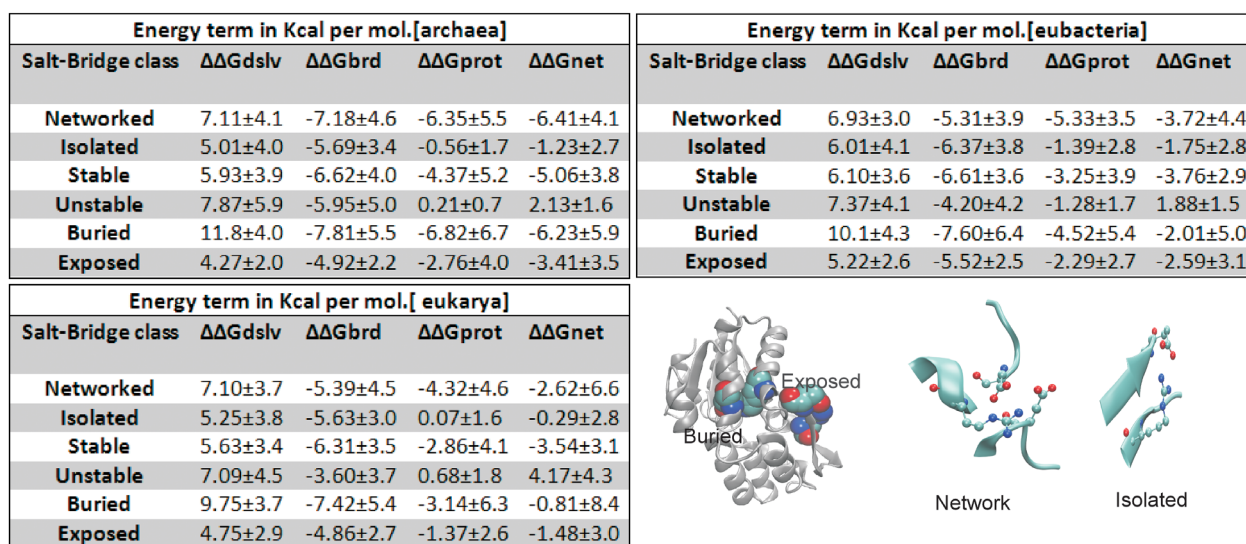
Energy term in Kcal per mol.[archaea]				
Salt-Bridge class	$\Delta\Delta G_{dslv}$	$\Delta\Delta G_{brd}$	$\Delta\Delta G_{prot}$	$\Delta\Delta G_{net}$
Networked	7.11±4.1	-7.18±4.6	-6.35±5.5	-6.41±4.1
Isolated	5.01±4.0	-5.69±3.4	-0.56±1.7	-1.23±2.7
Stable	5.93±3.9	-6.62±4.0	-4.37±5.2	-5.06±3.8
Unstable	7.87±5.9	-5.95±5.0	0.21±0.7	2.13±1.6
Buried	11.8±4.0	-7.81±5.5	-6.82±6.7	-6.23±5.9
Exposed	4.27±2.0	-4.92±2.2	-2.76±4.0	-3.41±3.5

Energy term in Kcal per mol.[eubacteria]				
Salt-Bridge class	$\Delta\Delta G_{dslv}$	$\Delta\Delta G_{brd}$	$\Delta\Delta G_{prot}$	$\Delta\Delta G_{net}$
Networked	6.93±3.0	-5.31±3.9	-5.33±3.5	-3.72±4.4
Isolated	6.01±4.1	-6.37±3.8	-1.39±2.8	-1.75±2.8
Stable	6.10±3.6	-6.61±3.6	-3.25±3.9	-3.76±2.9
Unstable	7.37±4.1	-4.20±4.2	-1.28±1.7	1.88±1.5
Buried	10.1±4.3	-7.60±6.4	-4.52±5.4	-2.01±5.0
Exposed	5.22±2.6	-5.52±2.5	-2.29±2.7	-2.59±3.1

Energy term in Kcal per mol.[eukarya]				
Salt-Bridge class	$\Delta\Delta G_{dslv}$	$\Delta\Delta G_{brd}$	$\Delta\Delta G_{prot}$	$\Delta\Delta G_{net}$
Networked	7.10±3.7	-5.39±4.5	-4.32±4.6	-2.62±6.6
Isolated	5.25±3.8	-5.63±3.0	0.07±1.6	-0.29±2.8
Stable	5.63±3.4	-6.31±3.5	-2.86±4.1	-3.54±3.1
Unstable	7.09±4.5	-3.60±3.7	0.68±1.8	4.17±4.3
Buried	9.75±3.7	-7.42±5.4	-3.14±6.3	-0.81±8.4
Exposed	4.75±2.9	-4.86±2.7	-1.37±2.6	-1.48±3.0

FIGURE 3 Comparison of average energy term in different salt-bridge categories among three domains of life. Stability of network forming salt-bridge in archaea is comparatively higher than the other domain. Whereas salt-bridge stability in bacterial protein core is relatively lower from archaea and higher than eukaryotic protein [Color figure can be viewed at wileyonlinelibrary.com]
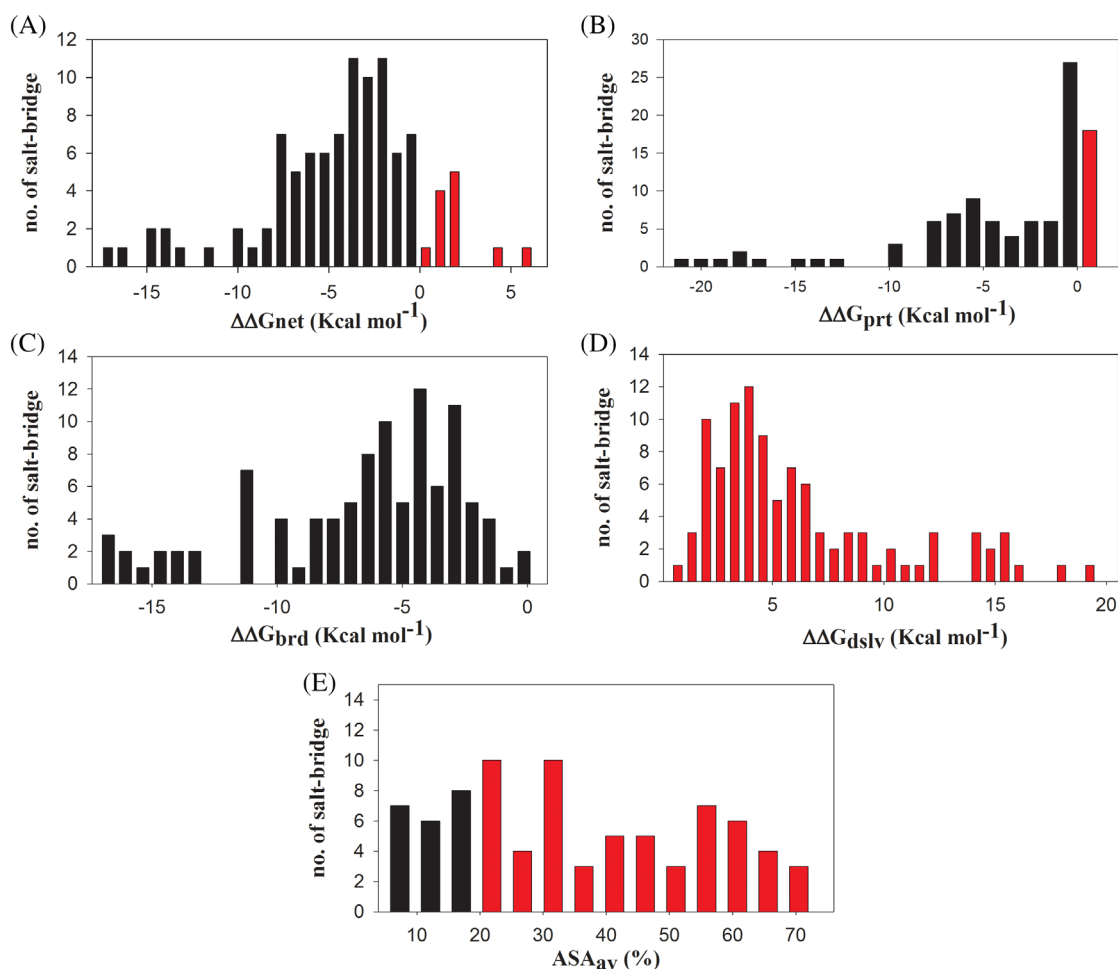
FIGURE 4 Histogram showing distribution of salt-bridge from archaeal protein as a function of A, $\Delta\Delta G_{net}$ kcal mol^{-1}; B, $\Delta\Delta G_{prt}$ kcal mol^{-1}; C, $\Delta\Delta G_{brd}$ kcal mol^{-1}; D, $\Delta\Delta G_{dslv}$ kcal mol^{-1}; and E, ASA$_{av}$ (%) [Color figure can be viewed at wileyonlinelibrary.com]

distribution of $\Delta\Delta G_{net}$ term containing both stable and unstable salt-bridges, majority of which are falling in stabilizing zone ($\Delta\Delta G_{net} < 0$) (Figure 4 and Supporting Information Figures S2 and S3). How much of the population of salt-bridges contribute to the stability of these three domains is yet to be unclear. Remarkably, it is seen that almost 90% of salt-bridges contribute to the stability in archaea whereas 78% in eubacteria and 67% give stability in eukarya, respectively. The profile also tell us that distribution of bridging energy term (Figure 4 and Supporting Information Figures S2C and S3C) and desolvation energy term (Figure 4 and Supporting Information Figures S2D and S3D) occupying stabilizing and destabilizing zone, respectively, while background terms (Figure 4 and Supporting Information Figures S2B and S3B) possesses both stabilizing and destabilizing population of salt-bridge.

3.5 | The relationship among different energy terms with average ASA

In Figure 5 and Supporting Information Figures S4 and S5, the correlation plot between ASA with energy term has been shown

for three domains. During the time of protein folding, salt-bridge's desolvation energy ($\Delta\Delta G_{dslv}$) is linear and has negative relationship with ASA. The relationship could be express in the equation of:

$$\text{Log}\,(\Delta\Delta G_{dslv}) = -0.01\text{ASA}_{av} + 1.3.$$

Correlation coefficient ($\Delta\Delta G_{dslv}$ vs ASA$_{av}$) of the fitted line is 0.46, 0.34 and 0.38, 0.62 in archaea, eubacteria, and eukarya, respectively. The linear correlation between $\Delta\Delta G_{dslv}$ and ASA$_{av}$ with negative slope indicates that the transfer of salt-bridge into the core of protein involves greater desolvation penalty. Figure 5 and Supporting Information Figures S4 and S5 show that $\Delta\Delta G_{dslv}$ is negatively correlated whereas $\Delta\Delta G_{prt}$ and $\Delta\Delta G_{brd}$ are positively correlated with ASA$_{av}$. Distribution of salt-bridge forming residues is very high below the average value of ASA$_{av}$ ($\leq$20%) in archaea and eukarya than eubacteria. But average total stability of buried salt-bridge is very high in archaea than eubacteria whereas salt-bridge stability is very low in the case of eukarya as compared to archaea and eubacteria (Figure 3), despite being more than one salt-bridge present in protein core.
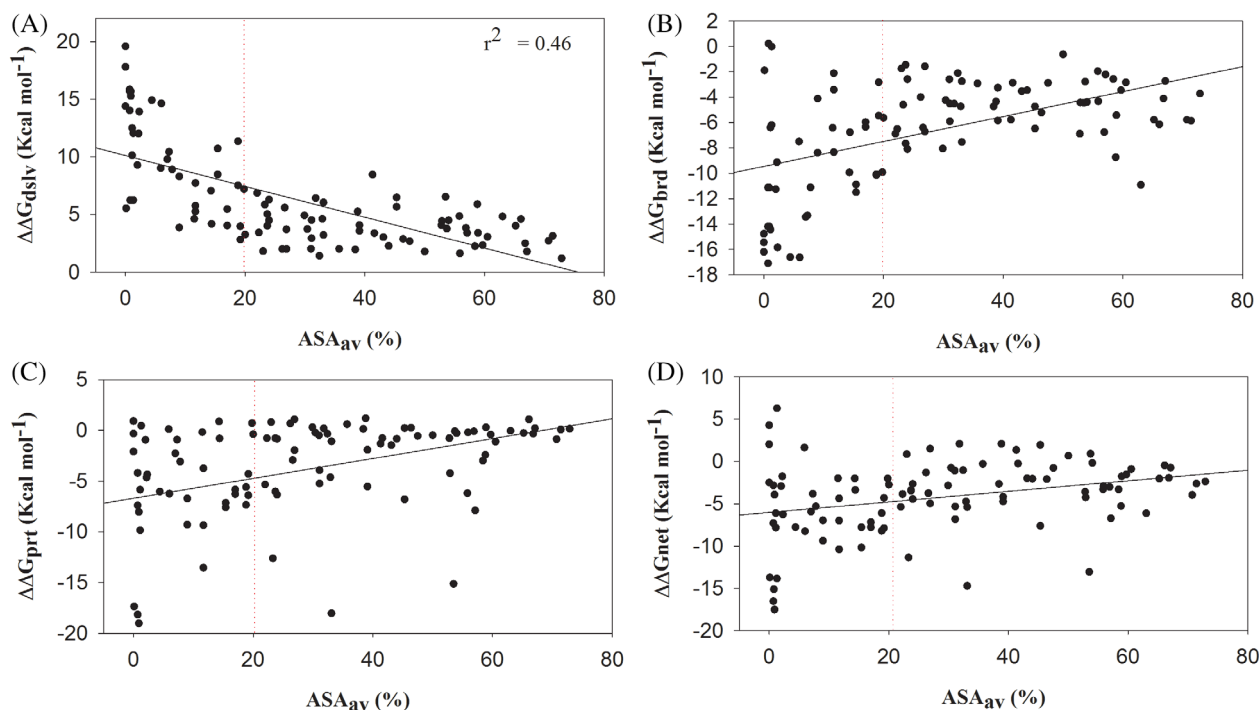
FIGURE 5   Correlation of ASA_{av} (%) (x-axis) with A, $\Delta\Delta G_{dslv}$ kcal mol^{-1}; B, $\Delta\Delta G_{brd}$ kcal mol^{-1}; C, $\Delta\Delta G_{prt}$ kcal mol^{-1}; and D, $\Delta\Delta G_{net}$ kcal mol^{-1} for 101 salt-bridges from archaeal hydrolase [Color figure can be viewed at wileyonlinelibrary.com]

3.6 | Network and isolated salt-bridge

Salt-bridge confers stability to the tertiary structure of protein.[25] However, the stability of the salt-bridge depends on its location, geometry, and the side-chain interaction within protein structure.[33] Isolated salt-bridge provides marginal stabilization and network salt-bridge causes cooperative stabilization. It is seen that 56% salt-bridges forming network salt-bridge and rest 44% are isolated type in archaea. While, 33% and 41% forming network salt-bridge in eubacteria and eukarya, respectively and rest are forming isolated type salt-bridge (Table 4).

4 | DISCUSSION

While the PBE and its solver method[30] is reliable for the computation of direct and indirect component energy terms and hence the net energetics of ion-pairs,[6,14] it is an approximate method. Computation either at low pH (<1.0) or without mobile ion concentration may produce erroneous results due to high charge density situations within protein molecules.[18] The latter effect would be more prominent in the case of halophilic protein like 4E19.pdb, which has been a representative in our archaeal dataset. Although we consider linear PBE in APBS using finite size of ions approximation as earlier,[5,14,23] all of our computations were performed at neutral pH (pH 7.0) in the presence of univalent (Na$^+$ and Cl$^-$) mobile ions (each at 0.2 M) to lessen the above-mentioned error-prone situations.[18]

Despite the similarity in topology, protein folding may be different for different domains. Apart from intrinsic codes, protein folding also depends on the surrounding environment. Notably in these domains of life, both intrinsic and extrinsic factors show wide variations in our homologous sets of protein. Like other noncovalent interactions, salt-bridge plays significant role in the stabilization of native protein. Secondary structure, which determines protein topology, is more conserved in evolution. Our observation in this context suggests that local salt-bridges of proteins tend to occupy much higher frequency for all domains. However and interestingly, higher frequency of long-range salt-bridges in archaea is indicative of a new design.

It is also interesting to note that the stability and propensity of salt-bridges are highly variable among domains of life. While 88% ($\Delta\Delta Gnet = -5.06 \pm 3.8$) of salt-bridges are highly stable in archaea, 78% ($\Delta\Delta Gnet = -3.7 \pm 2.9$) and 67% ($\Delta\Delta Gnet = -3.54 \pm 3.1$) of salt-bridges in eubacteria and eukarya have been stabilizing type. The higher stability in the former seems to be due to preference for specific pair types of salt-bridge. The rest salt-bridge is destabilizing type. The instability of this salt-bridge population is coming due to weaker contribution of background and bridge energy. In different salt-bridge categories, while least variation is notable in the case of bridge energy, for background energy such observation has been dramatic. Notably, background energy is more sensitive to microenvironment of protein among all component terms and thus the best representative of intrinsic codes of protein. In this context, excessive charges and dipole constitute local environment around individual salt-bridge partners, which determine the magnitude of their contribution to net stability.

Unlike bridge and background energy terms, desolvation is negatively correlated with the average ASA. It is not at all surprising as desolvation process is thermodynamically uphill toward the core of protein.[5,14] Further, within buried boundary (ie, ASA ≤ 20%), local environmental effects seem to supersede the linear gradient of dielectric constant and charge screening on protein surface (dielectric constant = 76) and interior (dielectric constant = 4).[15] Strong background energy over bridge energy in archaea, as well as large spread of individual salt-bridge energy of these terms against ASA$_{av}$ (Figure 5), might indicate the involvement of additional nonlinear factor.[5]

However, in our observation, not all buried salt-bridge are stable in protein core. While majority of (100%) buried salt-bridges are stable in archaea. Where, eubacteria and eukarya show comparatively lower population of stable buried salt-bridge (ie, 36 and 72, respectively). Now the question is how buried salt-bridge overcomes the desolvation penalty and gain net stability of salt-bridge. Due to large desolvation penalty, buried salt-bridge suffers very much and entropic cost is minimum whereas in local salt-bridge interaction gives favorable conditions on enthalpy terms. In our data set out of 13 buried salt-bridges, 13 (100) formed network salt-bridge in archaea, whereas, 4 (36%) of 11 and 8 (72%) of 11 buried salt-bridges in case of eubacteria and eukarya, respectively. In this context, it appears that the recruitment of excess network salt-bridges is related to overcoming the desolvation penalty and thus, to gain average net stability with respect to buried salt-bridge in archaea (Supporting Information Table S4). Overall based on the energy perspective, we can say that network salt-bridge stability is more than that of isolated salt-bridge (Figure 3). In archaea, network salt-bridge contributes higher stability and promotes cooperative interaction among charge residues within protein structure for stabilization at adverse environment.

Taken together, we have shown that the contribution of salt-bridges is more prominent in archaea than eubacteria and eukarya. Overall, our comparative analyses of salt-bridges and their energy terms give us a new insight into the constancy of protein tertiary structure among domains of life.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Saumen Datta 🄳 https://orcid.org/0000-0002-3638-9383

REFERENCES

1. Tadeo X, Lopez-Mendez B, Trigueros T, Lain A, Castano D, Millet O. Structural basis for the aminoacid composition of proteins from halophilic archaea. *PLoS Biol*. 2009;7(12):e1000257.

2. Frolow F, Harel M, Sussman JL, Mevarech M, Shoham M. Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. *Nat Struct Biol*. 1996;3(5): 452-458.

3. Dill KA. Dominant forces in protein folding. *Biochemistry*. 1990;29 (31):7133-7155.

4. Pace CN. Conformational stability of globular proteins. *Trends Biochem Sci*. 1990;15(1):14-17.

5. Roy C, Datta S. ASBAAC: automated salt-bridge and aromatic-aromatic calculator. *Bioinformation*. 2018;14(4):164-166.

6. Nayek A, Sen Gupta PS, Banerjee S, Mondal B, Bandyopadhyay AK. Salt-bridge energetics in halophilic proteins. *PLoS One*. 2014;9(4): e93862. https://doi.org/10.1371/journal.pone.0093862.

7. Dym O, Mevarech M, Sussman JL. Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. *Science*. 1995;267(5202):1344-1346.

8. Bandyopadhyay AK, Krishnamoorthy G, Padhy LC, Sonawat HM. Kinetics of salt-dependent unfolding of [2Fe-2S] ferredoxin of *Halobacterium salinarum*. *Extremophiles*. 2007;11(4):615-625.

9. Kumar S, Tsai CJ, Nussinov R. Factors enhancing protein thermostability. *Protein Eng*. 2000;13(3):179-191.

10. Sun DP, Sauer U, Nicholson H, Matthews BW. Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry*. 1991;30(29):7142-7153.

11. Hendsch ZS, Tidor B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci*. 1994;3(2):211-226.

12. Horovitz A, Fersht AR. Co-operative interactions during protein folding. *J Mol Biol*. 1992;224(3):733-740.

13. Marqusee S, Sauer RT. Contributions of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in lambda repressor. *Protein Sci*. 1994;3(12):2217-2225.

14. Lounnas V, Wade RC. Exceptionally stable salt bridges in cytochrome P450cam have functional roles. *Biochemistry*. 1997;36(18):5402-5417.

15. Kumar S, Nussinov R. Salt bridge stability in monomeric proteins. *J Mol Biol*. 1999;293(5):1241-1255.

16. Dao-pin S, Anderson DE, Baase WA, Dahlquist FW, Matthews BW. Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. *Biochemistry*. 1991;30(49):11521-11529.

17. Waldburger CD, Schildbach JF, Sauer RT. Are buried salt bridges important for protein stability and conformational specificity? *Nat Struct Biol*. 1995;2(2):122-128.

18. Barril X, Aleman C, Orozco M, Luque FJ. Salt bridge interactions: stability of the ionic and neutral complexes in the gas phase, in solution, and in proteins. *Proteins*. 1998;32(1):67-79.

19. Elcock AH, McCammon JA. Electrostatic contributions to the stability of halophilic proteins. *J Mol Biol*. 1998;280(4):731-748.

20. Kumar S, Ma B, Tsai CJ, Nussinov R. Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. *Proteins*. 2000;38(4):368-383.

21. Kumar S, Nussinov R. How do thermophilic proteins deal with heat? *Cell Mol Life Sci*. 2001;58(9):1216-1233.

22. Li L, Li C, Sarkar S, et al. DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys*. 2012;5(9):2046-1682.

23. Guest WC, Cashman NR, Plotkin SS. Electrostatics in the stability and misfolding of the prion protein: salt bridges, self energy, and solvation. *Biochem Cell Biol*. 2010;88(2):371-381.

24. Hendsch ZS, Sindelar CV, Tidor B. Parameter dependence in continuum electrostatic calculations: a study using protein salt bridges. *J Phys Chem B*. 1998;102(22):4404-4410.

25. Barlow DJ, Thornton JM. Ion-pairs in proteins. *J Mol Biol*. 1983;168 (4):867-885.

26. Gupta PS, Nayek A, Banerjee S, et al. SBION2: analyses of salt bridges from multiple structure files, version 2. *Bioinformation*. 2015;11(1):39-42.

27. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26(16):1781-1802.

28. Pedretti A, Villa L, Vistoli G. VEGA—an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. *J Comput Aided Mol Des*. 2004;18(3):167-173.

29. Fleming PJ, Gong H, Rose GD. Secondary structure determines protein topology. *Protein Sci*. 2006;15(8):1829-1834.

30. Nayek A, Gupta PSS, Banerjee S, et al. ADSBET2: automated determination of salt-bridge energy-terms version 2. *Bioinformation*. 2015;11(8):413-415.

31. Madigan D, Ryan PB, Schuemie M, et al. Evaluating the impact of database heterogeneity on observational study results. *Am J Epidemiol*. 2013;178(4):645-651.

32. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*. 1971;55(3):379-400.

33. Albeck S, Unger R, Schreiber G. Evaluation of direct and cooperative contributions towards the strength of buried hydrogen bonds and salt bridges. *J Mol Biol*. 2000;298(3):503-520.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.