Characterization of a putative antimicrobial peptide transporter in Gram-negative bacteria via *in vitro* and computational strategies

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Pratik Dasgupta
(176106022)

Under supervision of

Prof. Shankar Prasad Kanaujia

Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India

April 2024

Dedicated to Dadu, Thamma, Monai, Dida, Ma, Baba, Riya, and my amazing family members

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

DECLARATION

I do hereby declare that the research findings of this thesis entitled "**Characterization of a putative antimicrobial peptide transporter in Gram-negative bacteria via *in vitro* and computational strategies**" are the result of research work carried out by me at the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, India, under the supervision of **Prof. Shankar Prasad Kanaujia**.

I also declare that the contents of this thesis have not been the basis for award of any other degree, diploma, fellowship or any other similar title of any University or Institution.

In keeping with the general practice of reporting scientific observations, due acknowledgments have been made, wherever the research findings of other researchers have been cited in this thesis.

Date: April 04, 2024

Place: IIT Guwahati

Pratik Dasgupta

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

CERTIFICATE

This is to certify that the work described in the thesis entitled "**Characterization of a putative antimicrobial peptide transporter in Gram-negative bacteria via *in vitro* and computational strategies**" is an authentic record of the results obtained from the research work carried out by **Pratik Dasgupta** in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Assam, India, under my supervision and this work in part or as a whole has not been submitted elsewhere for the award of any other degree.

Shankar Prasad Kanaujia

Date: April 04, 2024
Place: IIT Guwahati

Prof. Shankar Prasad Kanaujia
(Thesis Supervisor)

As my steps inch closer to the finishing line, my heart fills with an immense sense of gratitude. For a journey like this can never be fulfilled alone.

First and foremost I would like to thank my Supervisor Prof. Shankar Prasad Kanaujia, for introducing me to the amazing world of protein structural biology. Despite my failures he has been very considerate and supportive through-out my journey. He was always present in my hour of need and has pushed me through this arduous journey. I shall forever be indebted by his immense contributions in whatever little I have accomplished.

I would like to thank my doctoral committee members Prof. Biman B. Mandal, Prof. Bhubaneshwar Mandal and Dr. Rajkumar P. Thummer, for their kind words of support in hard times and insightful suggestions which helped in enhancing the quality of this thesis.

I would also like to cease the opportunity to acknowledge all the support provided by the past and present heads of the department specifically Prof. Kannan Pakshirajan, Prof. Latha Rangan, and Prof. Rakhi Chaturvedi. Their tireless efforts in providing the necessary departmental facilities have been instrumental in facilitating my research endeavors.

I am filled with gratitude towards the Department of Biosciences and Bioengineering at IIT Guwahati for providing me with invaluable infrastructural support and fostering an enabling academic environment. I am also grateful to the Central Instruments Facility (CIF), IIT Guwahati, for granting me access to sophisticated equipment, including the X-ray diffractometer (XRD), isothermal titration calorimetry (ITC) and matrix-assisted laser desorption ionization–time of flight (MALDI-ToF), which were essential to my research work.

I am immensely grateful to the Ministry of Education, Government of India, for their generous financial support during my Ph.D. Program.

I express my heartfelt gratitude to the past and present members of the Structural and Computational Biology Laboratory (SCBL) for their support and contribution towards creating a conducive work environment. I am particularly grateful to Dr. Prerana Gogoi, Dr. Monika Chandravanshi, Dr. Suraj Kumar Mandal, Dr. Angshu Dutta, and Dr. Sayan Saha my senior fellows, for their invaluable intellectual contributions to my work. Their consistent support and encouragement have been instrumental in the successful completion of my work, and I remain deeply indebted to them for sharing their extensive experience. Their contributions has been pivotal in advancing my research, and I feel honored to have had the privilege of working with them. I would also like to express my sincere gratitude to my fellow lab members, Kalyan Ghosh, Ritu Tripathi, Arpana Gupta, Shimpy Nigam, Aditi Dey, Monisha Dhar, Ankita Mallick, Ankit, and Aarthi Murali for their invaluable cooperation and timely assistance, which have been essential to the progress of my work. Their contributions have been invaluable to the success of my work, and I am truly grateful for their support. I would like to specially mention Kavya Vinil for her help in the data collection and preliminary analysis of the *sap* operon entries. Lastly, I would like to acknowledge and extend my gratitude to Dr. Ankita Gupta, Prerana, Reshma, Smriti, Shreya, Sakshi, Tejasvi, Smith, Ankita, Arpita, Arnab, Gautam, Aman, Sanjay, Harsh, Jayesh, and Dhruv for all their support.

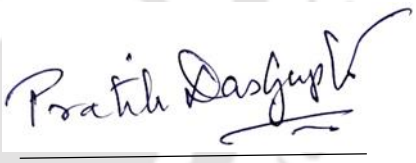
I am grateful for the support and encouragement provided by my esteemed friends and peers, who have stood by me through my professional and personal voyage. I would like to extend my sincerest appreciation to Barlina Konwar, Muktashree Saha, Arupam Patra, Arman Mohanty, Alok Senapati, Alok Pandey, Ashutosh Bandhopadhyay, Prabir Das, Ananya Chouhan, Dr. Saddam Hussain, Vineet Anand, Dr. Vinay Bachu, Saswat Hota, Surbhi Kumari, Shibam Dey, and Prattusha Khan for their unwavering support during challenging times. I would also like to convey my heartfelt thanks to all my friends outside the campus. My heart goes out to Uday Pandey, Subhashree Barik, Greeshma Sateeshan, Aishwarya Chamling, Rakesh Banerjee, Ujjaini Baruah, Abilasha, Banu Priya, Jasmine, and Rajat Verma for being ever so inspiring.

Acknowledgements

I am deeply indebted for the selfless nature of love, support and strength conveyed by my family throughout the journey. I shall remain forever grateful to Ma (Mrs Manidipa Das Gupta), Baba (Mr. Prasad Das Gupta), and my sweet little sister (Ms. Meghna Das Gupta). I would also like to convey my love and immense gratefulness to Dadu, Thamma, Monai, Dida, Moshai, Ama pishi, Gathu, Bonena, Adal, Tomtom, Giggles, Buku pishi, Gublu moshai, Mesho, Mashi, Choto mesho, Hamunai, Junie di, Jamai babu, Papa da, Atigae, Tito da, Annindita boudi, Bagai da, Rohana di, Tiku da, Neil, Tintoo, Mohor, Jeena, Goju, Pupi, Arjob, Meghu, Reyaan, Roshi, and Amaya for their indelible support.

I would like to take this opportunity to express my sincere gratitude to everyone who contributed to this project. If I have inadvertently omitted anyone from my thank you note, I sincerely apologize and assure you that your efforts are equally valued.

In conclusion, I express my sincere gratitude to the Almighty God for his blessing.



Pratik Dasgupta

Table of Contents

Abstract

Abstract

Humans, on a daily basis, are challenged by a plethora of pathogens scavenging for a suitable environment to create a niche for them. To hinder the pathogens from forming a favorable niche, the human host puts forth a robust wall in the form of the human immune system. The human immune system is a multilayered conglomeration of multiple mechanisms deployed to suppress invading pathogens. The human immune system comprises three definitive layers, namely, the anatomical and physiological barriers, the innate immune system, and the adaptive immune system. The pathogens first face anatomical and physiological barriers, such as tears, saliva, the epidermal layer (skin), low pH in the stomach, etc. As the pathogens escape the first layer, they are attacked by the components of the innate immune system. The innate immune system is further subdivided into cellular and humoral innate immune systems. The cellular counterpart of the innate immune system comprises various agranulocyte cells, such as macrophages, and granulocytes, such as basophils. These cells are suggested to phagocytize the invading pathogens and further process and represent their antigens to activate downstream immunogenic reactions. The humoral innate immune system comprises molecules such as antimicrobial peptides (AMPs), c-reactive proteins, etc., which are secreted by the aforementioned cells and help in clearing up the invading pathogens. Both layers of the human immune system are a type of non-targeted mechanism that withhold the invading pathogen, buying some time for the third targeted layer of the human immune system, known as the adaptive immune system. This penultimate layer of the human immune system is further subdivided into the cellular and humoral immune systems. The cellular adaptive immune system consists of cells such as B- and T-cells, while the humoral counterpart houses the antibodies produced by the B-cells. The B-cells are suggested to recognize the presented antigens and further produce antibodies that interact with the antigens on the pathogen, as well as target the complement proteins for their directed lysis. The T-cell, on the contrary, does not produce antibodies; instead, it manipulates the pathogenic cells to undergo apoptotic cell death. The adaptive immune response takes a few days to be initiated, and hence, until then, the hosts are at the mercy of their first two layers for preliminary clearance of the invading pathogens. A key member

that marks their presence in the first two layers of the immune system is the AMPs.

The AMPs are short peptides that are reported to target the bacterial membrane and lyse the cells. The interaction between the AMPs and the bacterial membrane is primarily charge-driven, wherein the cationic AMPs interact with the negatively charged bacterial membranes. The AMPs either disrupt the bacterial membrane homeostasis or get internalized and target crucial internal mechanisms, leading to the death of the bacterium. The AMPs are reported to possess multiple secondary structural elements, which dictate their diversity and also vary in their residue compositions and mode of action. Owing to these differences, AMPs target a wide range of pathogens with varying mechanisms, making them one of the most lethal attributes of the human immune system.

To neutralize the effects of the AMPs, multiple Gram-negative pathogens have strategized various mechanisms, such as employing outer membrane proteases to cleave the AMPs, modifying the charge of the outer membrane, shielding the bacterial outer membrane with an additional layer, and downregulating the production of AMPs. When all the aforementioned strategies fail to neutralize the AMPs, the bacterial systems employ certain transporters to export or import the AMPs out of the bacterial systems. Amongst the transporters, the ABC transporter superfamily is quite prevalent in garnering resistance against the action of the AMPs, and one such eminent member of the ABC superfamily is the sensitivity to the antimicrobial peptide (Sap) transport system.

The Sap transport system is an ABC importer possessing a substrate-binding protein (SBP) SapA, two transmembrane domains (TMDs) SapBC, and two nucleotide-binding domains (NBDs) SapDF. The components of the Sap transporter are suggested to uptake the AMPs from the periplasmic space and import them to the bacterial cytoplasm, where they are proteolytically degraded via the bacterial proteases. The Sap transport system was first identified in protamine-sensitive mutants of *Salmonella typhimurium*. Following this, the *Haemophilus influenzae* Sap (*Hi*Sap) transport system was also reported and widely studied. Studies of the *Hi*Sap transport system established the interaction

of *Hi*SapA with human LL-37 and β-defensins. Further, with the aid of elegant molecular biological strategies, the operonic arrangement of the *Hi*Sap transport component was validated. The action of the bacterial proteases post-uptake of the AMPs was also validated using the *Hi*Sap transport system. Apart from importing AMPs, the Sap transport system has also been suggested to be involved in heme and oligopeptide uptake, hence highlighting the multifunctional attribute of the Sap transport system.

Despite the wide number of studies pertaining to the Sap transport system, certain reports divert the functional purpose of the Sap transport system. The *Escherichia coli* Sap (*Ec*Sap) transport system was suggested to function as a putrescine exporter, wherein *Ec*SapBCDF functions in concert, while *Ec*SapA was banished from the system and suggested to be an orphan SBP. Hence, this deviation observed in the *Ec*Sap transport system embroiled its functional prospects into an enigmatic status. This piqued our interest in exploring and enlightening the functional prospects of the *Ec*Sap transport system. In this study, with the aid of computational and biophysical strategies, the components of the *Ec*Sap transporter, primarily *Ec*SapA, were studied.

To understand the orphan SBP-like status of *Ec*SapA, we initiated our analysis by employing data mining strategies to collect all available Sap components. A total of 421 Sap transport systems belonging to unique bacterial species were enlisted. Among the 421 entries, 352 (consensus) were observed to follow the reported genetic trend, and the remaining 69 (non-consensus) were observed to demonstrate variations. Further, the genetic neighborhood analysis of the consensus operons unraveled the conservation of the Psp system and enoyl reductase in the upstream and downstream of the operon, respectively. The *Ec*sap operon was observed to be in accordance with all the trends observed in the consensus operons, and hence, the results suggest *Ec*SapA to be a part of the *Ec*Sap transport system. Further, with the aid of the non-consensus entries and phylogenetic strategies, our results demonstrate a probable evolutionary trend of *sap* operon formation.

Having established that *Ec*SapA is part of the *Ec*Sap transport system, the functional prospects of *Ec*SapA were delved into. Sequence and structure-based homology suggested the dipeptide-binding possibility of *Ec*SapA. Hence, a virtual screening experiment was set up with *Ec*SapA, wherein 400 dipeptides were docked to identify the probable ligands of *Ec*SapA. The results suggested the dipeptide-binding propensity of *Ec*SapA. Interestingly, the varying binding poses of the dipeptide indicated a wider binding site of *Ec*SapA, which was confirmed with the aid of volume-based analysis. Further, the binding-site charge and volume-based analysis suggested the AMP-binding possibility of *Ec*SapA. The AMP-binding propensity of *Ec*SapA was further established through protein-protein docking experiments, which revealed protegrin-1 and protamine-binding propensities of *Ec*SapA. Further, the molecular docking strategies highlighted *Ec*SapA's heme-binding capacity, emphasizing its multifunctionality. The study also delves into the TMD components of the *Ec*Sap transport system, wherein computational insights suggest an importer-like function of the *Ec*Sap transport system.

As the molecular docking studies unraveled the multifunctional prospects of *Ec*SapA, we further employed Molecular Dynamics (MD) simulations on all the obtained docked poses to estimate their validity. The MD simulation studies unraveled a dipeptide-binding selectivity in *Ec*SapA, wherein two dipeptides were gradually observed to be removed from the binding site of *Ec*SapA. Apart from the two, the other four dipeptides were observed to be maintained throughout the 1 µs trajectory. Further, the results demonstrated an initial and a final dipeptide binding site in *Ec*SapA, wherein the dipeptides are observed to dock in the initial site and translate into the final site. The MD simulation data also establishes the protegrin-1, protamine, and heme binding propensities of *Ec*SapA. In-depth inspection unraveled a segregated binding site of *Ec*SapA corresponding to its wide array of ligands. Further, *Ec*SapA was observed to demonstrate varied domain movements based on the size of the ligand so as to accommodate the wide range of ligands.

Finally, *Ec*SapA was studied with the aid of a series of biophysical strategies to further characterize and corroborate our computational insights. The results

indicated the sub-cellular co-localization of *Ec*SapA to be of key importance for adept folding of the protein. Further, the two disulfide bonds present in *Ec*SapA were observed to be a prerequisite for their structural stability. Various biophysical strategies demonstrate the heme-binding propensity of *Ec*SapA. Alongside, differential scanning fluorescence (DSF)-based analysis suggests the protamine-binding propensity of *Ec*SapA, in accordance with prior reports.

Thus, in summary, the outcomes of this study enlighten the enigmatic status of the *Ec*Sap transport system with the aid of elegant computational and biophysical strategies. The results encompass a first report on the distribution and evolutionary trend of the *sap* operon. Further, computational and biophysical strategies establish the multifunctional attributes of *Ec*SapA. Hence, proposing the crucial role of the *Ec*Sap transport system in the pathogenicity and sustenance of *E. coli*.

ABBREVIATIONS

ABC	ATP-binding cassette
ACP	Enoyl-acyl carrier protein
AMP	Antimicrobial peptide
APBS	Adaptive Poisson-Boltzmann Solver
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
CD	Circular dichroism
CH	Coupling helix
CM	Cell membrane
CTD	C-terminal domain
CV	Column-volume
CyH	cytoplasmic helix
DAS	Difference absorption spectroscopy
Ddp	D,D-dipeptide
DLS	Dynamic light scattering
Dpp	Dipeptide-binding protein
DSF	Differential scanning fluorimetry
EMSA	Electrophoretic mobility shift assay
FabI	Enoyl-ACP reductase
FPLC	Fast protein liquid chromatography
GsiA	Glutathione-binding protein
IM	Inner membrane
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Luria-Bertani
LGA	Lamarckian genetic algorithm
MALDI-ToF	Matrix-associated laser desorption/ionization-time of flight
MFS	Major facilitator superfamily
ML	Maximum likelihood
MSA	Multiple sequence alignment
N.D.	Not detected

NBD	Nucleotide-binding domain
NCBI	National Center for Biotechnology Information
NEB	New England Biology
NFW	Nuclease-free water
Nik	Nickel-binding protein
Ni-NTA	Ni-nitrilotriacetic acid
NJ	Neighbor-joining
NTD	N-terminal domain
MD	Molecular dynamics
OM	Outer membrane
OppA	Oligopeptide-binding proteins
ORF	Open reading frame
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PH	Periplasmic helix
PMSF	Phenylmethylsulfonyl fluoride
PS	Periplasmic space
Psp	Phage shock proteins
QC	Query coverage
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SAP	Sensitivity to antimicrobial peptide
SBP	Substrate (or solute)-binding protein
SI	Sequence identity
TMD	Transmembrane domain
TMH	Transmembrane helix
β-ME	β-mercaptoethanol

NOTATIONS

~	Approximately
%	Percent
Å	Angstrom
a.u	Arbitrary unit
bp	Base pair
°C	Degree Celsius
F_U	Fraction of unfolded molecule
g	Relative Centrifugal Force (RCF)
K_a	Association constant
kb	kilobase
kcal	Kilocalories
K_d	Dissociation constant
Da	Dalton
K_{av}	Distribution coefficient
kDa	kilodalton
L	Litre
M	Molar
mdeg	Change in ellipticity
mM	Millimolar
mL	Millilitre
n	Binding stoichiometry
OD	Optical density
RLU	Relative luminescence unit
s	Second
T_m	Melting temperature
ΔG	Change in free energy
ΔH	Change in enthalpy
ΔS	Change in entropy
V_o	Void volume
V_e	Elution volume
V_t	Total volume
Y_{obs}	CD singal at given temperature

Y_U	CD signal observed in unfolded protein
λ	Wavelength
µg	Microgram
µL	Microlitre
µM	Micromolar

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1. Introduction

1.1. The human immune system

Innumerous pathogens are omnipresent around human hosts, which always pose a risk of contracting diseases. Prior to any drugs, a robust system termed the human immune system acts as a wall to protect the host from any pathogenic attacks (Su *et al.,* 2014). The human immune system is a multilayered machinery wherein the system challenges pathogens and debilitates its capacity to form a definitive niche within the human host, which acts as a prerequisite for causing diseases (Zheng *et al.,* 2020). The human immune system can be broadly classified into three different parts: (1) Anatomical and physiological barrier, (2) Innate immune system, and (3) Adaptive immune system (Turvey and Broide, 2010).

The pathogen invading the human host is under attack by the layered host immune response. The human immune system has three distinct layers, which are put up against bacterial pathogens one by one. The first layer is termed the anatomical and physiological barrier, which consists of the skin, sweat, tears, etc. (Müller *et al.,* 2019). The epidermal layer of the human host hinders the entrance of the pathogens and is further challenged by the low pH of the human skin (Uberoi *et al.,* 2021). The human host further targets the incoming pathogens with the help of sweat and tears, which consists of certain peptide-like molecules termed antimicrobial peptides (AMPs), which attack the bacterial membrane and nullify their capacity to form a niche (McDermott, 2013; Wang *et al.,* 2016a). Further, the lungs demonstrate a phenomenon coined as ciliary clearance, wherein bacterial pathogens are expelled out of the body as a result of ciliary mobility (Joskova *et al.,* 2020). Also, the duodenal mucosa and pH hinder the formation of pathogenic niches within the human host (Figure 1.1) (Dupont *et al.,* 2014; Yamamura *et al.,* 2023).

The second layer is termed the innate immune system, which is further broadly classified into cellular and humoral immune systems (Shishido *et al.,* 2012). The innate cellular immune system is made up of white blood cells, including various mononuclear agranulocytes, such as macrophages and natural killer cells, as

well as various polymorphonuclear granulocytes, like neutrophils and basophils. These cells have been reported to be involved in the untargeted clearance of foreign material attacking human hosts (Sollome and Fry, 2015). The agranulocytes phagocytize the foreign material and present their antigens to elicit an immune response and generate immunological memory (Marshall *et al.,* 2018). The granulocytes, apart from phagocytosis, also contain granules consisting of lytic enzymes, which are released on the foreign body and disrupt their cellular homeostasis, ultimately leading to their clearance (Marshall *et al.,* 2018). Further, the humoral innate immune system comprises certain independent molecules, which encompass AMPs, c-reactive proteins, etc., which are produced by the aforementioned cells like macrophages, neutrophils, etc. (Figure 1.1) (De Santis *et al.,* 2020). These molecules either target the pathogen membrane homeostasis or enable activation of a directed immune response against the foreign molecule, leading to its clearance (Sproston and Ashworth, 2018; Benfield and Henriques, 2020).

The third and the penultimate barrier faced by pathogens is the adaptive immune system. The adaptive immune system, akin to the innate immune system, is further subdivided into the cellular and humoral immune systems. The cellular adaptive immune system comprises B and T-cells. The humoral adaptive immune system comprises antibodies produced via the B-cells, which interact with the pathogen surface and direct the complement proteins to these targets (Figure 1.1) (Killick *et al.,* 2018; Pollard and Bijker, 2021). The T-cells, on the contrary, recognize the antigens on the pathogen surface and manipulate the target cell to undergo apoptotic death (Killick *et al.,* 2018). A unique feature of the adaptive immune system is the presence of memory B and T-cells, which remember the antigen that resulted in the activation of the adaptive immune system and hence elicit a direct attack by the adaptive immune system on re-exposure of the same antigen (Farber *et al.,* 2014; Inoue *et al.,* 2018).

Amongst the three layers of immune responses elicited by the human immune system, the adaptive immune system takes several days (4-5 days) to initiate, and hence, the host is left at the mercy of the first two layers of immune responses for the preliminary clearance of pathogens (Mak *et al.,* 2013). A key member

present in the interface of both the layers is the AMPs, which enable the clearance of bacterial pathogens and hence are key defense machinery employed by the human hosts.

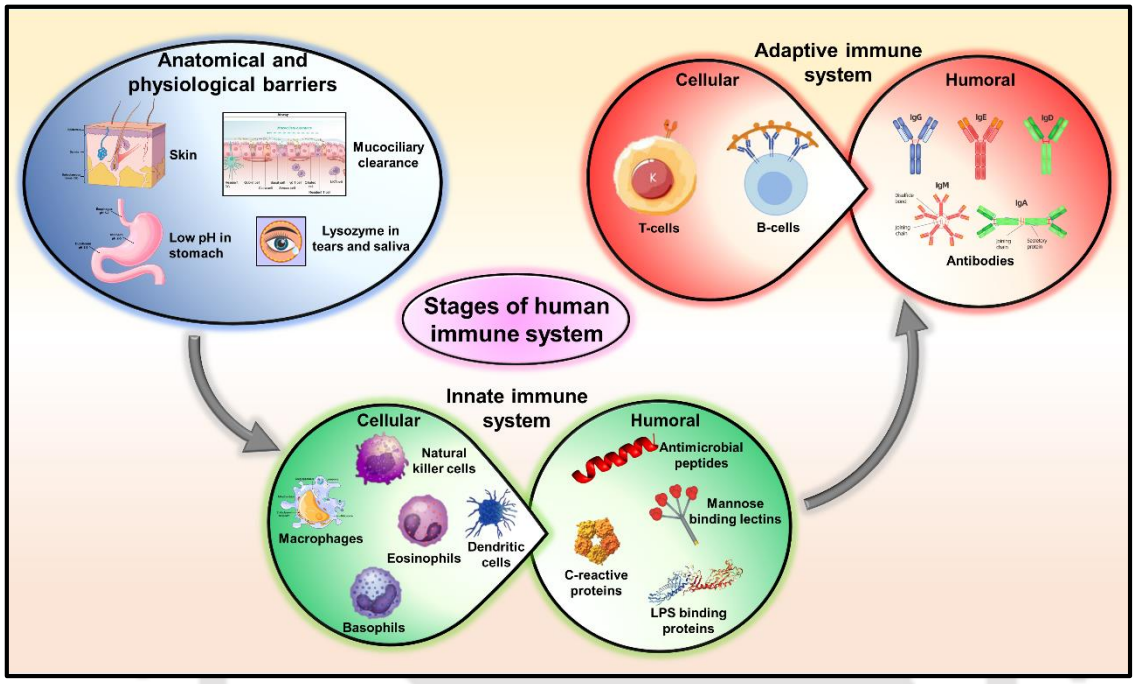


Figure 1.1. Stages of the human immune system. A schematic representation of the three layers of the human system has been depicted, wherein the three components are (1) anatomical and physiological barriers (blue), (2) the innate immune system (green), and (3) the adaptive immune system (red). In (2) and (3), the immune system was further divided into the cellular and humoral systems. The order of appearance of the layers of the immune system has been depicted in the form of grey-curved arrows. The components of each of the layers of the immune system have been pictorially represented.

1.2. Antimicrobial peptides (AMPs)

AMPs are peptides generally ranging from 12 to 50 amino acids in length (Ciulla and Gelain, 2023). The AMPs are mostly positively-charged molecules, which have been widely reported to form charge-based contacts with the negatively-charged bacterial membrane (Hollmann *et al.,* 2018). The interaction of AMPs with the bacterial membrane initiates the process of pathogen clearance by mostly disrupting the membrane homeostasis or by internalizing and attacking

crucial internal targets (Duarte-Mata and Salinas-Carmona, 2023). These AMPs, in many cases, are produced in an inactive form, wherein the C-terminal active AMP is expressed alongside an N-terminal signal peptide and pro-peptide (Rončević *et al.,* 2019). The AMPs are activated via post-translational modifications, wherein the signal peptide and the pro-peptide are cleaved off with the aid of signal peptidase and protease, respectively, rendering the AMP in active form (Mabrouk, 2022). The AMPs are obtained from multiple sources and widely vary from one another on the basis of structure and amino acid compositions, based on which they are often categorized (Huan *et al.,* 2020; Mabrouk, 2022).

1.2.1. Classification of AMPs

1.2.1.1. Structure-based classification of AMPs

The AMPs are classified into five distinct classes based on their secondary structural components, as follows: (1) linear α-helical AMPs; (2) β-sheet consisting AMPs; (3) AMPs composed of α-helices and β-strands; (4) loop-based/ extended peptides; and (5) cyclic or unusual AMPs (Table 1.1) (Huan *et al.,* 2020).

1.2.1.1.1. Linear α-helical AMPs

These AMPs, which are primarily constituted of a single helix, are the most well-studied and the largest group (Figure 1.2) (Koehbach and Craik, 2019). These AMPs are reported to be constituted of residues such as alanine, leucine, and lysine, which stabilize their helical conformation (Mabrouk, 2022). Various noteworthy AMPs, such as LL-37 cathelicidins, magainins, melittin, and cecropins, belong to this class (Mackler and Kreil, 1977; Steiner *et al.,* 1981; Zasloff, 1987; Xhindoli *et al.,* 2016). While multiple AMPs retain their helical conformation in solutions, certain AMPs of this class have been reported to enhance their α-helical properties in the presence of the membrane-targeted by them (Porcelli *et al.,* 2013).

1.2.1.1.2. β-sheet consisting AMPs

The AMPs of this class consist of a minimum of two β-strands and contain cysteine residues, which are reported to form disulfide bonds and stabilize the

AMPs (Figure 1.2) (Getahun *et al.,* 2022). Based on the directionality of the disulfide bonds formed, the members of this class are further divided into two subclasses: (1) β-hairpin peptides and (2) α-defensin peptides (Koehbach and Craik, 2019). A few examples of the β-hairpin peptides are protegrin-1, thanatin, gomesin, and tachyplesin, wherein the number of disulfide bonds varies in this group of AMPs (Miyata *et al.,* 1989; Kokryakov *et al.,* 1993; Mandard *et al.,* 1998; Mandard *et al.,* 2002). Defensin peptides (α-, β-, and θ-defensins) lie in the interphase of this class and the forthcoming class (α/β mix) of peptides (Shafee *et al.,* 2017). The α-defensins are the class of defensins that belong to this structural class of AMPs. The α-defensin AMPs consist of three antiparallel β-strands that are connected via trans-oriented disulfide bonds (i.e., wherein one single β-strand is connected with two other elements via two disulfides pointed in opposite directions) (Shafee *et al.,* 2017). The β-strand consisting of AMPs is stable, and unlike the α-helix-containing AMPs, these AMPs have not been reported to undergo conformational change upon contacting their target membranes (Kumar *et al.,* 2018).

1.2.1.1.3. AMPs composed of α-helices and β-strands

A section of the defensin AMPs possesses structural features of this class of AMPs, wherein, alongside β-strands, the AMP also consists of α helix (Figure 1.2). In this class, the β-strands of the defensin AMPs are linked via disulfide bonds to helical elements (Koehbach and Craik, 2019). The human β-defensins hail from this group, wherein the β-strand forms a trans-orientation disulfide bond with an α-helix. In contrast, the β-defensins obtained from plants and invertebrates form cis-orientation disulfide bonds (i.e., wherein the β-strand forms two disulfide bonds in the same direction with an α-helix) (Koehbach, 2017). Various other AMPs from scorpions (defensins) and plants (α-1-purothionin) also possess a combination of α-helix and β-strands (Zhu *et al.,* 2014). This class of AMPs has been reported to possess high numbers of proline and arginine residues (Nayab *et al.,* 2022).

1.2.1.1.4. Linear AMPs

Certain AMPs are reported to be devoid of secondary structural elements like α-helices and β-sheets in their free-floating state and even in the presence of

membranes (Koehbach and Craik, 2019). Linear AMPs are reported to be rich in certain residues like histidine, arginine, tryptophan, and proline (Butler *et al.,* 2013). Certain examples of linear AMPs include indolicidin, histatin, and abaecin (Oppenheim *et al.,* 1988; Casteels *et al.,* 1990; Selsted *et al.,* 1992). The proline-arginine (PR) AMPs also possess a linear extended structure and belong to this class of AMPs. Certain examples of PR AMPs are mammalian PR-39 and drosocin, produced by *Drosophila melanogaster* (Graf *et al.,* 2017). Despite belonging to this class of AMPs, the PR-39 is also reported to form a polyproline-II (PP-II) helix (Cabiaux *et al.,* 1994). Similarly, disordered cationic peptides like protamines from humans and bulls have also been recently observed to form helices and hairpin loops when simulated against water. However, the protamine from salmon remained disordered throughout the trajectory, hence suggesting the varying structural attributes of these AMPs (Shadman *et al.,* 2022).

1.2.1.1.5. Cyclic or other unusual peptide topologies

Recently, in an eloquent review, apart from the above-mentioned classes, a fifth class of AMPs has been added, which accounts for various AMPs possessing cyclic and unusual structures (Koehbach and Craik, 2019). One of the key members of this class is the θ-defensins (e.g., RTD-1) (Figure 1.2) (Conibear *et al.,* 2012). Various other AMPs have been observed to possess cyclic and unusual structures; selected few examples of them are as follows: Microcin J25 (bacteria), Subtilosin-A (bacteria), and Kalata B1 (plant) (Saether *et al.,* 1995; Rosengren *et al.,* 2003; Kawulka *et al.,* 2004). The members of this class are reported to demonstrate high stability and are widely tweaked for chemical engineering applications (Koehbach and Craik, 2019).

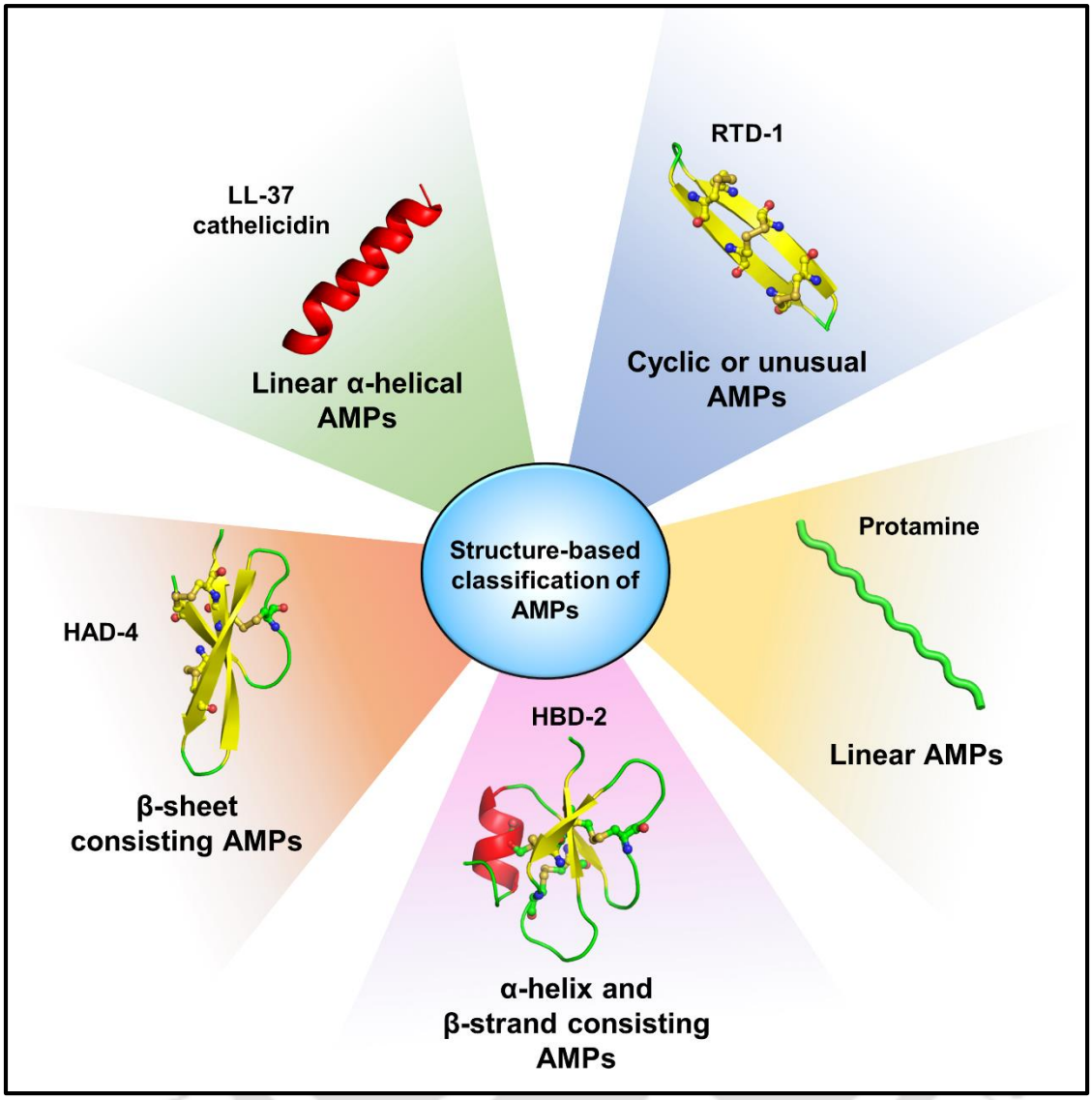


Figure 1.2. Structure-based classification of AMPs. A few examples of AMPs have been represented in the form of cartoons and highlighted in various colors based on the classes to which they belong. Wherein, the LL-37 cathelicidin (UniProt id: P49913) represents the linear α-helical AMPs (green background); human α-defensin-4 (HAD-4, UniProt id: P12838) represents the β-sheet containing AMPs (orange background); human β-defensin 2 (HBD-2, UniProt id: O15263) represents α-helix and β-strands containing AMPs (pink background); Protamine (UniProt id: P69014) represents linear AMPs (yellow background); and rhesus θ-defensin-1 (RTD-1, UniProt id: P82271) represents the cyclic or unusual AMPs (blue background). Wherever present, the disulfide bonds in the AMPs have been represented in the form of a yellow ball-and-stick model. The α-helices, β-strands, and loops have been represented in red, yellow, and green colors, respectively.

1.2.1.2. Amino acid-based classification of AMPs

AMPs are observed to be rich in certain residues, based on which their functionality and structural attributes are dictated. Based on the dominant residues composing the AMPs, they are classified into various classes of AMPs, such as proline-rich AMPs (PrAMPs), tryptophan- and arginine-rich AMPs, histidine-rich AMPs, and glycine-rich AMPs (Huan *et al.,* 2020). Apart from PrAMPs, all other classes of AMPs are capable of interacting with the membrane via hydrophobic and charge-based contacts, which enables the AMP to permeate through the membrane, ultimately leading to membrane rupturing (Dong *et al.,* 2019; Walrant *et al.,* 2020; Singh *et al.,* 2023). PrAMPs are reported to target internal targets like ribosomes, wherein the AMPs are reported to internalize via the membranes through the aid of bacterial transporters such as SbmA (Mattiuzzo *et al.,* 2007; Seefeldt *et al.,* 2015).

1.2.2. Mechanism of action of AMPs

The AMP molecules majorly exert their antibacterial attributes primarily via two strategies. The first strategy targets the bacterial membrane and disrupts the homeostasis of the membrane, while in the second strategy, AMPs cross the membrane and target essential internal machinery of the bacterial system, leading to their death (Figure 1.3 and Table 1.1) (Zhang *et al.,* 2021).

1.2.2.1. Membrane-targeted action mechanism

The AMPs target the membranes and act upon them primarily with the aid of three distinct mechanisms: the toroidal pore model, the barrel-stave model, and the carpet model (Huan *et al.,* 2020). The toroidal pore model forms a pore on the bacterial membrane, wherein the AMP molecules are vertically inserted within the membrane and interact with the lipid head group, making the lipid molecules bend to accommodate the AMPs (Matsuzaki *et al.,* 1998; Yang *et al.,* 2001). The insertion of the AMPs within the membrane leads to the formation of a ring hole-like structure, which is reported to have a diameter of 10-20 Å (Matsuzaki *et al.,* 1996). The pores formed on the membrane lead to membrane depolarisation, ultimately leading to cell death (Omardien *et al.,* 2018). Certain well-studied examples of AMPs forming toroidal pores are melittin, magainin, and aurein 2.2 (Matsuzaki *et al.,* 1998; Yang *et al.,* 2001; Cheng *et al.,* 2009).

The second model demonstrated by the AMPs is the barrel-stave model, wherein the AMPs form barrel-shaped channels on the bacterial membrane (Jean-François *et al.,* 2008). In this case, the AMPs aggregate and penetrate the entire length of the bacterial membrane, and the channel formed leads to cytoplasmic leakage, followed by cell death due to rupturing of the membrane (Lohner and Prossnigg, 2009). The major difference from the toroidal pore model is the pattern of interaction with the lipid bilayer. Unlike the toroidal pore model, the barrel-stave model does not exclusively interact with the lipid head group, and no apparent bending of the lipid molecules is observed (Pandidan and Mechler, 2021). The α-helical AMP alamethicin and the β-sheet containing AMP protegrin-1 are reported to form barrel-stave model-based channels, wherein the AMPs are reported to oligomerize and form stable barrel-shaped channels which disrupt the bacterial membrane homeostasis (He *et al.,* 1996; Lipkin and Lazaridis, 2015).

In the third mechanism, the AMPs are positioned parallel to the membrane and cover the membrane akin to a carpet. Hence, this mechanism has been coined as the carpet model. Akin to detergents when the AMP concentration crosses the critical-peptide concentration, the AMPs disrupt the membrane by forming peptide-coated lipid micelles (Oren and Shai, 1998). The parallel orientation of the AMPs is reported to affect the fluidity of the membrane, leading to displacement of the lipids and subsequent disruption of the membrane (Powers and Hancock, 2003; Pandidan and Mechler, 2021). Examples of AMPs demonstrating the carpet model hail from the α-helix containing AMPs (Human LL-37 cathelicidin and mammalian cecropin P1) and also from β-sheet containing AMPs (arenicin) (Oren *et al.,* 1999; Shenkarev *et al.,* 2011; Lyu *et al.,* 2019).

1.2.2.2. Non-membrane targets of AMPs

Even though the primary target of multiple AMPs is the bacterial membrane, various AMPs are reported to inhibit certain internal targets crucial for the sustenance of the bacterial system (Brogden, 2005). The AMPs that attack these internal targets enter the cells via endocytosis or with the aid of certain transporters (Huan *et al.,* 2020). The internal targets of the AMPs majorly focus on the central dogma of the target organism, wherein the AMP inhibits the replication, transcription, or translation of the organism, ultimately leading to its

death (Figure 1.3) (Matthyssen *et al.,* 2022). For instance, the AMP Indolicidin binds to DNA molecules, inhibits DNA synthesis, and has been reported to inhibit the function of the DNA topoisomerase I (Subbalakshmi and Sitaram, 1998). Further, certain AMPs like PR-39, $Bac7_{1-35}$, and Tur1A target the ribosome and inhibit protein synthesis in bacterial systems (Mardirossian *et al.,* 2014; Graf *et al.,* 2017; Mardirossian *et al.,* 2018). The peptide TC24 has a dual functionality wherein it ruptures the bacterial cell membrane and further degrades its RNA and DNA molecules (He *et al.,* 2017).

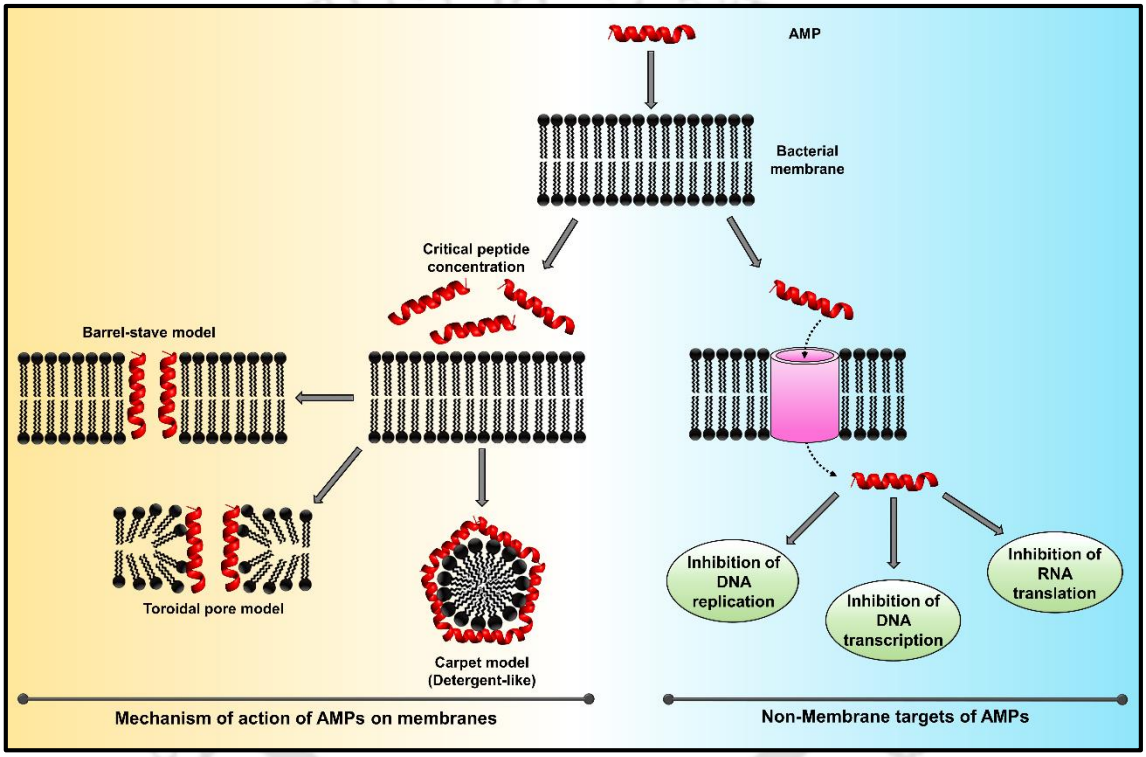


Figure 1.3. Schematic representation of the mechanism of action of AMPs. The AMPs (red) gain proximity to the bacterial membrane and either disrupt the membrane (left panel) or get internalized with the aid of certain bacterial transporters (right panel). The bacterial membrane is disrupted by the AMPs via three types of machinery: the barrel-stave model, the toroidal model, and the carpet model, represented in the left panel of the figure. The right panel of the figure demonstrates the internal targets (DNA replication, transcription, and RNA translation) of the AMPs post-transport.

Table 1.1. Mechanism of action of selected AMPs from the five structural classes of AMPs.

S. No.	AMP	Source	Mechanism of action
Linear α-helical AMPs			
1.	LL-37 cathelicidin	Human	Carpet model (Majewska *et al.,* 2021).
2.	Cecropin	Insects	Carpet model (Ziaja *et al.,* 2020).
3.	Melitin	Honey bee	Toroidal pore model (Pandidan and Mechler, 2019).
β-sheet containing AMPs			
1.	Protegrin-1	Pig	Barrel-stave and toroidal pore model (Bolintineanu *et al.,* 2012; Capone *et al.,* 2010).
2.	α-defensins	Human	Transmembrane pore formation and inhibition of cell-wall biosynthesis via binding to lipid II (White *et al.,* 1995; de Leeuw *et al.,* 2010).
3.	Tachyplesin-1	Horse-shoe crab	Toroidal pore model (Imura *et al.,* 2007).
α-helix and β-strands containing AMPs			
1.	β-defensins	Human	Carpet model (Dhople *et al.,* 2006).
2.	α-1-purothionin	Wheat	Membrane depolarization via channels (Llanos *et al.,* 2006).
Linear extended AMPs			
1.	Drosocin	*Drosophila sp.*	Binds to the ribosome (Krizsan *et al.,* 2015).
2.	Indolicidin	Cattles	Binds to DNA (Le *et al.,* 2017).
3.	PR-39	Pig	Binds to the ribosome (Graf *et al.,* 2017).
Cyclic or unusual AMPs			
1.	Nisin-A	*Lactococcus lactis*	Binds to lipid II and inhibits cell-wall synthesis (Hsu *et al.,* 2004).
2.	Kalata B1	*Oldenlandia affinis*	Formation of membrane pore (Huang *et al.,* 2009).

1.2.3. Membrane composition-based interaction of AMPs

The membrane is the site of interaction for all the AMPs. Further, the membrane-AMP interactions are primarily-based on electrostatic interactions (Lei *et al.,* 2019). The composition of the eukaryotic membrane and the prokaryotic membrane vary to a great extent. The mammalian membrane is dominated by zwitterionic lipids such as phosphatidylcholine, phosphatidylethanolamine, and sphingomyelins (Raghava *et al.,* 2011; Greco *et al.,* 2020). The bacterial membrane generally contains zwitterionic lipids such as phosphatidylethanolamine mixed with negatively-charged lipids such as phosphatidylglycerol (Raghava *et al.,* 2011; Sohlenkamp and Geiger, 2016). Hence, the absence of negatively-charged lipids in the eukaryotic membranes protects them from the attack of the cationic AMPs and makes the bacterial cell a viable target for their attack (Figure 1.4). The presence of cholesterol stabilizes the eukaryotic membrane and further averts the action of cationic AMPs on the eukaryotic cells (Greco *et al.,* 2020; Mabrouk, 2022). Hence, this difference in membrane lipid composition between the cells enables successful targeting of the AMPs against the bacterial cells.

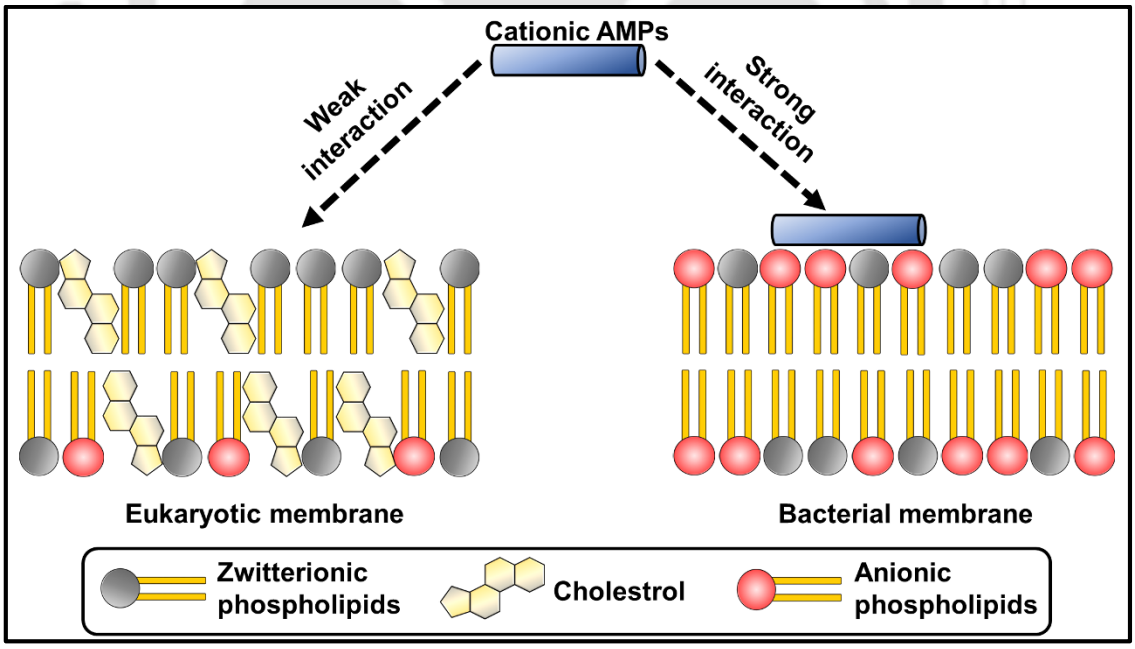


Figure 1.4. Schematic representation of AMP interaction with eukaryotic and bacterial membranes. The cationic peptide (blue) is observed to have a strong hydrophobic and charge-based interaction with the bacterial membrane (right panel) and a weak hydrophobic interaction with the eukaryotic membrane

(left panel). The details of certain symbols utilized in the figure have been elaborated in the box inset (figure redrawn from Lei *et al.,* 2019).

1.3. Resistance machinery employed by Gram-negative bacteria

Gram-negative bacteria have developed multiple mechanisms to evade the lethal action of the AMPs. These mechanisms can be broadly classified into five distinct classes as follows: (1) Proteolytic cleavage of AMPs; (2) shielding the surface of the bacterial cells; (3) modification of outer membrane; (4) altering AMP expression in the host; (5) export or import of AMPs via transporters (Gruenheid and Le Moual, 2012) (Figure 1.5).

1.3.1. Proteolytic cleavage of AMPs

In Gram-negative bacterial systems, certain proteases are reported to be present on the cell surface (Figure 1.5). These proteases are reported to act on the attacking AMPs and degrade them; certain instances of such proteases encompass *Escherichia coli* OmpT, *Salmonella typhimurium* PgtE, *Citrobacter rodentium* CroP, and *Yersinia pestis* Pla (Stumpe *et al.,* 1998; Guina *et al.,* 2000; Galvan *et al.,* 2008; Le Sage *et al.,* 2009). All these proteases are examples of aspartate proteases belonging to the omptin family and have been reported to cleave AMPs such as LL-37 (OmpT and Pla), CRAMP (CroP), protamine (OmpT), and C18G (PgtE) (Stumpe *et al.,* 1998; Guina *et al.,* 2000; Galvan *et al.,* 2008; Le Sage *et al.,* 2009; Thomassin *et al.,* 2012). Further, certain other proteases like *Proteus mirabilis* ZapA, which is a zinc-metalloprotease is reported to act upon AMPs such as LL-37, β-defensin 1, and protegrin-1 (Belas *et al.,* 2004). Similarly, the zinc-metalloproteases ZmpA and ZmpB of *Burkholderia cenocepacia* have been reported to act upon LL-37 and β-defensin 1 (Kooi and Sokol, 2009). The chronic periodontal disease-causing pathogen *Porphyromonas gingivalis* evades the attack of LL-37 and β-defensin 3, with the aid of three proteases termed as gingipains (Gutner *et al.,* 2009; Maisetta *et al.,* 2011). Few reports have also suggested that the cytoplasmic proteases of *Haemophilus influenzae* are involved in the degradation of AMPs that are transported to the cytosol (Shelton *et al.,* 2011).

1.3.2. Shielding the surface of the bacterial cells

Various Gram-negative bacterial species possess an additional layer above the outer membrane, which in turn acts as a shield against the AMPs (Figure 1.5). This layer is generally constituted of polysaccharides and enables the formation of capsules, slime layers, and biofilms (Mirghani *et al.,* 2022). The layers formed have a sieving effect wherein these reduce the access of the AMPs to the bacterial membranes (Blair *et al.,* 2022). The capsular polysaccharides of *Klebsiella pneumoniae* and *Pseudomonas aeruginosa* are reported to induce aggregation of the AMPs, which leads to electrostatic trapping of the AMPs, hindering their movement (Chan *et al.,* 2004; Campos *et al.,* 2004; Foschiatti *et al.,* 2009). The *E. coli* polysaccharide intercellular adhesin (PIA) involved in biofilm formation is also reported to provide resistance against dermicidin, LL-37, and β-defensin 3 (Wang *et al.,* 2004). Another strategy employed by *P. aeruginosa* involves the proteolytic cleavage of the proteoglycan decorin present in the host cells, which releases negatively charged elements (dermatan sulfate), which in turn inactivates certain AMPs (Schmidtchen *et al.,* 2001). Further, in the case of *E. coli* on the surface, certain structures termed curli fimbriae are reported to interact with LL-37 and hence protect the bacteria from their attack (Kai-Larsen *et al.,* 2010).

1.3.3. Modification of outer membrane

The modification of lipid A and core moieties of Gram-negative bacteria can take place either during lipopolysaccharide synthesis and its transport to the outer membrane or within the outer membrane itself (Gruenheid and Le Moual, 2012). The lipopolysaccharide modifications are often regulated by environmental stimuli via two-component signaling systems. These modifications have been reported to be involved in virulence, resistance against AMPs, and modulate TLR4-mediated inflammatory response (Miller *et al.,* 2005). The lipopolysaccharide modifications, specifically in lipid A, have been reported to influence the virulence of the bacteria in a positive way, as this modification reinforces the outer membrane permeability and neutralizes the negative charge of the lipopolysaccharides, preventing interaction with AMPs (Figure 1.5) (Gruenheid and Le Moual, 2012). The role of lipopolysaccharide modification in AMP resistance is reported in multiple instances of *S. typhimurium*, *P.*

aeruginosa, *Yersinia sp.*, *E. coli*, and *Neisseria sp.* (Richards *et al.,* 2010). Further, in *S. typhimurium,* the outer membrane enzyme PagP transfers the palmitoyl group from phospholipids onto the lipid A, producing a hepta-acylated lipid A, which confers resistance against AMPs such as C18G and protegrin, reinforcing the outer membrane barrier (Guo *et al.,* 1998). In addition to these modifications, the *arnBCADTEF* operon confers resistance to polymyxin B by synthesizing and transferring the positively-charged L-Ara4N to mask the negatively-charged 4′-phosphate of lipid A. This modification takes place in the periplasmic side of the outer membrane prior to being transported to the extracellular side of the outer membrane (Gunn, 2008).

1.3.4. Altering AMP expression in the host

Along with the aforementioned strategies, certain Gram-negative bacterial systems have been observed to down-regulate the expression of the host cells (Figure 1.5). Certain notable instances comprise *Shigella spp*., wherein the bacteria uses its type III secretion system along with the transcriptional regulator mxiE to suppress the expression of certain AMPs such as β-defensins and LL-37 produced by intestinal epithelial cells (Islam *et al.,* 2001; Sperandio *et al.,* 2008). Similarly, *S. typhimurium* is also reported to employ the type III secretion system and PhoP to reduce the expression of cryptdin in mice (Salzman *et al.,* 2003). Further, certain bacterial species, such as *Vibrio cholerae* and *E. coli*, produce exotoxins that downregulate the expression of LL-37 and β-defensin-1 produced by the host cells (Chakraborty *et al.,* 2008).

1.3.5. Export or import of AMPs via transporters

When all the above-discussed mechanisms fail to neutralize the action of the AMPs, they form pores and start accumulating above the bacterial membrane. Various transporters are recruited to efflux out the AMPs or import them to the cytoplasm to be proteolytically degraded (Figure 1.5) (Blair *et al.,* 2022). The transporters involved in the clearance of the AMPs hail from various transporter families, such as the resistance-nodulation-division (RND) transporters, the major facilitator superfamily (MFS), ATP-binding cassette (ABC) transporters, and K^+ transporters (SKT) (Table 1.2). The members of the RND and MFS transporters are reported to be involved in exporting the AMPs from the periplasm-inner

membrane interphase (Blair *et al.,* 2014). While ABC transporters are primarily involved in the uptake of AMPs, a recent report suggested a member of this class be involved in exporting AMPs as well (Gruenheid and Le Moual, 2012; Honeycutt *et al.,* 2020).

The RND efflux pumps function with the aid of proton motive force (PMF), which relies on the pH gradient across the bacterial inner membrane (Martins *et al.,* 2009). One example of the RND transporter is the MtrCDE efflux pump. The loss of the MtrCDE efflux pump of *Neisseria gonorrheae* makes the bacterium more susceptible to LL-37 and protegrin-1 (Shafer *et al.,* 1998). The MtrC component of the efflux pump of *Haemophilus ducreyi* is reported to be crucial for resistance against the AMPs LL-37 and β-defensins (Rinker *et al.,* 2011). Further, the AcrAB-TolC efflux pumps are also an example of RND transporters and have been observed to render *K. pneumoniae* resistant against the AMPs polymyxin B, α- and β-defensins (Padilla *et al.,* 2010). Hence, all the studies establish the significance of the RND transporters in rendering the bacterial system resistant to certain AMPs. Further, the members of the MFS transporters (another PMF-driven pump), such as EmrAB-TolC and RosAB of *E. coli* and *Yersinia enterocolitica,* respectively, have been reported to render the pathogens resistant against certain AMPs (Bengoechea and Skurnik, 2000; Weatherspoon-Griffin et al., 2014). Further, it is proposed that the MFS transporters function in clearing the toxic molecules in the bacterial inner membrane by transporting them into the periplasm, wherein the RND efflux pumps, in a relay-like manner, get rid of the toxic molecules accumulated in the periplasm by pumping them out of the bacterial cell (Tal and Schuldiner, 2009). Interestingly, reports have also suggested the involvement of the *Vibrio vulnificus* TrkA-TrkH (K^+ uptake) system to be involved in resistance against the AMPs polymyxin B and protamine (Chen *et al.,* 2004). However, the mechanism of action of the Trk system against the AMPs remains obscure.

Apart from exporters, various importers have also been reported to be involved in contributing to the resistance against the AMPs. The AMP molecules that evade the action of the exporter system are imported with the aid of certain importers, which are reported to hail from the ABC transporter superfamily (Parra-

Lopez *et al.,* 1993; Eswarappa *et al.,* 2008). One example is the Yej transport system, which has been reported to provide resistance in *S. typhimurium* against a plethora of AMPs such as protamine, melittin, polymyxin B, and β-defensins 1 and 2 (Eswarappa *et al.,* 2008). Further, another ABC importer termed the Sap transport system has been reported to internalize the AMPs akin to the Yej transport system (Parra-Lopez *et al.,* 1993). The Sap transport system has been observed to be present in a wide number of Gram-negative pathogens and has been reported to enhance their pathogenicity within the host (Hsu *et al.,* 2019). Further, a recent report suggests the ABC transporter MacAB-TolC of *Salmonella sp.* to be involved in AMP export (Honeycutt *et al.,* 2020). Hence, it suggests that members of the ABC transport system be involved in the import and export of the AMP molecules, substantiating their prevalence.

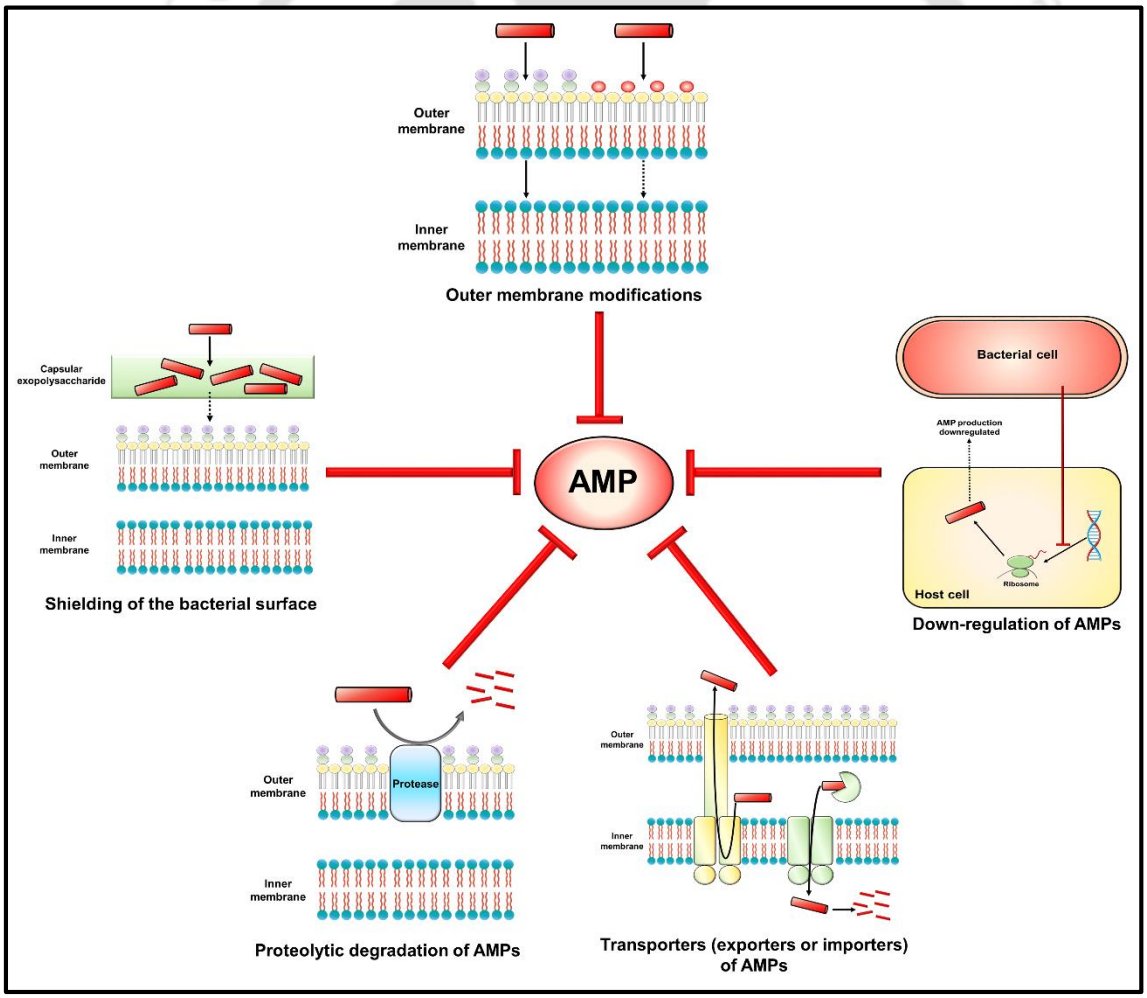


Figure 1.5. Schematic representation of Gram-negative resistance machinery against AMPs. The figure demonstrates various mechanisms of

resistance against AMPs (red), such as proteolytically degrading AMPs, shielding of the bacterial membranes to hinder AMP access, modifying the bacterial outer membrane, down-regulating the expression of the AMPs, and employing transporters to export or import the AMP molecules. The red blunt arrows arising from each mechanism represent the inhibition of AMPs. In the figure, all the solid arrows represent strong penetration or transport of AMPs, while all the broken arrows represent weak penetration or transport of AMPs.

Table 1.2. Transporter families involved in resistance against AMPs in Gram-negative bacteria. PG-1: protegrin-1, CO: colistin, PXB: polymyxin B, HNP-1: human α-defensin-1, HBD: human β-defensin, PRO: protamine, MEL: melittin, SN-1: Snakin-1, and ATH-1: α-thionin.

Transporter family	Example	Organism	Ligand(s)	Reference(s)
RND	MtrCDE	*N. gonorrhoeae*	LL-37, PG-1, CO, and PXB	Shafer *et al.*, 1998; Chitsaz *et al.*, 2019
		H. ducreyi	LL-37, HBD-2, -3, and -4	Rinker *et al.*, 2011
	AcrAB-TolC	*K. pneumoniae*	PXB, HNP-1, HBD-1, and -2	Padilla *et al.*, 2010
	VexAB–TolC	*V. cholerae*	PXB	Bina *et al.*, 2008
MFS	ErmAB-TolC	*E. coli*	PRO	Weatherspoon-Griffin *et al.*, 2014
	RosAB	*Y. enterocolitica*	PXB, CP1, and MEL	Bengoechea and Skurnik, 2000
SKT	TrkG/H-TrkA	*V. vulnificus*	PRO and PXB	Chen *et al.*, 2004

ABC	SapABCDF(Z)*	*S. typhimurium*	PRO, MEL	Parra-Lopez *et al.,* 1993
		H. influenzae	LL-37 and HBD-3	Rivera *et al.,* 2024
		K. pneumoniae	LL-37	Hsu *et al.,* 2019
		H. ducreyi	LL-37	Mount *et al.,* 2010
		Dickeya dadantii	SN-1 and ATH	López-Solanilla *et al.,* 1998
	YejABEF	*Salmonella sp.*	PXB, MEL, PRO, HBD-1, and -2	Eswarappa *et al.,* 2008
	MacAB-TolC	*Salmonella sp.*	C18G	Honeycutt *et al.,* 2020

*SapZ is only present in the case of *H. influenzae* and absent in other examples.

1.4. ATP-binding cassette (ABC) transporters

The ABC transporters are a ubiquitous class of transporters located in the inner membrane of Gram-negative bacteria and are powered by the hydrolysis of adenosine triphosphates (ATP) molecules for the transport of molecules across the membrane (Rees *et al.,* 2009). The members of this class are well reported to be unidirectional, i.e., they either export or import molecules across the membrane, based on which the members of this group are sub-divided as exporters and importers (Figure 1.6A-1.6B) (Saurin *et al.,* 1999). The ABC exporters are ubiquitous across all domains of life (eukarya, archaea, and bacteria) and are utilized to efflux molecules such as toxins, drugs, and peptides from the cells (Wilkens, 2015). The ABC exporters are constituted of four functional domains, comprising of two transmembrane domains (TMDs; TMD1 and TMD2) and two nucleotide-binding domains (NBDs; NBD1 and NBD2) (Figure 1.6A). The TMDs and NBDs either form homodimers (i.e., two copies of

the same TMD or NBD dimerize) or heterodimers (i.e., two different TMDs or NBDs dimerize). The dimerized TMDs and NBDs further oligomerize and form the functional unit of the ABC exporter. Contrary to the exporters, the ABC importers are employed to sequester of essential nutrients, such as peptides, metals, sugars, etc., for the bacterial system. They are predominantly observed in prokaryotes; however, certain instances have been reported in eukaryotes as well (Shitan *et al.,* 2003; Terasaka *et al.,* 2005). Along with the four functional sub-units, the ABC importers require an additional sub-unit known as the substrate (solute)-binding protein(s) (SBPs), which sequesters ligands and delivers them to the TMDs (Figure 1.6B). These SBPs also act as the key determinant regarding the specificity of the transporter it is associated with (Maqbool *et al.,* 2015).

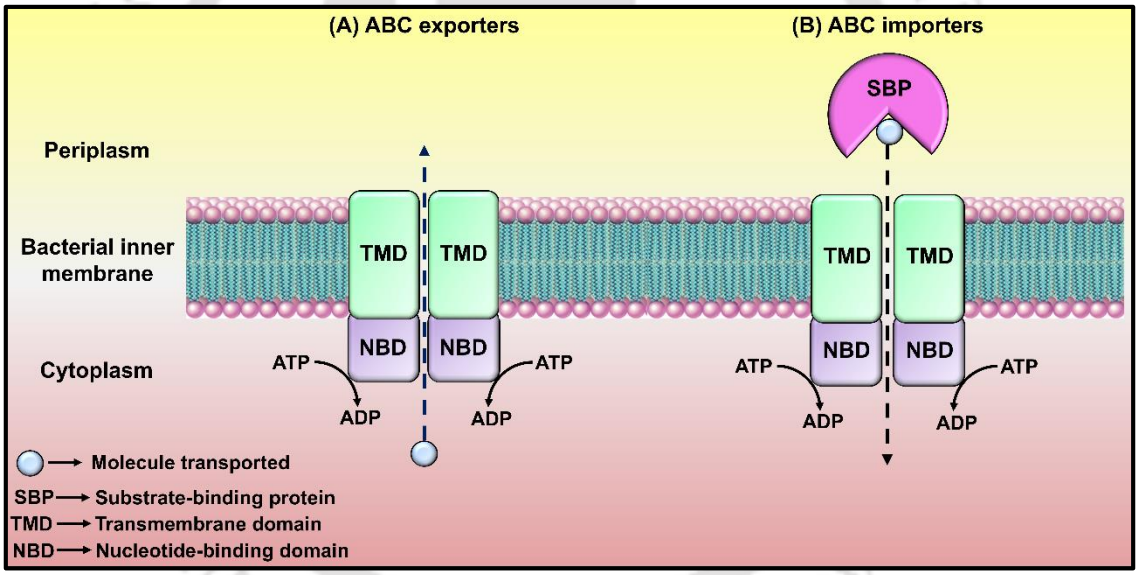


Figure 1.6. Schematic representation of ABC transporters. (A) ABC exporter is composed of TMDs (green) and NBDs (purple). (B) ABC importer consists of an additional component in the form of a SBP (magenta), along with the TMDs and NBDs. The abbreviations utilized in the figure have been elaborated in the lower left panel.

1.4.1. Transmembrane domains (TMDs)

The TMDs, often referred to as 'permeases', are membrane-spanning proteins of the ABC transporters that form portals within the membranes and allow selective passage of diverse molecules. The TMDs are majorly composed of long transmembrane helices (TMH), which span across the membrane. Interestingly, despite the absence of any conservation in the primary structure of the TMDs, they are largely conserved based on their tertiary structure. The variation in the TMDs at the sequence level could be attributed to the wide range of ligands transported through them. The number of TMHs of the ABC importers and the ABC exporters has been observed to vary, wherein the ABC importers constitute 12-20 THMs, and the ABC exporters house 10-12 TMHs (Hollenstein *et al.,* 2007). The variation in the TMH arrangement was recently utilized for a systematic classification of the TMDs into seven distinct classes based on the types of TMD folds and the number of TMHs (Table 1.3) (Thomas *et al.,* 2020). The types I-III belong to the ABC importers, while the remaining variations (types IV-VII) comprise exporters (Table 1.3). Further, the TMD of the Mla system (MlaE) did not fit into any of the types, and hence, recently, an additional type VIII was added to the TMD classification to incorporate the same (Table 1.3) (Ekiert *et al.,* 2022). The TMDs are reported to possess a short cytoplasmic helix termed the coupling helix (CH). The TMDs are reported to make contact with the NBDs with the aid of the CHs. The number and position are also observed to vary amongst the ABC importers and exporters and have been enlisted in Table 1.3. The CHs are reported to couple the conformational shifts of the NBDs to the TMDs and, hence, in turn, enable the transport of the ligand (Rice *et al.,* 2014).

Table 1.3. Classification of prokaryotic TMDs based on their folds.

TMD fold	TMH arrangement	No. of CH/ TMHs connected by CHs	Example (PDB id)
Type I	(5-6/8*) + (5-6)	2/ TMH6-7 and TMH4-5	$MalFGK_2$-MalE (2R6G)
Type II	10 + 10	2/ $(TMH6-7)_2$[#]	$BtuC_2D_2$-BtuF (1L7V)

Type III	4-8 (TMD) + 6-7 (S-component)	2/ TMH4-5	EcfTAA´-FolT (5D3M)
Type IV	6 + 6	4/ $(TMH2-3)_2$ and $(TMH4-5)_2$	TmrAB (5MKK)
Type V	6 + 6	2/ $(TMH2-3)_2$	TarGH (6JBH)
Type VI	6 + 6	2/ $(TMH2-3)_2$	$LptB_2FG$ (5X5Y)
Type VII	4 + 4	4/ $(TMH2-3)_2$ and $(TMH4)_2$	MacB (5GKO)
Type VIII	5 + 5	2/ $(TMH2-3)_2$	$MlaF_2E_2D_6B_2$ (6XBD)

*Only in the case of the maltose transporter, the TMD component MalF is observed to consist of eight TMHs. The remaining members of this class follow the mentioned distribution.

$()_2$ the parenthesis followed by the subscript two represents the same TMHs for both the TMDs.

1.4.2. Nucleotide-binding domains (NBDs)

The NBDs are the motor units of the ABC transport system and are responsible for facilitating the transport process by coupling with ATP hydrolysis for the transport of the ligand (Linton, 2007). The NBDs are reported to have high levels of sequence similarity and to possess multiple conserved motifs, which act as a hallmark of the ABC transporter superfamily. Multiple tertiary structures available for NBDs reveal a RecA-like domain and a helical domain (Figure 1.7A). The RecA domains of the NBDs are reported to harbor the highly conserved Walker A and Walker B motifs, while the helical domain possesses the ABC signature motif (LSGGQ), which is reported to be unique in NBDs (Figure 1.7B) (Higgins *et al.,* 1982; Higgins, 1992). The NBDs have been reported to form a symmetric dimeric conformation to be functionally active, and the dimerization of the two sub-units is suggested to be nucleotide-driven (Figure 1.7C) (Chen *et al.,* 2003). The structure of NBDs co-crystallized with ATP has revealed the involvement of both the protomers in holding the ATP molecules, wherein the one ATP molecule interacts with the Walker A motif of NBD1 and the Signature loop of the NBD2,

vice-versa for the second ATP (Figure 1.7D) (Oswald *et al.,* 2006). Walker B motif is reported to be essential for the ATPase activity of the NBD. The Walker B motif consists of a glutamate residue, which acts as a nucleophile and completes the catalysis. Further, the D-loop present in the NBDs has been reported to be involved in interaction with the TMDs and NBDs (Zaitseva *et al.,* 2005; Ambudkar *et al.,* 2006).

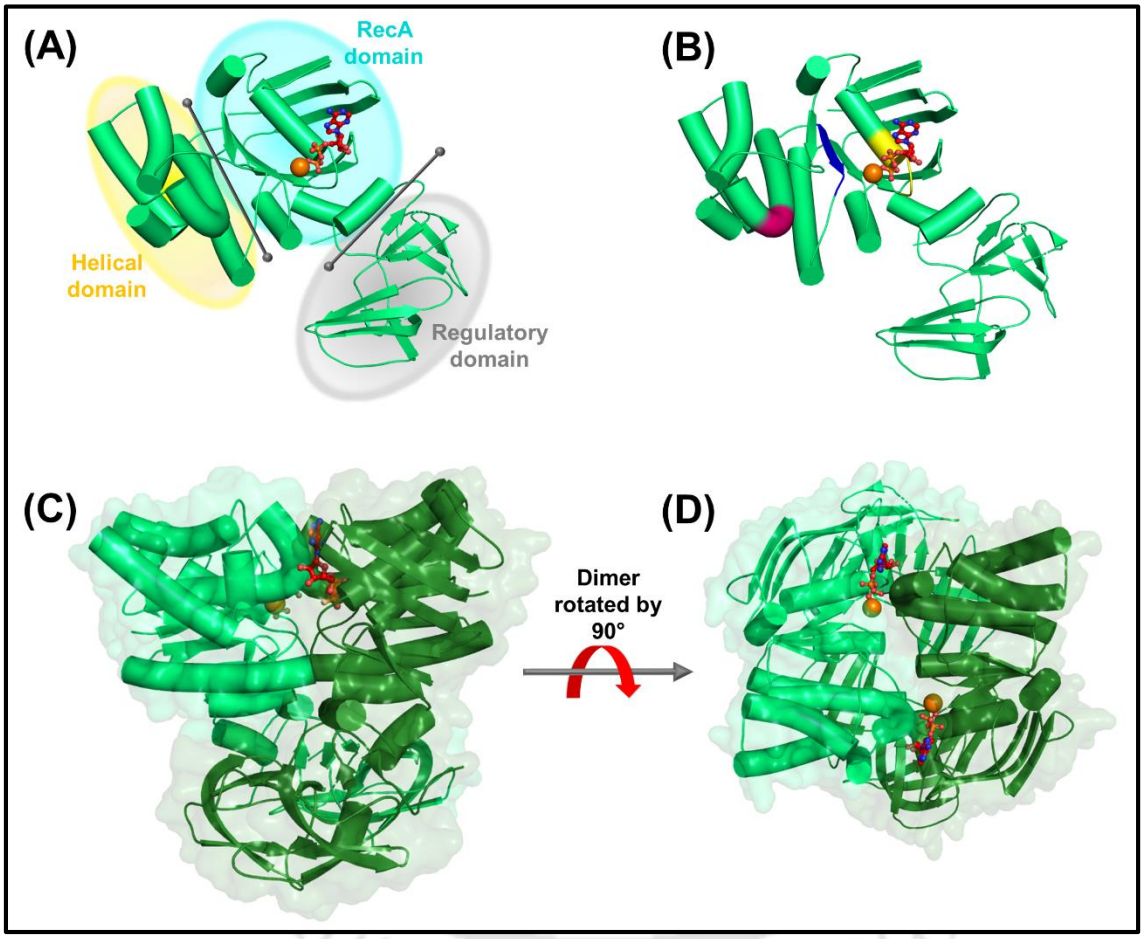


Figure 1.7. Overall domain architecture and ATP-binding of NBDs. (A) Domain arrangement of the monomeric unit of the maltose transport system NBD MalK (PDB id: 3PUV). MalK is divided into three distinct domains: helical (yellow), RecA (cyan), and regulatory domain (an additional domain present in MalK, grey). (B) Conserved motif distribution in the monomeric unit of the NBD MalK. Walker A, B, and the ABC signature motif have been represented in yellow, blue, and pink, respectively. (C) A vertical view of the dimeric conformation of the NBD MalK, wherein the two protomers of NBD1 and NBD2 have been represented in light green and dark green, respectively. (D) Horizontal view of the dimeric

conformation of MalK after rotating 90°. The binding site of ATPMg^{2+} is visualized at the interface of NBD1 (light green) and NBD2 (dark green). In all the cases, the bound ATP molecule (red) has been represented with the help of a ball-and-stick model, and the Mg^{2+} (orange) is represented in the form of a sphere.

1.4.3. Substrate (solute)-binding protein (SBP)

The ABC importers in prokaryotes consist of an extra-cytoplasmic protein termed the SBP; it is essential for sequestering the ligands and carrying them to the membrane components (TMD-NBD hetero-complex) for their transport. The SBPs are responsible for the specificity and the selectivity of the ABC importers (Maqbool *et al.,* 2015). The SBPs in the Gram-negative bacteria are free-floating in the periplasmic space and hence are also referred to as periplasmic-binding proteins (PBPs). Structurally, the SBPs are divided into N- and C-terminal domains (NTD and CTD), which are α/β domains, wherein the β-sheets are flanked by α-helices. The NTD and CTD are linked by the hinge region, which, in accordance with its name, allows the conformational transitions in the NTD and CTD of the SBP, enabling the entrapment of the ligand (Fukami-Kobayashi *et al.,* 1999). The SBPs are observed to exist in two different forms, namely the 'open unliganded state' and the 'closed liganded state' (Figure 1.8A). In the open unliganded state, the NTD and the CTD are very flexible, and free rotations are observable around the hinge region (Figure 1.8B). However, upon ligand binding, the SBP attains closed conformation wherein the hinge region stabilizes the NTD and CTD (Figure 1.8B) (Tang *et al.,* 2007). During the transition from the open to a closed state, the hinge region allows NTD and the CTD to move towards each other, leading to the closed conformation. This simultaneous movement of both domains to capture the ligand has been coined as the 'Venus fly-trap' mechanism (Figure 1.8C) (Mao *et al.,* 1982). Along with the Venus fly-trap mechanism, recently, the SBPs have been reported to perform multiple ligand-binding mechanisms such as 'asymmetric domain movement' (NTD demonstrates greater movement than CTD), 'one domain movement' (only the NTD moves and the CTD is static), and the 'subdomain movement' (subdomains of NTD (N1 and

N2) and CTD (C1 and C2) demonstrate different movements) (Pandey *et al.,* 2016; Chandravanshi *et al.,* 2020a,b).

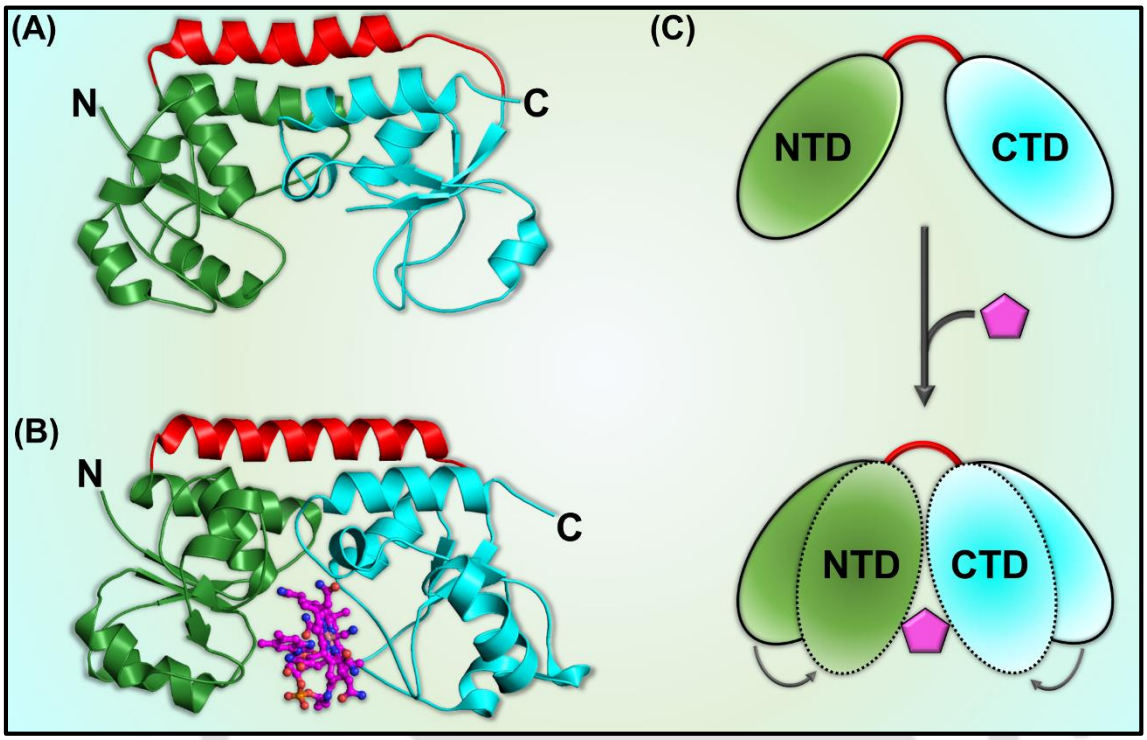


Figure 1.8. Structural topology and binding mechanism of SBP. The tertiary structure of the SBP Vitamin B_{12}-binding protein in (A) ligand-free (PDB id: 1N4D) and (B) ligand-bound (PDB id: 1N4A) conformations. The bound ligand (magenta) has been represented in the form of a ball-and-stick model. The NTD, CTD, and hinge regions of the SBP have been represented in dark green, cyan, and red, respectively. (C) Schematic representation of the Venus fly-trap mechanism, wherein the NTD, CTD, and hinge region have been represented in dark green, cyan, and red, respectively. The ligand (magenta) has been represented in the shape of a pentagon. The conformational changes in the NTD and CTD on ligand binding have been represented in the form of dotted lines.

1.4.3.1. SBP classification

Based on their topology, the SBPs were initially classified into three different classes (I, II, and III) (Fukami-Kobayashi *et al.,* 1999; Lee *et al.,* 1999). Amongst the three classes, classes I and II were observed to differ based on their hinge region and their β-sheet arrangement. Class I follows the β_2-β_1-β_3-β_4-β_5

arrangement, while class II follows the β_2-β_1-β_3-β_n-β_4 (n is the β-strand which crosses over between NTD and CTD) arrangement (Fukami-Kobayashi *et al.,* 1999). Further, the hinge of class I SBPs is constituted of three connecting strands, while class II is made of two connecting strands. The hinge region of the class III SBPs was observed to consist of one α-helix, which functions as the hinge of these SBPs (Karpowich *et al.,* 2003).

Following this classification, multiple SBPs were structurally elucidated, and multiple structures were available from different classes of SBPs. Interestingly, despite poor sequence similarity between these SBPs, they were observed to highly conserve their three-dimensional structural folds (Berntsson *et al.,* 2010). Hence, this led to a new classification wherein the SBPs were classified based on their structural folds into six different classes (A-F) (Figure 1.9) (Berntsson *et al.,* 2010). Within a few years, an updated classification of SBPs was proposed, wherein a total of 504 structures were collected and classified into seven distinct classes (A-G) (Figure 1.9) (Scheepers *et al.,* 2016). Further, in the updated classification, certain clusters such as A, B, D, E, and F were classified into sub-clusters. All the clusters (A-G) consisted of topologically unique SBPs, wherein the variation in the hinge region is identified as the major differentiating feature for each cluster (Figure 1.9). Recently, the SBP classification was again updated, wherein 764 unique entries were inspected, resulting in new sub-clusters and a novel cluster H (Figure 1.9) (Chandravanshi *et al.,* 2021).

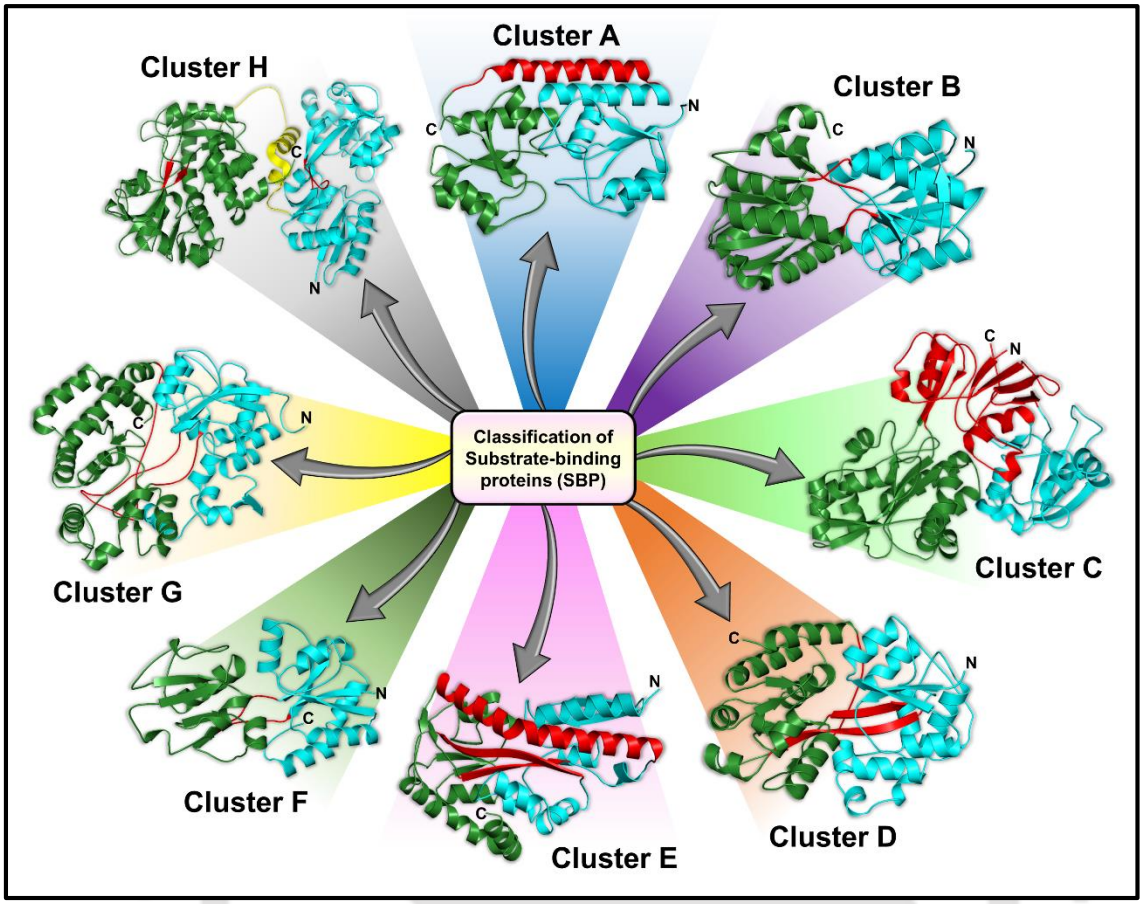


Figure 1.9. Classification of SBPs based on structure. The differences in the structural attributes of the SBPs, primarily observed in their respective hinge regions, allowed them to be classified into eight distinct clusters as follows: Cluster A (blue, PDB id: 1PSZ), cluster B (purple, PDB id: 4WZZ), cluster C (light green, PDB id: 4QFK), cluster D (orange, PDB id: 3TTM), cluster E (pink, PDB id: 4O94), cluster F (dark green, PDB id: 4QHQ), cluster G (yellow, PDB id: 1Y3P), and cluster H (grey, PDB id: 1JNF). In all the SBPs, the NTD and CTD are represented in cyan and dark green, respectively, while the hinge region that connects them is represented in red. The additional linker peptide, which connects the NTD with the CTD in cluster H, is represented in yellow.

1.5. ABC transporters involved in resistance against AMPs

The ABC transporters of Gram-negative pathogens have been suggested to play a pivotal role in their resistance against the AMPs. An instance of the ABC transport system is the MacAB-TolC exporter, which has been suggested to confer resistance against the peptide C18G in *Salmonella sp.* (Honeycutt *et al.,*

2020). The ABC exporter forms a tripartite assembly between MacAB and TolC, wherein MacB hydrolyses ATP, which enables the transport of the molecules from the periplasm to the extracellular space with the aid of MacA and TolC (Crow *et al.,* 2017). The most prevalent role of ABC transporters in AMP resistance is pertaining to the import of AMPs. Two ABC import systems, namely, the Yej and Sap transport systems, have been reported to function as AMP import systems (Parra-Lopez *et al.,* 1993; Eswarappa *et al.,* 2008).

1.5.1. Yej transport system

Akin to ABC importers, the Yej ABC transport system consists of an SBP (YejA), two TMDs (YejB and E), and a NBD (YejF) (Eswarappa *et al.,* 2008). On deletion of the components of the Yej system (Δ*yejE* and Δ*yejABEF*) in *Brucella melitensis*, it rendered the organism sensitive to the action of polymyxin B in contrast to the wildtype (Table 1.2). Further, both the mutants were observed to demonstrate retarded invasion and replication within macrophages (Wang *et al.,* 2016b). Moreover, a knockout mutation of the YejF component of the *Salmonella sp.* Yej transport system further substantiated the significance of this system, wherein the mutant was observed to demonstrate enhanced sensitivity against polymyxin B, melittin, protamine, and HBD-1 and -2 (Table 1.2) (Eswarappa *et al.*, 2008). However, despite the reports, a proper idea of the fate of the AMPs proposed to be internalized via the Yej transport system remains vague.

1.5.2. Sap transport system

This ABC importer was first observed in *S. typhimurium* wherein, from random mutant library-based screening, certain mutants were observed to be susceptible to the AMP protamine. On further analysis of the susceptible mutants, components from an ABC importer were observed and hence were coined as sensitivity to antimicrobial peptides (Sap) transport system (Groisman *et al.,* 1992; Parra-Lopez *et al.,* 1993). The *S. typhimurium* Sap (*St*Sap) transport system has been reported to possess an operonic arrangement, where the authors observed the system to consist of an SBP component (SapA), two TMDs (SapBC), and two NBDs (SapDF) (Parra-Lopez *et al.,* 1993). A potential mechanism was also proposed suggesting that *St*SapA will directly bind the AMPs and ferry them to its membrane components (*St*SapBCDF). It was further

proposed that the internalized AMP may be subjected to proteolytic cleavage via the bacterial proteases (Parra-Lopez *et al.,* 1993).

The genetic arrangement of the non-typeable *H. influenzae* Sap (*Hi*Sap) transport system was determined wherein, along with the other components, an additional component *Hi*SapZ, was reported to belong to the *Hisap* operon. Further, results demonstrated the active participation of the *Hi*sap operon in the process of non-typeable *H. influenzae* (NTHI) infection in the host cell and suggested the *Hi*SapA component necessary for sustenance within the host (Mason *et al.,* 2005). Further, utilizing biochemical strategies, the interaction of *Hi*SapA with chinchilla β-defensin-1 was established (Mason *et al.,* 2006). Thereafter, based on multiple strategies, Mason and co-workers successfully provided direct evidence of human β-defensin-3 and LL-37 uptake via the Sap transport system. The authors utilized cellular fractionation to analyze the uptake of the targeted AMPs. Further, the authors observed that the uptake was inhibited when *Hi*SapA was mutated, which signifies that the *Hi*Sap transport system was involved in the uptake of AMPs (Shelton *et al.,* 2011). Moreover, via multiple strategies, the significance of *Hi*SapBC in the functioning of the system was established. The proteolytic cleavage of the internalized AMPs via elegant biochemical strategies was also successfully demonstrated in the study (Shelton *et al.,* 2011).

The Sap transport system of *H. ducreyi* (*Hd*Sap) demonstrated selectivity wherein the system was reported to impart resistance only against the AMP LL-37 and not against the β-defensins. Further, an interesting aspect observed in the *Hdsap* operon was the absence of the NBD component SapF from the operon, suggesting the complex to function even in the absence of the SapF component (Mount *et al.,* 2010). Apart from *H. ducreyi* multiple other Gram-negative bacteria such as *K. pneumoniae*, *D. dadantii*, and *Actinobacillus pleuropneumoniae* have been reported to possess the Sap transport system and with the aid of which they are reported to gain resistance against LL-37, α-thionin, and PR-39 respectively (López-Solanilla *et al.,* 1998; Xie *et al.,* 2017; Hsu *et al.,* 2019). Hence, all the literature suggests the direct involvement of the Sap transport system in providing resistance against AMPs and, in turn, enhancing the sustenance of the bacterial pathogens within the host.

1.5.2.1. Multifunctional attribute of the Sap transport system

The Sap transport system has been suggested to possess multiple functional aspects apart from imparting resistance against AMPs. Alongside the feature of providing resistance against AMPs, the *Hi*Sap transport system has been attributed to be involved in the uptake of heme, and its NBD component (*Hi*SapD) is reported to help TrkH-TrkA in the uptake of K^+ ions (Mason *et al.,* 2006*,* 2011). In the case of *K. pneumoniae,* the Sap transport system has been suggested to play a key role in the interaction and adhesion of *K. pneumoniae* in the host systems (Hsu *et al.,* 2019). The *Aliivibrio fischeri* Sap (*Af*Sap) transport system has been implicated as being involved in quorum sensing and has also been suggested to play a role in bioluminescence regulation. Interestingly, the *Af*Sap transport system has not been suggested to be involved in providing resistance against any AMP, unlike other Sap transport systems (Chen *et al.,* 2000). Apart from heme and K^+ ion uptake, the Sap transport system of *Shewanella oneindensis* (*So*Sap) has been reported to uptake oligopeptides and is suggested to fulfill the nutritional needs of the bacterium (Liang *et al.,* 2019). Akin to *Hi*SapD, the NBD component (SapD) of the *E. coli* Sap transport system (*Ec*Sap) is observed to be involved in K^+ ion uptake (Harms *et al.,* 2001). Surprisingly, the membrane components (*Ec*SapBCDF) of the *Ec*Sap transport system have been observed to be involved in the export of putrescine. The *Ec*SapBCDF components were reported to function in concert as an ABC exporter, and *Ec*SapA was banished from being a part of the *Ecsap* operon and was suggested to be an orphan SBP (Sugiyama *et al.*, 2016). Therefore, the literature suggests multifunctional aspects of the Sap transport system and implies that the Sap transport system is of great prevalence in each case. However, the functional aspects of the *Ec*Sap transport system, primarily that of *Ec*SapA, remain an enigma that needs to be unraveled.

1.6. Rationale of the study

The literature pertaining to the Sap transport system establishes its role in resistance against AMPs. However, certain studies suggest that the system moonlights and performs multiple crucial strategies for the bacterial systems, such as heme, K^+ ions, and oligopeptide uptake. Surprisingly, the *Ec*Sap transport system is reported to function as a putrescine exporter, banishing the

SBP component SapA as unrelated. Further, the crystal structure of *Hi*SapA was elucidated and was observed to endogenously bind with an RNA molecule at a location distant from the putative binding-site of *Hi*SapA. The authors suggested the binding-site of *Hi*SapA to be very constricted and not suitable for AMP-binding. They challenged the earlier reports of AMP-binding with *Hi*SapA, asserting that the interactions observed were likely nonspecific, resulting from the binding of AMP to the RNA molecule bound to *Hi*SapA (Lukacik *et al.,* 2021). Hence, the reports pertaining to the *Hi*Sap and *Ec*Sap transport systems diverted the functional prospects of the Sap transport system into ambiguity.

Recently, the three-dimensional structure of *Hi*SapA in open conformation was reported, wherein the protein was devoid of any endogenously bound RNA molecule. Further, the authors did not observe the binding-site of *Hi*SapA constricted and, via biophysical and computational strategies, confirmed the interaction of human β-defensins with *Hi*SapA (Rivera *et al.,* 2024). Hence, the recent report enlightened the functional prospects of the *Hi*Sap transport system. However, the functional attributes of the *Ec*Sap transport system yet remain untapped, with multiple unanswered questions (Figure 1.10).

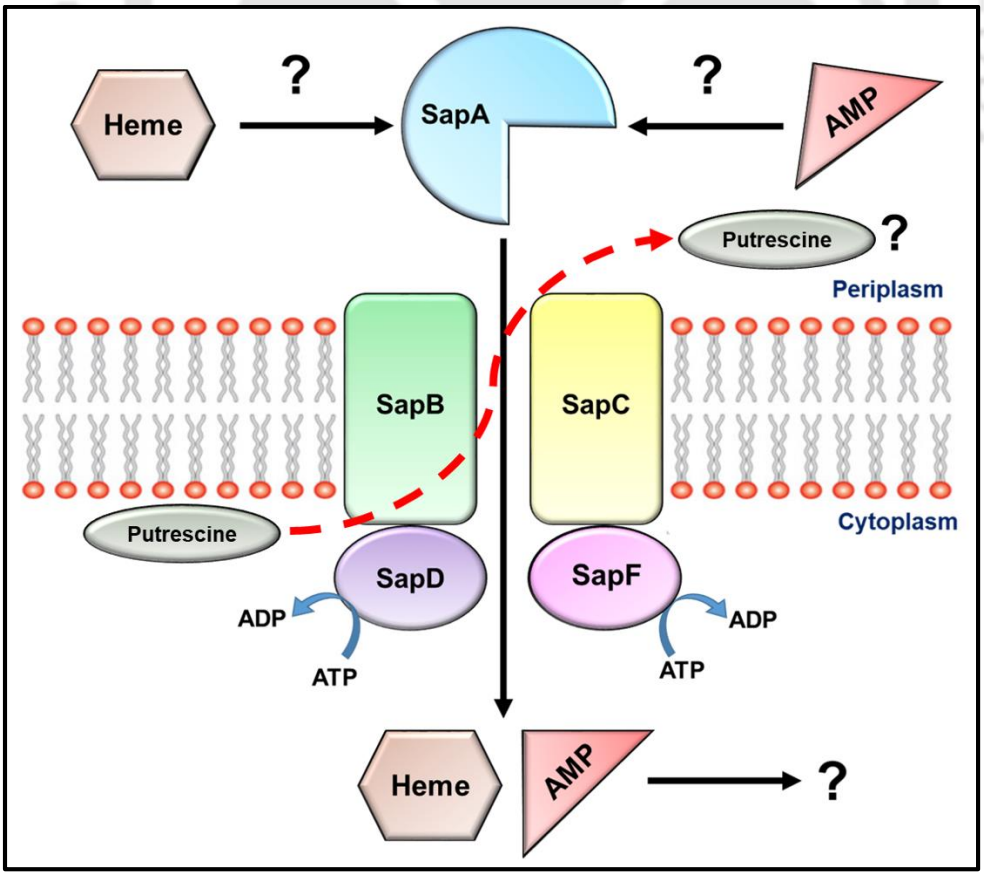


Figure 1.10. Schematic representation of the *Ec*Sap transport system. The putative functions implicated in the system have been represented, followed by interrogation marks that highlight the ambiguity surrounding the system.

1.7. Objectives

To gain insight into the enigmatic nature of the *Ec*Sap transport system, we have devised a set of objectives as follows:

1. Elucidation of the structural aspects of SapA from the *E. coli* Sap transport system.

2. Functional and mechanistic insights into the *E. coli* Sap transport system.

To address the above-stated objectives, the following strategies would be applied:

a. *In silico* characterization of all the Sap transporter proteins of *E. coli*.
b. Molecular cloning, overexpression, and protein purification of *Ec*SapA.
c. Crystallization of *Ec*SapA.
d. Biophysical characterization of *Ec*SapA to determine the specific ligand and their respective kinetics.
e. Molecular dynamics simulations of *Ec*SapA to unravel the mechanistic insights into the binding of the ligands.

Chapter 2

Materials and Methodology

General outline of Chapter 2

This chapter elaborates upon the robust computational strategies devised to elucidate the putative functional prospects of the *Ec*Sap transport system. Further, the chapter also delves into the methods employed for experimental validation of the outputs of the computational strategies. A general framework of the methodologies employed has been represented in the form of a flowchart in Figure 2.1.

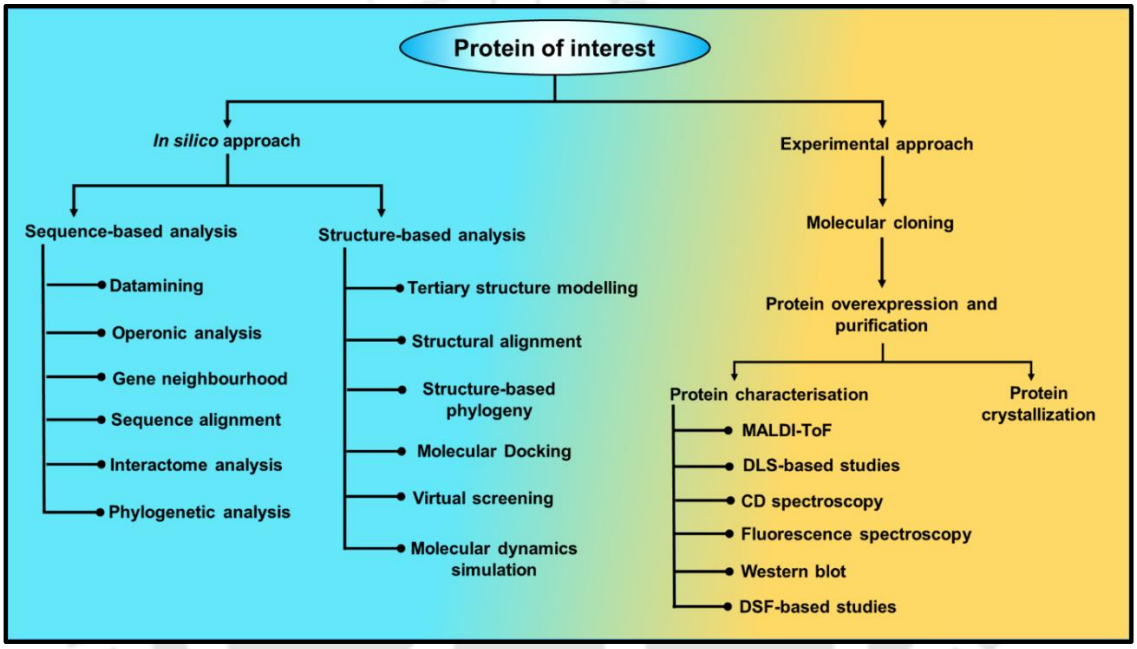


Figure 2.1. A flowchart of the methodologies employed in the thesis to accomplish the proposed objectives.

2.1. Materials

2.1.1. Reagents

The essential enzymes for molecular cloning (restriction enzymes, shrimp alkaline phosphatase, and T_4 DNA ligase) were retrieved from New England Biolabs. Further, various kits required for molecular cloning were sourced from various reputable suppliers, such as plasmid isolation kits from GCC Biotech (India) and QIAGEN (Germany), PCR clean-up kits, and gel extraction kits from QIAGEN (Germany). To purify our protein of interest, Ni-NTA (nickel-nitrilotriacetic acid) agarose affinity resin procured from QIAGEN (Germany) was

packed in a Pierce centrifuge column procured from Thermo Fischer Scientific (USA). The packed column yielded purified protein fractions, which were then pooled and dialyzed utilizing a semipermeable dialysis membrane obtained from HiMedia (India). The dialyzed protein fractions were further concentrated utilizing Vivaspin Turbo 15 protein concentrators from Sartorius (Germany). Various crystallization buffers, microbatch, and hanging-drop plates, along with other chemicals and tools pertinent to protein crystallization, were procured from Hampton Research (USA) and Molecular Dimensions (UK). For Western blotting experiments, the primary antibody (Anti-His_6) and secondary antibody (Anti-rabbit anti-His_6) were procured from Sigma Aldrich (USA). In addition, nitrocellulose membrane (0.45 µm) and the remaining required setup were purchased from Bio-Rad (USA). Further, the dye SYPRO orange required for ligand-binding studies was procured from Thermo Fischer Scientific (USA).

2.1.2. Ligand molecules

The ligands hemin and protamine sulfate were procured from Sigma Aldrich (USA).

2.2. Methods

2.2.1. Sequence- and structure-based analysis

Various web tools and strategies that enabled our computational endeavor have been elaborated upon in the forthcoming sections.

2.2.1.1. Data-mining of Sap components and *sap* operon analysis

All the available Sap transporter components were retrieved by querying the UniProt KB database (The UniProt Consortium, 2023). The entries obtained were inspected for redundancy, and the unique entries were further shortlisted. The obtained components were utilized to scrutinize their operonic arrangement utilizing the graphic view available in the nucleotide database of NCBI (National Center for Biotechnology Information). The operonic arrangements were tabulated and segregated into classes as per arrangement. The intergenic distance of two adjacent genes (e.g., A and B) were estimated using the formula: [((Genetic starting coordinate of B)-(Genetic ending coordinate of A))-1]. The

formula ascertains the presence and conservation of overlaps amongst the genes, which is a key attribute of the operons (Assaf *et al.,* 2021). Further, the genetic neighbourhood of the data-mined operons was inspected, and five genes upstream and downstream were noted with the help of the nucleotide database of NCBI. Further, to validate the existence of certain operons, web tools Operon Mapper (Taboada *et al.,* 2018) and Operon Finder (Tomar *et al.,* 2023) were utilized. Whenever required, the promoter of the operons was predicted using the web tool BPROM (Softberry) (Salamov and Solovyevand, 2011).

2.2.1.2. Sequence-based analysis

2.2.1.2.1. Homologous protein identification and sequence alignment

From the data-mined dataset, the *Ec*Sap transporter components were shortlisted for further analysis. The sequences of our protein of interest were utilized to identify homologous proteins. The basic local alignment search tool (BLAST; Altschul *et al.,* 1990) program was used to identify the homologous entries, wherein the sequences are queried against the non-redundant (nr), UniProt/SwissProt, and protein data bank (PDB) databases. The parameters utilized for protein BLAST are as follows: General parameters (max target, 100; expected threshold, 10; and word size, 6) and scoring parameters (BLOSSUM62 scoring matrix and Gap costs, existence: 11 and extension: 1). The outputs of the homology search are enlisted and analyzed based on sequence identity and query coverage to gauge the similarity amongst the sequences. Further, the homologs are subjected to Pairwise and multiple sequence alignment (PSA and MSA) with the help of EMBOSS Needle (Rice *et al.*, 2000) and Clustal Omega (Sievers and Higgins, 2014), respectively, keeping the default parameters. Whenever required, the MSA outputs were rendered using the web tool ESPript 3.0 (Easy Sequencing in PostScript; Gouet *et al.,* 2003).

2.2.1.2.2. Protein-protein interactome analysis

For the functioning of the Sap transport system, all its components need to interact with each other. A protein-protein interactome analysis was utilized to map the interactions amongst the Sap components. The proteins of interest were subjected to protein-interactome analysis utilizing the database STRING v12 with a default set of parameters (Szklarczyk *et al.,* 2023). The STRING database

analyses the interactions based on eight parameters, which encompass available experimental data, database-based knowledge, genetic neighborhood, gene fusions, gene co-occurrence, text mining, co-expression of genes, and protein homology. Based on these parameters, the interactions are scored from 0 to 1. The higher the score, the more relevant the prediction. The protein-protein interactome map generated from the STRING database is further rendered utilizing Cytoscape 3.8.1 to enhance the visual clarity of the interactome network (Shannon *et al.,* 2003).

2.2.1.2.3. Phylogenetic tree-based analysis

All the existing proteins are the outcome of numerous events of evolution. Phylogeny enables us to map the lines of evolutionary descent of most evolutionary events nucleating from a common ancestor. The evolutionary relatedness of the *Ec*Sap components with their homologs and all other data-mined Sap entries were mapped with the help of phylogenetic strategies. The phylogenetic trees, whenever required, were generated using the MEGA X (Molecular Evolution and Genetic Analysis, Kumar *et al.,* 2018). The following four steps were followed for the generation of a phylogenetic tree:

(1) All the sequences selected for phylogenetic analysis were retrieved from suitable databases and enlisted.

(2) These sequences were then aligned via ClustalW embedded in MEGA X using a default set of parameters, wherein the gap opening and closing penalties were set to 10 and 0.2, respectively.

(3) Utilizing the aligned sequences, a tree was generated using the maximum likelihood method present in the MEGA X software. The bootstrapping method was utilized to estimate the significance of the tree generated, wherein 1000 bootstrap replicates were generated. The bootstrapping replicates provide confidence in the formed clusters.

(4) For enhanced visual clarity, the phylogenetic trees generated were visualized in the web tool iTOL v6.3 (Letunic and Bork, 2021).

2.2.1.3. Structure-based analysis

2.2.1.3.1. Tertiary structure prediction

The tertiary structure of our protein of interest was predicted using servers employing homology/ threading/ *ab initio* modeling strategies. The tertiary structure prediction tools RaptorX (Källberg *et al.,* 2012), SWISS-MODEL (Biasini *et al.,* 2014), and Phyre 2.0 (Kelley *et al.,* 2015) were utilized for homology-based modeling, wherein the crystal structure of a related protein is utilized by the server to model the available protein sequence. Further, the web-server ITASSER (Yang *et al.,* 2015), which employs threading-based modeling, was also utilized. In threading/ fold recognition modeling, sequences similar to the query protein are searched in a protein structure database. Similar sequence-possessing structures are listed as templates in the databases. The template structures are then used to thread together the predicted tertiary structure of the query protein. The most recent modeling sever AlphaFold (Varadi *et al.,* 2022) has also been utilized, which employs *ab initio* modeling strategies combined with homology modeling. The AlphaFold server employs deep learning techniques that enable them to predict the tertiary structure of the protein solely based on its amino acid sequence and protein folding principles. However, Alphafold's smart algorithm enables them to utilize the principles of homology modeling if the template structure is available. All the models generated were refined utilizing the web tool GalaxyRefine (Ko *et al.,* 2012). Further, the geometrical and stereo-chemical parameters of the initial models and the refined models were validated with the help of the PDBsum database (Laskowski *et al.,* 2018). The validation of the parameters led us to the models with the optimal geometrical and stereo-chemical properties, which were finalized for further analysis. Hence, the rigorous tertiary structure prediction strategy employed here enabled us to finalize the optimal theoretical models of our protein of interest.

2.2.1.3.2. Data retrieval of AMPs

The AMP molecules, as discussed in Chapter 1, possess distinct folds akin to proteins. Hence, the tertiary structures of AMPs, whenever available, were retrieved from the Protein Data Bank (PDB, Berman *et al.,* 2000). However, in cases where only the sequences of the AMPs are available, the tool Avagadro

(Hanwell *et al.,* 2012) was employed, wherein the models of the AMPs were generated. The sequence of the AMPs was retrieved from the UniProt KB database (The UniProt Consortium, 2023).

2.2.1.3.3. Structural comparison with homologous proteins

The structural homologs enable insights into the structural conservation of patterns that aid in predicting the putative functional attribute of the query protein. The structural homologs of the *Ec*Sap components were retrieved from the PDB database (Berman *et al.,* 2000). The calculation of the root mean square deviation (rmsd) provides a measurable attribute to compare the similarities between the two structures, wherein the lower the value, the more similar the structures. Hence, the rmsd amongst the homologs and our protein of interest were estimated utilizing PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC) and the program 3dSS (Sumathi *et al.,* 2006).

2.2.1.3.4. Structure-based classification of SBPs

The three-dimensional structures of multiple SBPs are available in the PDB database. These SBPs share low sequence similarity but high structural similarity amongst each other. The slight differences in the architecture of these proteins were utilized to classify the SBPs into eight distinct clusters (A-H) (Chandravanshi *et al.,* 2021). A similar structure-based classification was performed to analyze the cluster to which *Ec*SapA belongs. The tertiary structures from all the sub-classes of SBPs were retrieved from the PDB database (Berman *et al.,* 2000). The rmsd amongst all the SBPs collected from the different clusters was estimated using a home-built Python script, which was run utilizing the program PyMOL. The script enlists the rmsd of all the SBPs in the form of a matrix. This matrix was further utilized to generate a structure-based phylogenetic tree to study the structural diversity amongst the SBPs being studied. The web tool dendroUPGMA (Garcia-Vallvé *et al.,* 1999) was utilized to build the structure-based phylogenetic tree. The matrix was arranged in a suitable format and introduced to the web tool, which converts the rmsd matrix into a Newick (.nwk) file format. The obtained .nwk file was visualized in MEGA X software to generate the phylogenetic tree (Kumar *et al.,* 2018). For better clarity, the generated tree

was visualized in the web tool iTOL v6.3, wherein the final tree was generated (Letunic and Bork, 2021).

2.2.1.3.5. Domain movement analysis of SapA proteins

Multiple SBPs demonstrate a high degree of movement amongst their N- and C-terminal domains (NTD and CTD). The movement leads to two distinct conformations, namely open and closed. Hence, the predicted tertiary structure of all the SapA proteins has been modeled in both open and closed conformations. Since the binding site of SapA is occluded in the closed conformation, the open conformation models would be required for certain analyses. The domain movement between the closed and open conformation of SapA was estimated utilizing the web tool DynDom (Taylor *et al.,* 2014). The DynDom server calculates the translational movement of the domains and also enlists the residues present in the hinge axes, which leads to hinge bending, ultimately leading to the open and closed conformations of the SBPs. Further, the server also predicts the rotational angle of the domains of the SBPs.

2.2.1.3.6. Electrostatic potential charge distribution of proteins

The electrostatic potential charge of a protein plays a key role in its functional as well as structural aspects. The charge of the binding site of targeted proteins might provide a clue about their putative ligand. Further, the surface charge of the target proteins might hint towards the plausible assembly machinery of key protein complexes. The electrostatic potential charge distribution of our protein of interest was calculated using the program Adaptive Poisson-Boltzmann Solver (APBS) embedded in PyMOL (Baker *et al.,* 2001). The calculated electrostatic potential charge distribution for each protein has been color-coded red (negatively-charged surface) and blue (positively-charged surface) for visual clarity. The intensity of the blue and red colors is on the basis of a scale ranging from +1 kT/e to -1 kT/e, respectively.

2.2.1.3.7. Binding-site volume estimation of SBPs

The domain movement of the NTD and CTD leads to the closed conformation of the SBPs. The void space encapsulating the ligand in the closed conformation

provides insight into the probable size of the putative ligand of the SBP. The binding site volume was calculated to estimate the available space in the binding site of *Ec*SapA and other peptide-binding SBPs. The binding-site volume of our target proteins was estimated utilizing the web tools POCASA 1.1 and CASTp 3.0 with a default probe radius of 1.4 Å and 2.0 Å, respectively (Yu *et al.,* 2010; Tian *et al.,* 2018). The output of both servers was averaged and represented in the form of histograms utilizing the software OriginLab (OriginLab Corporation, USA). All the models were visualized and analyzed using the program PyMOL.

2.2.1.3.8. Topological arrangement of TMDs

The TMDs of ABC importers consist of membrane pass helices, also known as transmembrane helices (TMHs). The arrangement and the number of the TMHs are utilized for the TMD classification of the ABC transporter superfamily. The arrangement of *Ec*SapBC in the bacterial inner membrane was analyzed using the PPM 3.0 web server integrated within the OPM (orientation of proteins in the membrane) database (Lomize *et al.,* 2022). The topological arrangement of *Ec*SapBC has been determined by utilizing the deep learning algorithm-based web server MemBrain 3.0 (Feng *et al.,* 2020).

2.2.2. Receptor-ligand interaction calculations

The affinity of various plausible ligands of our selected proteins was mapped with the aid of molecular docking using the program AutoDock v4.2 (Morris *et al.,* 2009). The three-dimensional atomic coordinates of the putative ligands, such as dipeptides and hemin, were retrieved from PubChem (Kim *et al.,* 2023) and the ChemSpider database (Pence and Williams, 2010). All the retrieved ligands were arranged to form a ligand library. Prior to each molecular docking calculation, hydrogen atoms were added to the protein. Subsequently, the partial charges were assigned to the proteins using the Gasteiger charge algorithm (Gasteiger and Marsili, 1980). For the blind docking simulation of all the 400 dipeptides, a grid box of grid size 126x126x126 with a grid spacing of 0.484 Å was generated, considering the center of mass of the protein as the grid center. For each molecular docking simulation, the protein molecule was kept rigid while the ligand was considered flexible along the rotatable bonds. For each molecular docking simulation, a total of 2000 runs of the Lamarckian genetic algorithm (LGA) were

performed. After each molecular docking calculation, the best-docked poses were analyzed, and the atomic interactions between the docked ligand and the protein were identified using the program Coot v0.9.4 (Emsley *et al.,* 2010). The molecular interactions were visualized using the program PyMOL.

To perform protein-protein docking of AMPs, an open conformation of the selected SapA proteins was predicted using various three-dimensional structure prediction programs, as mentioned previously. The degree of movement and the flexible domain in the open and closed conformations of the SapA proteins were estimated using the web tool DynDom (Taylor *et al.,* 2014). The web tool ClusPro v2.0 was utilized for the protein-protein docking calculations (Desta *et al.,* 2020) with the default set of parameters. The SapA proteins were considered as receptors, while the AMPs were considered as ligands. For each docking simulation, the highest-ranked clusters were considered for further analysis. The molecular interactions of the protein-ligand complexes were analyzed utilizing the PDBsum database (Laskowski *et al.,* 2018). The figures were generated using the program PyMOL.

2.2.3. Molecular dynamics (MD) simulation

MD simulations are widely utilized *in silico* strategy that attempts to mimic the biological environment of biomolecules to map their distinctive motions. The simulations enable the capturing of the conformational dynamicity and landscape of the proteins of interest in atomic details (Hollingsworth and Dror, 2018). GROningen MAchine for Chemical Simulations (GROMACS 2021.1) (Abraham *et al.,* 2015) was employed to perform all the MD simulation experiments. GROMACS utilizes Newtonian equations of motion for simulating biomolecular systems consisting of proteins, nucleic acids, and lipids. The steps of simulating a protein system involve a series of steps wherein the topological parameters of the protein are generated from the PDB file. Further, in the case of complexes, the ligand topologies were generated using the CGenFF server (Vanommeslaeghe and MacKerell, 2012). Post-generation of topology, the protein/ complex was centered within a dodecahedron box that was filled with transferable intermolecular potential with a 3-points (TIP3P) water model (Jorgensen *et al.*, 1983). In all the cases, the system was simulated using the

CHARMM36 all-atom force field (Huang and MacKerell, 2013). The system was neutralized by adding sodium or chloride atoms whenever needed. The periodic boundary condition (PBC) was applied in all three spatial directions to avoid the edge effect of the boxes during simulation. All the systems were equilibrated in two parts, viz. NVT (Isochoric-isothermal) and NPT (isobaric-isothermal). Both the equilibration steps (NVT and NPT) were generated one after the other for 1000 ps at 310 K. The equilibrated system was simulated for 1 µs. The simulated outputs were further analyzed utilizing home-built shell scripts and programs available in GROMACS. The graphs generated were visualized via the program Xmgrace (Turner, 2005).

2.2.4. Recombinant construct for protein overexpression

The most lucrative strategy in molecular biology is the process of molecular cloning, which is an ensemble of experimental protocols enabling a combination of distinct nucleotide fragments to generate a novel recombinant DNA molecule. To initiate the molecular cloning process, the desired genome housing our gene of interest was manually isolated. The gene of interest was amplified from the isolated genome. Following this, the amplified gene of interest and a suitable vector were double digested via the restriction endonucleases. Further, the digested vector and the gene were ligated with the help of the ligase enzyme. The recombinant construct generated post-ligation was replicated by introducing the ligated product into a host to produce multiple copies of the clone (Figure 2.2.). Each step of molecular cloning will be elaborated on in the forthcoming sections.

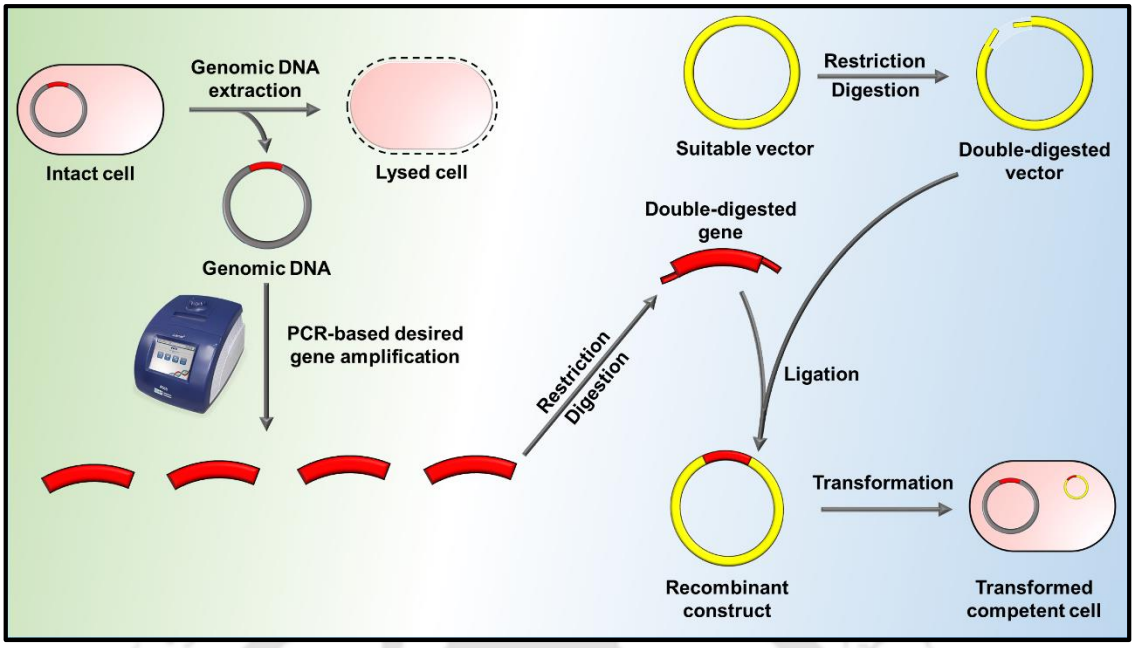


Figure 2.2. Schematic representation for molecular cloning.

2.2.4.1. Competent cell preparation for cloning of gene and protein overexpression

Various strains of *Escherichia coli* have been used to prepare competent cells for molecular cloning and protein overexpression. The cells were rendered competent using the calcium chloride ($CaCl_2$)-based chemical method. Fresh bacterial cells from their log phase were resuspended in chilled 100 mM $CaCl_2$ solution and incubated for 30 min at 4°C. Post incubation, these cells were resuspended in a solution consisting of 100 mM $CaCl_2$ and 15% Glycerol. The resuspended homogenate consisting of the competent cells was subjected to flash freezing via liquid nitrogen and stored at -80°C for future use. The strains of *E. coli* utilized in this study are manipulated for enhanced expression of desired proteins, wherein they possess a chromosomal copy of the T_7 RNA polymerase, which is under the regulation of the T_7 promoter. For overexpression of our desired proteins, the expression cells are induced using Isopropyl β-d-1-thiogalactopyranoside (IPTG), which is a lactose analog and binds to the repressor.

The below-described strains of *E. coli* have been utilized in this study to perform molecular cloning, protein overexpression, and solubility optimization.

- DH5α competent cells: Possess high transformation efficiency, hence employed for enhancing the copy numbers of the cloned plasmid.
- BL21(DE3) competent cells: These cells were utilized for overexpression of desired proteins using an inducer (IPTG), wherein the gene is regulated by the T_7 promoter.

2.2.4.2. Gene amplification

To amplify our desired gene of interest, a set of primers was designed considering parameters such as restriction digestion sites, GC content, and melting temperature. The web tool OligoAnalyzer (IDT technologies, USA) was utilized to analyze and verify all the parameters of the primers designed. Further, to amplify our gene of interest, the genome of *E. coli* BL21(DE3) was manualy isolated. The isolated genome was used as a template and was subjected to a polymerase chain reaction (PCR).

The steps of the PCR reaction encompassed an initial denaturation at 95°C for 10 min, followed by three subsequent steps: 1) denaturation (95°C, 45 sec), 2) annealing (at the T_m of the primers, 1 min), and 3) extension (72°C, 45 sec). These three steps were set up in a loop of 30 cycles. After the completion of 30 cycles, a final extension step was performed at 72°C for 10 min. The amplification of the genes was inspected in 0.8% agarose gel. The amplified products obtained contain additional primer additions alongside the gene of interest.

2.2.4.3. Double digestion of desired vector and gene of interest

The directional cloning strategy was employed, wherein the amplified gene product and the desirable vector (pET28a(+)) were double digested using suitable restriction enzymes. The restriction enzymes (such as *Nde*I, *Xho*I, and *Bam*HI), which generate sticky ends at the 5′ and 3′ of the gene, were selected. Both the vector and the gene of interest were separately incubated with the same set of restriction enzymes and incubated at 37°C for 3 hours. Post the incubation, the vector was subjected to treatment via alkaline phosphatase, wherein the vector reaction mixture was further incubated for 1 hour at 37°C. The alkaline phosphatase removes the 5′-phosphate group of the vector, hence avoiding self-

ligation of the vector. After the completion of incubation, both the samples were heated at an adequate temperature in accordance with the enzyme for 20 mins to inactivate the restriction enzymes. Further, the double-digested vector and the gene were purified using gel extraction (QIAGEN, Germany) and PCR purification kit (QIAGEN, Germany), respectively.

2.2.4.4. Ligation of the gene of interest and clone confirmation

The final step in molecular cloning is a ligation reaction, which is performed to insert our gene of interest into its compatible digested vector pET28a(+). The two fragments of DNA are ligated by a reaction catalyzed by the enzyme T_4 DNA ligase. The enzyme forms covalent bonds between the insert (gene of interest) and the vector. To enable this ligation reaction, the vector and the insert possessing sticky ends were incubated at varying vector: insert ratios (1:1, 1:2, 1:3, 1:4, and 1:5) along with T_4 DNA ligase for 16 hours at 16°C. Post incubation, the ligation mix was utilized to transform *E. coli* DH5α competent cells. The *E. coli* DH5α cells were utilized to amplify the ligated recombinant clone for further use. These transformed cells were plated onto Luria Bertani (LB) agar plates supplemented with antibiotics. The colonies obtained were further screened by inoculating the obtained colonies into LB broth supplemented with the suitable antibiotic and incubating the culture at 37°C for 12 hours. Further, this culture was utilized to isolate the cloned plasmids. The cloned plasmids isolated were further subjected to double digestion using the same set of restriction enzymes as previously used and incubated at 37°C for 2 hours. The digested products were analyzed using 0.8% agarose gel and confirmed. For further confirmation, the isolated cloned plasmid was subjected to DNA sequencing.

2.2.5. Protein overexpression, solubility and purification

2.2.5.1. Overexpression of selected protein

The obtained clone in pET28a(+) of our protein of interest was used to transform *E. coli* BL21(DE3) expression cells. The transformed cells were plated onto LB agar supplemented with a suitable antibiotic. The colonies obtained were inoculated in LB broth supplemented with a suitable antibiotic and were used to analyze the overexpression of our protein of interest. The overexpression analysis of our protein of interest was optimized at 37°C, wherein different

samples were retrieved post-induction at different time intervals (1 to 7 hours). Further, IPTG is used as an inducer, which leads to cell death at higher concentrations; hence, IPTG concentrations (0.1, 0.5, and 1.0 mM) were also optimized. Once the optimum temperature, expression time, IPTG concentration, and expression host were optimized, the overexpressed protein was analyzed for its solubility.

2.2.5.2. Solubility of selected protein

The *E. coli* BL21(DE3) cells were transformed with our clone construct. The transformed cells were cultured in optimal conditions determined through prior overexpression optimizations. After the desired duration of incubation, post-induction, the cells were harvested by centrifugation and resuspended in a resuspension buffer. The homogenized solution was subjected to sonication with 30% amplitude for 15 min, with a sonication cycle of 2 s on and 6 s off. Post-sonication, the cell lysate was centrifuged at ~18500g for 45 min to separate the soluble and insoluble fractions. The soluble and insoluble fractions obtained were further analyzed with the aid of 10% SDS-PAGE. When our protein of interest is observed in the soluble fraction in SDS-PAGE analysis, they are subjected to protein purification using chromatographic strategies.

2.2.5.3. Affinity chromatography for protein purification

In our cloning strategies, we have incorporated a C-terminal hexa-histidine tag (6X-His) into our recombinant construct. This 6X-His tag, which is present with our protein of interest, enables it to form coordination with various transition metals (Co^{2+}, Ni^{2+}, Cu^{2+}, and Zn^{2+}). This property of the interaction of the 6X-His tag with transition metals is utilized in the immobilized metal affinity chromatography (IMAC) method. In this study, we have utilized Ni^{2+}-charged beads, wherein the Ni^{2+} ions are coated onto a bead. Hence, when the proteins are passed through these beads, the proteins possessing the 6X-His tag get immobilized and separated from other non-specific proteins. A pierce centrifuge column was packed with Ni-NTA (nitrilotriacetic acid) beads and equilibrated with the desired buffer. Post-equilibration, the soluble fraction containing our protein of interest was allowed to bind to the Ni-NTA resins. Following this, the non-specific proteins were washed off with wash buffers consisting of varying

concentrations of imidazole (10 mM to 50 mM). Along with imidazole, other reagents such as salt (NaCl), protease inhibitors like phenylmethanesulphonyl fluorides (PMSF), and reducing agents (β-mercaptoethanol) were added to the wash buffers to reduce non-specific host proteins. After adequate volumes of wash, our protein of interest was eluted out at higher concentrations of imidazole (250 mM to 400 mM) present in the elution buffer. The high concentration of imidazole disrupts the coordination between Ni^{2+} and the 6X-His tag present in the bead and our protein of interest, respectively. The imidazole competitively binds the Ni^{2+} beads and hence leads our target protein (with the 6X-His tag) to elute out. The quality and purity of the eluted proteins were inspected with the help of 10% SDS-PAGE. The elutes obtained were pooled together and dialyzed to remove imidazole from the protein. Imidazole was removed in a gradient manner and gradually removed completely from the protein. The dialyzed elutes were concentrated via the protein concentrators to a desired concentration, which was further utilized for the characterization of the protein.

2.2.6. Crystallization trials of *Ec*SapA

Two major approaches were used to obtain *Ec*SapA protein crystals: (1) vapor diffusion and (2) microbatch-under-oil. To obtain diffractable crystal, initial screening was performed at two different temperatures (4 and 20 ºC) in both microbatch-under-oil and hanging drop methods. Once the initial hits were obtained, further optimization was performed by separating the two major events of crystallization, i.e., nucleation and crystal growth, via microseeding methods and adding multiple additives (polyethylene glycol, low melting agar, etc.). In microseeding, submicroscopic crystals were prepared and used as seeds to be introduced into equilibrated new drops. Further, the conditions in which the crystals were obtained have been optimized by varying the concentration of the components and further adding additives to enhance the crystal growth rate. Once the single and suitable crystals were obtained, they were exposed to a monochromatic X-ray beam to get the diffraction pattern. However, despite obtaining multiple mountable crystals, we failed to obtain a well-diffracting crystal for *Ec*SapA despite robust optimizations.

2.2.7. Protein characterization of *Ec*SapA

2.2.7.1. Matrix-assisted laser desorption/ ionization-time of flight (MALDI-ToF)

MALDI-ToF is a key strategy utilized to estimate the accurate molecular weight of proteins. Purified proteins are mixed with a suitable matrix in varying protein: matrix ratios (1:1, 1:2, and 1:3). The protein matrix mix is then spotted onto the ground steel plate and ionized with the help of a laser. The laser charges the matrix, which in turn enables the protein coated with it to travel toward the oppositely charged detector. The time of flight of the protein molecule to reach the detector is monitored, and it enables the system to detect the molecular weight of the sample being tested. Our protein of interest has a molecular weight higher than 20 kDa, and hence, the most suitable matrix is Sinapinic acid. The Sinapinic acid crystals were solubilized in a solution of 30% acetonitrile and 0.1% trifluoroacetic acid mixed in a 30:70 ratio, respectively, to attain a final matrix concentration of 10 mg/mL. Further, the Sinapinic acid matrix was subjected to sonication for 30 mins to dissolve the undissolved crystals of Sinapinic acid. The matrix post-sonication is centrifuged at 12000g for 20 min prior to use. The obtained matrix was mixed with varying concentrations of *Ec*SapA (5 µM, 10 µM, 15 µM, and 20 µM) and was mixed with the freshly prepared matrix in a protein: matrix ratio of 1:1, 1:2, and 1:3. These samples were allowed to air dry on a ground steel plate, and further, the dried spots were ionized in a Bruker autoflex speed (Bruker Daltonics, Germany) machine. The software Flex control was utilized to analyze of the raw data.

2.2.7.2. Size exclusion chromatography (SEC)

SEC is a widely utilized strategy that employs the varying migration attributes of molecules through gel matrices tightly packed within a column. The smaller molecules are capable of passing through the beads and, therefore, migrate at a slower rate than larger molecules. This enables efficient segregation of molecules based on size. The proteins are detected based on their absorbance at 280 nm (A_{280}) as they elute in a peak mode through the column. Consequently, when discrete populations of proteins existing in diverse oligomeric states traverse the column, the outcome manifests as multiple peaks, each indicative of a distinct oligomeric state. To inspect the homogeneity of *Ec*SapA, it was injected into a

HiLoad™ 26/600 Superdex™ 200 pg column (GE Healthcare, USA), which has been attached to an ÄKTA purifier (GE Healthcare, USA) for SEC-based analysis. Further, apart from the homogeneity of the sample, SEC-based analysis has also been employed to estimate the size of *Ec*SapA. To estimate the molecular weight of *Ec*SapA, the elution volume (V_e) of *Ec*SapA was utilized to estimate the distribution coefficient (K_{av}) of *Ec*SapA using the following equation:

$$K_{av} = \frac{V_e - Vo}{V_t - Vo}$$

Where V_o is the void volume of the column used, and V_t denotes the total column volume. The molecular weight of *Ec*SapA was estimated by plotting the *Ec*SapA K_{av} onto a standard graph, which is plotted in a K_{av} (distribution of standard proteins) versus Log M_w (the log of molecular weights of standard proteins) graph.

2.2.7.3. Native-polyacrylamide gel electrophoresis (PAGE)

The PAGE in its native form enables proteins to migrate solely based on their native charge, which the proteins possess at pH 8.3. This strategy, unlike SDS-PAGE, is not adept at estimating molecular weight or checking the purity of a sample. Rather, native PAGE provides insight into the homogeneity and oligomeric nature of any protein sample. Purified *Ec*SapA was mixed with bromophenol blue consisting loading dye (devoid of SDS) and was loaded onto an 8% native-PAGE and allowed to migrate for 2 hours at 100 V. Post-run, the gel was stained using Coomassie brilliant blue dye, and the bands were visualized.

2.2.7.4. Dynamic light scattering (DLS)

The homogeneity and quality of the protein were inspected with the help of DLS. DLS enables the detection of aggregates in purified protein samples. The DLS data unravels the hydrodynamic diameter based on the scattering of the light and hence estimates the homogeneity of any given sample. In this study, all the DLS-based studies were performed with the aid of LITESIZER 500 (Anton Paar, Austria). The protein sample to be analyzed is diluted to 2 µM in nuclease-free water. The freshly diluted sample is utilized to estimate the hydrodynamic

diameter of the protein. Protein samples tend to aggregate with increasing temperatures. Hence, to estimate the stability of the purified protein, it was subjected to increasing temperatures (20°C to 90°C), and the DLS data was recorded with a temperature interval of 10°C. The recorded data was further visualized and analyzed with the help of the program Origin v9.6.

2.2.7.5. Circular dichroism (CD)

CD spectroscopy is one of the most commonly utilized biophysical strategies that enlightens the secondary structural composition of any protein. All the CD-based studies were performed in a JASCO J-1500 spectrophotometer (JASCO, Japan), wherein the samples were added into a quartz cuvette with an optical path length of 1 mm and scanned in the far-UV range (190 nm to 260 nm). The scan in the far-UV range encompasses various signature peaks demonstrating the presence of protein secondary structural elements like α-helices, β-sheets, and random coils. In this study, CD spectroscopy was employed to unravel the secondary structural composition and thermal stability of *Ec*SapA. The purified *Ec*SapA was diluted to 3 µM in NFW prior to the scan. The Peltier present in the JASCO J-1500 spectrophotometer enabled us to increase the temperature from 20°C to 90°C with a heat elevation rate of 2°C min^{-1}. The variations in the secondary structural elements with the elevation of temperature were analyzed. Further absorbance at 222 nm was mapped at each temperature and further utilized to estimate the melting temperature of *Ec*SapA.

2.2.7.6. Differential absorption spectroscopy (DAS)

To analyze the interaction of a porphyrin ring-containing molecule (Hemin) with *Ec*SapA, a difference absorption spectroscopy (DAS)-based experiment was performed. The shift in the porphyrin ring spectra on interaction with protein is observed in DAS. The hemin stock solution was prepared by dissolving in 100 mM NaOH. The stock solution was further diluted to a working condition using the *Ec*SapA dialysate buffer. For DAS-based analysis, an equimolar concentration (25 µM) of hemin was mixed with *Ec*SapA and was incubated at room temperature for 15 mins. The samples were scanned from 300 nm to 600 nm with the help of an infinite M200PRO plate reader (TECAN, Switzerland). A blank experiment was also performed simultaneously, wherein the spectra of

hemin solely were recorded. The final data was obtained by subtracting the free hemin spectra from the protein-hemin spectra.

2.2.7.7. Electrophoretic mobility shift assay (EMSA) coupled Western blot

EMSA-based studies inspect the variation in the migration of protein molecules in the presence of a ligand molecule. The interaction between the protein and ligand might enhance the migration or retard the migration of the protein through the native PAGE, hence suggesting an interaction between the two. In this study, 80 µM of *Ec*SapA was mixed with varying concentrations of hemin (50 µM to 300 µM) and incubated at room temperature for 15 mins. The incubated samples were loaded onto an 8% native PAGE and allowed to develop for 2 hours at 100 V. The gel developed was visualized by staining it with Coomassie brilliant blue stain.

A Western blotting experiment was devised to confirm the bands observed in the above gel. Western blotting or Immunoblotting enables specific detection of proteins immobilized on a nitrocellulose membrane with the aid of antibodies. Two sorts of antibodies are employed wherein the first antibody, known as the primary antibody, detects protein-specific epitopes and binds to them. Further, the secondary antibody detects the primary antibody and binds to it. These bound antibodies may be then visualized via multiple strategies such as chemiluminescence, fluorescence, and autoradiography. In this study, the proteins are transferred from a native PAGE. Hence, the gel was incubated in the transfer buffer supplemented with 0.04% SDS in a rocking condition. Post incubation, the gel was sandwiched with the nitrocellulose membrane with the help of filter papers pre-soaked in the buffer. The sandwich was introduced into a cassette, within which voltage was applied, and the membrane was kept towards the positive electrode and the gel towards the negative electrode. Hence, a charge-based transfer was facilitated at 4°C for 24 hours at 40V. Post-transfer, the blank regions without the proteins were blocked with the help of skimmed milk and then incubated with the primary antibody (mouse anti-His_6) for 16 hours at 4°C. Post incubation, the membrane was washed with tris buffer saline (TBS) supplemented with 0.1% Tween 20 to wash away unbound primary antibodies. After adequate rounds of washing, the membrane was incubated with the

secondary antibody (rabbit anti-mouse IgG) for 1 hour at room temperature, followed by washing with TBS supplemented with 0.1% Tween 20. After adequate washing, the bands in the membrane were visualized via chemiluminescence. The substrate of horseradish peroxidase (HRP) conjugated to the secondary antibody was introduced onto the membrane, and the bands were detected with the help of imaging equipment. The primary and secondary antibodies were used at 1:3000 and 1:5000 dilution, respectively. Further, the software ImageJ (Schneider *et al.,* 2012) was utilized whenever the bands obtained in the immunoblot were densitometrically analyzed.

2.2.7.8. Fluorescence spectroscopy

All protein molecules possessing buried aromatic amino acids, such as tryptophan, tyrosine, and phenylalanine, possess the property of intrinsic fluorescence. This intrinsic fluorescence property of proteins is used for analyzing the structural alterations in proteins and for binding studies with ligands. A widely known quencher of fluorescent signals is water molecules, and hence, whenever the buried aromatic residues get exposed to water, it leads to an alteration in the fluorescence spectra of the protein. In this study, the variation in the structural environment of 5 µM of *Ec*SapA was mapped in the presence of increasing concentrations of urea (1 M to 8 M) and the reducing agent β-mercaptoethanol (5 mM). The samples were excited at 285 nm, and the emission was recorded from 300 nm to 600 nm.

Further aromatic amino acids present in the binding site of *Ec*SapA interact with their ligand. Hence, the presence of the ligand would lead to a variation in the environment of the aromatic amino acids. Thus, leading to a difference in their fluorescence spectra. This attribute was utilized in this study to perform a hemin-binding study of *Ec*SapA, wherein 5 µM of *Ec*SapA was incubated for 30 mins with incremental concentrations of hemin (2 µM-30 µM) at room temperature. The fluorescence spectra were recorded following the above method. The raw data obtained were further analyzed with the help of the software Origin v9.6 to estimate the binding kinetics.

2.2.7.9. Differential scanning fluorimetry (DSF) analysis

The thermal denaturation of proteins provides multiple insights into the protein. The DSF-based study utilized this particular attribute of proteins to map the binding of compounds. The study is performed with the aid of the SYPRO orange dye (Thermo Fischer Scientific, USA), wherein the dye interacts with the exposed hydrophobic patches of the denatured protein and fluoresces. The thermal stability of the protein molecules is expected to increase in the presence of their cognate ligands. The analysis was performed in an Applied Biosystem 7500 real-time PCR system (Thermo Fischer Scientific, USA). The reaction was set up wherein 5X SYPRO orange dye was mixed with 10 µM *Ec*SapA. To the protein-dye mix, 100 µM protamine sulfate was added and incubated for 15 mins. Following incubation, the samples were introduced into the instrument and further analyzed. The data obtained were plotted on a relative fluorescence unit (a.u) versus temperature (°C) graph, and the shift in the T_m in each case was monitored.

Chapter 3

Sequence-based analysis of *sap* operonic arrangement

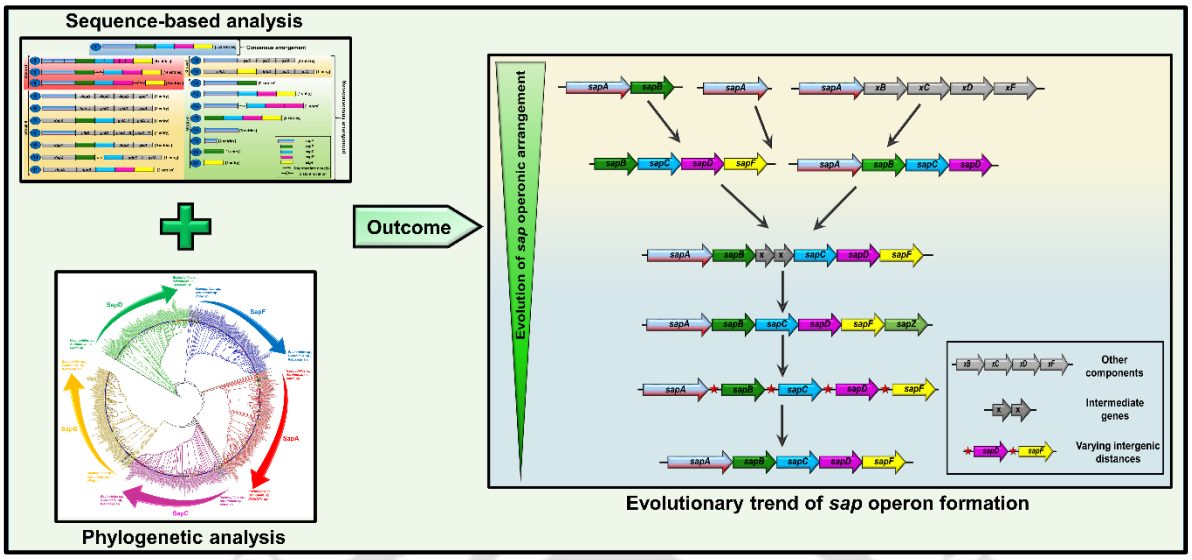
This chapter has been published as:

1. **Dasgupta P**, Vinil K, and Kanaujia SP (2024). Evolutionary trends indicate a coherent organization of *sap* operons. *Res. Microbiol.*, 175(8), 104228.

Abstract

Human hosts possess a complex network of immune responses that safeguards them against hordes of microbial pathogens. One such machinery is the production of antimicrobial peptides (AMPs), which target the pathogen cell membranes and inhibit them from inhabiting the hosts. However, pathogens have been found to employ systems that encounter these host-produced AMPs. One such system is known as the Sap (sensitivity to antimicrobial peptides) transporter, which is responsible for the uptake of AMPs inside the microbial cell for its subsequent proteolytic degradation. The Sap transporters comprise five subunits encoded by genes as part of an operon at the genetic level. The Sap transporter is prevalent in Gram-negative pathogens. Despite its ubiquitous nature, its subunits are not found to be in tandem with many organisms. In this study, a total of 421 Sap transporters were analyzed for their operonic arrangement. Out of 421, a total of 352 operons were found to be in consensus arrangement, while the remaining 69 operons show a varying arrangement of genes. Further, the analysis of the intergenic distance between the subunits of the *sap* operon suggests a signature pattern with *sapAB* (-4), *sapBC* (-14), *sapCD* (-1), and *sapDF* (-4 to 1). An evolutionary analysis of these operons favors the consensus arrangement of the Sap transporter systems, substantiating its prevalence in most of the Gram-negative pathogens. Overall, this study provides insight into bacterial evolution favoring the maintenance of the apt genetic organization of essential pathogenic factors for their sustenance within the host.

Graphical abstract



3.1. Introduction

The human immune system is a multilayer system comprising components viable for clearing any potent pathogen from establishing an unsolicited niche. The immune system can be broadly divided into three distinct layers, viz. the anatomical barrier, the innate immune system, and the adaptive immune system as the first, second, and third line of defense, respectively (Turvey and Broide, 2010). At the interphase of the first and second lines of defense, antimicrobial peptides (AMPs) exist as a key component of defense against microbes (Wiesner and Vilcinskas, 2010). AMPs are a class of peptides with varying lengths in the range of 12 to 50 amino acids (Ciulla and Gelain, 2023). AMPs are known to kill pathogens by disrupting their cell membrane homeostasis. The mode of action of AMPs varies, including the disruption of the membrane homeostasis, penetration, and inhibition of essential cellular machinery, ultimately leading to the death of the pathogen (Seyfi *et al.,* 2020).

However, several Gram-negative bacterial pathogens have evolved various mechanisms to evade the action of AMPs and thus become resistant to the host defense system. Among the various strategies adapted by microbial pathogens, modification of outer membrane charge, presence of an additional capsular layer, downregulation of AMP production, proteolytic degradation of AMPs by surface

proteases, and efflux or influx of AMPs across the bacterial membrane via transporters are prevalent (Gruenheid and Le Moual, 2012). Amongst the various transporters involved in the resistance against AMPs, members of the ATP-binding cassette (ABC) superfamily have been reported to be crucial. ABC transporters are classified into exporters and importers depending upon the direction of the substrate being translocated. Both the ABC exporters and importers possess two transmembrane domains (TMDs, which enable the transport of the molecule across the membrane) and two nucleotide-binding domains (NBDs, which energize the transport via ATP hydrolysis). However, ABC importers require an additional component termed substrate-binding protein (SBP, which provides selectivity to the importer system and hence enables the selectively permeable nature of the bacterial membrane) (Rice *et al.,* 2014). The ABC exporters efflux out the AMPs, avoiding their effect on the cellular processes and/or hindering their accumulation near the cell membrane. On the other hand, ABC importers uptake AMPs inside the bacterial cell to ultimately degrade them by proteolytic activities (Joo *et al.,* 2016).

In *Salmonella enterica* serovar Typhimurium, for the first time, the components of an ABC importer were reported to provide resistance towards the AMP protamine (Parra-Lopez *et al.,* 1993). The system involved in this process was subsequently termed the Sap (sensitivity to antimicrobial peptides) transport system. The Sap transporter consists of five components, viz. SapA (an SBP), SapBC (TMDs), and SapDF (NBDs). Further, the authors analyzed the genetic arrangement of the Sap transport system and reported that the genes involved in the system are found in tandem, similar to other ABC transport systems (Parra-Lopez *et al.,* 1993). Further, in-depth studies of the non-typeable *Haemophilus influenzae* (NTHI) Sap transport system established the interaction of the SBP component SapA with human-derived AMPs such as cathelicidin LL-37 and β-defensin 3. Also, the study established the crucial role of SapBC in the uptake of the AMPs and further proteolytic degradation via bacterial proteases (Shelton *et al.,* 2011). Interestingly, other studies have reported the role of NTHI SapA in the uptake of heme molecules (Mason *et al.,* 2011). Similarly, NTHI SapD has been reported to be involved in potassium ion uptake by energizing the Trk system of NTHI (Mason *et al.,* 2006). Utilizing various molecular biological experiments, the

polycistronic and tandem arrangement of the NTHI Sap transporter components (SapABCDFZ) has been reported (Mason *et al.,* 2005). Several Gram-negative organisms such as *Dickeya dadantii, Aliivibrio fischeri, Escherichia coli*, *Proteus mirabilis*, *Klebsiella pneumoniae,* and *Shewanella oneidensis* were reported to possess the Sap transporter system with their Sap genetic arrangement akin to *S. typhimurium* and NTHI (López-Solanilla *et al.,* 1998; Chen *et al.,* 2000; Harms *et al.,* 2001; McCoy *et al.,* 2001; Hsu *et al.,* 2019; Liang *et al.,* 2019). However, several Sap transporter systems from different organisms have been observed to defy the canonical tandem arrangement of their components. For example, the component SapZ has not been observed to be present in all the Sap-containing organisms. The functional aspects of SapZ remain ambiguous, and its relation to the Sap transport system remains obscure. Further, in *Haemophilus ducreyi,* although four components (SapABCD) are found to be in tandem, the component SapF is found at a distant genetic location. Interestingly, despite the varying genetic arrangement, the Sap transport system of *H. ducreyi* was reported to provide resistance against cathelicidin LL-37 (Mount *et al.,* 2010). Notably, the Sap transport system of *E. coli* (*Ec*SapBCDF) has been reported to function as a putrescine exporter, and its SapA (*Ec*SapA) has an independent function (Sugiyama *et al.,* 2016). Further, an *in silico* study showed the transcription of *Ec*SapA and *Ec*SapBCDF under different promoters, suggesting a non-operonic arrangement of the system (Sugiyama *et al.,* 2016).

Such variations observed in the genetic arrangements of various *sap* operons intrigued us to obtain an in-depth understanding of the conservation of the Sap transport system. The conservation of the genetic arrangement of any operon indicates the evolutionary status of its genes. The crucial operons are found to undergo accelerated positive evolution, maintaining their genetic arrangement. On the other hand, the redundant and unrelated genes undergo deceleration, losing the race of evolution (Price *et al.,* 2006). Thus, we aspired to map the evolutionary trend for the Sap transporters. Hence, in this study, we data-mined all the Sap transport systems available in various databases and analyzed their genetic arrangements using *in silico* approaches.

3.2. Methodology

3.2.1. Data retrieval and analysis

All the existing *sap* operon components were extracted using keywords such as "Sap OR SapA OR SapB OR SapC OR SapD OR SapF" from the UniprotKB database (The UniProt Consortium, 2023). The search led to the accumulation of a large number of entries, which were downloaded and maintained in the form of a spreadsheet. Subsequently, these entries were manually filtered for non-redundant Sap transporter components from the unique genus and species, enlisting one component of each unique Sap transporter. The operonic arrangement of the collected entries was scrutinized utilizing the nucleotide database of NCBI (National Center for Biotechnology Information). The operons containing all five Sap components in tandem were referred to as 'consensus', while others as 'non-consensus'. To understand the operonic demarcation and its functional co-relatedness, five neighbouring genes immediately upstream and downstream of the consensus *sap* operon were also collected. The genetic coordinate, ORF (open reading frame) id, protein name, strand polarity, and intergenic distance between the genes of the consensus *sap* operons were collected for further analysis. The intergenic distance between two tandem genes (let us say A and B) was calculated using the following formula: [((Genetic starting coordinate of B) – (Genetic ending coordinate of A)) – 1]. To further confirm the identification of consensus *sap* operons, the web servers Operon Mapper (Taboada *et al.,* 2018) and Operon Finder (Tomar *et al.,* 2023) were utilized for the prediction of *Ecsap* operon. Further, the program BPROM (Softberry) was utilized to predict of promoter sites near the *Ecsap* operon (Salamov and Solovyevand, 2011). For this, the nucleotide sequences of the *Ecsap* operon, along with five neighboring genes upstream and downstream, were extracted from the nucleotide database of NCBI and subjected to the prediction of operons. For the identification and confirmation of non-consensus Sap entries, sequence-based homology analysis was performed using the web tool BLAST (Altschul *et al.,* 1990). Pair-wise sequence alignment was performed using the web tools EMBOSS Needle (Rice *et al.,* 2000) and BLAST. Multiple sequence alignment was performed using the program Clustal Omega (Sievers and Higgins, 2014). The details of the domain arrangement of entries, whenever required, were

extracted from the InterPro database (Mitchell *et al.,* 2019). The protein-protein interaction network of the entries was generated using the STRING v12.0 database, using a default set of parameters, and the top 10 interactors were considered for further analysis (Szklarczyk *et al.,* 2023). For visual clarity, the protein-protein interaction networks were rendered using the tool Cytoscape 3.8.1 (Shannon *et al.,* 2003). The tertiary structure models of entries, wherever required, were retrieved from the AlphaFold database (Varadi *et al.,* 2022) and visualized using the program PyMOL (PyMOL Molecular Graphics System, Schrodinger, LLC). The evolutionary trend of the *sap* operons was analyzed by generating phylogenetic trees using the statistical method of maximum likelihood available in the program MEGAX (Kumar *et al.,* 2018). All the phylogenetic trees generated were further validated by 1000 bootstrap replicates for analysis. The web-based tool iTOL v6.3 was utilized for the visualization of phylogenetic trees (Letunic and Bork, 2021).

3.3. Results

3.3.1. The repertoire of the Sap transporters

The search using the keywords mentioned in the methodology section led to the accumulation of 210749 entries. The redundant entries were removed by considering the unique genera and species, enlisting one component of each unique Sap transporter. This resulted in a total of 421 unique Sap transporter entries. Out of these 421, a total of 352 entries (~84%) contained all five *sap* components in tandem and thus are consensus operons, while the remaining 69 entries (~16%) were found to lack one or other components and thus are considered as non-consensus operons (Figure 3.1A). The consensus arrangement of *sap* operon was observed in the case of 83 unique genera majorly belonging to the class of γ-proteobacteria (Figure A1). In accordance with prior data, the results show that out of 421 organisms containing the *sap* operon, a total of 413 entries (~98%) belong to the Gram-negative bacteria. The remaining seven and one entries(y) belong(s) to the Gram-positive bacteria and eukaryotes, respectively (Figure 3.1B). Interestingly, out of 421, 228 (~54%) and 131 (~31%) entries belong to pathogenic and non-pathogenic organisms, respectively. The remaining 62 (~15%) entries cannot be classified into either type (Figure 3.1C).

The genetic coordinate, ORF (open reading frame) id, protein name, strand polarity, and intergenic distance between the genes of the consensus *sap* operons were collected for further analysis.

3.3.2. The genetic arrangement of *sap* operons reveals a conserved pattern of organization

The varying operonic arrangement of various *sap* operons intrigued us enough to embark on an analysis of the operonic arrangement of the Sap transport systems available across all domains of life. The analysis was initiated with the *E. coli sap* (*Ecsap*) operon. In *E. coli*, all the components (*Ec*SapABCDF) of the *sap* operon are found to be in tandem with either overlaps or small intergenic distances. This is in contrast to the previous report, where *Ec*SapA was suggested to be an independent component of the *Ecsap* operon (Sugiyama *et al.,* 2016). This was further verified using various operon prediction tools, which predicted all five components (*Ec*SapABCDF) of the *Ec*Sap transporter to be in the same operon. Additionally, BPROM, a bacterial promoter prediction tool (Salamov and Solovyevand, 2011), further predicted a Pribnow box upstream to the *Ec*SapA, indicating the operonic arrangement of the *Ecsap* operon components (Figure 3.1D). Interestingly, *Ec*SapBCDF has previously been suggested to function as a putrescine exporter, considering *Ec*SapA as an independent component regulated by a different promoter (Sugiyama *et al.,* 2016). However, our initial operon/promoter prediction and *sap* operonic trends (consensus operons) suggest all the components of the *Ecsap* operon (*EcsapABCDF*) to be transcribed via the same promoter (Figures 3.1A, 3.1D). Further, on visual inspection, the *Ecsap* operon was observed to follow a similar genetic arrangement as the consensus *sap* operons, hence substantiating the operonic nature of the *Ec*Sap system.

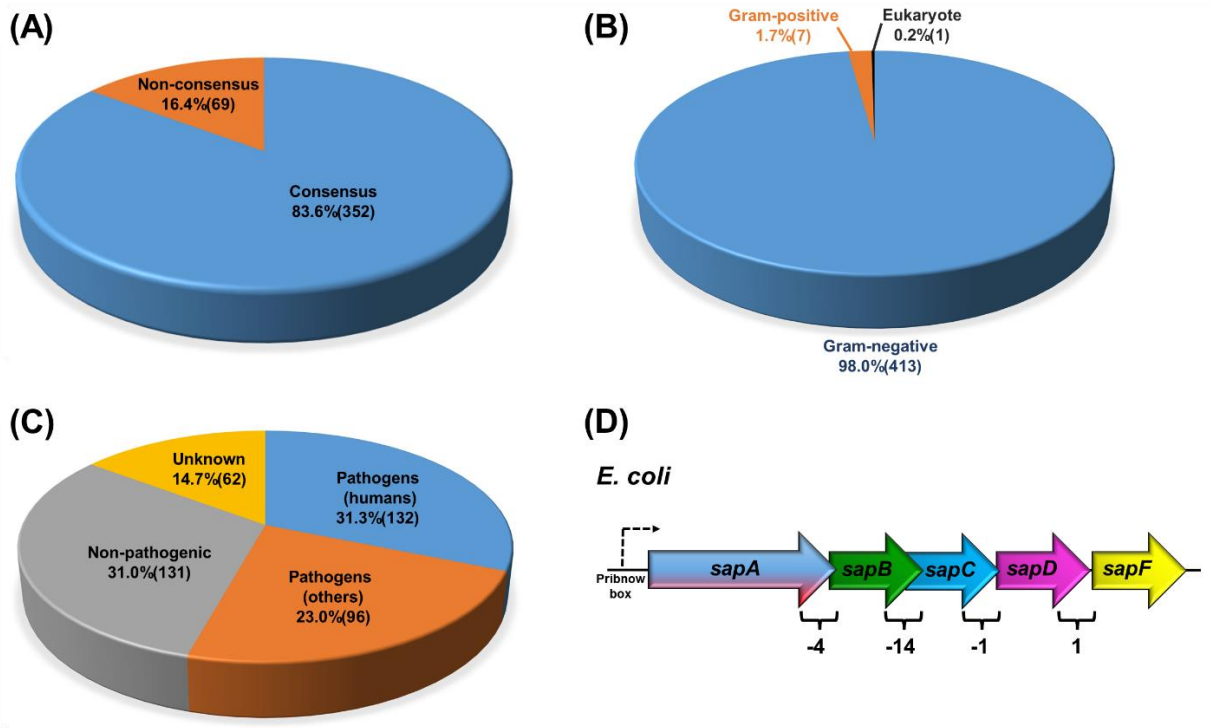


Figure 3.1. The distribution and genetic operonic construct of the Sap transporter components. The distribution of *sap* operons amongst (A) varying genetic arrangements (consensus/non-consensus), (B) domains of life (Gram-negative/Gram-positive/Eukaryotic), and (C) pathogenic/non-pathogenic organisms. The distribution in each case has been represented in the form of pie charts, and the details of each section of the pie have been mentioned within each section. (D) The predicted operonic arrangement of *Ec*Sap transporter components is wherein the gene name of each component is mentioned within the arrows. The dashed arrow upstream of *Ec*SapA represents the predicted Pribnow box. The numbers mentioned below in the braced brackets indicate the intergenic distances.

3.3.3. The consensus arrangement of the *sap* operon reveals a fingerprint-like intergenic distance pattern

One of the key features of an operonic arrangement is the conservation of the genetic arrangement and the presence of overlap amongst genes present in tandem (Assaf *et al.,* 2021). Although most of the reported *sap* operons display similar attributes, few *sap* operons lack these features. The intergenic distances

between each adjacent gene of the operon were calculated to analyze the conservation of the genetic arrangement. Interestingly, the intergenic distances amongst the components of the *sap* operon were found to be conserved to a great extent. The intergenic distance between *sapAB*, *sapBC*, *sapCD,* and *sapDF* was found to be -4 (~66%), -14 (~94%), -1 (~80%), and -4 to 1 (~87%), respectively (Figure 3.2).

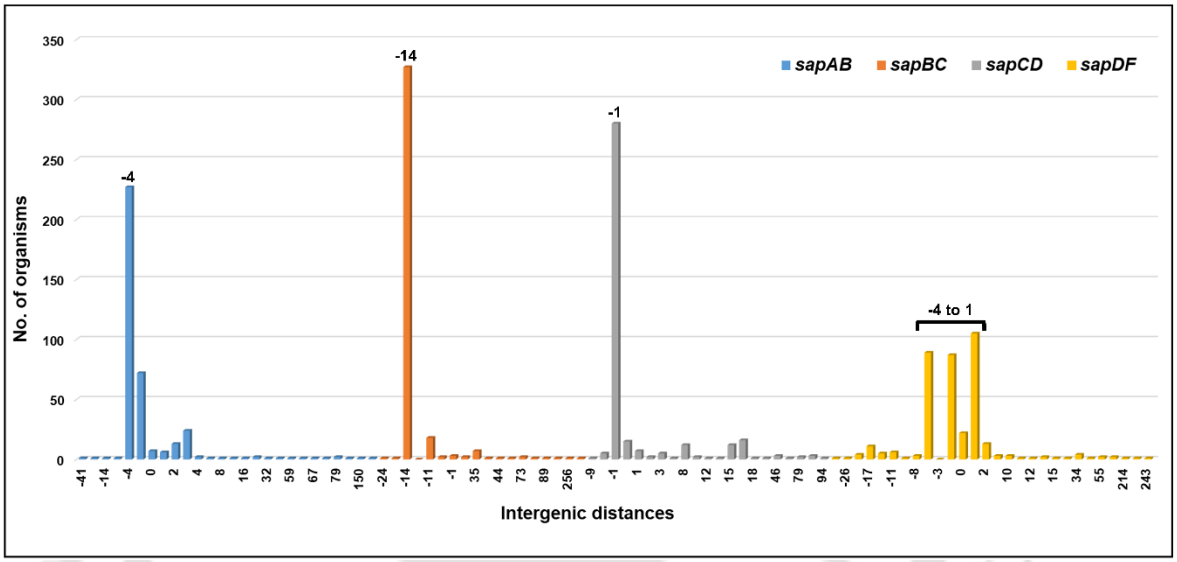


Figure 3.2. Intergenic distance distribution of consensus *sap* operons. The X-axis represents the varying intergenic distances present in between the components, while the Y-axis represents the distribution of the intergenic distances amongst the organisms possessing consensus arrangement of *sap* operons.

To further ascertain the signature intergenic distance pattern observed in the case of *sap* operons, we investigated all the ABC importers of *E. coli* strain K-12 available in the TransportDB 2.0 database (Elbourne *et al.,* 2017). The results revealed that out of the total 40 entries of *E. coli* ABC importers, none of the entries have an intergenic distance pattern similar to that of the *sap* operon (Table A1). This suggests that the intergenic distance pattern of the *sap* operon is unique in nature and can aid in identifying unknown *sap* operons.

Interestingly, the analysis of the gene neighborhood of the consensus operons further established the boundaries of the *sap* operonic system, having a substantial genetic distance between SapA & its upstream gene and SapF & its downstream gene. However, in *Haemophilus influenzae*, the Sap transport system, which contains an additional component, SapZ, lacks this feature. Notably, previous studies have established the *Hi*Sap system (*Hi*SapABCDFZ) to be transcribed in the form of a polycistronic mRNA (Mason *et al.,* 2005). Thus, the downstream genes of *sapF* of all 352 consensus operons were investigated to identify any hypothetical/putative membrane protein with a minimal intergenic distance. The analysis revealed that out of 352, a total of 16 consensus operons, including *H. influenzae*, contain an additional gene downstream of *sapF*, satiating the requirements stated above (Table A2). To annotate these downstream genes, they were subjected to a sequence-based homology search. The results suggest that none of these proteins show similarity with *Hi*SapZ. These observations suggest the exclusivity of each of the 16 genes to their respective operons and hence indicate their alien gene-like nature, implying random gene insertion events of operonic evolution (Fondi *et al.,* 2009). In summary, except for the 16 consensus *sap* operons, all other consensus *sap* entries possess substantial intergenic distances at both the termini, which hints towards a probable evolutionary trend in preserving the Sap transport system.

3.3.4. The gene neighborhood of *sap* operons hints towards a role in inner membrane maintenance

Previously, the Sap transport system has been demonstrated to be involved in the AMP as well as heme/dipeptide uptake (Mason *et al.,* 2011; Liang *et al.,* 2019). An analysis of the neighboring genes (five each of upstream and downstream) of the *sap* operon revealed the presence of two systems of phage shock proteins (Psp) and enoyl-acyl carrier protein (ACP) reductase in the upstream and downstream, respectively (Figure 3.3A). Interestingly, out of 352 entries, a total of 307 (~87%) contain the Psp system upstream of consensus *sap* operons (Figure 3.3B). The Psp system is reported to play a major role in the maintenance of bacterial inner membrane homeostasis (Flores-Kim and Darwin, 2016). The genetic arrangement of the Psp systems is in the form of an operon, wherein the Psp components (PspABCD) are present in tandem. Additionally, the

Psp system also possesses a transcriptional regulator PspF immediately upstream of the *psp* operon. (Figure 3.3A) (Flores-Kim and Darwin, 2016).

Further, an enzyme enoyl-ACP reductase (FabI) was observed to be present downstream of the 176 (~50%) *sap* operon entries (Figure 3.3A). The enzyme FabI is known to be involved in the fatty acyl-chain elongation step, which is essential for the formation of phospholipids present in the inner membrane of bacteria (Zhang and Rock, 2008). Notably, the presence of the enzyme FabI downstream of the *sap* operon was mostly observed in pathogenic bacteria (Figure 3.3C). This suggests compliance with the modes of action of the Sap transport system and Psp system, helping enhance the pathogenicity of the organism. Thus, the conservation of these proteins in the upstream and downstream of the *sap* operon suggests an involvement of the Sap transport system in the maintenance of the inner membrane homeostasis by avoiding the accumulation of AMPs in the bacterial periplasmic space (Shelton *et al.,* 2011).

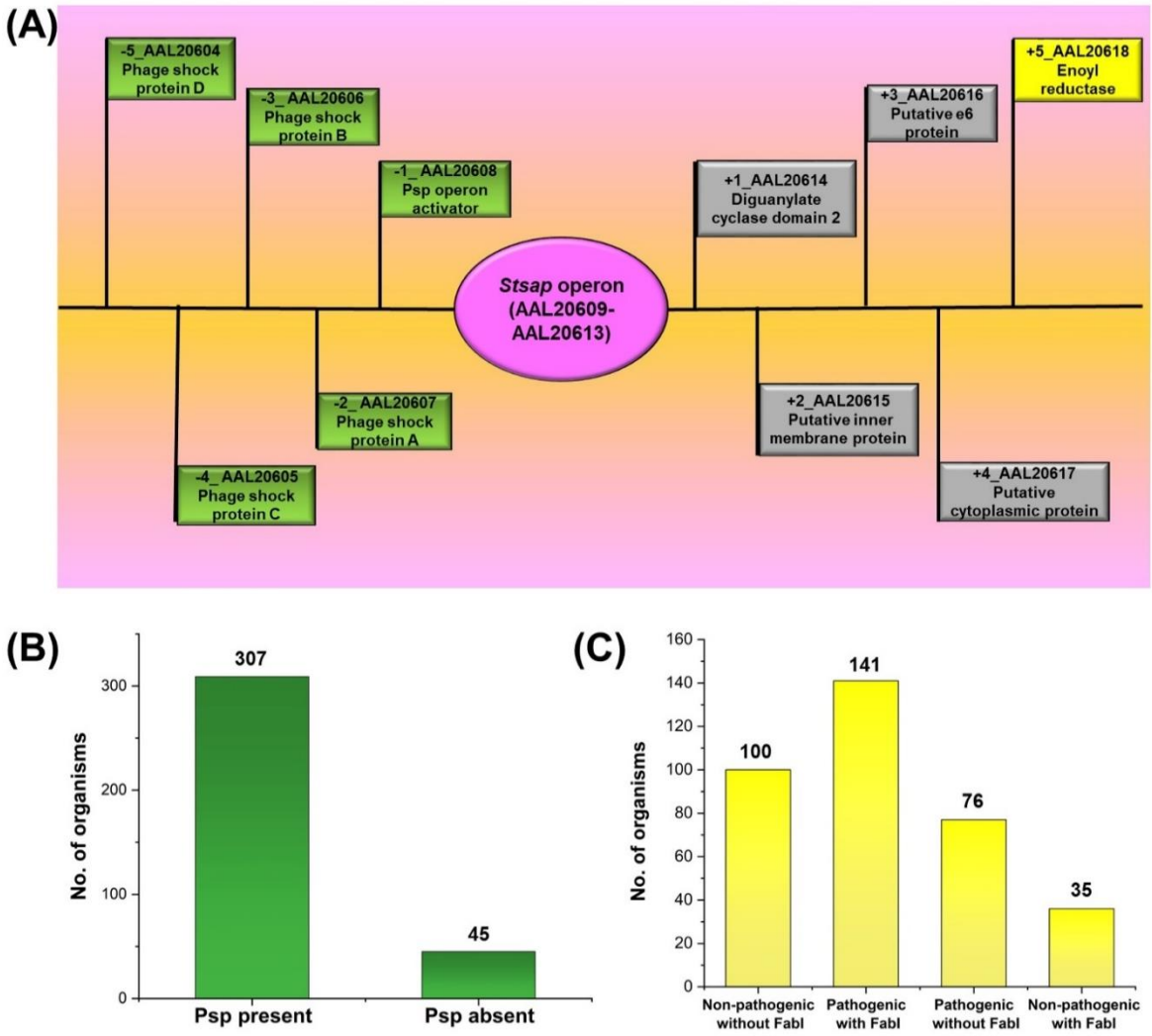


Figure 3.3. Gene-neighborhood conservation of *sap* operons. (A) Schematic representation of *Salmonella enterica* serovar Typhimurium *sap* (*Stsap*) operon (pink oval) gene neighborhood. The *psp* genes (green rectangle) and *fabI* (yellow rectangle) are present upstream and downstream of the *sap* operon, respectively. Other genes have been represented as grey rectangles. (B) The distribution of the *psp* genes present in the upstream of the *sap* operons of different organisms. (C) The distribution of the presence of *fabI* in the *sap* operon neighborhood of different organisms has been represented as histograms.

3.3.5. Varying genetic arrangements of non-consensus *sap* operons

Out of 421 *sap* operons, a total of 69 entries were found to be non-consensus; they do not contain all the five *sap* operon genes in tandem. Thus, to understand the genetic arrangement, these 69 non-consensus *sap* operons were further investigated. The analysis revealed that they are arranged in a varying manner, including the absence of certain *sap* operon components, the presence of intermediate genes in between the *sap* operon, the presence of fragmented *sap* operon components, distant genetic locations of *sap* operon components and/or the presence of other proteins apart from *sap* operon components (Figure 3.4 and Table A3). Thus, these 69 non-consensus entries were broadly classified into three distinct groups: (1) Group I consists of *sap* operons with fragmented or intermediate genes, (2) Group II consists of other components in place of *sap* operon components, and (3) Group III consists of missing *sap* genes (Figure 3.4). Such a distinct pattern of the genetic arrangement in the non-consensus entries suggests an evolutionary selection pressure. Thus, to obtain an in-depth insight into these variations in the genetic arrangement of the *sap* operons, the individual groups were further investigated.

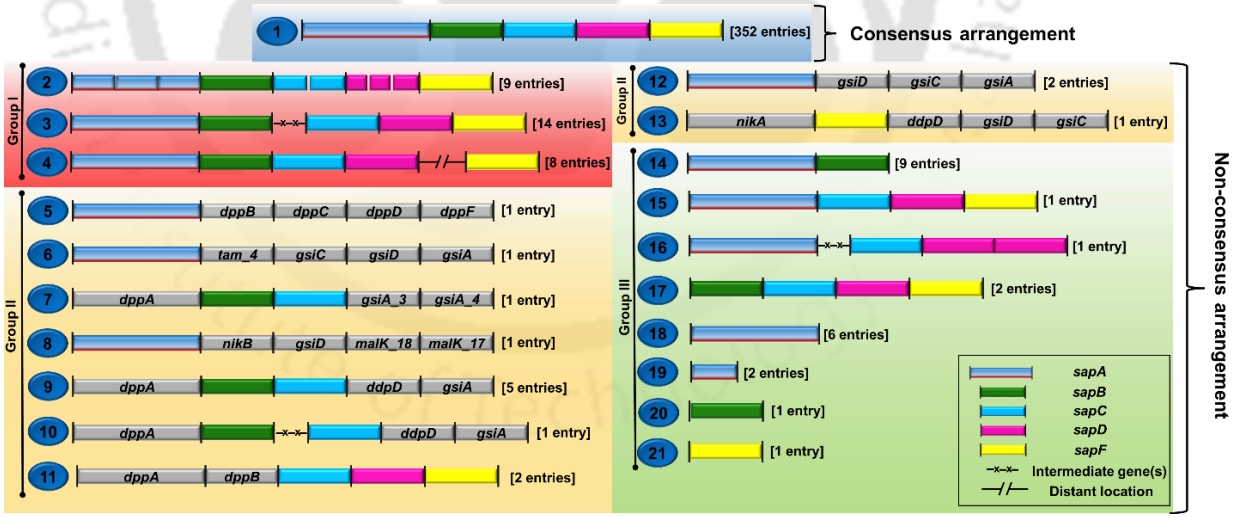


Figure 3.4. A schematic representation of the various genetic architectures of *sap* operons. The consensus arrangement (blue shade) consists of all five components in tandem, whereas in the case of the non-consensus entries, the genetic arrangement of the *sap* components was observed to vary. The non-consensus entries have been demarcated into three types, viz., Group I, fragmented or intermediate genes (red shade); Group II, the presence of other

components (yellow shade); and Group III, missing genes (green shade). The *sap* operon components have been depicted in the form of solid blocks, wherein *sapA* (cyan block), *sapB* (green block), *sapC* (light blue block), *sapD* (magenta block), and *sapF* (yellow block) have been depicted with varying colors. Other non-*sap* operon components have been depicted in the form of grey blocks with the gene name mentioned within the blocks. The details of the non-consensus entries have been provided in Table A3. The bottom right corner box elaborates on various abbreviations utilized in the figure.

3.3.5.1. Group I: *sap* operons with fragmented or intermediate genes

The members of Group I were observed to include those operons with (1) fragmented *sap* components, (2) the presence of an unrelated gene(s) between the *sap* components, and (3) the presence of a *sap* component at a distant location. In the first instance, it is tempting to speculate that the *sap* components might have been fragmented due to a genome sequencing error. This is because the gene fragments are in tandem with the canonical operonic arrangement patterns. Hence, all nine entries containing fragments of the *sap* components were analyzed and compared with their consensus counterparts. The pair-wise sequence alignment of these fragments revealed their similarity with their respective Sap components with complete query coverage. Further, these results hint at the possibility of the individual proteins oligomerizing to form the complete Sap component. For instance, the *sap* operon of *Vibrio ponticus* (*Vpsap*) contains three fragments *VpsapA_1*, *VpsapA_2,* and *VpsapA_3* of *sapA* and two fragments *VpsapC_1* and *VpsapC_2* of *sapC* (Figure 3.5A). Interestingly, these fragments, when concatenated, covered the entire length of the respective Sap component. Further, these fragments (*Vp*SapA_1-3 and *Vp*SapC_1-2) form a clade with their respective SapA and SapC components from *E. coli*, *S. typhimurium*, and *H. influenzae* (Figure 3.5B). Thus, the sequence-based similarity and phylogenetic analyses suggest that these fragments ultimately form a complex upon protein folding and perform their cognate function. To further establish this, a protein-protein interactome analysis of *Vp*SapB against all the proteins of *V. ponticus* was performed. Expectedly, the results revealed the interactions of these fragments with all other Sap components (Figure 3.5C and

Table A4). However, notably, these fragments (*Vp*SapA_1-3 and *Vp*SapC_1-2) do not seem to interact with each other, suggesting their non-oligomerization to form a single complete component (Figure 3.5C and Table A4). Like *V. ponticus*, all other eight entries possessing fragmented *sap* operonic components were also found to exhibit similar results. Hence, preliminary analysis suggests that the fragmented *sap* components observed might be either an error in genome sequencing or a misannotation of the genetic frame.

In the cases where the *sap* operons are disrupted by the presence of an unrelated intermediate gene, the protein-protein interactome analysis revealed interaction amongst the *sap* components. For instance, in *Pantoea conspicua,* the *sap* operon (*Pcsap*) is disrupted by the presence of two intermediate genes (an integrase and a hypothetical protein) in between *PcsapB* and *PcsapC* (Figure 3.5D). The protein-protein interactome analysis revealed interaction among its components (*Pc*SapABCDF), suggesting that the Sap system can perform its function despite its operonic disruption (Figure 3.5E). Like *P. conspicua*, five remaining entries having insertion of unrelated genes in its operon exhibit similar results (Table A4). In conclusion, the occurrence of such variations might be an instance of primitive operon formation stages maintained within certain extant bacterial species.

In the cases where a *sap* component is found at a distant location on the genome, the operons were investigated for their functionality. This particular case is observed in *H. ducreyi*, where SapF (*Hd*SapF) is observed at a distant genetic location (coordinate range: 688854-689648) far from its other Sap components (*Hd*SapABCD, coordinate range: 996781-1001361) (Figure 3.5F). Previously, *H. ducreyi* has been reported to be resistant to cathelicidin LL-37, likely utilizing its Sap transport system (Mount *et al.,* 2010). Thus, to corroborate these findings, a protein-protein interactome analysis was performed. The results revealed that *Hd*SapF interacts with all other Sap components (*Hd*SapABCD) despite their distant genetic location (Figure 3.5G). Similarly, in the case of six other *sap* operon entries, wherein a component is located distantly, the protein-protein interactome results revealed a similar interaction. Precisely, in the case of six entries, the SapF components were revealed at a distant genetic location akin to

*Hd*sap operon (Table A4). The Sap transporter systems mentioned above were observed to interact with each other despite deviation from the consensus arrangement. This suggests that the systems may be functionally active despite the genetic variations observed. Hence, the occurrence of the variations discussed may be an out-turn of preliminary evolutionary events for the formation of the *sap* operon.

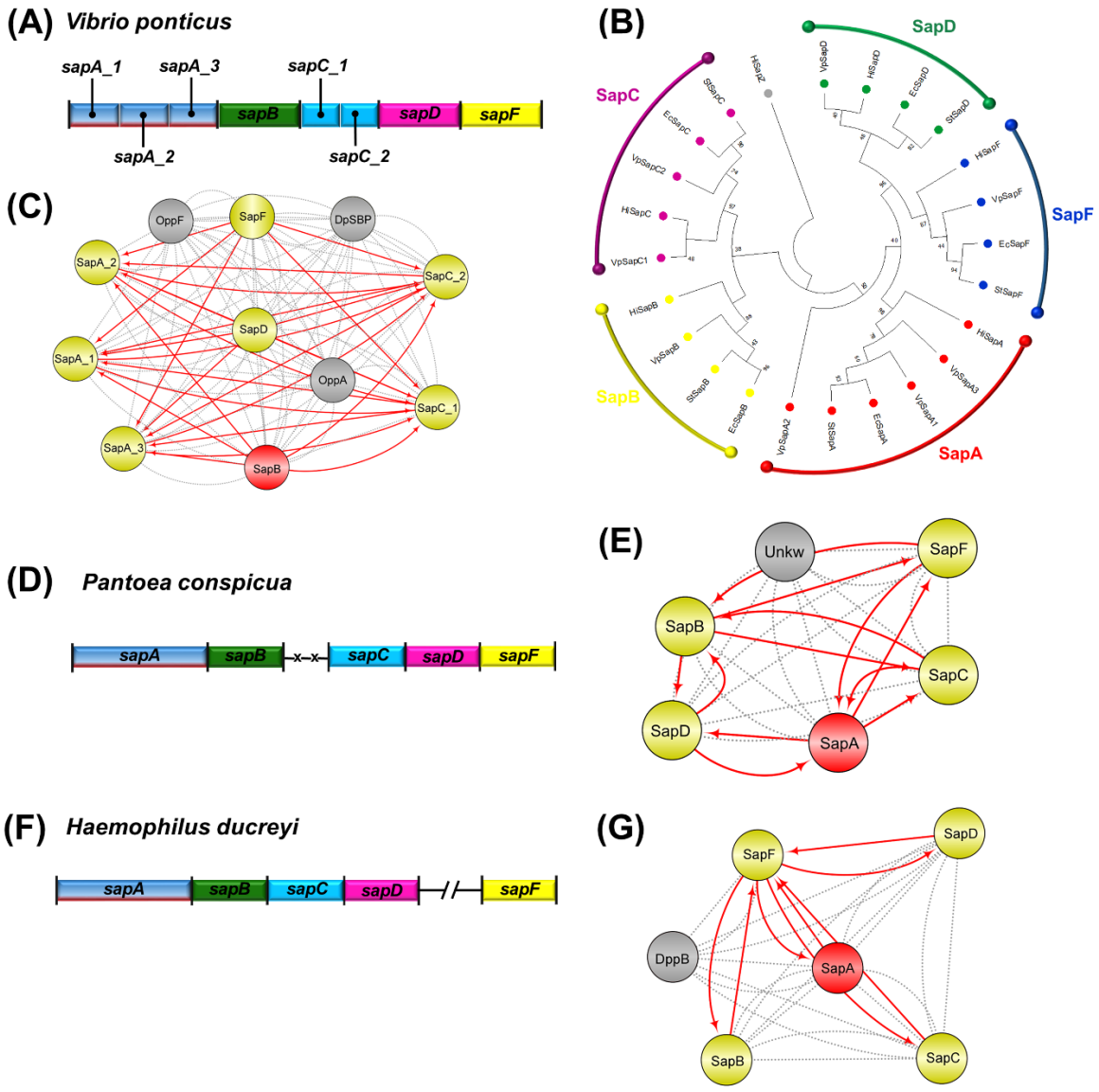


Figure 3.5. Non-consensus arrangement of *sap* operons. (A) Operonic arrangement of *sap* operon in *V. ponticus*. The solid blocks represent the *sap* operon components (*sapA*, cyan block; *sapB*, green block; *sapC*, light blue block; *sapD*, magenta block; and *sapF*, yellow block). The names of the genes have been mentioned within the solid blocks. (B) The proteins used to generate the

phylogenetic tree are the components of the Sap transporter system of *H. influenzae* (*Hi*SapA, P45285; *Hi*SapB, P45286; *Hi*SapC, P45287; *Hi*SapD, P45288; *Hi*SapF, P45289 and *Hi*SapZ, P45290), *S. typhimurium* (*St*SapA, P36634; *St*SapB, P0A2J3; *St*SapC, P0A2J5; *St*SapD, P36636 and *St*SapF, P36638), *E. coli* (*Ec*SapA, Q47622; *Ec*SapB, P0AGH3; *Ec*SapC, P0AGH5; *Ec*SapD, P0AAH4 and *Ec*SapF, P0AAH8), and *V. ponticus* (*Vp*SapA1, A0A090PE13; *Vp*SapA2, A0A090PX00; *Vp*SapA3, A0A090PC72; *Vp*SapB, A0A090PBV1; *Vp*SapC1, A0A090PF22; *Vp*SapC2, A0A090PE10; *Vp*SapD, A0A090PWZ5 and *Vp*SapF, A0A090PC65). The UniProt id of the proteins used has been mentioned within parenthesis. (C) The protein-protein interaction network of the *sap* operon components of *V. ponticus* wherein the query protein (red), the interacting Sap components (yellow), and other proteins (grey) have been represented in the form of circles with the protein name mentioned within. The protein-protein interaction networks amongst the *sap* operon components are denoted as solid-red lines, while those of other proteins are broken-grey lines. (D) Operonic arrangements of *sap* operons in *P. conspicua* (*Pcsap*), details as given in (A). The symbol (-x-x-) within the *Pcsap* operon is representative of the intermediate genes present. (E) The details of the protein-protein interaction network for *P. conspicua* are given in (C). (F) Operonic arrangements of *sap* operons in *H. ducreyi* (*Hdsap*), details are given in (A). The symbol (-//-) in between *HdsapD* and *HdsapF* represents the distant location of *HdsapF*. (G) The details of the protein-protein interaction network for *H. ducreyi* are given in (C). Further details pertaining to the protein-protein interaction network have been tabulated in Table A4.

3.3.5.2. Group II: the presence of other components in place of *sap* operon components

Group II consists of entries where one or other ABC transporter subunits substitute(s) *sap* components in the operon. The investigation of these substituents revealed that most of them belong to the ABC transporters such as Dpp (dipeptide), Ddp (D,D-dipeptide), and Gsi (glutathione) transporters, which mostly account for peptide importers. This observation might be due to either the preliminary stages of operon formation in evolutionarily primitive organisms or a

misannotation of the genetic frames. To verify, a sequence-based homology analysis of all 15 entries was performed. The results of our sequence-based homology revealed that 7 out of 15 entries were misannotated as other ABC components, but instead, they demonstrated higher sequence identity with their corresponding Sap counterparts (Table A3). The remaining eight entries seem to be correctly annotated and may be present due to evolutionary selection (Table A3). To further confirm this hypothesis, a protein-protein interactome analysis of one misannotated entry (*Providencia rustigianii*) and four correctly annotated entries was performed. In both cases, the *sap* components were observed to interact with the genes situated within their operon (Tables A3-A4). Hence, the eight entries wherein the *sap* components are interspersed with other ABC components suggest the sharing of ABC components. This sharing machinery is well-reported and prevalent among numerous ABC systems (Chandravanshi *et al.,* 2019). Both sequence-based homology and protein-protein interactome analysis suggest the eight correctly annotated entries to be probable instances of evolutionary selection.

3.3.5.3. Group III: missing *sap* genes

Group III comprises incomplete *sap* operons where one or the other *sap* component(s) is(are) missing. The data collected showed a total of 23 such entries. The plausible reason for such an arrangement of *sap* components has not been reported in the literature to date. To ascertain further the reliability of these data, a protein-protein interactome analysis for these entries was performed. Notably, out of the 23 entries, protein-protein interactome information of only six entries was available in the databases. The analysis revealed that none of these entries show any missing Sap transporter component, substantiating their absence from their genome (Table A4). Such types of incomplete and randomized presence of the *sap* operon components may contribute to an evolutionary pattern tilting towards the formation or degeneration of an operonic construct.

3.3.6. Sap components of Gram-positive bacteria, archaea, and eukaryotes display variations in the genetic and structural features

The Sap transport system is reported to be ubiquitous amongst Gram-negative bacteria; however, in this study, a few Gram-positive bacteria, archaea, and eukaryotes were found to contain the Sap transport system. The Sap components from these organisms seem to differ both at the sequence and structure levels from that of their Gram-negative counterparts. For example, the SapA of *Solibacillus isronensis* (*Si*SapA) and *Staphylococcus schweitzeri* (*Ss*SapA) contain the N- and C-terminus in Domain III instead of Domain I as found in *Ec*SapA. This causes the Domain III of *Si*SapA and *Ss*SapA to be divided into two subdomains as in the Domain I of *Ec*SapA (Figures 3.6A-3.6C). The remaining *sap* operons of Gram-positive bacteria belong to Group III of non-consensus entries and are found to be similar to the Gram-negative Sap components. Apart from Gram-positive bacteria, an archeal orphan SapA was observed in the case of *Thermoplasmatales archaeon* (*Ta*SapA). *Ta*SapA demonstrated structural similarity with *Ec*SapA (Figures 3.6A,3.6D). However, sequence-based homology suggested that *Ta*SapA is similar to nickel-binding proteins rather than the typical SapA, which binds to AMPs and/or heme molecules.

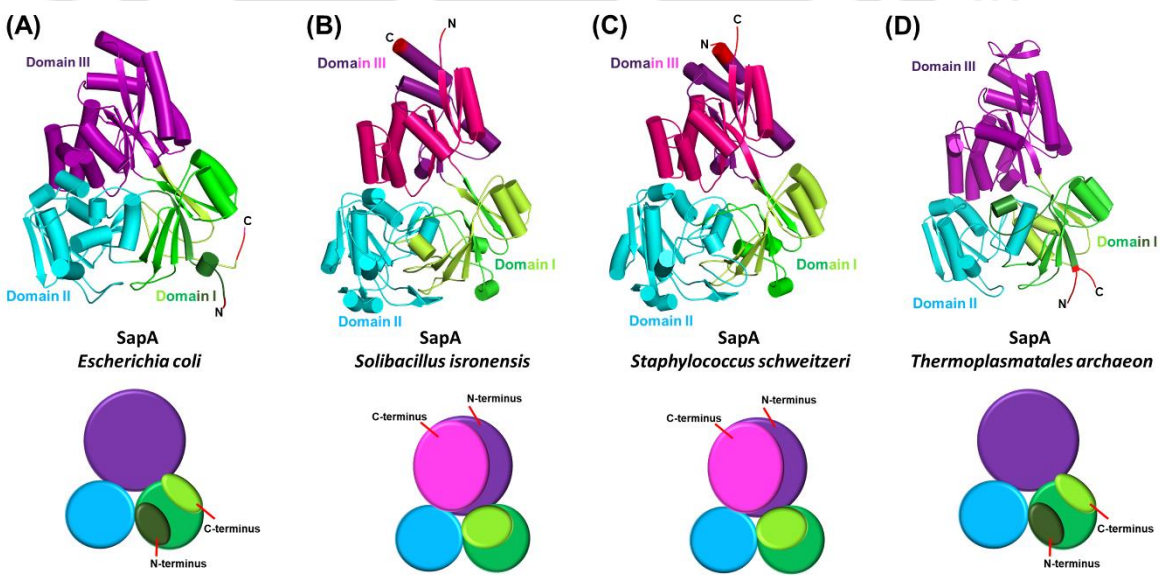


Figure 3.6. Structural comparison of Gram-negative, Gram-positive, and archaeal SapA. (A) Cartoon representation of the predicted three-dimensional model of *E. coli* SapA consisting of three domains, viz., Domain I (green), Domain II (cyan), and Domain III (magenta). The N- and C-termini (red) of *Ec*SapA have been observed in Domain I. (B and C) Cartoon representation of predicted three-

dimensional models of *S. isronensis* and *S. schweitzeri* SapA consisting of three domains, Domain I (green), Domain II (cyan) and Domain III (magenta). Both the N- and C-termini (red) are observed in Domain III, contrary to the case of *Ec*SapA. (D) Cartoon representation of the predicted three-dimensional model of *T. archaeon* SapA (*Ta*SapA) consisting of three domains: Domain I (green), Domain II (cyan), and Domain III (magenta). The N- and C-termini (red) of *Ta*SapA are observed in Domain I.

Interestingly, apart from Gram-positive bacteria and archaea, the Sap transport system was identified in *Beauveria bassiana*, an entomopathogenic fungi used as a biological control agent against mosquitoes and several other insects (Wang *et al.,* 2021). The entry of *B. bassiana* possessing the *sap* operon was noted to be a natural isolate within the Genome database in NCBI. In *B. bassiana*, the *sap* operon (*Bbsap*) consists of *sapA-hg-sapC-sapD* (*hg*: hypothetical gene) genes in tandem, lacking *sapB* and *sapF* (Figure 3.7A). Although *Bb*SapC is of similar length to that of Gram-negative homologs, *Bb*SapA (655 aa) and *Bb*SapD (790 aa) are longer than their *Ec*SapA (547 aa) and *Ec*SapD (390 aa) counterparts. However, interestingly, phylogenetic analysis revealed that the *Bb*Sap components clade with their respective homologous Sap components (Figure 3.7B).

In Gram-negative bacteria, SBPs (e.g., SapA) are located in the periplasm. They are translocated to the periplasm with the help of the signal peptide of 13-30 amino acid residues (Von Heijne, 1985). Interestingly, *Bb*SapA does not contain any signal peptide, as predicted by most of the web tools. However, as per the InterPro database, it contains an N-terminal extension (1-64 aa) which interacts with σ^{54} (Figure 3.7C). Notably, proteins containing σ^{54}-interacting domains are reported to be involved in two-component signal transduction (Albright *et al.,* 1989). A sequence-based homology search of the N-terminal extension of *Bb*SapA suggests a similarity with the C-terminal of PspF (a transcriptional regulator) from *Klebsiella michiganensis*. Upon further investigation, the protein PspF was found to contain a σ^{54}-interacting domain along with an ATPase domain.

Further, the protein *Bb*SapD possesses an extended C-terminal in comparison to other SapD counterparts. Sequence-based homology analysis of *Bb*SapD revealed its closest orthologs, *Ec*SapDF, and *Ec*FabI. A multiple sequence alignment of these three proteins exhibits that the N- (1-293 aa), central (294-527 aa), and C-terminal (528-789 aa) regions of *Bb*SapD are aligned with *Ec*SapD, *Ec*SapF, and *Ec*FabI, respectively (Figure A2). Further, on superimposing the predicted tertiary structure models of *Ec*SapDF and *Ec*FabI individually with *Bb*SapD, each model (*Ec*SapDF and *Ec*FabI) aligns at distinct regions similar to the aforementioned trend with a root mean square deviation (RMSD) of 0.858 Å, 0.425 Å, and 0.338 Å, respectively (Figure 3.7D). This suggests that the protein *Bb*SapD possesses the functional property of all three proteins, SapDF and FabI, indicating direct evidence of the correlation between FabI and Sap transporter function. However, interestingly, *B. bassiana* does not contain the SapB subunit, thus making the operon incomplete. Although a hypothetical gene exists between *Bb*SapA and *Bb*SapC, it shows no similarity to the SapB proteins. Thus, the role of the hypothetical gene remains to be characterized. In conclusion, all the genetic variations observed amongst the non-consensus entries hinted at a probable evolutionary trend towards the formation or degeneration or random genetic shuffling of the *sap* operon.

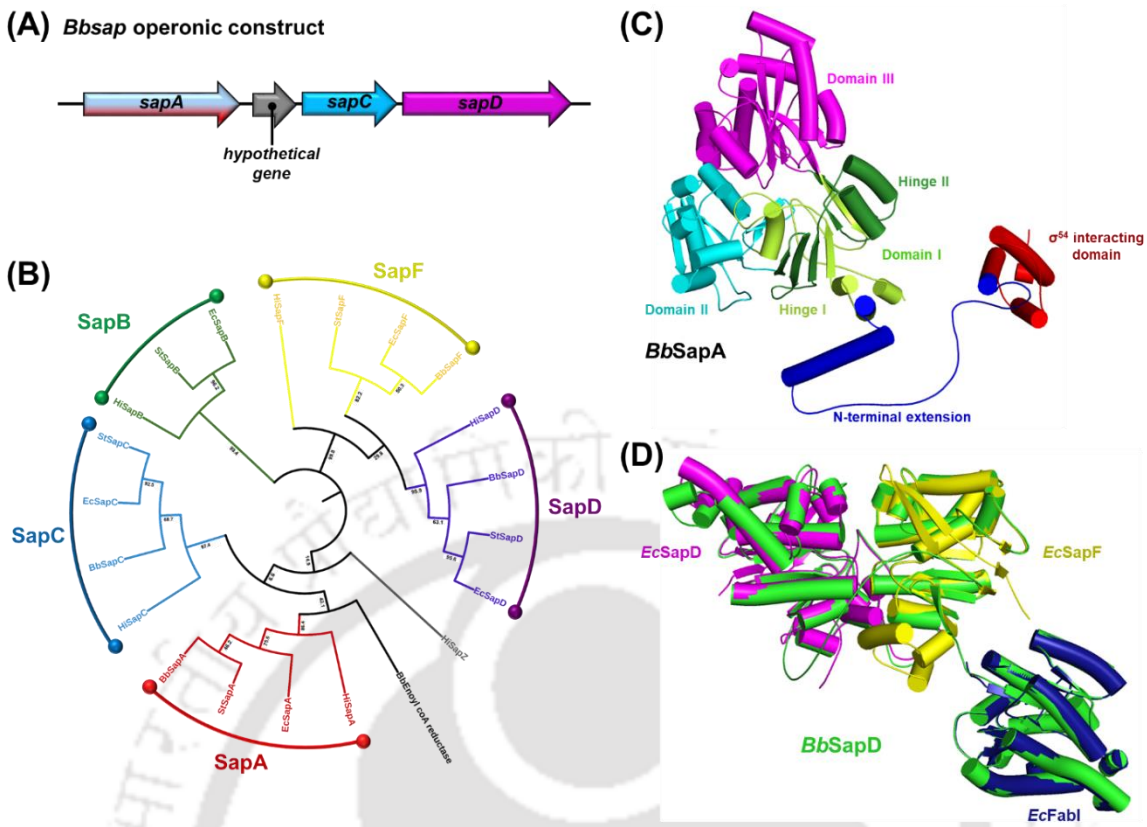


Figure 3.7. *B. bassiana* Sap transport system. (A) The operonic construct of *B. bassiana* wherein the gene names are mentioned within the arrows. (B) The proteins used to generate the phylogenetic tree are the components of the Sap transporter system of *H. influenzae* (*Hi*SapA, P45285; *Hi*SapB, P45286; *Hi*SapC, P45287; *Hi*SapD, P45288; *Hi*SapF, P45289 and *Hi*SapZ, P45290), *S. typhimurium* (*St*SapA, P36634; *St*SapB, P0A2J3; *St*SapC, P0A2J5; *St*SapD, P36636 and *St*SapF, P36638), *E. coli* (*Ec*SapA, Q47622; *Ec*SapB, P0AGH3; *Ec*SapC, P0AGH5; *Ec*SapD, P0AAH4 and *Ec*SapF, P0AAH8), and *B. bassiana* (*Bb*SapA, A0A0A2VIV3; *Bb*SapC, A0A0A2VIC0 and *Bb*SapD, A0A0A2VES7). The UniProt id of the proteins used has been mentioned within parenthesis. The *Bb*SapD frame was divided into three parts based on the MSA (Figure A2): *Bb*SapD, *Bb*SapF, and *Bb*FabI. All the Sap transporter components were observed to form distinct clades, and all the *Bb*Sap components were observed to clade with their respective Sap counterparts. (C) Cartoon representation of the three-dimensional structure model of *Bb*SapA. The predicted model consists of three domains: Domain I (green), Domain II (cyan), and Domain III (magenta). Domain I consists of two hinges, viz., Hinge I (limon), which connects Domain I to Domain II & III, and Hinge II (dark green), which connects Domain II & III. The

*Bb*SapA was also observed to possess an N-terminal extension (blue) and a σ^{54}-interacting domain (red). (D) Cartoon representation of the predicted model of *Bb*SapD (green) superimposed to the predicted three-dimensional structure models of *Ec*SapD (magenta), *Ec*SapF (yellow), and *Ec*FabI (blue), retrieved from the AlphaFold database (Varadi *et al.,* 2022).

3.3.7. Evolutionary analysis suggests conservation of *sap* operonic arrangement

A phylogeny-based analysis was performed to study the evolutionary conservation of *sap* operon, considering the 16S rRNA (18S rRNA for *B. bassiana*) sequence of all the 421 entries collected in this study. The results obtained exhibited an expected evolutionary trend in the class of γ-proteobacteria wherein the genera of *Escherichia* and *Salmonella* could be observed to form a separate clade from those of *Haemophilus*, *Vibrio,* and *Aeromonas* (Figure 3.8). The result hints at an evolutionarily less conserved status of the genera of *Aeromonas*, wherein it is observed to clade in the proximity to the out-group, corroborating its *sap* operonic layout but with varied intergenic distances, suggesting plausible stages of operon compaction (Figure A3). Similar trends were observed in the cases of *Vibrio* sp., *Photobacterium* sp., and *Photorhabdus* sp. Although these genera demonstrated higher conservation of the intergenic distances than *Aeromonas* sp., they were observed to clade separately from the consensus *sap* operons. Further, the *sap* operons with conserved intergenic distances were mostly observed to form a separate clade of conserved entries than those with varying intergenic distances (Figure 3.8). The clades distinctly observed in the phylogeny arise in a gradual manner, hence highlighting the distinction amongst the entries, which surprisingly goes hand in hand with the genetic architecture of the *sap* operons possessed by them. Further, compared to the consensus entries, the non-consensus *sap* operons are observed to be primarily cladding near the out-group. Further, both the consensus and the non-consensus entries are observed to form distinct clades, wherein the consensus entries are distal from the out-group, while the non-consensus entries are proximal to the out-group (Figure 3.8). Interestingly, Groups II and III non-consensus *sap* operons are observed to be more proximal to the out-group than

those of Group I (Figure 3.8). Hence, the 16S rRNA-based evolutionary tree revealed a probable positive trend for the evolutionary conservation of the consensus *sap* operonic arrangement.

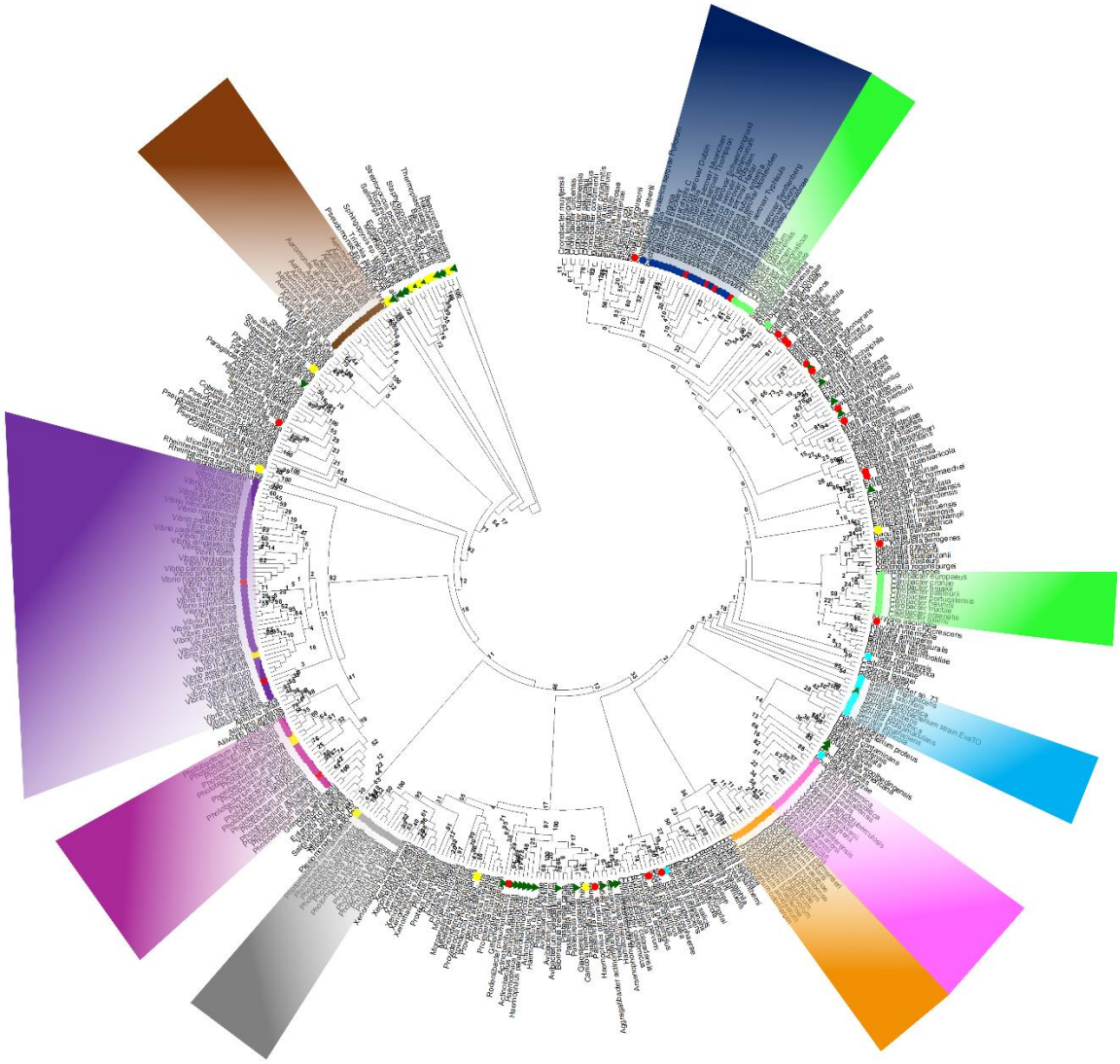


Figure 3.8. 16S rRNA-based phylogeny of all 421 organisms. The genera consisting of more than ten entries have been highlighted in the phylogenetic tree with distinct color codes, e.g., *Salmonella* sp. (dark blue), *Citrobacter* sp. (light green), *Serratia* sp. (light blue), *Yersinia* sp. (pink), *Pectobacterium* sp. (orange), *Photorhabdus* sp. (grey), *Photobacterium* sp. (magenta), *Vibrio* sp. (purple), and *Aeromonas* sp. (brown). The non-consensus entries have been broadly demarcated as fragmented/intermediate genes (Group I, red circle), the presence of other components (Group II, yellow square), and missing Sap components (Group III, green triangle).

To further establish the observed evolutionary trend, a phylogenetic analysis considering all the five *sap* components was performed. Concatenation of the operons was avoided as this would lead to a bias towards the operons from Group III of the non-consensus entries. The goal was to observe the distinction among the individual Sap proteins and how much they varied from their non-consensus counterparts. This included all five *sap* components of 83 unique genera, along with all the *sap* components of the 69 non-consensus entries. Additionally, the proteins unrelated to the Sap transport systems observed in the case of Group II non-consensus entries, such as Dpp, Gsi, Nik, Mal, and Ddp from *E. coli,* were also included. The results revealed that all the Sap components form five distinct clades, while the other *E. coli* ABC importer components formed out-groups in these clades (Figure 3.9). Interestingly, akin to the 16S rRNA-based phylogenetic analysis, in this case, the Sap components of the consensus entries were observed to clade together, while the non-consensus entries were limited majorly near the out-group of each distinct clade (Figure 3.9). Amongst the non-consensus entries, most of the Group II members clade with the other ABC importer components, substantiating that they function differently than the Sap transport system. Further, the Sap components of the Groups II & III non-consensus entries were also observed majorly in proximity to the out-group, forming a separate clade from the consensus entries of each distinct clade (Figure 3.9). Interestingly, the Group I entries were interspersed across the phylogenetic tree, suggesting its proximity to the consensus arrangement of the *sap* operon (Figure 3.9). In summary, the trend of the maintenance of the *sap* operonic arrangement is observed to be similar to that of the phylogenetic analysis using 16S rRNA sequences. In both analyses, the *sap* components of *Haemophilus sp*., *Vibrio sp.*, and *Aeromonas sp.* are proximal to the out-group, whereas *Escherichia sp.*, *Salmonella sp.,* and *Klebsiella sp.* are farther from the out-group and interspersed within the evolutionarily conserved *sap* operonic components (Figure 3.9). A similar trend in both analyses establishes the evolutionary trend favoring the maintenance of the consensus *sap* operonic arrangement.

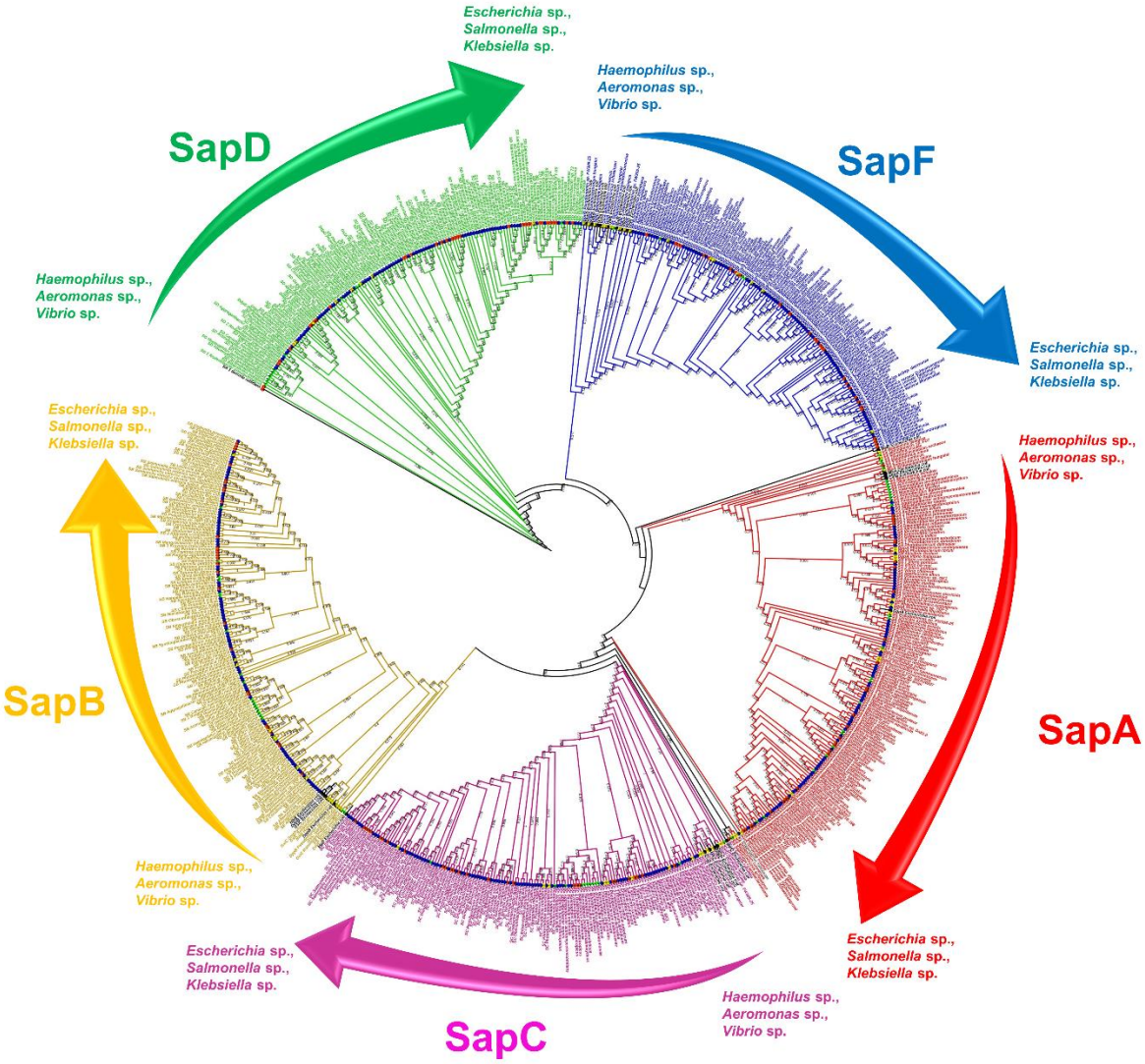


Figure 3.9. Phylogenetic classification of *sap* operon components. Five distinct clades for each component were observed, wherein the clades corresponding to SapA, SapB, SapC, SapD, and SapF are been displayed in red, yellow, magenta, green, and blue, respectively. Apart from 87 consensus entries (blue hexagon), the tree consists of 68 non-consensus entries along with their components. The non-consensus entries have been broadly demarcated as fragmented/intermediate genes (Group I, red circle), the presence of other components (Group II, yellow square), and missing Sap components (Group III, green triangle).

3.4. Discussion

The human immune system is a robust machinery deployed against a horde of pathogenic attacks (Chaplin, 2010). It is complex and layered broadly into three parts, viz., the anatomical barrier, the innate, and the adaptive systems (Turvey and Broide, 2010). One of the key components of the innate immune system is the antimicrobial peptides (AMPs), which interact with the pathogen cell membrane and lead to their death by forming pores and disrupting membrane homeostasis (Mookherjee *et al.,* 2020). However, Gram-negative bacterial pathogens have developed a series of strategies to negate the effect of AMPs. One such strategy is the Sap transport system, which belongs to the ABC transporter superfamily, consisting of an SBP (SapA), two TMDs (SapBC), and two NBDs (SapDF). The Sap transport system is ubiquitous among Gram-negative bacteria and has been studied to a large extent (Hsu *et al.,* 2019). The *sap* operons in *S. typhimurium* and *H. influenzae*, well-studied model organisms, consist of all the components (SapABCDF and SapABCDFZ, respectively) in tandem. Notably, the functional significance of the unique SapZ component is not very clear (Shelton *et al.,* 2011). Intriguingly, in *H. ducreyi*, the genetic arrangement of the Sap transporter system is altered, with the component SapF being located at a distant location on the genome. Still, the *H. ducreyi* Sap (*Hd*Sap) transporter has been reported to confer resistance against cathelicidin LL-37 (Mount *et al.,* 2010). Further variation was observed in the case of the *Ecsap* operon, wherein a previous study suggested *Ec*SapBCDF function as a putrescine exporter instead of AMP importer and further denounced *Ec*SapA from the operon, as it was predicted to have an independent promoter site (Sugiyama *et al.,* 2016).

These variations in the genetic arrangement of the Sap transport system intrigued us and made us investigate the evolutionary trends of the *sap* operons. The *sap* operons from various organisms available in the databases were data mined and scrutinized for any conservation in the *sap* operonic arrangement. The results demonstrated the presence of *sap* operons across 421 unique genera and species, of which 352 entries were found to retain the genetic arrangement of the *sap* operon and were referred to as consensus *sap* operons. The remaining 69

entries were referred to as non-consensus *sap* operons. Further, the results of this study revealed the presence of all the components of the *Ecsap* operon (*EcsapABCDF*), including *EcsapA*, in tandem. The presence of a Pribnow box upstream to *EcsapA* and the absence of a promoter between *EcsapA* and *Ec*sapB further confirmed the component *EcsapA* to be part of the operon. Alongside, the *Ecsap* operon genetic arrangement was similar to the consensus entries and hence indicated all five components (*EcsapABCDF*) to form a single operon, in contrast to the previous report (Sugiyama *et al.,* 2016).

The intergenic distance analysis of the consensus entries revealed a fingerprint-like pattern with -4 (*sapAB*), -14 (*sapBC*), -1 (*sapCD*), and -4 to 1 (*sapDF*). These data suggest the presence of overlaps among adjacent genes, which is a characteristic of operons (Assaf *et al.,* 2021). Interestingly, high conservation of the intergenic distance of *sapAB* indicates that the protein SapA is a part of the same operon. This conserved arrangement is observable in the case of the *Ecsap* operon, hence further confirming *EcsapA* to be a part of the operon (Figure 3.1D). Although most of the entries contain five *sap* components (*sapABCDF*), an additional component, *sapZ* is observed only in the *Hi*Sap system. The exclusivity of *sapZ* to the *Hisap* operon may be linked to the existence of an alien gene during its formative stages, implying that the *Hisap* operon might be less evolved than other consensus entries (Fondi *et al.,* 2009).

The gene neighborhood of the consensus *sap* operon revealed the presence of the Psp system upstream and FabI downstream. Both systems are implicated to be essential for the maintenance of membrane homeostasis. The Psp system rescues the membrane in situations of stress and helps in membrane repair, and FabI is an essential component required for the final step of lipid elongation, a prerequisite for membrane formation (Zhang and Rock, 2008; McDonald *et al.,* 2015). The presence of these two systems in the gene neighborhood establishes the functional role of the Sap transporter in stabilizing the membrane by importing the AMPs and hindering their accumulation. Further, the *sap* operon of *Beauveria bassiana* (*Bbsap*), a fungal system, establishes the correlation of the three systems Psp-Sap-FabI. The *Bb*SapA contains an extended N-terminus homologous to the PspF protein, and the *Bb*SapD contains three different

domains corresponding to the *Bb*SapD, *Bb*SapF, and *Bb*FabI. The occurrence of these three systems (Psp-Sap-FabI) in tandem indicates their putative role in the form of a pathogenicity island, enabling the sustenance of the bacterium within the host, as all the three systems have been well implicated in the pathogenicity of the bacterium (Shelton *et al.,* 2011; Darwin, 2013; Katiki *et al.,* 2022). Further, the directionality of occurrence of the three systems on the genome leads to the hypothesis that the Psp system may detect AMP-induced stress in the membranes and activate the Sap transport system to remove the AMPs and prevent membrane disruption. Subsequently, the FabI system is activated, which aids in revamping membrane homeostasis. However, in order to prove the validity of the aforementioned hypothesis, further experimental validation is required.

As the genetic arrangement and operonic neighborhood hint towards the major role of consensus *sap* operonic arrangement in bacterial pathogenesis, the conservation of such a system of importance is expected. However, the results of this study revealed a total of 69 entries with a non-consensus *sap* operonic arrangement. The non-consensus entries contained genes having (1) Group I, *sap* operons with fragmented or intermediate genes; (2) Group II, the presence of other components in place of *sap* operon components; and (3) Group III, missing *sap* genes. The variations observed amongst the Group I, II, and III entries entice us to speculate them to be instances of gradual formation of the *sap* operon. The Group I entries appear to be distributed across the phylogenetic tree, suggesting certain members are similar to the consensus *sap* operons while others are similar to the non-consensus entries. The sequence-based homology, protein-protein interactome, and phylogenetic analyses revealed nine Group I entries (fragmented frames of *sap* components) and seven Group II entries were observed to be out-turn of genome sequencing/annotation errors, respectively. The remaining non-consensus entries might be instances of preliminary stages of operon formation, which have somehow been maintained within certain extant bacterial species.

The evolutionary distribution of the 16S rRNA tree generated was in accordance with previous reports of γ-proteobacteria (Brandis, 2021). Also, the evolutionary trend observable in the tree was in accordance with the hypothesized

evolutionary flow of formation of the *sap* operon. Notably, the entries possessing non-consensus *sap* operons were majorly observed in proximity to the out-group, while the consensus entries were observed to be present in clades farther from the out-group. Within the consensus entries, the *sap* operons having conserved intergenic distances were observed to be more conserved and claded together, indicative of the evolutionary favoritism of the consensus arrangement of the *sap* operons. The phylogenetic analysis using only the consensus and non-consensus *sap* operon components corroborated with the previous phylogenetic analysis and revealed that all the consensus *sap* operon entries were evolutionarily favored, while the non-consensus entries were sparsely present and majorly claded in proximity to the out-group.

The evolutionary conservation of essential genetic elements has been recently reported, wherein, in the course of formation, the essential genetic element undergoes duplication/deletion events (Nguyen *et al.,* 2019). Further, such conserved systems have been suggested to predate their last common ancestor, hence suggesting a minimalistic change in their operonic arrangement (Nguyen *et al.,* 2019). Therefore, the high conservation of the *sap* operon might be a sign of its functional necessity for the sustenance of bacterial pathogens. Previous experimental studies have also established the essentiality of the Sap transport system in the sustenance and pathogenicity of the bacterium and, hence, are in accordance with the evolutionary trait described here (Mason *et al.,* 2005; Mason *et al.,* 2006; Rodríguez-Arce *et al.,* 2019). Further, certain instances of variations observed may be correlated to *sap* operon formations, which may have been sustained within certain extant bacterial specimens owing to the evolutionary conservation of this particular system. Hence, keeping in consideration the evolutionary conservation and the instances of slight variations of the *sap* operons observed, we may hypothesize the formation of *sap* operons to have followed the “piece-wise” model (Figure 3.10). Post-formation of the consensus operon, it was propagated and maintained in multiple extant bacterial systems, establishing its essentiality in the sustenance of bacterial pathogens. (Fondi *et al.,* 2009). Hence, the outcomes of this study clearly indicate an evolutionary favoritism towards the conservation of the *sap* operon and its significance.

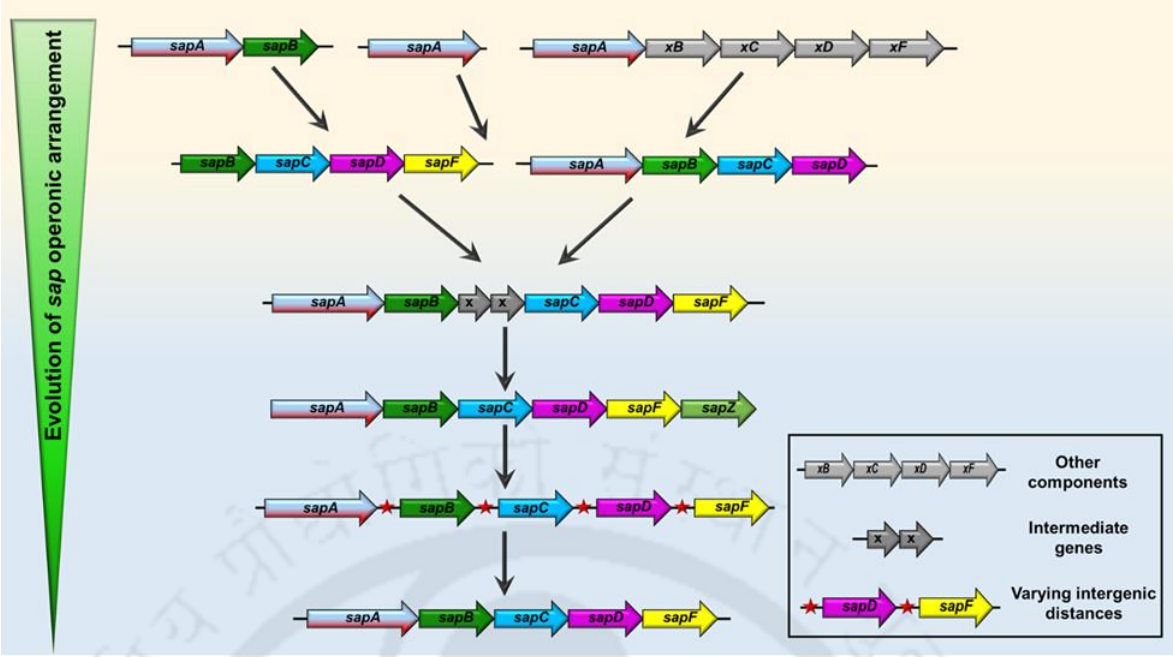


Figure 3.10. Schematic representation of the hypothetical evolutionary progression of the *sap* operon. The genetic arrangement of *sap* components is displayed as an arrow in *sapA* (blue), *sapB* (green), *sapC* (light blue), *sapD* (magenta), and *sapF* (yellow). The gene names have been labeled inside the arrows. The bottom right corner box elaborates on various abbreviations utilized in the figure.

3.5. Conclusion

This study unravels the evolutionary conservation of the Sap transport system and suggests a plausible path of its formative stages leading to its existence. It accounts for a comprehensive conglomeration of the conservation and predominance of the Sap transport system amongst the Gram-negative bacterial pathogens. The data analysis suggested extensive conservation of the *sap* operon genetic arrangement and revealed a fingerprint-like pattern amongst the intergenic distances of the adjacent genes of the system. Further, the gene neighborhood of the system suggested functional correlation with their neighbors (Psp and FabI systems) and further substantiated the role of the Sap transporter in AMP uptake. Apart from the conserved entries, certain instances of the genetic arrangement were observed wherein the Sap transport system completely strayed away from the conserved genetic arrangement. Such systems have been

termed non-consensus entries. The non-consensus entries might depict the rudiments of certain initial stages of operon formation sustained by extant bacterial species, which enlightened the plausible evolutionary path of the *sap* operon. This study can serve as a template for contemplating the significance of various systems by tracing their evolutionary paths and conservation. Leading to the discovery of numerous systems of prevalence in bacterial sustenance and pathogenesis. These systems may be pinned down as lucrative targets for broad-range drug design against a horde of potent pathogens.

Appendix A. Supplementary data

Supplementary Figures: A1-A3

Supplementary Tables: A1-A4

Chapter 4

In silico characterization of *Ec*Sap transport system

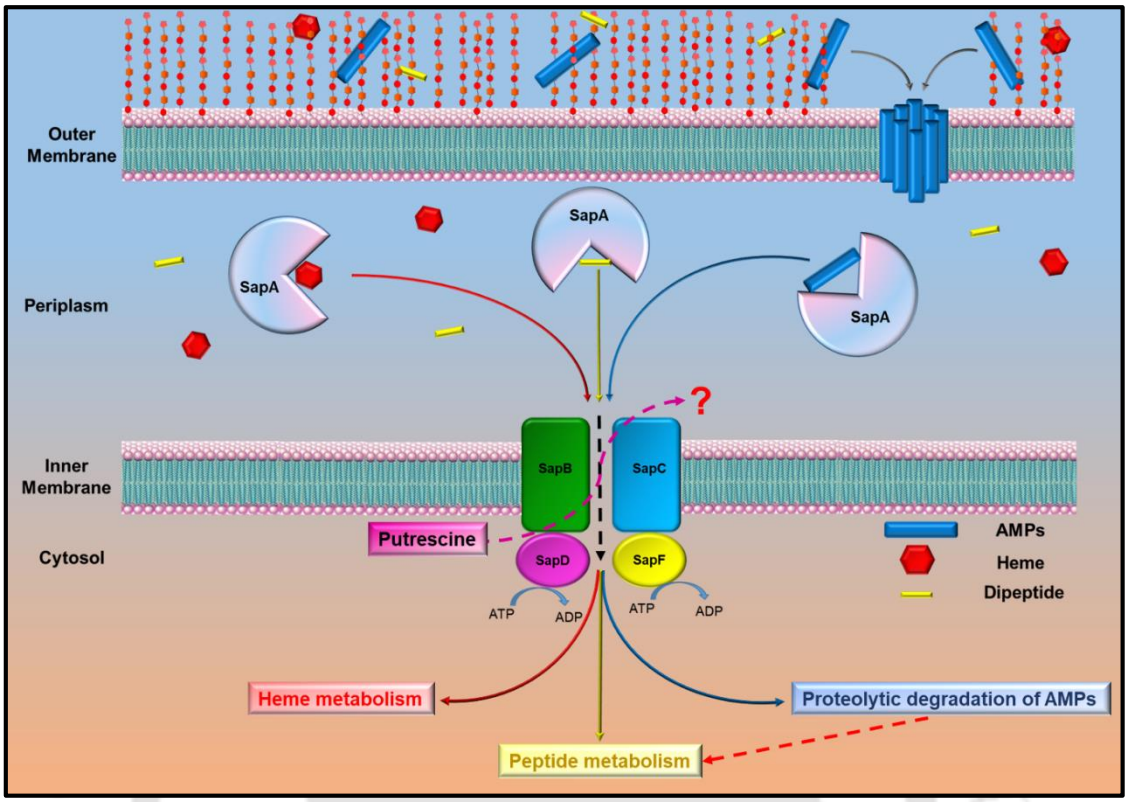
The manuscript for this chapter is under review as:

1. **Dasgupta P** and Kanaujia SP. Enlightening the multifarious attributes of the *Escherichia coli* Sap transport system: A computational perspective.

Abstract

Antimicrobial peptides (AMPs) are majorly utilized by the hosts to safeguard against a horde of bacterial pathogens. AMPs help in the clearance of bacterial pathogens primarily by disrupting their membrane homeostasis. However, most Gram-negative pathogens have adapted machinery that enables resistance against AMPs. One such machinery is the sensitivity to the antimicrobial peptide (Sap) transport system. The Sap system belongs to the ATP-binding cassette (ABC) transport system and consists of five components, viz. SapABCDF. It uptakes AMPs inside the cell that are proteolytically degraded by proteases. In contrast, in *Escherichia coli*, the Sap (*Ec*Sap) transport system was suggested as a putrescine exporter. In this study, efforts were made to scrutinize the enigmatic aspects of the functional role of the *Ec*Sap transport system using an *in silico* approach. The results of this study suggest that the protein *Ec*SapA can bind to dipeptides having aromatic amino acids. Further, it can bind to longer peptides such as tri-, tetra-, and so on, including AMPs. AMPs such as protamine and protegrin-1 show binding to the protein *Ec*SapA. In addition, the molecule heme shows binding affinity towards the protein *Ec*SapA. In summary, *Ec*SapA seems to be involved in the uptake of a wide range of molecules, such as dipeptides, AMPs, and heme. The results of this study can be utilized to design inhibitors targeting the protein SapA.

Graphical Abstract



4.1. Introduction

The human body resorts to the innate and adaptive immune systems to clear any foreign object that poses a threat. Broadly, the human immune system can be classified into three levels: the anatomical barrier, the innate immune system, and the adaptive immune system (Turvey *et al.,* 2010). Out of these, the innate and adaptive immune systems are further divided into cellular and humoral immune systems. In the cellular innate immune system, various cells such as natural killer cells, mast cells, and several classes of phagocytic cells are involved. Whereas, in the humoral innate immune system, a variety of molecules like complement proteins, lipopolysaccharide-binding proteins, antimicrobial peptides (AMPs), etc. play a crucial role (Turvey *et al.,* 2010). AMPs are one of the first lines of defense employed by the human immune system and have been reported to be biocidal to a range of pathogens infecting the host (Zhang and Gallo, 2016). AMPs are known to be produced by Paneth cells in the intestinal mucus and by neutrophils in the skin and blood plasma (Oppenheim *et al.,* 2003; Muniz *et al.,* 2012). AMPs tend to interact with the bacterial cell membrane and perturb its homeostasis,

ultimately leading to the death of the pathogen (Zhang and Gallo, 2016). The prevalence of AMPs across the human body prompted the pathogens to establish systems and mechanisms enabling them to escape from AMPs (Gruenheid and Le Moual, 2012). The resistance mechanisms include the modification of bacterial cell outer membrane, down-regulating AMP production by the host, shielding of the bacterial surface, proteolytic degradation of AMPs, and efflux or influx of AMPs (Gruenheid and Le Moual, 2012).

One of the influx mechanisms was first reported in *Salmonella typhimurium* and was termed as Sap (sensitivity to antimicrobial peptides) transport system, which provided resistance to exposure to the AMP protamine (Parra-Lopez *et al.,* 1993). Subsequently, the Sap transport system was classified as an ATP-binding cassette (ABC) importer, consisting of a substrate-binding domain (SBP, SapA), two transmembrane domains (TMDs, SapBC), and two nucleotide-binding domains (NBDs, SapDF). The Sap transport system has been reported to be ubiquitous amongst the class of Gammaproteobacteria and has been implicated to be essential for their pathogenesis (Parra-Lopez *et al.,* 1993; Hsu *et al.,* 2019). Several studies on the Sap transport system from non-typeable *Haemophilus influenzae* (*Hi*Sap) have corroborated with the previous studies and demonstrated the import of the host-induced AMPs and their degradation proteolytically (Shelton *et al.,* 2011). The Sap transport systems of *Proteus mirabilis* (*Pm*Sap) and *Dickeya dadantii* (*Dd*Sap) have also been reported to be involved in the uptake of AMPs such as protegrin-1 and α-1-purothionin, respectively (Lopez-Solanilla *et al.,* 1998; McCoy *et al.,* 2001). Notably, few studies have also suggested the involvement of certain components of *Hi*Sap in the uptake of potassium ions and heme molecules from the periplasmic space into the cytoplasm (Mason *et al.,* 2006*,* 2011). Further, diversity in the function of the Sap transport system was reported in *Klebsiella pneumoniae* and *Allivibrio fischeri*, wherein the Sap transport system has been implied to be responsible for adhesion to host cells and capsular formations apart from resistance against AMPs (Chen *et al.,* 2000; Hsu *et al.,* 2019). In a recent study, the Sap transport system of *Shewanella oneindensis* (*So*Sap) was reported to be involved in the uptake of oligopeptides and is the first report implying its role in the nutritional fulfillment of the bacterium (Liang *et al.,* 2019). In contrast, a previous study on

the Sap system of *Escherichia coli* (*Ec*Sap) suggested that the components SapBCDF are involved in the export of putrescine (Sugiyama *et al.,* 2016). Also, the study reported that the substrate-binding component SapA (*Ec*SapA) was not part of the *sap* operon (Sugiyama *et al.,* 2016). In contrast to the previous reports, the results elaborated in Chapter 3 have shown the protein *Ec*SapA to be part of the *sap* operon and suggested all five components function in synchrony.

This functional ambiguity of the *Ec*Sap system intrigued us to inspect their functional prospects in detail. Therefore, in this study, robust *in silico* strategies have been employed, which established the multispecific nature of *Ec*SapA, wherein it's observed to bind dipeptides, AMPs, and heme. Further, the study also delves into the importer-like attributes of the *Ec*Sap transport system.

4.2. Methodology

4.2.1. Data collection and sequence analysis

The amino acid sequences of all the Sap components were retrieved from the UniProtKB database and were cross-validated with the GenBank database (The UniProt Consortium, 2023). Wherever required, pair-wise sequence alignment of proteins was performed using the web tool EMBOSS Needle (Rice *et al.*, 2000). The homologs of these proteins were identified using the web tool BLAST (Altschul *et al.,* 1990). Evolutionary analysis of the proteins of interest was performed by generating a phylogenetic tree employing the maximum likelihood method embedded in the program MEGAX (Kumar *et al.*, 2018). Phylogenetic trees were visualized and rendered using the web tool iTOL v6.3 (Letunic and Bork, 2021).

4.2.2. Structure analysis

All the three-dimensional structures of proteins, wherever required, were predicted utilizing various programs such as RaptorX, SWISS-MODEL, Phyre2, I-TASSER, and AlphaFold (Källberg *et al.,* 2012; Biasini *et al.,* 2014; Kelley *et al.,* 2015; Yang *et al.,* 2015; Varadi *et al.,* 2022). The predicted models were further refined and minimized energetically using the program GalaxyRefine (Ko *et al.,* 2012). The stereo-chemical properties of the predicted models were validated

using the PDBsum database (Laskowski *et al.,* 2018). Wherever available, the tertiary structures of the homologs and members from all the sub-classes of SBPs were retrieved from the Protein Data Bank (PDB). The pair-wise structural distance matrices were generated by calculating the root mean square deviation between structures using a home-built Python script in the program PyMOL v2.4.1 (Molecular Graphics System, Schrödinger, LLC). These distance matrices were utilized to generate a structure-based phylogenetic tree using the web tool DendroUPGMA (Garcia-Vallvé *et al.,* 1999). The phylogenetic trees were further rendered using the web tool iTOL v6.3 (Letunic and Bork, 2021). The volume of the ligand-binding site of the proteins was computed using the web tools POCASA 1.1. and CASTp 3.0 (Yu *et al.,* 2010; Tian *et al.,* 2018). The average volume of the ligand-binding site obtained employing both programs was used for further analysis. The structures were visualized, and the figures were generated using the program PyMOL. The 3-dimensional atomic coordinates of the antimicrobial peptides such as LL-37 cathelicidin (PDB id: 2K6O), β-defensin 2 (PDB id: 6CS9), protegrin-1 (PDB id: 1ZY6), and α-1-purothionin (PDB id: 2PLH) were retrieved from the PDB database. For protamine, the 3-dimensional atomic coordinates were generated using the polypeptide chain (UniProt id: P69014) employing the web tool Avogadro (Hanwell *et al.,* 2012). The electrostatic potential charge distribution on the proteins was generated using the ABPS tool embedded in PyMOL. The structural arrangement of *Ec*SapBC in the bacterial inner membrane was generated using the web server PPM 3.0, available at the OPM (orientation of proteins in the membrane) database (Lomize *et al.,* 2022). The topology of the *Ec*SapBC complex was estimated utilizing the deep learning algorithm-based web server MemBrain 3.0 (Feng *et al.,* 2020).

4.2.3. Molecular docking and virtual screening calculations

The binding affinities of ligands to the proteins *Ps*DppA, *Ec*SapA, and *Hi*SapA were estimated using the freely available program AutoDock v4.2 (Morris *et al.,* 2009). Further, the virtual screening of all 400 dipeptides against *Ps*DppA and *Ec*SapA was performed using the program Raccoon, available in the program AutoDock (Forli *et al.,* 2016). The 3-dimensional atomic coordinates of all 400 dipeptides were retrieved either from the PubChem (397) or ChemSpider database (3) (Kim *et al.,* 2023; Pence and Williams, 2010). Prior to each

molecular docking calculation, hydrogen atoms were added to the protein. Subsequently, the partial charges were assigned to the proteins using the Gasteiger charge algorithm (Morris *et al.,* 2009). For the blind docking simulation of all the 400 dipeptides, a grid box of grid size 126x126x126 with a grid spacing of 0.484 Å was generated, considering the center of mass of the protein as the grid center. For each molecular docking simulation, the protein molecule was kept rigid while the ligand was considered flexible along the rotatable bonds. For each molecular docking simulation, a total of 2000 runs of the Lamarckian genetic algorithm (LGA) were performed. After each molecular docking calculation, the best-docked poses were analyzed, and the atomic interactions between the docked ligand and the protein were identified using the program Coot v0.9.4 (Emsley *et al.,* 2010). The molecular interactions were visualized using the program PyMOL.

4.2.4. Protein-protein docking

To perform protein-protein docking of AMPs, an open conformation of the selected SapA proteins was predicted using various three-dimensional structure prediction programs, as mentioned previously. The degree of movement and the flexible domain in the open and closed conformations of the SapA proteins were estimated using the web tool DynDom (Taylor *et al.,* 2014). The web tool ClusPro v2.0 was utilized for the protein-protein docking calculations (Desta *et al.,* 2020) with the default set of parameters. The SapA proteins were considered as receptors, while the AMPs were considered as ligands. For each docking simulation, the highest-ranked clusters were considered for further analysis. The molecular interactions of the protein-ligand complexes were analyzed utilizing the PDBsum database. The figures were generated using the program PyMOL.

4.3. Results

4.3.1. *Ec*SapA homologs unravel its evolutionary proximity to the dipeptide-binding (DppA) class of proteins

Sequence-based homology search revealed that *Ec*SapA is similar to the class of dipeptide-binding proteins (DppA), nickel-binding protein (NikA), Glutathione-binding protein (GsiA), and oligopeptide-binding proteins (OppA). A phylogenetic

analysis of *Ec*SapA, along with its homologs, suggested its closeness towards DppA of *Pseudoaltermonas sp.* strain SM9913 (*Ps*DppA) (Figure 4.1A). *Ps*DppA is known to bind dipeptides of varying properties (Li *et al.,* 2015). The peptides in the binding site of *Ps*DppA are stabilized by both main- and side-chain atoms (Li *et al.,* 2015). A pair-wise sequence alignment of *Ec*SapA and *Ps*DppA reveals partial conservation of the ligand-binding residues, including Gly50, Ala434, and Asp436 (Figure 4.1B).

The conservation of these residues was further verified at the structural level. The predicted model of *Ec*SapA shows that it is composed of three domains (I-III), akin to the members of the Cluster C SBPs. The domain I of *Ec*SapA is split into three distinct regions: Ala22-Phe65 (hinge I), Pro217-Pro293 (hinge II), and Arg512-Pro547 (hinge I), while domains II and III span over Tyr66-Gly216 and Gly294-Leu511, respectively (Figure 4.1C), as observed in the cluster C SBPs (Chandravanshi *et al.,* 2021). A structural comparison of *Ps*DppA (PDB id: 4QFL) and *Ec*SapA (model) reveals their similarity (rmsd: 1.064 Å). Further, the main-chain interacting residues Thr48, Gly50, Arg383, Tyr385, Ala434, and Asp436 are spatially conserved, while the side-chain interacting residues show no conservation (Figure 4.1D-4.1E). A structure-based phylogenetic analysis revealed that *Ec*SapA is distinctly cladded with the cluster C SBPs (Figure 4.1F). The cluster C SBPs have further been sub-classified into five sub-clusters (I-V) based on their cognate ligands (Fukamizo *et al.,* 2019). The structure-based phylogenetic analysis of *Ec*SapA hints at its closeness to the sub-cluster C-II, specific to dipeptides (Figure 4.1G). These results suggest that *Ec*SapA can bind dipeptides. However, the partial conservation of the ligand-binding residues hints at its flexibility toward different combinations of dipeptides as well as peptides of higher lengths, in contrast to *Ps*DppA.

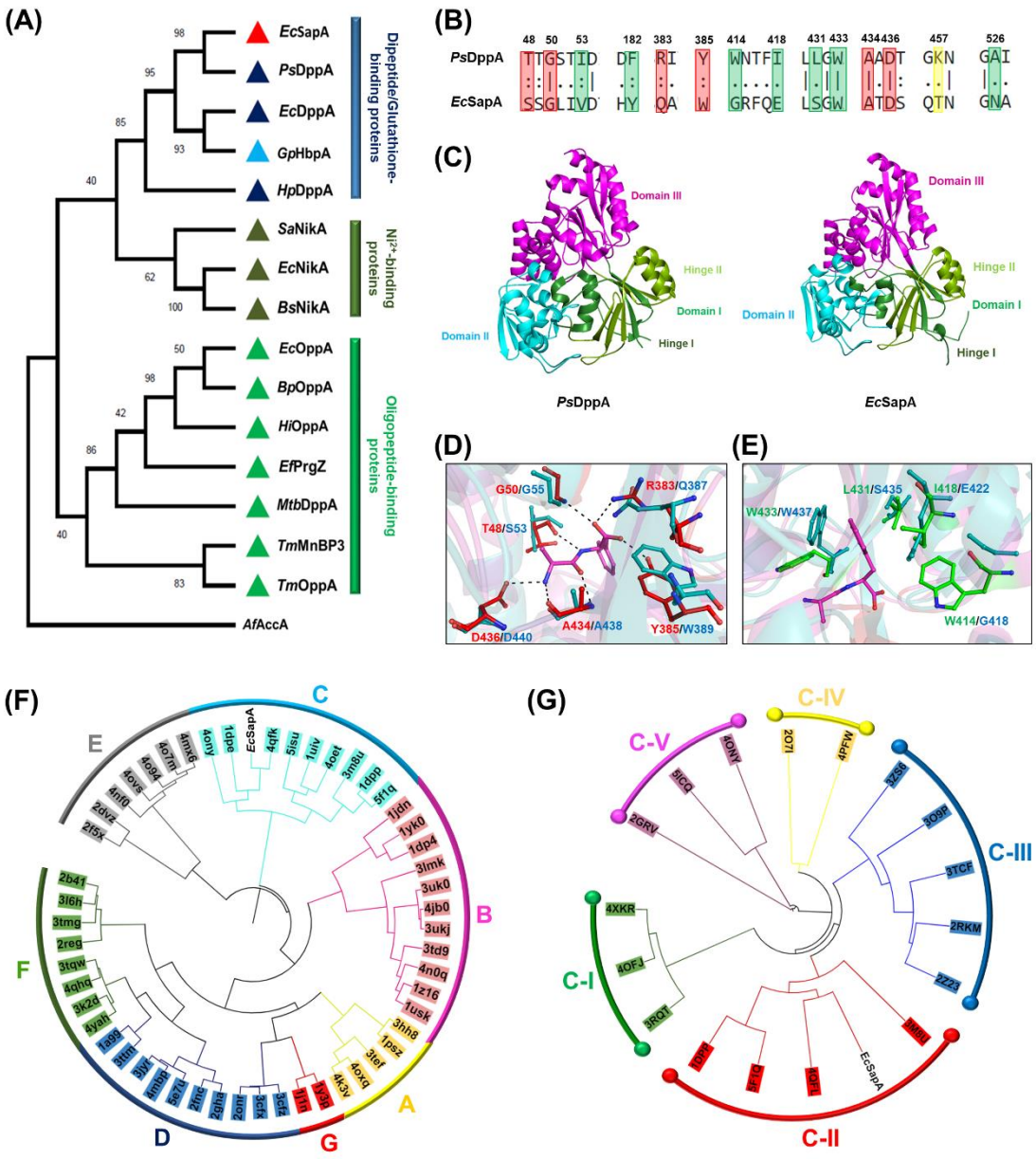


Figure 4.1. Sequence and structure-based comparative study of *Ec*SapA and *Ps*DppA binding site. (A) The proteins used to generate the phylogenetic tree are from *Pseudoalteromonas sp.* (*Ps*DppA, A7Y7W1), *E. coli* (*Ec*DppA, P23847), *Glasserella parasuis* (*Gp*HbpA, B8F653), *Helicobacter pylori* (*Hp*DppA, A0A1U9ISL1), *Staphylococcus aureus* (*Sa*NikA, Q2G2P5), *E. coli* (*Ec*NikA, P33590), *Brucella suis* (*Bs*NikA, Q9AL82), *E. coli* (*Ec*OppA, P23843), *Burkholderia pseudomallei* (*Bp*OppA, Q63ID0), *H. influenzae* (*Hi*OppA, Q4QLH0), *Enterococcus faecalis* (*Ef*PrgZ, Q51643), *Mycobacterium tuberculosis* (*Mtb*DppA, I6X811), *Thermatoga maritima* (*Tm*MnBP3, Q9X0V0; *Tm*OppA, Q9WXR2), and *Agrobacterium fabrum* (*Af*AccA, Q52012); The UniProt id of the

proteins used have been mentioned within parenthesis. *Ec*SapA (red triangle) was observed to clade with dipeptide (blue triangle)/glutathione (cyan triangle)-binding proteins, whereas the Ni^{2+}-binding (dark green triangle) proteins and the oligopeptide-binding (light green triangle) protein were observed to form a separate clade. (B) Pair-wise sequence alignment of *Ps*DppA with *Ec*SapA. The residues reported to interact with the main chain of the ligand have been enclosed in red boxes, while the residues that make hydrophobic interactions and salt bridges with the side chain of the ligand are enclosed in green and yellow boxes, respectively. (C) Cartoon representation of the structure of *Ps*DppA and three-dimensional prediction model of *Ec*SapA. Both structures consist of three domains: Domain I (green), Domain II (cyan), and Domain III (magenta). Domain I consists of two hinges: Hinge I (dark green), which connects Domain I to Domain II and III, and Hinge II (light green), which connects Domain II and Domain III. (D) Superimposition of the binding-site pocket of Ala-Phe (AF) bound *Ps*DppA with *Ec*SapA. The binding-site residues of *Ps*DppA interacting with the main chain of AF (magenta ball and stick) and their corresponding residues in *Ec*SapA are represented in red and cyan ball-and-stick models, respectively. (E) Amino acid residues of *Ps*DppA form hydrophobic interactions with the side chain of AF and their superimposition with *Ec*SapA. The residues interacting with the side chain of AF (magenta ball and stick model) in *Ps*DppA and *Ec*SapA are shown in green and cyan ball and stick model, respectively. (F) Structural distance-based phylogeny of SBPs, wherein the proteins used have been denoted by their PDB ids, and the different clusters of SBPs have been earmarked by varying color highlights. *Ec*SapA was observed to clade along with cluster C SBPs. (G) *Ec*SapA was observed to clade within sub-cluster C-II (red) of dipeptide-binding proteins alongside four more sub-clusters: C-I (green) for metal-binding, C-III (blue) for oligopeptide-binding proteins, C-IV (yellow) for oligosaccharide-binding proteins, and C-V (magenta) for varying ligands. The proteins utilized for the analysis have been denoted in the form of PDB ids.

4.3.2. The protein *Ec*SapA can bind dipeptides

Although the sequence and structure-based homology results suggest *Ec*SapA to be a dipeptide-binding protein, these do not specify the cognate dipeptide(s).

To identify the putative dipeptide(s) that might have specificities toward *Ec*SapA, molecular-docking-based virtual screening calculations were performed. For methodological verification, the protein *Ps*DppA was used as a positive control. A total of 400 possible dipeptides were docked against *Ps*DppA. The previously reported eight cognate dipeptides for *Ps*DppA were found to dock with decent binding energy (-10.26 to -8.71 kcal mol^{-1}) and in similar orientations (Li *et al.,* 2015) (Figure 4.2 and Table B1). As most dipeptides are stabilized by main-chain atoms, the protein *Ps*DppA shows multi-specificity (Li *et al.,* 2015). Interestingly, in addition to the known eight cognate dipeptides, another 25 dipeptides bind *Ps*DppA, having a binding energy greater than or equal to -8.71 kcal mol^{-1} (Table B2).

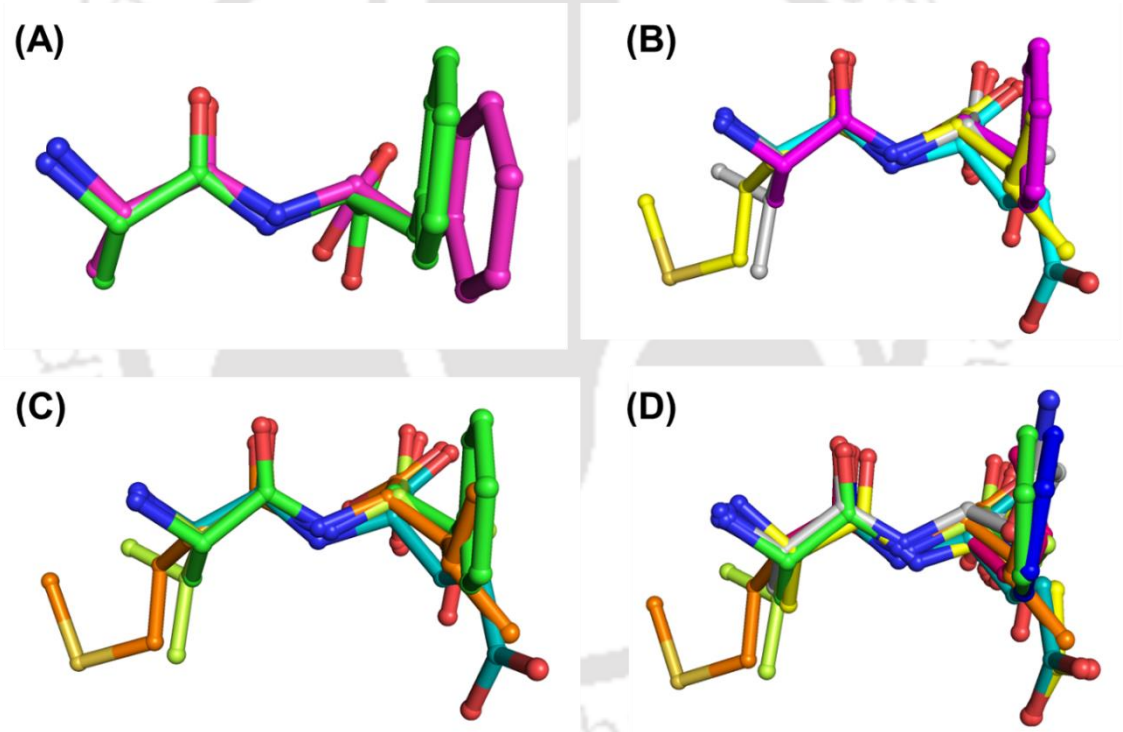


Figure 4.2. *Ps*DppA dipeptide docking studies. (A) The docked AF (green ball and stick) was observed to be perfectly aligned with the crystal structure AF (4QFL, magenta ball-and-stick). (B) Overlayed ligands of *Ps*DppA crystal structures bound to AF, GE (4QFN, cyan ball-and-stick), ML (4QFP, yellow ball-and-stick), and VT (4QPO, grey ball-and-stick). (C) Molecular docking corroborated the crystal data wherein AF (green ball and stick), GE (teal ball and stick), ML (orange ball-and-stick), and VT (limon ball and stick) were observed to align in a similar manner. (D) Alignment of all eight cognate ligands of *Ps*DppA in

accordance to crystal data wherein AF (green ball and stick), GE (teal ball and stick), ML (orange ball and stick), VT (limon ball and stick), GF (blue ball and stick), GL (pink ball and stick), AE (yellow ball and stick) and AQ (grey ball and stick).

In a similar manner, all the 400 dipeptides were docked to *Ec*SapA. The dipeptides exhibiting a binding affinity better than -8.0 kcal mol^{-1} were considered cognate ligands. Three dipeptides, Trp-Phe, Phe-Trp, and Phe-Tyr, docked in the putative ligand-binding site of the protein *Ec*SapA with a binding energy of -8.59 kcal mol^{-1}, -8.41 kcal mol^{-1}, and -8.26 kcal mol^{-1}, respectively (Figure 4.3A-4.3B and Table B1). Notably, three other dipeptides, Trp-Pro, Trp-Ile, and Tyr-Ile, docked at a site different than the putative ligand-binding site with a binding energy of -8.54 kcal mol^{-1}, -8.2 kcal mol^{-1}, and -8.01 kcal mol^{-1}, respectively (Figure 4.3A, 4.3C and Table B1). Hereafter, the putative ligand-binding site of the protein *Ec*SapA will be referred to as A and the other additional site as B. Notably, several dipeptides that showed binding to the protein are hydrophobic in nature, consisting of aromatic amino acids, and interact primarily through the main-chain atoms (Figure B1 and Table B1). Interestingly, the presence of an additional binding site B alongside A suggests a larger binding site of *Ec*SapA (Figure 4.3A).

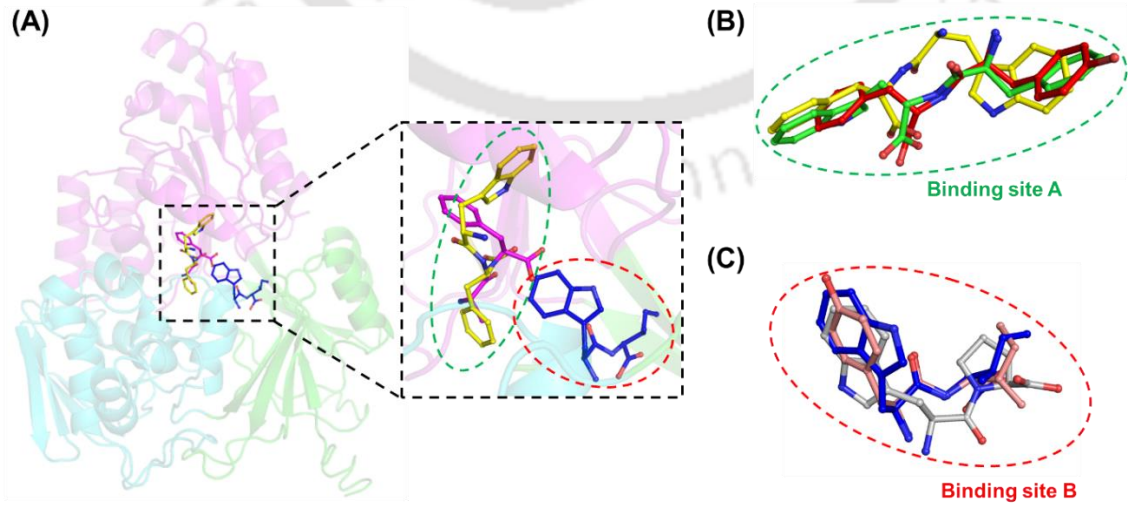


Figure 4.3. The dipeptide-binding propensity of *Ec*SapA. (A) Trp-Phe (yellow ball and stick) was observed to dock in binding-site A (green dashed circle),

similar to the pose of Ala-Phe (magenta ball and stick) structure of *Ps*DppA overlayed with *Ec*SapA. Trp-Ile (blue ball and stick) was observed to dock at an alternative region earmarked as binding site B (red dashed circle). (B) Phe-Trp (green ball and stick) and Tyr-Phe (red ball and stick) were observed docking at binding site A. (C) Trp-Pro (gray ball and stick) and Tyr-Ile (pink ball-and-stick model) were docked at binding site B, suggesting a larger binding site for *Ec*SapA.

4.3.3. The binding site of *Ec*SapA expands upon ligand binding

The molecular docking of dipeptides to *Ec*SapA hinted at an expansion of the binding site in comparison to that of *Ps*DppA. The binding-site volume of the closed conformations of *Ps*DppA and *Ec*SapA are 265.1 $Å^3$ and 634.5 $Å^3$, respectively (Figure 4.4A-4.4B). These were further compared with the closed conformations of known peptide-binding proteins of the cluster C SBPs (Figure 4.4A). The binding-site volume of *Ec*SapA was similar to *Mtb*DppA, which is reported to be a tetrapeptide-binding protein and hence corroborating with possibly a wider binding site of *Ec*SapA (Mitra *et al.,* 2019). Intriguingly, the binding-site volume of *Ps*DppA was similar to that of the tripeptide-binding protein *Yersinia pestis* OppA (*Yp*OppA); hence, this indicated plausible tripeptide-binding capacity.

To identify the tripeptides binding to *Ps*DppA, a total of 40 tripeptides were designed by incorporating all the possible 20 amino acids at the 2^{nd} position in the dipeptides Ala-Phe and Ala-Glu. The molecular docking results revealed that three tripeptides, Ala-Pro-Phe, Ala-Gly-Phe, and Ala-Ala-Phe, bind to the protein *Ps*DppA at the site A with a binding energy of -9.91, -9.5, and -8.9 kcal mol^{-1} (Figure B2 and Table B1). Notably, all these tripeptides bound to the protein in a similar manner as the dipeptides, and the additional residue was accommodated in the extra space available within the binding site. Previous reports have indicated the promiscuous boundaries of several peptide-binding proteins, such as *E. coli* DppA (*Ec*DppA) and *Helicobacter pylori* DppA (*Hp*DppA), capable of binding polypeptides (Smith *et al.,* 1999; Rahman *et al.,* 2019; Fernando *et al.,* 2022). The high similarity between *Ec*SapA and the DppA class of proteins

suggests a high possibility for the promiscuous nature of *Ec*SapA, and further, the larger binding site volume complements the notion of plausible AMP binding.

As the binding-site volume of *Ec*SapA is similar to that of *Mtb*DppA, hence its cognate tetrapeptide Ser-Ser-Val-Thr (SSVT) was docked with *Ec*SapA. The tetrapeptide SSVT binds to *Ec*SapA, covering both sides A and B with a binding energy of -4.95 kcal mol^{-1} (Figure B3A and Table B1). Further, three dipeptides (Ser-Ser, Ser-Val, and Val-Thr) derived from the tetrapeptide SSVT were also docked with the protein *Ec*SapA. Interestingly, the dipeptide Ser-Ser docked at the binding site A with a binding energy of -4.68 kcal mol^{-1}, while the other two dipeptides, Ser-Val and Val-Thr, docked at the binding site B with a binding energy of -5.74 kcal mol^{-1} and -5.86 kcal mol^{-1}, respectively (Figure B3A). Notably, the docked poses of the dipeptides are in accordance with that of the tetrapeptide SSVT. As a negative control, the protein *Ps*DppA was also docked with the tetrapeptide SSVT. The results reveal that *Ps*DppA does not show any binding to the tetrapeptide (Figure B3B).

In *Mtb*DppA, two residues, Trp442 and Asp445, are found to be spatially conserved across the class of DppA proteins and act as a signature for the class (Mitra *et al.,* 2019). Both at the sequence and structure levels, *Ec*SapA shows high similarity with the DppA class of proteins. However, the *Ec*SapA residues Trp437 and Asp440 were observed to have varied spatial orientation in comparison to *Mtb*DppA (Trp442 and Asp445), *Ps*DppA (Trp433 and Asp436), and *Ec*DppA (Trp405 and Asp408) residues (Figure 4.4C). In the case of *Ec*SapA, the distance between Trp437 and Asp440 is 7 Å, whereas the distance between Trp405 and Asp408 of *Ec*DppA was 3.2 Å, and *Mtb*DppA and *Ps*DppA residue distances were similar range with that of *Ec*DppA (Figure 4.4C). Further, the spatial orientation of Trp437 and Asp440 of *Ec*SapA were similar to the residues of *Ll*OppA (Trp473 and Ala476), which is reported to bind up to 35 amino acid long peptides (Figure 4.4A, 4.4C) (Berntsson *et al.,* 2009). Hence, the spatial arrangement of Trp437 and Asp440 reveals a wider binding site of *Ec*SapA and hints towards potential AMP-binding capacity.

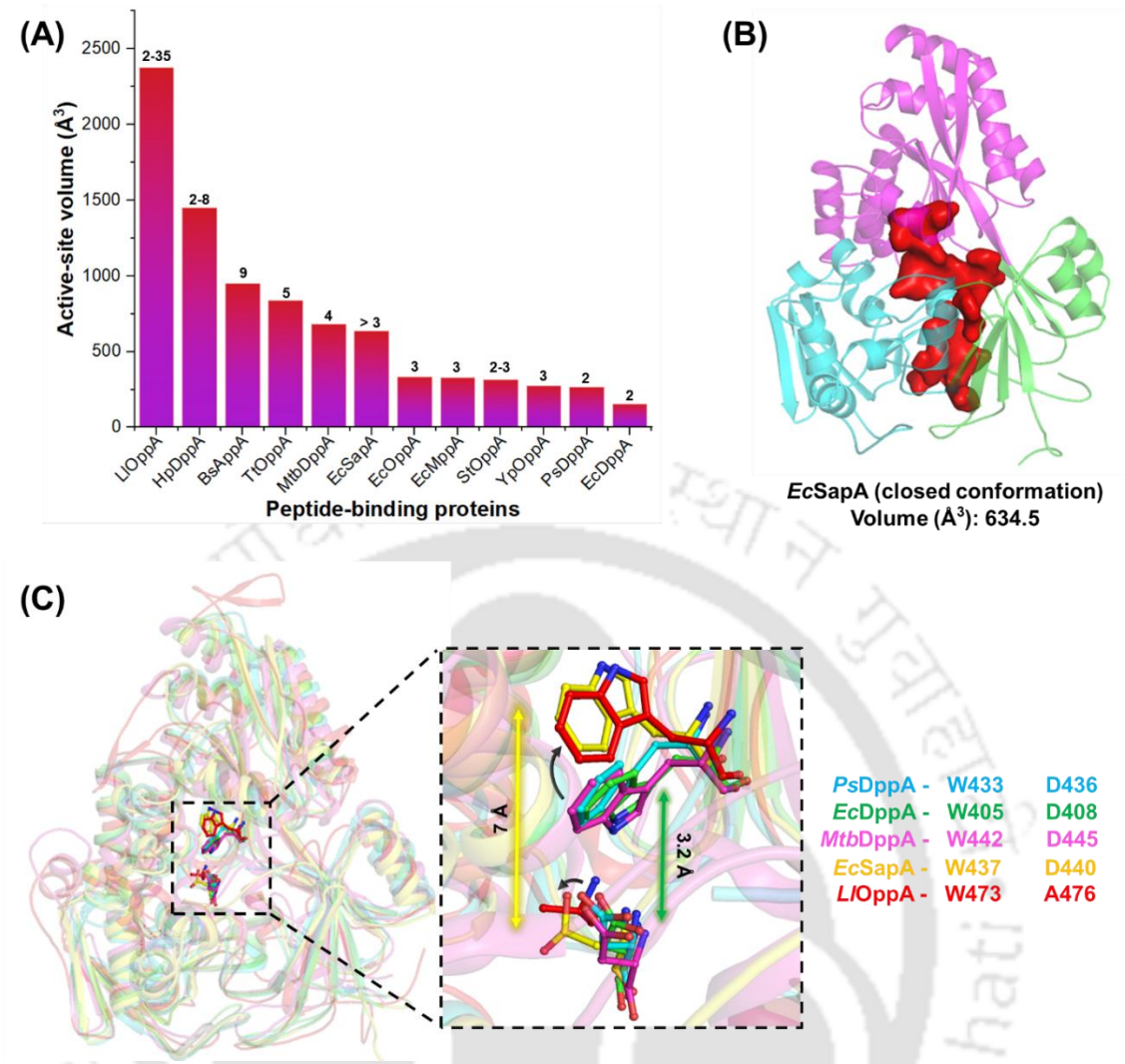


Figure 4.4. Comparative analysis of *Ec*SapA binding site with various peptide-binding proteins. (A) Histogram-based representation of the binding-site volume calculation of *Ec*SapA closed conformation model alongside the structure of various peptide-binding proteins in bound form. The proteins used to calculate the binding-site volume are from *Lactococcus lactis* (*Ll*OppA, A2RJ53), *Helicobacter pylori* (*Hp*DppA, A0A3Q9Y393), *Bacillus subtilis* (*Bs*AppA, P42061), *Thermus thermophilus* (*Tt*OppA, Q5SHU6), *Mycobacterium tuberculosis* (*Mtb*DppA, I6X811), *E. coli* (*Ec*OppA, P23843), *E. coli* (*Ec*MppA, P77348), *S. typhimurium* (*St*OppA, P06202), *Yersinia pestis* (*Yp*OppA, A0A0H2W5P5), *Pseudoalteromonas sp.* (*Ps*DppA, A7Y7W1) and *E. coli* (*Ec*DppA, P23847); The UniProt id of the proteins used has been mentioned within parenthesis. (B) Cartoon representation of *Ec*SapA wherein the ligand binding cavity has been shown using red surface representation. (C) Superimposition of *Ec*SapA (yellow) along with *Ps*DppA (cyan), *Ec*DppA (green), *Mtb*DppA (magenta), and *Ll*OppA

(red) revealed a similar binding-site pocket. Two residues, tryptophan and aspartate, were observed to be conserved, and the conserved residue numbers have been mentioned in parentheses. The orientation of Trp437 and Asp440 of *Ec*SapA is in accordance with the orientation of Trp473 and Ala476 of *Ll*OppA. The overlayed proteins have been represented using cartoon representation, the dashed box represents the zoomed active site, and the conserved residues have been represented with ball and stick models. The distance between the Trp437 and Asp440 of *Ec*SapA has been mentioned adjacent to the yellow double-headed arrow, and the distance between Trp405 and Asp408 of *Ec*DppA has been mentioned adjacent to the green double-headed arrow.

4.3.4. The electrostatic potential charge distribution of *Ec*SapA favors AMP binding

AMPs are mostly rich in positively-charged residues and thus tend to primarily interact with negatively-charged surfaces of the membrane. Such electrostatic potential charge-based interaction has given rise to the hypothesis that the binding of AMPs necessitates the presence of a negatively charged binding-site core. Hence, an investigation of the electrostatic potential charge distribution of the binding site of *Ec*SapA might hint towards the kinds of AMPs that bind to it. However, the lack of AMP-binding SBPs as homologs of *Ec*SapA limits such investigation. Thus, the tertiary structure of a total of six well-reported SapA proteins, viz. *St*SapA, *Hi*SapA, *Kp*SapA, *Af*SapA, *Pm*SapA, and *Dd*SapA were modeled to comparatively analyze the binding-site charge distribution (Parra-Lopez *et al.*, 1993; Lopez-Solanilla *et al.,* 1998; Chen *et al.,* 2000; McCoy *et al.,* 2001; Shelton *et al.,* 2011; Hsu *et al.,* 2019). All these SapA proteins belong to cluster C SBPs, which have been implicated in utilizing the ‘one or two domain movement’ mechanism of ligand binding (Chandravanshi *et al.,* 2021). In this mechanism, domain III (flexible domain) moves towards domain II with the help of domain I, consisting of two subdomains (hinge I and II). Domain I enables the movement of Domain III, leading to the occlusion of the available binding site (Figure 4.5). The movement of domain III leads to two distinct conformations, viz. open and closed, of SapA proteins. As the closed state cannot be utilized for the rigid docking calculations, all seven models were generated in an open

conformation. The open conformation was confirmed by analyzing the degree of domain movement in comparison to the closed models of each SapA protein. In each case, a high degree of movement (32.2° to 48.8°) was observed, as seen in *Ps*DppA (38.3°) (Figure 4.5 and Table B3).

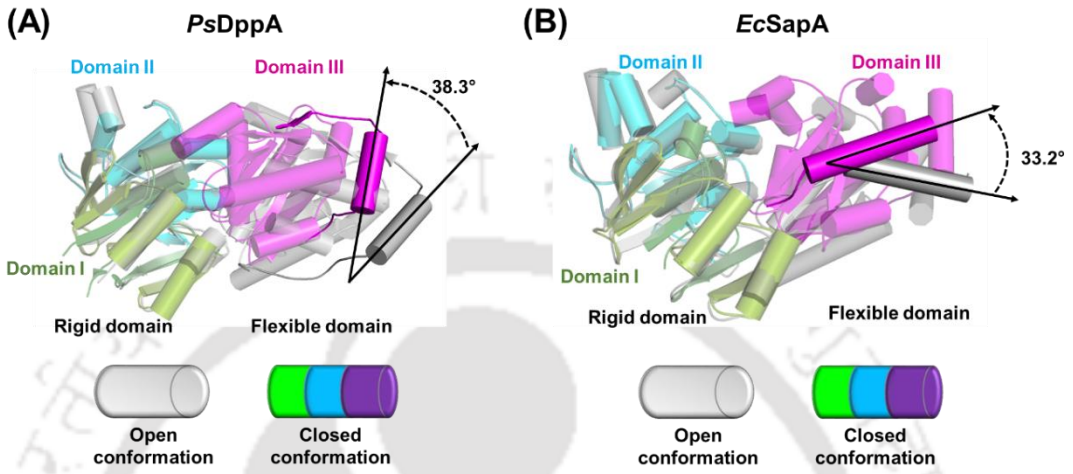


Figure 4.5. Domain movement in *Ec*SapA. (A) On superposing the open conformation of *Ps*DppA (PDB id: 4QFK) with the closed conformation of *Ps*DppA (PDB id: 4QFL), a significant movement is observed in the Domain III of the closed conformation (magenta) with respect to the open conformation (grey). The angle of movement has been denoted by a dashed arrow. (B) Significant domain movement of *Ec*SapA open and closed models could be observed, wherein prominent movement was observed in Domain III of the closed conformation in contrast to the open conformation.

The electrostatic potential charge distribution of AMPs depicted them to be rich in positively charged residues in accordance with our expectations (Figure 4.6A). Further, the binding-site core of *Ec*SapA is rich in negatively-charged residues, in addition to a few positively-charged residues (Figure 4.6B). The electrostatic potential charge distribution of the binding site of *Ec*SapA is similar to all other SapA proteins, except *Af*SapA, which consists of completely negatively-charged residues (Figure 4.6B-4.6H). The presence of a few positively-charged residues within the binding site of SapA may be postulated to provide an environment for accommodating the sporadically present negatively-charged residues in AMPs

(Figure 4.6A-4.6G). The results indicate that *Ec*SapA provides evidence for its charge-based AMP-binding capacity.

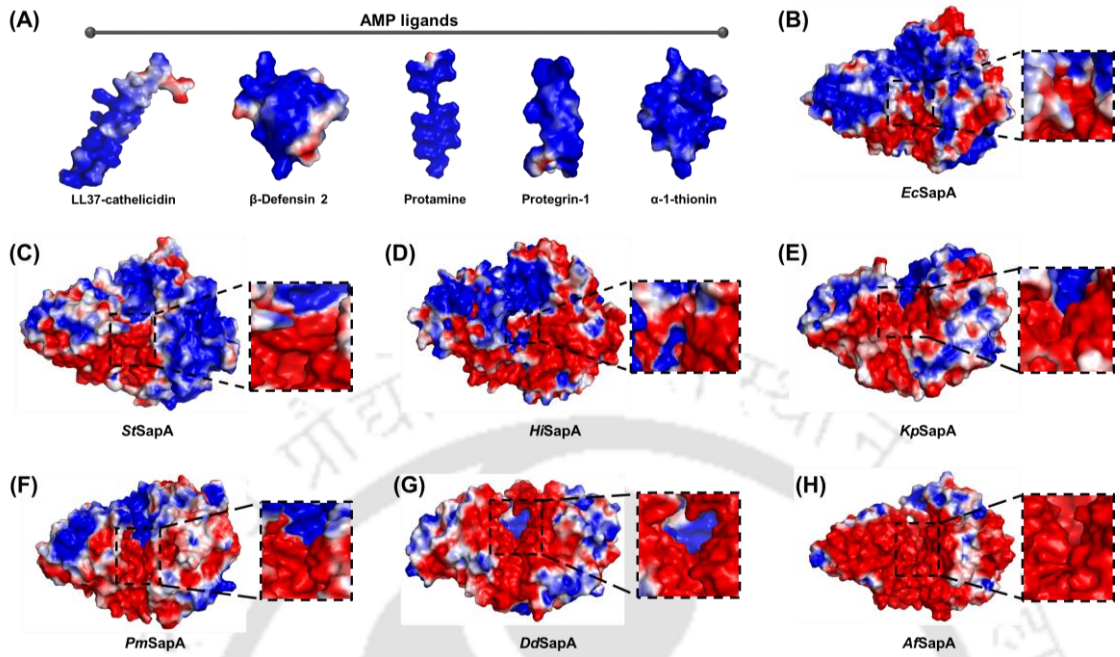


Figure 4.6. Electrostatic potential charge distribution of AMP ligands and selected SapA proteins. (A) Electrostatic charge distribution of AMP ligands was reported to interact with SapA. Apart from (B) *E. coli* SapA (*Ec*SapA, Q47622), the protein models used for comparison of binding-site electrostatic charge distribution were taken from (C) *S. typhimurium* SapA (*St*SapA, P36634), (D) *H. influenzae* SapA (*Hi*SapA, P45285), (E) *Klebsiella pneumoniae* SapA (*Kp*SapA, W8UWE0), (F) *Proteus mirabilis* (*Pm*SapA, B4EWK8), (G) *Dickeya dadantii* SapA (*Dd*SapA, E0SFZ5), and (H) *Aliivibrio fischeri* SapA (*Af*SapA, Q5E0R7). The UniProt id of the proteins used in the study has been mentioned in parentheses. The positively-charged and negatively-charged regions have been demarcated by blue surface (+1 kT/e) and red surface (-1 kT/e), respectively.

4.3.5. Protein-protein docking suggests AMP binding to SapA

Although the binding-site architecture and the potential charge distribution of *Ec*SapA hint at its potential binding capability to AMPs, no report to date is available in the context of *Ec*SapA bound to AMPs. However, *St*SapA has been implicated as being responsible for resistance against the AMP protamine (Parra-Lopez *et al.,* 1993). Notably, the protein *Ec*SapA shares a high similarity with

*St*SapA (query coverage (QC): 100% and sequence identity (SI): 90%). Thus, it is imperative to speculate that the protein *Ec*SapA can bind to protamine. As no structural insights of SapA bound to AMPs are available in the literature, protein-protein docking calculations were performed with five AMPs, viz. protamine, LL-37 cathelicidin, β-defensin 2, protegrin-1, and α-1-purothionine (Figure 4.7A) and SapA proteins, viz. *St*SapA, *Hi*SapA, *Kp*SapA, *Pm*SapA, *Dd*SapA, *Af*SapA, and *Ec*SapA.

The protein-protein docking results revealed that *Hi*SapA can bind to LL-37 cathelicidin and β-Defensin 2, in corroboration with earlier reports (Mason *et al.,* 2011). Further, *Hi*SapA also demonstrates its ability to bind protegrin-1 and protamine (Figure 4.7B). Interestingly, *Hi*SapA failed to bind α-1-purothionine. This demonstrates its selectivity towards animal-produced AMPs (LL-37 cathelicidin, β-defensin 2, protegrin-1, and protamine) rather than plant-based AMPs (α-1-purothionine). Similar results were obtained for *Kp*SapA, which showed binding to animal-based AMPs (LL-37 cathelicidin, protegrin-1, and protamine) but failed to bind β-defensin 2 and α-1-purothionine (Figure 4.7C). These results corroborated and validated the protein-protein docking methodology, as *Kp*SapA has been reported to bind LL-37 cathelicidin rather than β-defensin 2 (Hsu *et al.,* 2019). In the cases of *Pm*SapA and *Dd*SapA, the AMPs protegrin-1 and α-1-purothionine could dock, in corroboration with previous reports (Lopez-Solanilla *et al.,* 1998; McCoy *et al.,* 2001) (Figure 4.7D-4.7E). The protein *St*SapA shows binding to its cognate AMP protamine (Parra-Lopez *et al.,* 1993) in addition to protegrin-1 (Figure 4.7F).

Since the results obtained for these proteins through protein-protein docking calculations corroborated the reported literature, the proteins *Af*SapA and *Ec*SapA were also subjected to docking with all five AMPs. Notably, akin to *St*SapA, both *Af*SapA and *Ec*SapA bind protamine and protegrin-1 (Figure 4.7G-4.7H).

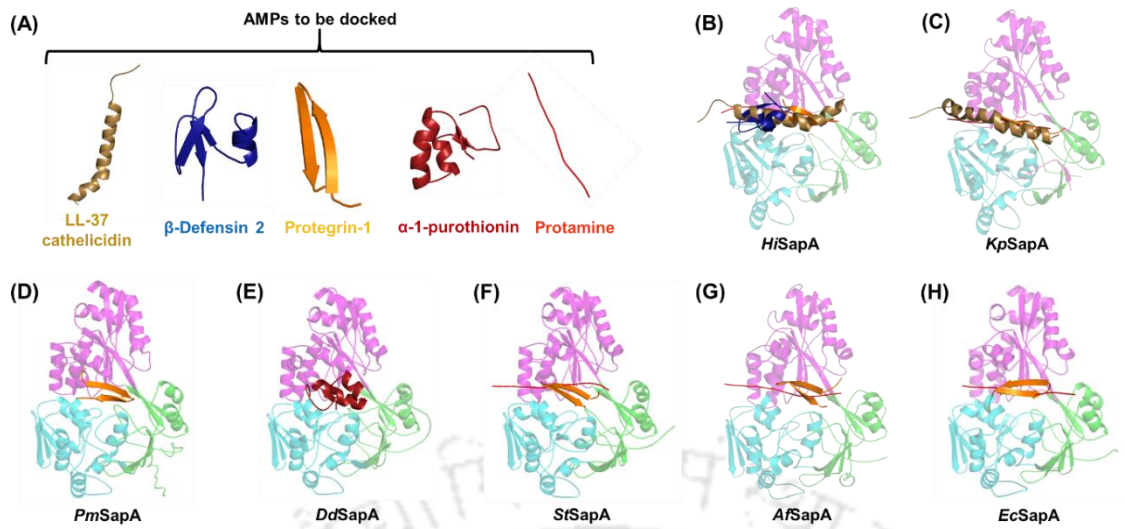


Figure 4.7. AMP docking studies with selected SapA proteins. (A) Cartoon representation of the AMPs to be docked against selected SapA proteins. The AMPs used are LL-37 cathelicidin (gold), β-Defensin 2 (blue), protegrin-1 (orange), α-1-purothionin (deep red) and protamine (light red). The cartoon representation of (B) *H. influenzae* SapA (*Hi*SapA, UniProt id: P45285), (C) *K. pneumoniae* SapA (*Kp*SapA, UniProt id: W8UWE0), (D) *P. mirabilis* SapA (*Pm*SapA, UniProt id: B4EWK8), (E) *D. dadantii* SapA (*Dd*SapA, UniProt id: E0SFZ5), (F) *S. typhimurium* SapA (*St*SapA, UniProt id: P36634), (G) *A. fischeri* SapA (*Af*SapA, UniProt id: P36634), and (H) *E. coli* SapA (*Ec*SapA, UniProt id: Q47622) depict docking of their probable AMP ligands within their binding-site core. The three distinct domains of all the SapA proteins have been color-coded as domain I (green), domain II (cyan), and domain III (magenta).

4.3.6. *Ec*SapA binds protamine and protegrin-1 via arginine anchoring machinery

The protein-protein docking calculations suggested that *Ec*SapA can bind protamine and protegrin-1 with a similar binding energy of -1009.1 kcal mol^{-1} and -1121.3 kcal mol^{-1}, respectively, much higher than that of dipeptides (-8.59 to -8.01 kcal mol^{-1}). An in-depth analysis of the interactions between the protein and AMPs revealed multiple types of contacts, including salt bridges, hydrogen bonds, and hydrophobic interactions. Protamine is stabilized by five distinct salt bridges formed between the residues Asp59, Asp91, Glu422, and Asp440 of *Ec*SapA and Arg3, Arg13, Arg4, Arg6, and Arg7 of the AMP (Figure 4.8A and

Table B4). This suggests that the interactions are primarily charge-specific, involving arginine residues of the AMP and the acidic residues of the protein *Ec*SapA.

In contrast, protegrin-1 is stabilized by two salt bridges between Asp59 and Asp440 of *Ec*SapA and Arg4 and Arg11 of protegrin-1 (Figure 4.8B and Table B4). In this case, the involvement of arginine residues in the binding of protegrin-1 to *Ec*SapA is comparatively less pronounced due to the low number of arginine residues in protegrin-1 (6) compared to protamine (10).

In addition to the salt bridges, a hydrogen bond between Arg385 and Ser435 of *Ec*SapA and Tyr7 and Arg9 of protegrin-1 are involved in the interaction (Figure 4.8B and Table B4). Further, a π-π stacking interaction between Trp437 of *Ec*SapA and Phe12 of protegrin-1 contributes to the stabilization of the AMP (Figure 4.8B). In summary, charge-specific interactions and the utilization of arginine residues as anchors in the AMP-binding machinery of *Ec*SapA are crucial.

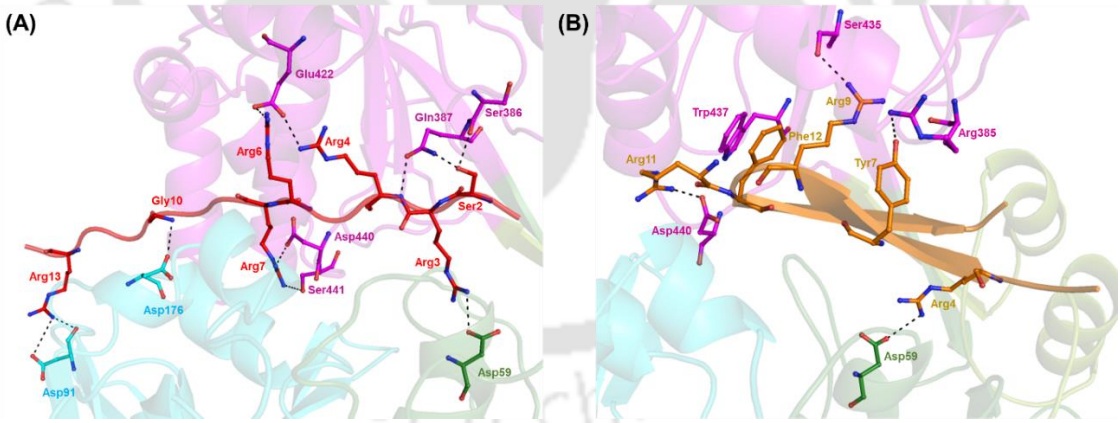


Figure 4.8. Arginine-based AMP-binding machinery of *Ec*SapA. Protein-protein docking analysis of *Ec*SapA with (A) protamine (red) and (B) protegrin-1 (orange). The interacting residues of protamine and protegrin-1 have been represented with red and orange ball-and-stick models, respectively. The interacting residues of *Ec*SapA have been represented with a ball-and-stick model and colored based on their distribution in domains viz., Domain I (green), Domain II (cyan), and Domain III (magenta).

4.3.7. *Ec*SapA can bind heme

In *H. influenzae*, the Sap transport system has been well-reported to be involved in heme uptake with the help of the SBP SapA (Tanaka and Pinkett, 2019; Rodríguez-Arce *et al.,* 2019; Lukacik *et al.,* 2021). Thus, molecular docking calculations were performed to investigate whether *Ec*SapA can bind to heme. As a positive control, *Hi*SapA was docked with a ferrated heme molecule, which showed that heme can bind *Hi*SapA with a binding energy of -10.98 kcal mol^{-1}. The heme molecule is coordinated by the residues Arg438 and Arg534 of *Hi*SapA (Figure 4.9A and Table B1). Although apart from salt bridges, various other interactions through the residues tyrosine/histidine to hold the bound ferrous atoms have been reported (Mandal *et al.,* 2019), no such interaction was observed with the bound ferrous atom in *Hi*SapA. However, the residues His41 and Tyr409 were observed to be in proximity to the docked heme molecule (Figure 4.9A). This is likely due to the artifact of the rigid docking methodology. In solution, these residues may be speculated to interact and stabilize the ferrous atom.

In a similar manner, a blind docking of heme was performed against *Ec*SapA. The results reveal that the heme molecule was docked in the binding site of *Ec*SapA, located at the interphase of domains I and III, akin to *Hi*SapA. Further, the propionate chains A and D of heme form salt bridges between Arg385 and Gln43. Although the binding energy (-8.35 kcal mol^{-1}) of heme with *Ec*SapA is lower than that with *Hi*SapA, the interactions are almost similar in both proteins (Figure 4.9B and Table B1). Notably, instead of histidine or tyrosine, as in *Hi*SapA, the residue Thr384 of *Ec*SapA interacts with the ferrous atom of heme. The residues Thr384 and Ser41 of *Ec*SapA are spatially located at the proximal positions of the ferrous atom of heme. Thus, these residues are likely to stabilize the ferrous atom of heme in *Ec*SapA (Figure 4.9B).

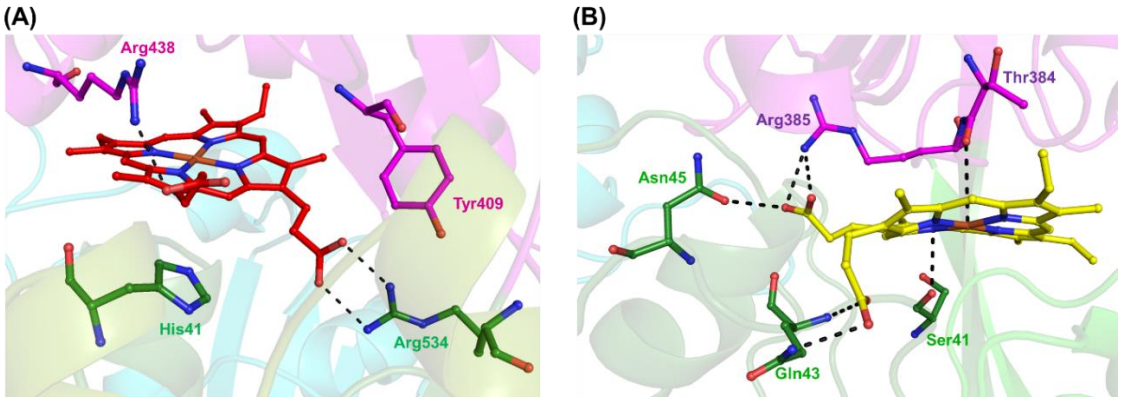


Figure 4.9. Heme docked poses of *Hi*SapA and *Ec*SapA. (A) Heme molecules (red ball-and-stick model) were observed to dock within the binding-site interface of *Hi*SapA, wherein the interacting residues have been represented with a ball-and-stick model. The color of the residues represents the domain they belong to, viz., domain I (green) and domain III (magenta). (B) Heme molecules (yellow ball-and-stick model) were observed docking in the binding-site interface of *Ec*SapA, wherein the interacting residues have been represented as ball-and-stick model. The color of the residues represents the domain they belong, and the domain color scheme is the same as in the case of *Hi*SapA.

4.3.8. Sequence and structural features suggest an importer-like function of *Ec*SapBC

In *E. coli*, the membrane component (SapBCDF) of the Sap transport system has been reported to function as a putrescine exporter, SapA being an orphan SBP unrelated to the Sap system (Sugiyama *et al.,* 2016). However, our results discussed in Chapter 3 suggested *sapA* to be a part of the *sap* operon. Further, the results suggested that the SBP component of the transporter, SapA, can function in synchrony with other components.

In this study, further exploration of the function of *Ec*SapA has been performed by identifying its putative cognate ligands. A sequence homology search and domain analysis of *Ec*SapBC revealed its similarity with the members of type-I ABC importers. Further, evolutionary analysis shows that both *Ec*SapBC and *Hi*SapBC clade alongside the members of type-I ABC importers (Figure 4.10A).

Interestingly, ABC importers and exporters are cladded into two distinct clusters, highlighting their disparity (Figure 4.10A).

Further, the tertiary structure models of *Ec*SapBC reveal that each subunit contains six transmembrane helices (TMHs), akin to ABC type-I importers (Figure 4.10B). Furthermore, akin to ABC type-I importers, both the components *Ec*SapBC possess a coupling helix (CH) in between the helices TMH4 and TMH5 (Figure 4.10B). In contrast, ABC exporters contain CH in between TMH2 and TMH3. Also, ABC exporters possess an additional helix, referred to as the "elbow helix" (Thomas *et al.,* 2020). These observations suggest that *Ec*SapBC is similar to importers rather than exporters.

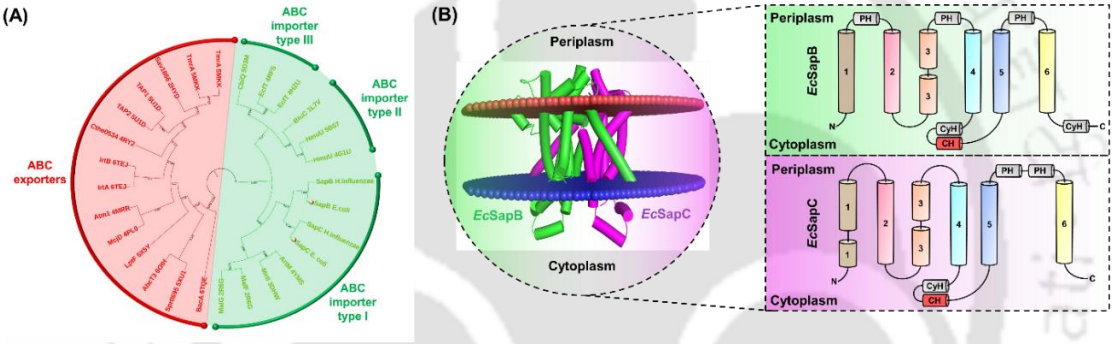


Figure 4.10. Sequence and structure-based analysis of *Ec*SapBC. (A) All the proteins utilized to generate the sequence-based phylogenetic tree have been referred to by gene name followed by their corresponding PDB id. The proteins *Ec*SapB (UniProt id: P0AGH3) and *Ec*SapC (UniProt id: P0AGH5) highlighted with golden sphere were observed to clade alongside members of ABC importer type I alongside *Hi*SapB (UniProt id: P45286) and *Hi*SapC (UniProt id: P45287). Further, ABC importers (green) and exporters (red) were observed to form two distinct clades. (B) The right panel consists of membrane-embedded cartoon representations of *Ec*SapB (green) and *Ec*SapC (magenta), wherein the upper leaflet of the inner membrane is represented as red spheres and the lower leaflet is represented as blue spheres. The left panel represents the topological arrangement of *Ec*SapB (green) and *Ec*SapC (magenta), representing six transmembrane helices (TMH1 brown, TMH2 red, TMH3 orange, TMH4 cyan, TMH5 blue, and TMH6 yellow) passing the bacterial inner membrane. Akin to type I ABC importers, in the case of both *Ec*SapBC, a coupling helix (CH, red)

was observed to be present in between TMH4 and TMH5. Several periplasmic helices (PH) and cytoplasmic helices (CyH) are also observable in the topology diagram of *Ec*SapBC.

4.3.9. Charges play a crucial role in the assembly of the *Ec*Sap transport system

For the Sap transport system to function as an importer, all five subunits (SapABCDF) need to assemble in the membrane. To investigate the possibility of the components forming a complex, their individual electrostatic potential charge distribution was calculated. In *Ec*SapA, the binding-site region in the closed conformation possesses two halves, which are positively- and negatively-charged regions. This complements the charge distribution on the periplasmic side of the *Ec*SapBC (Figure 4.11). The NBDs (*Ec*SapDF) are reported to interact with TMDs (*Ec*SapBC) through the CH (Dawson and Locher, 2006). Further, a complementary electrostatic potential charge distribution is observed on the NBDs facing the CHs (Figure 4.11).

The analysis also suggests that the complex *Ec*SapBCDF is more stable than *Ec*SapABC due to the absence of any domain movement in the NBDs. The transient binding of *Ec*SapA to the TMD components (*Ec*SapBC) is likely achieved by the disruption of charges upon domain opening (Figure B4). This, most likely, helps transport the substrate across the membrane components.

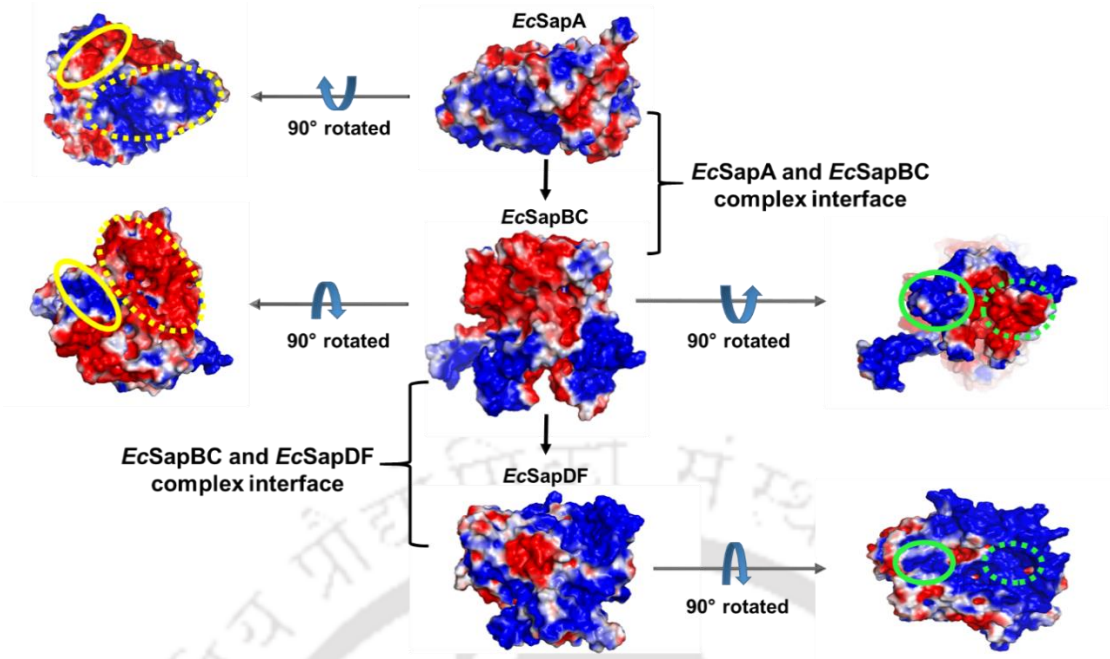


Figure 4.11. Charge-based assembly analysis of *Ec*SapABCDF. Electrostatic potential charge distribution of the complex components (*Ec*SapA, *Ec*SapBC, and *Ec*SapDF) revealed the mechanism of assembly of the complex system. The central panel represents the probable assembly order of the *Ec*SapABCDF complex. The left panel represents the interacting interface of *Ec*SapA onto *Ec*SapBC, represented by dashed and solid yellow ovals on each counterpart. The right panel represents the region of contact of *Ec*SapDF with *Ec*SapBC, represented by dashed and solid green circles on each counterpart. The matching circular patterns represent the region of contact for the complex formation. The positively-charged and negatively-charged regions have been demarcated by blue surface (+1 kT/e) and red surface (-1 kT/e), respectively.

4.4. Discussion

SapA is the periplasmic component of the Sap transport system and is responsible for ferrying AMPs across the bacterial inner membrane via their corresponding TMDs (SapBC) and NBDs (SapDF) (Shelton *et al.,* 2011). Post internalization, the AMPs are reported to be rendered inactive via bacterial proteases (Shelton *et al.,* 2011). Kurihara and coworkers experimentally suggested *Ec*SapBCDF to be involved in putrescine export and, via basic *in silico* predictions, suggested *Ec*SapA to be independent of *Ec*Sap transporter

(Sugiyama *et al.,* 2016). However, on the contrary, our results discussed in Chapter 3 have established *EcsapA* to be a part of the *Ecsap* operonic arrangement. Having established the correlation of the *sap* genes, this study delves into the functional aspects of the *Ec*Sap transport system employing multiple computational strategies.

Sequence-based homology search unraveled the similarity of *Ec*SapA along with the class of dipeptide-binding proteins. This observation was in accordance with prior reports wherein *Hi*SapA was also observed to possess similarities with dipeptide-binding proteins (Lukacik *et al.,* 2021). *Ps*DppA was observed to be the closest homolog of *Ec*SapA. However, on analysis of the binding-site residue conservation amongst *Ps*DppA and *Ec*SapA, partial conservation was observed, indicative of a distinct peptide-binding possibility of *Ec*SapA. Further, the molecular docking studies established the dipeptide-binding capacity of *Ec*SapA, which favored aromatic and hydrophobic residues. Akin to *Ps*DppA, *Ec*SapA was also observed to majorly interact with the backbone atoms of the bound dipeptides. Apart from *Ec*SapA, molecular docking studies also unearthed 25 different dipeptide ligands of *Ps*DppA apart from its eight cognate ligands and, hence, corroborated with the promiscuous nature of this class of peptide-binding proteins (Li *et al.,* 2015; Fernando *et al.,* 2022). The promiscuous nature of this class of proteins may be a prerequisite to fulfill the peptide-based nutrition demand of the bacterial systems. Further, *Ec*SapA demonstrated a bias for aromatic residues, hence hinting towards AMP-binding possibility as AMPs are well reported to be rich in aromatic and basic residues (Huan *et al.,* 2020). The docked poses of the dipeptides revealed two different binding sites with *Ec*SapA, which have been coined as binding sites A and B. The presence of two binding sites within *Ec*SapA suggested a broader binding site capable of accommodating polypeptides. Apart from a broader binding site, the promiscuous nature of *Ec*SapA further enables the binding of peptides of larger sizes akin to AMPs.

Further, the comparative analysis of the binding-site volume of multiple peptide-binding proteins revealed the *Ec*SapA binding-site volume to be similar to that of *Mtb*DppA, which is a tetrapeptide-binding protein. The cognate ligand of *Mtb*DppA SSVT was observed to dock in the interphasic region of binding sites

A and B of *Ec*SapA with a weak binding energy (-4.95 kcal mol^{-1}). The low binding energy suggests scope for plausible transient binding of SSVT to *Ec*SapA. The docked pose of SSVT in *Ec*SapA indicated binding sites A and B to combinatorially function as a single binding site of *Ec*SapA. Further, the spatial arrangement of the peripheral binding-site residues Trp437 and Asp440 of *Ec*SapA were observed to vary from other DppA class of proteins but were similar to the spatial arrangement of Trp473 and Ala476 of *Ll*OppA, which has been reported to bind peptides up to 35 amino acids long (Berntsson *et al.,* 2009). The similar orientation of *Ec*SapA residues with *Ll*OppA is indicative of a wider binding site and, hence, indicative of AMP-binding possibility. Apart from aromatic residues, AMPs are also rich in basic residues such as arginine and lysine, hence imparting AMPs with a net positive charge, enabling charge-based interaction of positively-charged AMPs and negatively-charged bacterial membranes (Huan *et al.,* 2020). In complementarity to the charge of the AMP, the binding-site core of *Ec*SapA was observed to be rich in negative charge. A dearth of AMP-binding proteins as homologs compelled us to analyze the binding site of well-reported SapA proteins (*St*SapA, *Hi*SapA, *Kp*SapA, *Pm*SapA, and *Dd*SapA) that have been experimentally suggested to interact with their cognate AMPs (Parra-Lopez et al., 1993; Lopez-Solanilla *et al.,* 1998; Chen *et al.,* 2000; McCoy *et al.,* 2001; Shelton *et al.,* 2011 and Hsu *et al.,* 2019). The binding site of all five SapA proteins was observed to be rich in negatively-charged residues, and hence, all SapA proteins revealed competence for binding positively-charged AMPs in accordance with prior reports.

Despite multiple hints of AMP-binding possibility with *Ec*SapA, certain studies suggest no plausible role of *Ec*Sap transporter in resistance against LL-37 cathelicidin (Sugiyama *et al.,* 2016). Further, Bakker and coworkers observed no surge in the expression of *Ec*Sap transporter on exposure to protamine and hence attributed the system to not being involved in the uptake of protamine (Harms *et al.,* 2001). However, in the present study, the *Ec*Sap transporter system has been observed to be involved in nutrient uptake in the form of dipeptides and, hence, implicated in a housekeeping function. The role of nutrient acquisition of Sap transporters has been observed in the case of *S. oneindensis,* and we see hints of the same in *Ec*Sap transporter (Liang *et al.,* 2019). Hence,

as the *Ec*Sap transporter is present in normalcy in the bacterial system, there may not be a surge in exposure to protamine as the system moonlights and performs multiple functions altogether. In this study, protein-protein docking studies established the interaction of protamine and protegrin-1 with *Ec*SapA, which, however, failed to interact with human-produced AMPs such as LL-37 cathelicidin and β-defensin 2. Hence, *Ec*SapA may provide resistance against antibacterial cationic peptides, imparting the bacterial system resilience and the capacity to sustain itself within the host.

Both protegrin-1 and protamine were observed to be accommodated within *Ec*SapA binding sites A and B and further extended through the peripheral residues of the binding site A (Trp437 and Asp440). The formation of salt bridges amongst the arginine residues and the acidic residues of *Ec*SapA were distinctly observable in both docked poses. In the case of protamine, the binding machinery was primarily based on charge-based interactions, in accordance with our prior analysis. Meanwhile, protegrin-1 formed both hydrophobic contacts and salt bridges. Hence, it would be suitable to state that one of the key factors holding the AMPs within the binding site of *Ec*SapA is the salt bridges formed between the protein and the AMP. Hence, our results clearly demonstrate the essential requirement of charge complementarity between the AMPs and AMP-binding proteins. Further, in the case of protamine, certain regions of the AMP were observed to protrude out of the *Ec*SapA binding site. This mode of binding has been previously hypothesized in the case of *Ll*OppA, taking reference to the mechanism of oligosaccharide binding via maltose-binding proteins, wherein the long oligosaccharides were observed to protrude out of the protein's binding site (Quiocho *et al.,* 1997; Berntsson *et al.,* 2009). Hence, this mode of binding enables the closure of the domain of the SBP, which is most likely a prerequisite for docking onto the TMDs for ligand transfer (Doeven *et al.,* 2008).

Alongside dipeptides and AMPs, molecular docking studies revealed the heme-docked poses of *Hi*SapA and *Ec*SapA, wherein the heme molecule was observed to be docked within the interphase of domains I and III in both cases. The heme molecule was observed to occupy the binding site B in *Ec*SapA, and hence, the binding sites of heme, dipeptides, and AMPs coincide with one another. This hints

towards competitive binding machinery for each ligand, wherein one is replaced by the other. Prior studies have reported competitive binding in the case of *Hi*SapA, wherein the heme molecule was observed to be replaced by highly charged AMPs while the dipeptides failed to replace heme (Mason *et al.,* 2011). Hence, it may be implied that *Ec*SapA has an essential role in nutrient uptake, resistance against AMPs, and heme sequestration, and it functions in a priority-based manner. When the bacterial system is under the threat of AMPs or needs to sequester iron to gain an advantage over the hosts, *Ec*SapA might prioritize the binding of AMPs/heme over dipeptides, in such a case. Hence, this follows the trend of binding affinity (positively charged AMPs > heme > dipeptides) depicted via *Hi*SapA in earlier reports (Mason *et al.,* 2011).

Having understood the functional aspects of the SBP *Ec*SapA, this study steers toward the putative machinery of assembly and uptake of the ligand bound to *Ec*SapA via *Ec*SapBCDF. Sequence and structure-based analysis revealed clear similarities between *Ec*SapBC and ABC type I importers wherein there are 6 TMHs in each component, and the CHs are present in between TMH4 and 5, akin to type I ABC importers (Thomas *et al.,* 2020). Hence, it further cements the synchronous function of the *Ec*Sap transport system components in uptaking a wide variety of ligands essential to providing resistance/nutrition to the bacterial system. Further, none of the structural features reported to be part of the ABC exporter were observed in the case of *Ec*SapBC, hence questioning the scope for the *Ec*Sap transporter to function as a putrescine exporter. However, since experimental evidence suggests the exporter function of *Ec*SapBCDF, it is not possible to negate the same, and hence, a plausible dual functionality may be attributed to this system (Sugiyama *et al.,* 2016). However, in the recent past, there have been instances of ABC transport systems functioning as both importers/exporters, like IrtAB and Mla systems (Arnold *et al.,* 2020; Low *et al.,* 2021). However, both IrtAB and Mla systems have been classified as type IV and VIII ABC exporters, respectively, and hence, there is a dearth of any dual functionality in any ABC transporter classified as an importer (Thomas *et al.,* 2020; Ekiert *et al.,* 2022). The topological arrangement of *Ec*SapBC is fully in accordance with ABC type I importers, and hence, the possibility of importer/exporter functionality of *Ec*SapBCDF appears bleak. However, further

experimental evidence is required to clarify the dual-functioning possibility of *Ec*SapBCDF.

The *Ec*Sap transport system further reveals a charge-based assembly machinery wherein *Ec*SapA (closed conformation) and *Ec*SapBC are observed to have complementary charge distribution, hence enabling the docking of *Ec*SapA onto *Ec*SapBC. However, in the open conformation of *Ec*SapA, this distribution gets distorted and, hence, disrupts the complex. This disruption of charge distribution further entices us to propose that the opening of *Ec*SapA may lead to the opening of *Ec*SapBC towards the periplasmic side and, in turn, lead to the transport of the bound ligand. Following the ligand transfer, *Ec*SapDF will hydrolyze adenosine triphosphate (ATP) to adenosine diphosphate (ADP), leading to the opening of *Ec*SapBC towards the cytoplasmic side and allowing the transfer of the ligand into the cytoplasm. This proposed machinery arising from the charge-based assembly of the *Ec*Sap transporter system is similar to the machinery reported to be followed by the ABC type I importers (Rice *et al.,* 2014). Hence, all the outcomes of this study are indicative of the multifarious attributes of the *Ec*Sap transporter, wherein the system is implicated in the uptake of multiple essential ligands crucial for the sustenance of the bacterial system within the host.

4.5. Conclusion

The current study delves into the underexplored stature of the *Ec*Sap transport system and attempts to decipher the plausible functional attributes of the system. A comprehensive *in silico* approach has been strategized in the study, which makes an effort to encompass and answer multiple queries attributed to this system. The study unravels and highlights the dipeptide-binding capability of *Ec*SapA. Further, the binding site of *Ec*SapA consists of signs indicative of a broader binding site and, hence, may be capable of accommodating AMPs. In accordance with the prior trend, *Ec*SapA was observed to dock protamine and protegrin-1 but failed to dock human-based AMPs. Further, via molecular docking strategies, the heme-binding propensity of *Ec*SapA was unveiled, underscoring the multifaceted role of *Ec*SapA. Alongside, the study also highlights the ABC importer-like attributes of *Ec*SapBC via sequence and structure-based analysis

and further suggests charge-based assembly machinery for the *Ec*SapABCDF complex, enabling the uptake of the substrate. The present study successfully suggests the role of the *Ec*Sap transport system in nutrient uptake, resistance against AMPs, and ferrous ion sequestration. Which, in turn, provides compelling evidence of its significant role in the sustenance and pathogenesis of the bacterial system within the human host. Consequently, our study successfully provides valuable insights for drug designing strategies against the ubiquitously present Sap transport system.

Appendix B. Supplementary data

Supplementary Figures: B1-B4

Supplementary Tables: B1-B4

Chapter 5

Molecular Dynamics simulation studies of *Ec*SapA

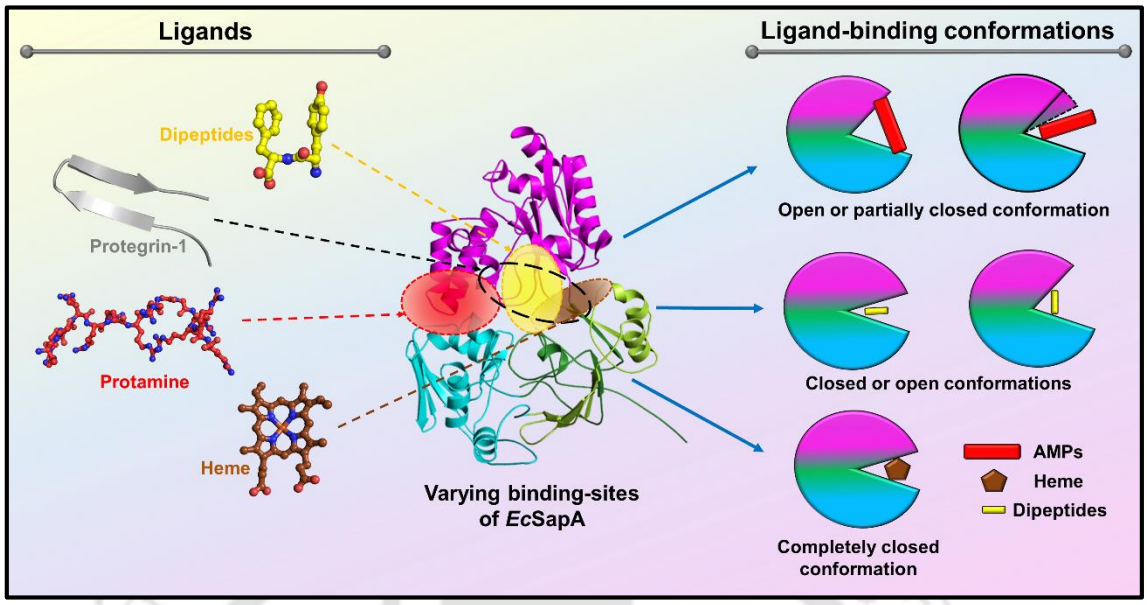
The manuscript for this chapter is under preparation as:

Dasgupta P and Kanaujia SP. Unlocking conformational dynamics of *Ec*SapA: Insights into the emphasis of ligand dependency using molecular dynamics simulations.

Abstract

The human host utilizes antimicrobial peptides (AMPs) to clear off a horde of pathogens. To overcome the action of AMPs, the Gram-negative bacterial pathogens have devised multiple types of machinery that hinder the action of the AMPs and enable the bacterial cell to form a suitable niche. An ABC importer termed as the sensitive to antimicrobial peptides (Sap) transport system is one such machinery reported to import the AMPs into the bacterial cytoplasm, wherein the AMPs are proteolytically degraded via bacterial proteases. Interestingly, the *Escherichia coli* Sap (*Ec*Sap) transport system was reported to function as a putrescine exporter; further, the SapA component was banished from the *sap* operon. However, our results discussed in Chapter 3 suggest otherwise, wherein robust computational strategies established *EcsapA* to be part of the *Ecsap* operon. Further, the outcomes of Chapter 4 suggested multifunctional prospects of *Ec*SapA in the uptake of an array of ligands (dipeptide, AMPs, and heme). In this study MD simulation was employed to establish the validity of the docked poses, as obtained in Chapter 4. The results demonstrated four dipeptides (Trp-Pro, Trp-Ile, Tyr-Ile, and Tyr-Phe), two AMPs, and heme to be retained within the *Ec*SapA binding-site throughout 1 µs time period of MD simulation. Further, the results also highlight ligand-specific domain movement to accommodate the varying ligands suggested for binding *Ec*SapA. Hence, this study enlightens the promiscuous nature of the *Ec*SapA binding site, making it a lucrative target for drugs.

Graphical abstract



5.1. Introduction

The host-pathogen tussle is a prerequisite prior to any pathogenic bacterial invasion (Mues and Chu, 2020). One of the key armors in the host defense machinery is the antimicrobial peptide (AMPs) (Wang, 2014). The AMPs obliterate the homeostasis of the bacterial membrane and hence hinder the formation of a niche for the pathogenic bacterial cells (Zhang *et al.,* 2021). However, to evade the action of the AMPs, the Gram-negative bacteria have developed robust mechanisms that neutralize the lytic effect of the AMPs. These types of machinery encompass multiple strategies such as forming a shield (capsule, slime layer, etc.) over the outer membrane, proteolytic cleavage of the AMPs via outer membrane proteases, modifying the charge of the bacterial outer membrane, and down-regulating the production of AMPs (Gruenheid and Le Moual, 2012). When the AMPs surpass all the aforementioned strategies, the bacteria resort to the transporters (exporters or importers) to clear off the AMPs (Blair *et al.,* 2022).

One such transporter is the sensitive to antimicrobial peptides (Sap) transport system first discovered in *Salmonella typhimurium*, from protamine-sensitive mutants. The *S. typhimurium* Sap (*St*Sap) transport system was reported to

possess five components comprising of a substrate-binding protein (SBP) SapA, two transmembrane domains (TMDs) SapBC, and two nucleotide-binding domains (NBDs) SapDF (Parra-Lopez *et al.,* 1993). Further, the *St*Sap transport system was hypothesized to import AMPs into the cytosol, wherein the AMPs are proteolytically degraded via bacterial proteases (Parra-Lopez *et al.,* 1993). The Sap transport system identified was observed to be ubiquitous amongst the Gram-negative pathogens (Parra-Lopez *et al.,* 1993; Hsu *et al.,* 2019). Amongst the pathogens, multiple studies have been reported pertaining to the non-typeable *Haemophilus influenzae* (NTHI) Sap (*Hi*Sap) transport system, wherein the *Hi*Sap transport system was implicated for AMP-uptake (Mason *et al.,* 2005, 2006, 2011). With the aid of elegant biochemical strategies, the uptake and proteolytic degradation of AMPs facilitated by the *Hi*Sap transport system was established (Shelton *et al.,* 2011).

Apart from binding AMPs, multiple studies have enlightened the multifunctional attribute of the Sap transport system, which encompasses heme uptake, oligopeptide uptake, and enabling K^+ ion uptake (SapD-Trk system complex) (Mason *et al.,* 2006, 2011; Liang *et al.,* 2019). Interestingly, our results discussed in Chapter 4 highlight the multifunctional attribute of *Ec*SapA, wherein docking results unraveled the dipeptide, AMP, and heme-binding propensities of *Ec*SapA. All the ligands were observed to bind *Ec*SapA in varying regions within the binding site. Hence, this piqued our interest in probing the structural variations demonstrated by *Ec*SapA to accommodate such an array of ligands and enlighten a probable binding mechanism pertaining to the various classes of ligands bound by *Ec*SapA.

In this study, all the docked poses of *Ec*SapA and their ligands (six dipeptides, AMPs, and heme) obtained in Chapter 4 were subjected to Molecular Dynamics (MD) simulations. The results of the MD simulations demonstrate distinct binding sites for varying ligands (dipeptides, AMPs, and heme). Further, the outcomes of this study also suggest ligand-specific domain movement of *Ec*SapA to accommodate varying ligands within the promiscuous and flexible binding site of *Ec*SapA.

5.2. Methodology

5.2.1. Retrieval of *Ec*SapA models and docked poses

The open and closed conformations of *Ec*SapA were modeled utilizing various programs such as SwissModel, RaptorX, Phyre2.0, ITASSER, and AlphaFold (Källberg *et al.,* 2012; Biasini *et al.,* 2014; Kelley *et al.,* 2015; Yang *et al.,* 2015; Varadi *et al.,* 2022). The retrieved models were refined with the aid of Galaxyweb to obtain refined structures (Ko *et al.,* 2012). Further, with the aid of the PDBsum database, the stereo-chemical properties of the models generated were validated (Laskowski *et al.,* 2018). Molecular docking studies discussed in Chapter 4 yielded six dipeptides (Phe-Trp, Trp-Phe, Trp-Ile, Trp-Pro, Tyr-Phe, and Tyr-Ile), two AMPs (protegrin-1 and protamine), and heme docked poses of *Ec*SapA. The optimal docked poses were selected, and their structural coordinates were saved for further MD simulations-based studies.

5.2.2. Molecular dynamics simulations

The GROMACS 2021.1. package was utilized to perform all the MD simulation-based studies (Abraham *et al.,* 2015). The systems to be simulated could be segregated into two distinct groups comprising ligand-free and ligand-bound forms of *Ec*SapA. The CHARMM36 all-atom force field was utilized in all the systems (Huang and MacKerell, 2013). The molecular topologies and coordinate files of the proteins were generated utilizing the program *gmx pdb2gmx* of GROMACS. Further, in the case of complexes, the topology and coordinate files of the ligands were generated with the aid of the CGenFF server (Vanommeslaeghe and MacKerell, 2012). Post-generation of topologies, the proteins are encased in a dodecahedron box with the aid of the module *editconf* present in the GROMACS suite, wherein the minimum distance between the solute and the edge of the box was maintained at 1 nm. The box consisting of the protein is then filled with transferable intermolecular potential with a a 3-points (TIP3P) water model with the aid of the module *genbox* present in GROMACS (Jorgensen *et al.*, 1983). Further, the net charge of the system was neutralized by adding chloride or sodium ions using the module *genion* present in GROMACS. The periodic boundary condition (PBC) was also applied in the system in all three directions to inhibit the edge effects. The energy minimization

of the systems generated was conducted using the steepest descent method with a maximum force cut-off of 1000 kJ/mol/nm. This energy minimization ensures that the structures in the system are relaxed and devoid of any steric clashes. The particle mesh Ewald (PME) method (Darden *et al*., 1993; Essmann *et al*., 1995) was employed for the calculation of long-range electrostatic interactions, and the short-range van der Waals interactions were calculated utilizing the Verlet neighbor list calculation with a cut-off of 1 nm. To constrain the bond lengths, the P-LINCS algorithm was used, and a time step of 2 fs was used to integrate the equations of motion (Hess, 2008). The first equilibration was executed under the canonical or the isothermal-isochoric NVT ensemble for 1000 ps keeping the reference temperature at 300 K. While the second was performed under isothermal-isobaric, NPT ensemble for 1000 ps, keeping the reference pressure of 1 bar. The reference temperature at 300 K was controlled by a velocity-rescaling thermostat with a coupling constant of 0.1 ps (Bussi *et al*., 2007). The Parrinello-Rahman barostat (Parrinello and Rahman, 1981) was utilized to control the reference pressure and the coupling constant, which were set to 100 kPa (1 bar) and 2 ps, respectively. Post adequate equilibration, all the systems were subjected to a production run for a time period of 1 µs for each system, resulting in a total of 11 µs of classical MD simulations (i.e., ~1 µs x 11 systems). For each run, the coordinates for systems were collected at every 10 ps for a total of 100,000 frames.

5.2.3. Simulation data analysis

The in-build GROMACS modules, such as *rms*, *rmsf*, *gyrate*, and *hbond,* were used to analyze MD trajectories. The program Xmgrace (Turner, 2005) was utilized to visualize the graphs generated from the trajectories, and for further analysis, the graphs were also visualized with the program Origin v9.6 (OriginLab Corporation, USA). Further three-dimensional coordinates of the frames from the trajectory were extracted using the module *trjconv*. The three-dimensional coordinates extracted were further visualized and analyzed with the aid of the software PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC).

5.3. Results

5.3.1. Conformational switching of *Ec*SapA is an outcome of external influences

The SBPs of ABC importers have been widely reported to demonstrate an open-to-closed conformation primarily owing to the selective/ combinatorial movement of the N-terminal domain (NTD) and the C-terminal domain (CTD) enabled by the hinge region (Chandravanshi *et al.,* 2021). The closed conformation of the SBP is predominantly suggested to be the holo conformation, wherein the ligand is bound to the SBP. Further, the apo conformation in SBPs corresponds to the open conformation, which is in the unliganded state (de Boer *et al.,* 2019). The results of Chapter 4 suggest that *Ec*SapA belongs to cluster C SBPs and has been observed to demonstrate both open and closed conformations with the aid of protein modeling strategies (Figure C1A-C1B). The SBPs of cluster C are reported to follow the same trend, wherein the apo and the holo state represent the open and close conformations (Chandravanshi *et al.,* 2021).

Hence, to map the transition of *Ec*SapA from apo to holo state or vice versa, the apo and holo (unliganded) conformations of *Ec*SapA were modeled and subjected to MD simulations. On analyzing the root mean square deviation (RMSD) progression, both states were observed to converge around ~300 ns (Figure 5.1A). However, despite similar trajectories, the apo state of *Ec*SapA was observed to demonstrate higher fluctuations in comparison to the holo (unliganded) state of *Ec*SapA (Figure 5.1A). Further, the root mean square fluctuations (RMSFs) revealed varying regions of fluctuations in the apo and holo conformations of *Ec*SapA. Overall, both the states (apo and holo) were observed to be mostly stable, with large fluctuations in the N- and C-termini of the protein (Figure 5.1B). However, some fluctuations observable in the apo state of *Ec*SapA are in the α-helix 13 present in domain III of the protein (Figure 5.1B and Figure C1A). Akin to the apo state, the holo (unliganded) state of *Ec*SapA is also observed to be mostly stable, with most fluctuations centered in the α-helix 9 and β-strand 12 region present in the hinge II (Figure 5.1B and Figure C1A). Interestingly, the region fluctuating in the apo state is stable in the holo (unliganded) state and vice-versa. Hence, the inversely related fluctuations are

an indication of the varied conformations, suggesting that neither of the forms (apo or holo) transitioned to the other form.

To further analyze the aforementioned transition, the radius of gyration (Rg) of the two forms was estimated across the trajectory. The initial Rg of the apo and the holo (unliganded) states were observed to be ~2.4 nm and ~2.5 nm, respectively (Figure 5.1C). With the progression of the trajectory, the Rg of the apo state was observed to increase from 2.5 nm to 2.6 nm, indicating further opening up of the apo state of *Ec*SapA. In the case of the holo (unliganded) state, the Rg was observed to fluctuate between 2.4 nm and 2.45 nm, suggesting no drastic conformational variation in the holo (unliganded) conformation in accordance with the RMSD results. To probe into the reasons for the fluctuations observed in the graphs, the structural coordinates of various frames (200 ns, 400 ns, 600 ns, 800 ns, and 1000 ns) were extracted from both trajectories. The structures obtained were overlaid and inspected for domain movements. In accordance with the Rg and RMSF data, the apo state reflected further opening up of *Ec*SapA, wherein the α-helix 13 (Domain III) was observed to translate, leading to the further opening of *Ec*SapA (Figure 5.1D). The holo (unliganded) state suggested a stable conformation, with some fluctuations pertaining to the hinge II (α-helix 9 and β-strand 12) region, akin to the prior observed Rg and RMSF data (Figure 5.1E).

Thus, the MD simulation data suggests that both the conformations (apo and holo) are stable in the free-floating state and revealed that there is no transition between the two states. Hence, it could be comprehended that the transition from open to closed conformation in *Ec*SapA may be dependent on external influences.

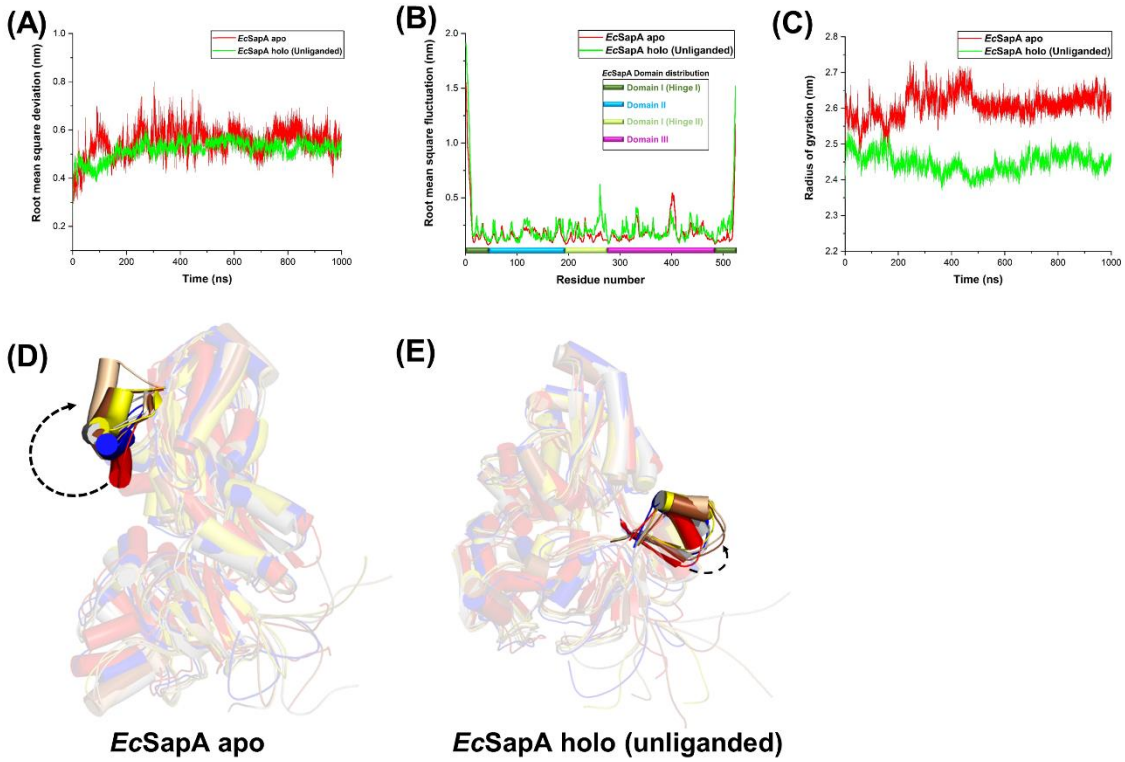


Figure 5.1. MD simulations-based studies of *Ec*SapA apo and holo (unliganded) conformations. The graphs of (A) RMSD, (B) RMSF, and (C) Rg were generated for the *Ec*SapA apo (red line) and *Ec*SapA holo (unliganded) (green line) conformations. Cartoon representation of tertiary structures of (D) *Ec*SapA apo and (E) *Ec*SapA holo (unliganded) extracted from the trajectories. In both cases, the structures have been color-coded based on their time points as follows: initial equilibrated structure (red), 200 ns (blue), 400 ns (grey), 600 ns (yellow), 800 ns (brown), and 1000 ns (wheat). The dashed arrows demonstrate the directionality of domain movement in the apo and holo (unliganded) conformations of *Ec*SapA.

5.3.2. MD simulations unravel the dipeptide selectivity of *Ec*SapA

Our results discussed in Chapter 4 have unraveled the dipeptide-binding propensity of *Ec*SapA. Six different dipeptides (Trp-Phe, Phe-Trp, Trp-Ile, Trp-Pro, Tyr-Phe, and Tyr-Ile) were observed to dock at two distinct locations referred to as binding sites A and B, wherein Trp-Phe, Phe-Trp, and Tyr-Phe were observed to dock within binding site A, and the remaining dipeptides docked in binding site B. All the dipeptide docked poses of *Ec*SapA were subjected to MD

simulation for a 1 µs time period. After the trajectory was generated, the RMSD plots of each *Ec*SapA-dipeptide complex were generated. The RMSD plots of Trp-Phe, Phe-Trp, and Tyr-Phe demonstrated no distinct fluctuations, and all three graphs were observed to converge at ~300 ns and demonstrate a similar trend akin to the holo (unliganded) state (Figure 5.2A). Both Trp-Ile and Trp-Pro were observed to converge at ~300 ns. However, both trajectories demonstrated distinct fluctuations in comparison to the apo and holo (unliganded) states. The initial observable fluctuations were observed to subside post-convergence. The most drastic fluctuation was observed in the case of Tyr-Ile, wherein the RMSD trajectory attained convergence much later at ~500 ns.

To understand the region of *Ec*SapA responsible for the fluctuations observable in the RMSD plots, all the *Ec*SapA-dipeptide complexes were subjected to RMSF-based analysis. Unlike the trends observed in the RMSD plots, the RMSF plots demonstrated three distinct trends. The first trend was observed to be similar for Phe-Trp and Trp-Phe, wherein the fluctuations were observed to be akin to the apo conformation of *Ec*SapA, hence suggesting the opening of *Ec*SapA in these two cases (Figure 5.2B). In the second trend, the dipeptides Trp-Ile and Trp-Pro were observed to stabilize fluctuations in the apo and holo (unliganded) conformations of *Ec*SapA (Figure 5.2B). Therefore, both ligands are observed to stabilize *Ec*SapA and potentially serve as candidates for the cognate ligands. Further, the third trend observed in the dipeptides Tyr-Phe and Tyr-Ile represents the fluctuations of the apo as well as the holo (unliganded) conformations of *Ec*SapA, suggesting the opening and closing of the protein (Figure 5.2B). Hence, unlike the RMSD plots, the RMSF plots provided insights into the movement of *Ec*SapA bound to dipeptides.

To further validate the RMSF plot data, the Rg of all the *Ec*SapA-dipeptide complexes was determined. The Rg graphs of Phe-Trp, Trp-Ile, and Trp-Pro were observed to be similar to the holo (unliganded) conformation; however, with progression, the protein is observed to partially open up in all three cases (Figure 5.2C). In the case of Trp-Phe and Tyr-Ile, the Rg graph suggests the protein transition from the closed state to the partially open state and then again back to the closed state (Figure 5.2C). Only in the case of Tyr-Phe was the Rg graph

observed to be in accordance with the holo (unliganded) conformation, hence suggesting stable conformation (Figure 5.2C). As the RMSD, RMSF, and Rg data depicted contradictory trends, the three-dimensional coordinates of various time frames were extracted for all six complexes and analyzed. The dipeptides Phe-Trp and Trp-Phe were observed to be displaced from the *Ec*SapA binding site by ~400 ns (Figure C2). Hence, the removal of these two dipeptides justifies the fluctuations observable in their RMSF and Rg plots. Further, the dipeptides Trp-Ile, Trp-Pro, Tyr-Phe, and Tyr-Ile were observed to be maintained within the *Ec*SapA binding site, hence suggesting them to be the cognate dipeptides of *Ec*SapA (Figure C2). However, the fluctuations observable in the RMSD, RMSF, and Rg graphs of Trp-Ile, Trp-Pro, and Tyr-Ile were unclear. On further inspection, a sliding machinery was observable wherein all three dipeptides initially docked in the binding-site B of *Ec*SapA however with the progression of the simulation, the dipeptides can be visualized to translate into the binding-site A of *Ec*SapA (Figure C2). To enable this translation, *Ec*SapA underwent transitions from closed to open to closed conformations. Hence, the structural alterations of *Ec*SapA to translate the dipeptides could be visualized in the form of fluctuations in the aforementioned graphs.

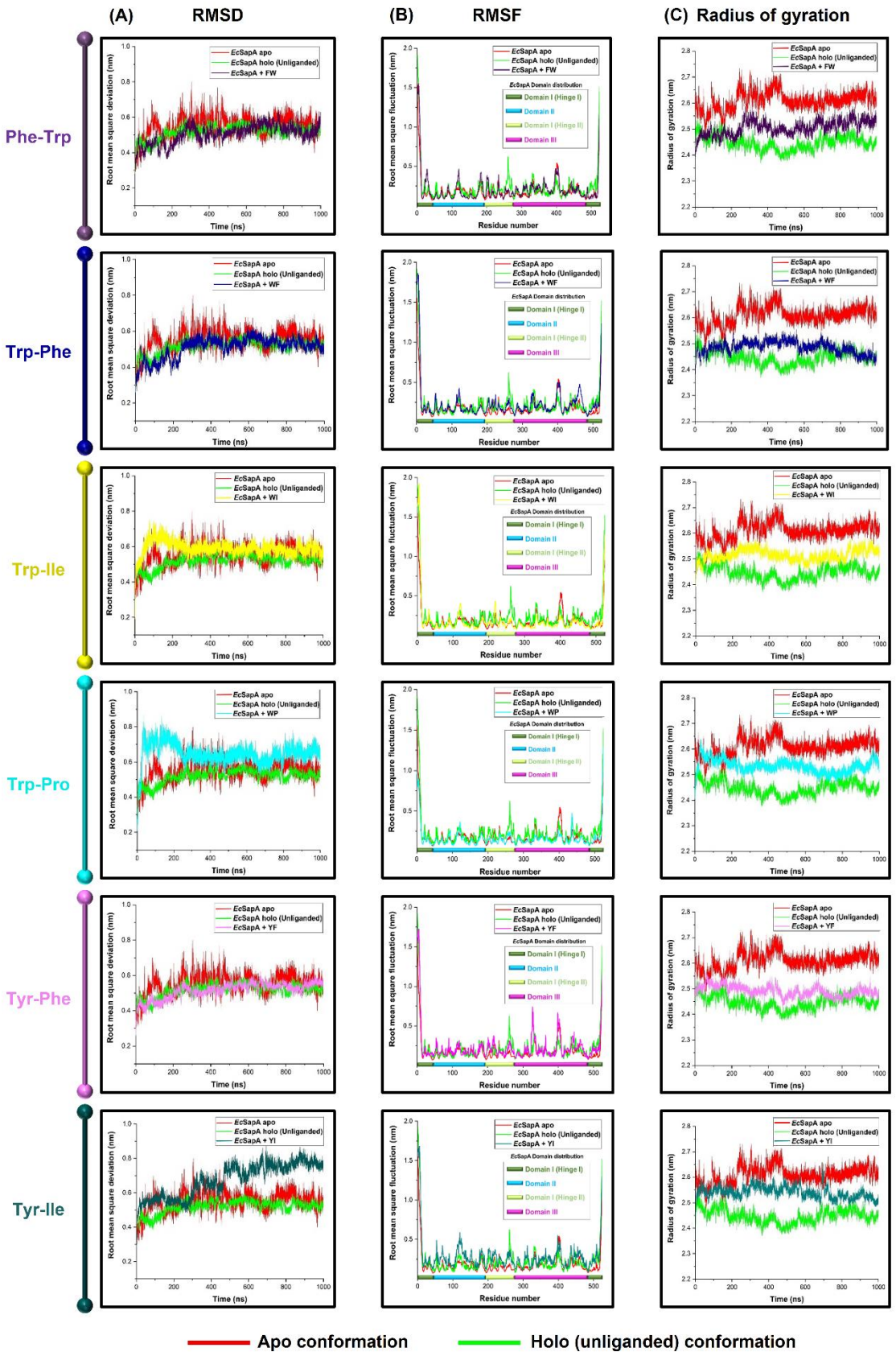


Figure 5.2. MD simulation analysis of *Ec*SapA-dipeptide complexes. The graphs of (A) RMSD, (B) RMSF, and (C) Rg were generated for all the complexes

of *Ec*SapA-dipeptide complexes. The outputs of the apo and holo (unliganded) states of *Ec*SapA have been represented in the form of red and green lines. The plots of each of the *Ec*SapA-dipeptide complexes have been represented as purple (Phe-Trp), blue (Trp-Phe), yellow (Trp-Ile), cyan (Trp-Pro), pink (Tyr-Phe), and teal (Tyr-Ile) lines. Each row of the panel represents the plots of each *Ec*SapA-dipeptide complex, with the name of the dipeptide mentioned in the farther left panel.

5.3.3. AMP-binding propensity of *Ec*SapA revealed a varied domain conformation

The protein-protein docking studies elaborated in Chapter 4 suggest the binding of two AMPs, namely, protegrin-1 and protamine. The AMP molecules were docked into the open conformation of *Ec*SapA since the closed conformation occludes the available binding site of *Ec*SapA (Figure C1B). To confirm the validity of the obtained complexes, both the *Ec*SapA-AMP complexes were subjected to MD simulations.

In the case of the *Ec*SapA-protegrin-1 complex, the RMSD graph was observed to continuously fluctuate throughout the 1 µs trajectory and failed to attain convergence (Figure 5.3A). To negate the possibility of computational artifacts, the complex was simulated in triplicates, and in each condition, a similar trajectory was observable. Further, the RMSF data suggested the fluctuations of the complex to be similar to that of the apo conformation, hence suggesting an open-liganded conformation of *Ec*SapA (Figure 5.3B). The Rg data, akin to the RMSF data, suggested *Ec*SapA to attain an open-liganded conformation throughout the trajectory (Figure 5.3C). To understand if protegrin-1 was being maintained within the binding site of *Ec*SapA, the structural coordinates of the complexes were extracted from multiple time points across the trajectory to map the maintenance of protegrin-1 within the binding site of *Ec*SapA. The structural coordinates revealed the opening up of *Ec*SapA within the first 100 ns, which corresponds to the initial spike observable in the RMSD and Rg graphs (5.3D). As the protein opens up, protegrin-1 reorients into a favorable orientation, wherein it is maintained from ~200 ns - ~700 ns (Figure 5.3D). Post 700 ns, another spike is

observable, wherein the conformation of the protein reorients, hence leading to a change in the location of protegrin-1 in the *Ec*SapA binding site (Figure 5.3D). However, despite the fluctuations, throughout the trajectory, *Ec*SapA was capable of maintaining protegrin-1 within its binding site, hence confirming its binding propensity for protegrin-1.

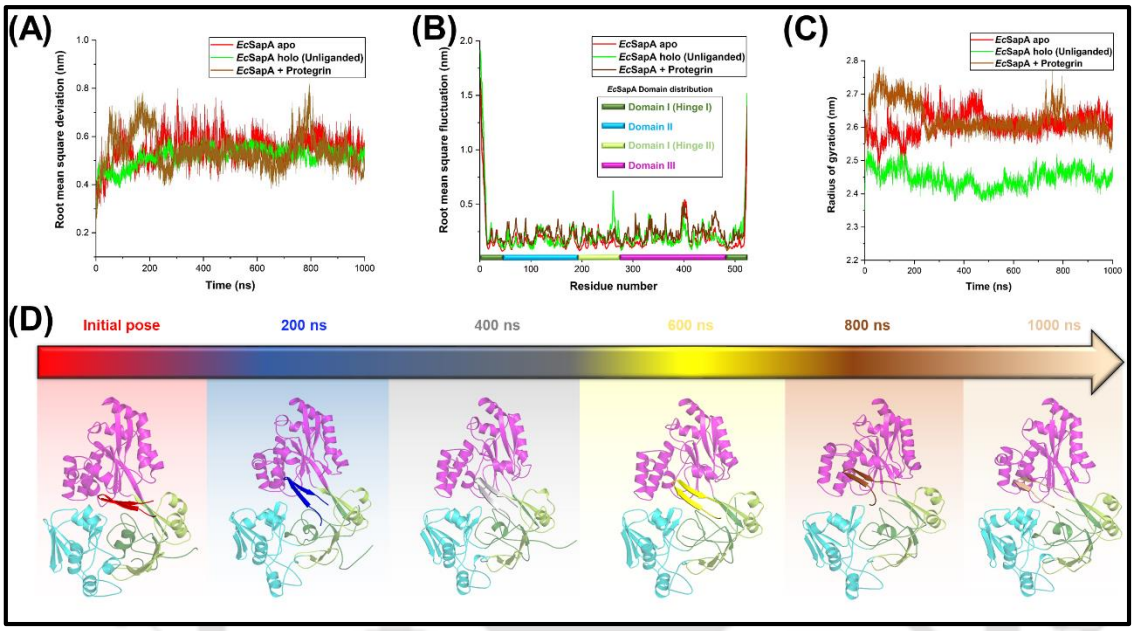


Figure 5.3. *Ec*SapA-protegrin-1 complex, MD simulation analysis. The graphs of (A) RMSD, (B) RMSF, and (C) Rg were generated for the *Ec*SapA-protegrin-1 complex. The outputs of the apo and holo (unliganded) states of *Ec*SapA have been represented in the form of red and green lines. The graphs of the *Ec*SapA-protegrin-1 (brown) complex were overlaid with the apo and holo (unliganded) state graphs. (D) The extracted frames of *Ec*SapA-protegrin-1 complex are as follows: initial equilibrated structure (red), 200 ns (blue), 400 ns (grey), 600 ns (yellow), 800 ns (brown), and 1000 ns (wheat) time points. Protegrin-1 and *Ec*SapA have been represented in the form of cartoons. *Ec*SapA has been color-coded on the basis of its distinct domains, which comprise domain I, which includes hinge I (dark green) and hinge II (light green), alongside domain II (cyan) and domain III (magenta).

In the case of the *Ec*SapA-protamine complex, the RMSD graphs demonstrated a stable complex with very few fluctuations and were observed to converge

around ~400 ns (Figure 5.4A). Further, the RMSF data for the complex was observed to demonstrate lesser fluctuations in comparison to the apo and holo (unliganded) states of *Ec*SapA (Figure 5.4B). Hence, indicating the binding of protamine renders structural stability to *Ec*SapA. Since the initial template of the *Ec*SapA-protamine complex was in the apo conformation, hence the Rg data initiated similar to the apo conformation; however, with the progression of the trajectory, *Ec*SapA was observed to undergo partial closure upon binding protamine (Figure 5.4C). To gain further clarity, the structural coordinates across various time points in the trajectory were extracted and analyzed. In the initial pose, protamine was observed to be docked within the interface of Domains I, II, and III (Figure 5.4D). However, with the progression of the trajectory, protamine was observed to develop a biased interaction with the domains II and III of *Ec*SapA, hence leading to partial closure of *Ec*SapA. However, despite the biased interactions, protamine was observed to be maintained in the interface throughout the trajectory, thus suggesting an exclusive protamine binding site in *Ec*SapA. Hence, MD simulation data validates the AMP (protegrin-1 and protamine) binding propensity of *Ec*SapA.

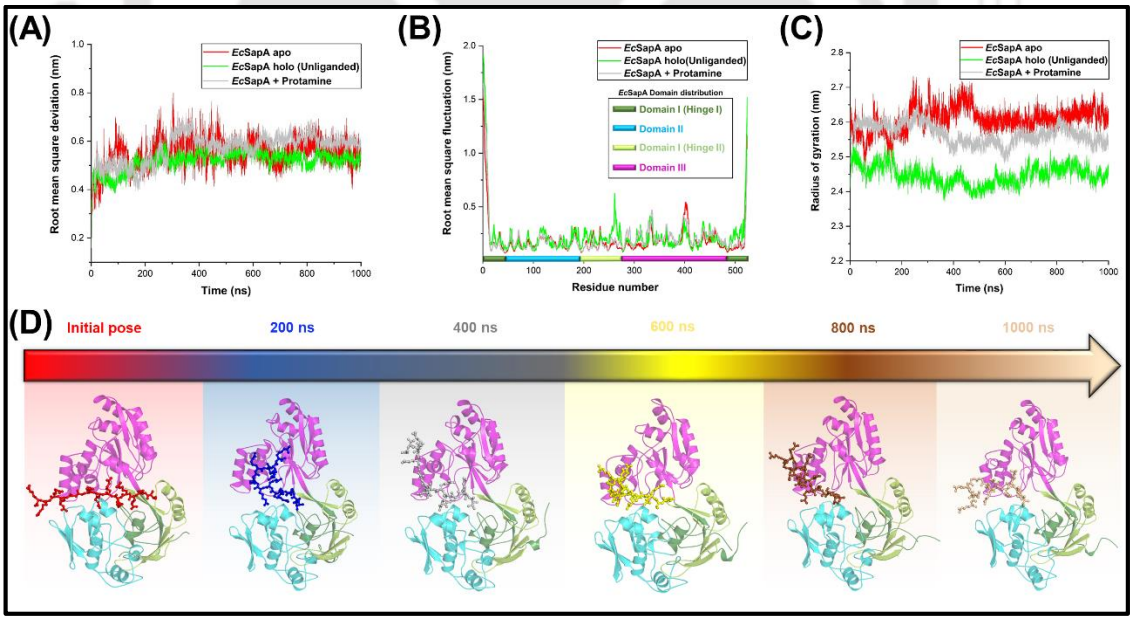


Figure 5.4. *Ec*SapA-protamine complex, MD simulation analysis. The graphs of (A) RMSD, (B) RMSF, and (C) Rg were generated for the *Ec*SapA-protamine complex. The outputs of the apo and holo (unliganded) states of *Ec*SapA have been represented in the form of red and green lines. The graphs of the *Ec*SapA-

protamine (grey) complex were overlaid with the apo and holo (unliganded) state graphs. (D) The extracted frames of the *Ec*SapA-protamine complex are as follows: initial equilibrated structure (red), 200 ns (blue), 400 ns (grey), 600 ns (yellow), 800 ns (brown), and 1000 ns (wheat) time points. Protamine is represented in the form of a ball-and-stick model, while *Ec*SapA is represented in the form of cartoons. *Ec*SapA has been color-coded on the basis of its distinct domains, which comprise domain I, which includes hinge I (dark green) and hinge II (light green), alongside domain II (cyan) and domain III (magenta).

5.3.4. *Ec*SapA has a unique heme binding site

Molecular docking studies elaborated on in Chapter 4 suggested *Ec*SapA has a heme-binding propensity. Hence, to establish the validity of the docked pose, the complex (*Ec*SapA-heme) was subjected to MD simulations for a time period of 1 µs. The RMSD graph suggests the trajectory to attain convergence around ~100 ns and was observed to demonstrate low fluctuations akin to the holo (unliganded) conformation of *Ec*SapA (Figure 5.5A). Further, the RMSF data suggested that heme stabilized the fluctuations observed in the apo and holo (unliganded) conformation (Figure 5.5B). The Rg data also suggests no domain movement in the *Ec*SapA-heme complex, as the data was similar to the trend of the holo (unliganded) conformation (Figure 5.5C). To corroborate the observed data, the structural coordinates of the *Ec*SapA-heme complex were extracted from various time points in the trajectory. The structures corroborated the trends observable in the graphs, thus suggesting that heme forms a very stable conformation with *Ec*SapA. The structural coordinates also reveal a specific binding site for heme at the interface of domains I and III (Figure 5.5D).

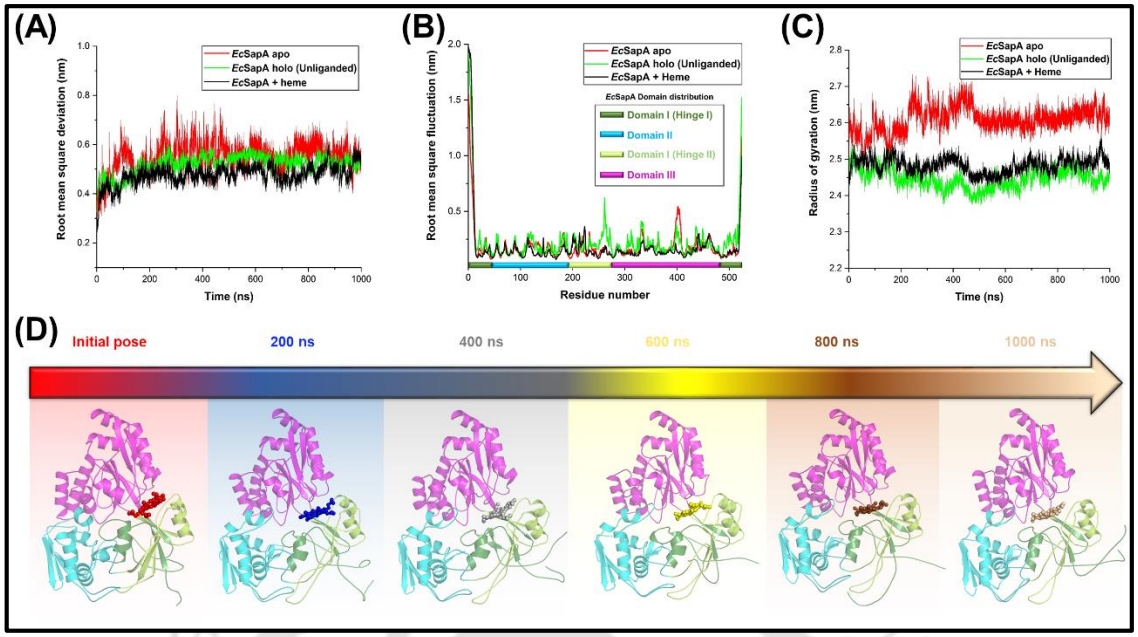


Figure 5.5. *Ec*SapA-heme complex, MD simulation analysis. The graphs of (A) RMSD, (B) RMSF, and (C) Rg were generated for the *Ec*SapA-heme complex. The outputs of the apo and holo (unliganded) states of *Ec*SapA have been represented in the form of red and green lines. The graphs of the *Ec*SapA-heme (grey) complex were overlaid with the apo and holo (unliganded) state graphs. (D) The extracted frames of the *Ec*SapA-heme complex are as follows: initial equilibrated structure (red), 200 ns (blue), 400 ns (grey), 600 ns (yellow), 800 ns (brown), and 1000 ns (wheat) time points. The heme molecule is represented by a ball-and-stick model, while *Ec*SapA as cartoons. *Ec*SapA has been color-coded on the basis of its distinct domains, which comprise domain I, which includes hinge I (dark green) and hinge II (light green), alongside domain II (cyan) and domain III (magenta).

5.3.5. Varied hydrogen bonding pattern of *Ec*SapA reveals ligand binding machinery

The maintenance of the suggested ligands via *Ec*SapA has been established in the prior sections. To further establish the interactions between *Ec*SapA and its cognate ligands, the maintenance of hydrogen bonds was analyzed for each complex across the trajectory. The number of hydrogen bonds of Trp-Phe and Phe-Trp was also calculated as a negative control to compare and contrast with other ligands. Intriguingly, the initial number of hydrogen bonds in the equilibrated

complex was observed to vary from the docked pose (Table C1). The hydrogen bond graphs further established the selectivity of ligands displayed by *Ec*SapA, wherein Phe-Trp and Trp-Phe were observed to have sparse hydrogen bonds, with an average number of ~1 hydrogen bond being maintained across the trajectory (Figure 5.6A-5.6B and Table C1). Further, the ligands Tyr-Phe and heme were observed to maintain distinct hydrogen bonding throughout the trajectory, with an average of ~2 hydrogen bonds (Figure 5.6C-5.6D and Table C1). The other three dipeptides (Tyr-Ile, Trp-Pro, and Trp-Ile) were observed to form stronger hydrogen bonds and maintained an average of ~3 hydrogen bonds in each case (Figure 5.6E-5.6G and Table C1).

The maximum number of hydrogen bonds was observable in the case of both the AMPs (protegrin-1 and protamine), wherein distinct hydrogen bonds were maintained across the trajectory (Figure 5.6H-5.6I). The average number of hydrogen bonds of protegrin-1 and protamine were observed to be ~10 and ~9, respectively (Table C1). Hence, the trends of hydrogen bond maintenance suggest the ligands (four dipeptides, two AMPs, and heme) have stable interactions with *Ec*SapA, validating them as cognate ligands.

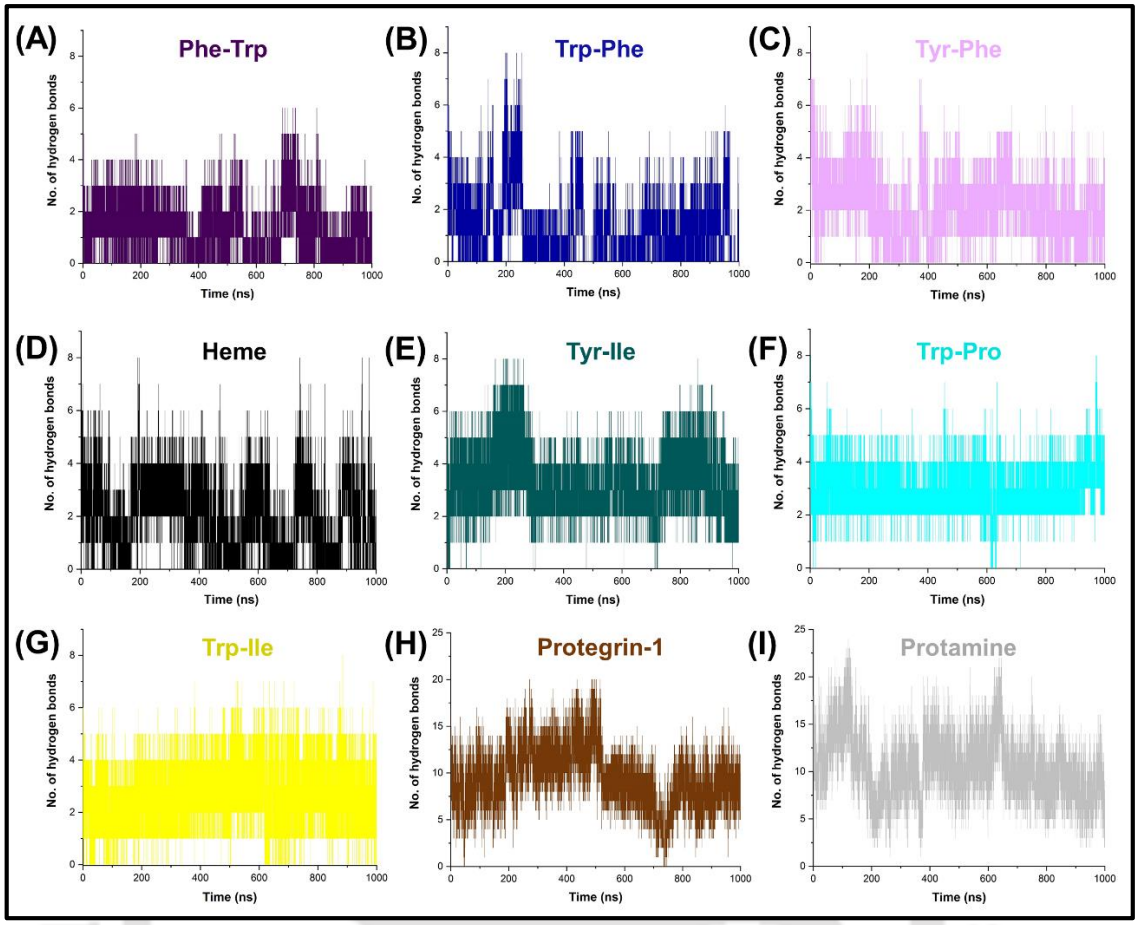


Figure 5.6. Comparative analysis of the number of hydrogen bonds of all *Ec*SapA complexes. The number of hydrogen bonds maintained across the trajectory has been represented in a graphical format with the time of simulation on the X-axis and the number of hydrogen bonds on the Y-axis. The graphs have been color-coded based on the ligand bound to *Ec*SapA as follows: (A) Phe-Trp (purple), (B) Trp-Phe (blue), (C) Tyr-Phe (pink), (D) Heme (black), (E) Tyr-Ile (teal), (F) Trp-Pro (cyan), (G) Trp-Ile (yellow), (H) Protegrin-1 (brown), and (I) Protamine (grey).

5.3.6. *Ec*SapA reveals a segregated binding site of varied ligands

The MD simulations revealed seven distinct ligands of *Ec*SapA, encompassing four dipeptides (Trp-Pro, Trp-Ile, Tyr-Phe, and Tyr-Ile), two AMPs (protegrin-1 and protamine) and heme. The tertiary structural coordinates of all the complexes suggest a differential binding site for each of the proposed ligands. The dipeptides are observed to possess two distinct binding sites: an initial binding site and a final binding site (Figure 5.7A). The peptides are observed to translate towards

the final binding site, which is present in the central interface of the domains I, II, and III of *Ec*SapA. Hence, the central core of *Ec*SapA may be referred to as the dipeptide-binding core (Figure 5.7A). Amongst the two AMPs, protamine is reported to possess more arginine residues (10) and hence is suggested to be highly charged in comparison to protegrin-1, which has a sparser number of arginine residues (6) and hence is less charged. Therefore, *Ec*SapA is observed to segregate its binding site to accommodate both types of AMPs, wherein it accommodates protamine within the interface of Domain II and III (Figure 5.7B). Meanwhile, protegrin-1 is housed in the central core of *Ec*SapA, overlapping with the dipeptide-binding site of *Ec*SapA (Figure 5.7C). The results further reveal a distinct binding site for heme molecule in *Ec*SapA that is present in the interface of domains I and III (Figure 5.7D). Hence, *Ec*SapA demonstrated trends similar to *Mtb*DppA wherein the dipeptide and the heme-binding sites are separate (Mitra *et al.,* 2019).

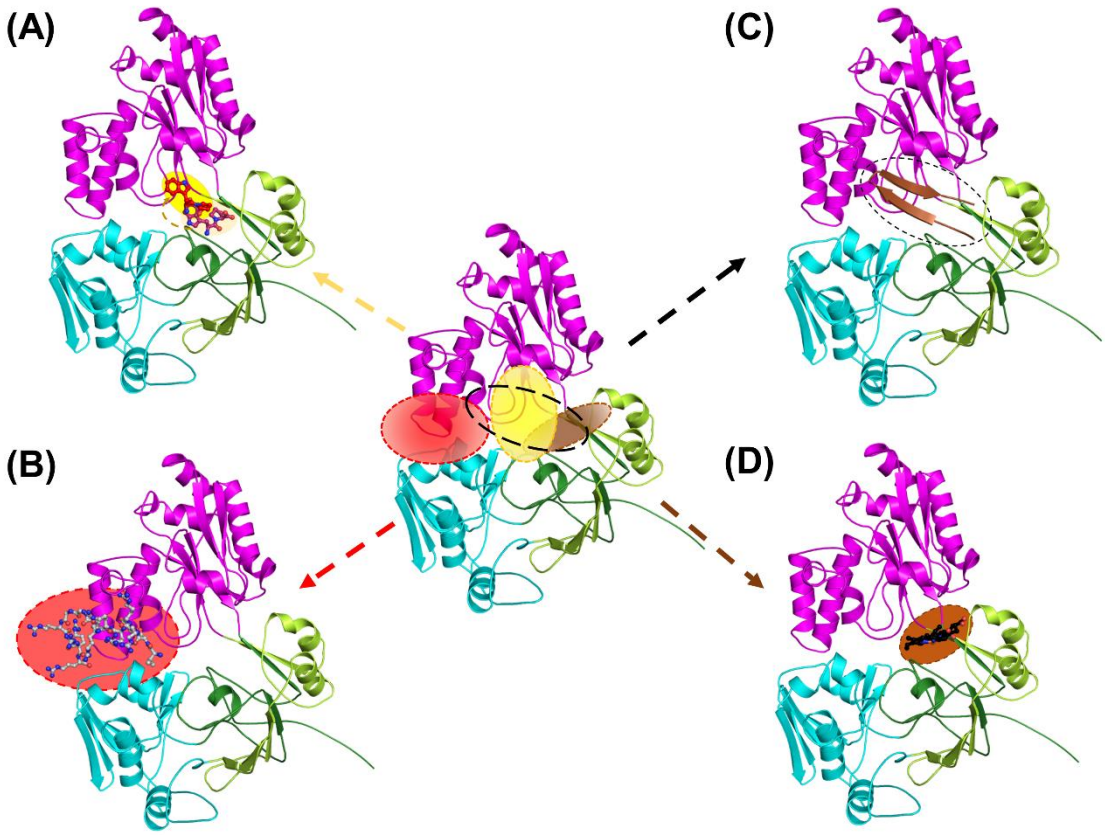


Figure 5.7. Segregated binding site of *Ec*SapA. Cartoon representation of the binding-site segregation of the *Ec*SapA binding site, wherein (A) represents the initial (pink ball-and-stick model, pale yellow shade) and the final (red ball-and-

stick model, yellow shade) binding site of Trp-Pro, (B) represents the protamine (grey ball-and-stick model, red shade) binding site, (C) represents the protegrin-1 (brown cartoon) binding site, and (D) represents the heme (black ball-and-stick model, brown shade) binding site.

5.3.7. *Ec*SapA demonstrates a varied degree of domain movement in accordance with the ligands

Apart from varied binding sites of *Ec*SapA for each ligand, the movement of *Ec*SapA domain III varies to accommodate the different ligands. For example, the complexes of *Ec*SapA with protegrin-1, Trp-Ile, and Trp-Pro have been observed to demonstrate an open-liganded form of *Ec*SapA, which appears to have no significant difference from the open-unliganded form (Figure 5.8A-5.8B). While the complexes of *Ec*SapA with heme and Tyr-Phe demonstrate a closed-liganded form (Figure 5.8C). Interestingly, apart from the above two conformations, a third different conformation was observable in the case of complexes of *Ec*SapA with protamine and Tyr-Ile, wherein *Ec*SapA is observed to be partially closed (Figure 5.8D). The varying degrees of domain movement to accommodate ligands of varying sizes have been suggested in other SBPs, and *Ec*SapA appears to be following a similar trend to accommodate a vast array of ligands (de Boer *et al.,* 2019; Rivera *et al.,* 2024).

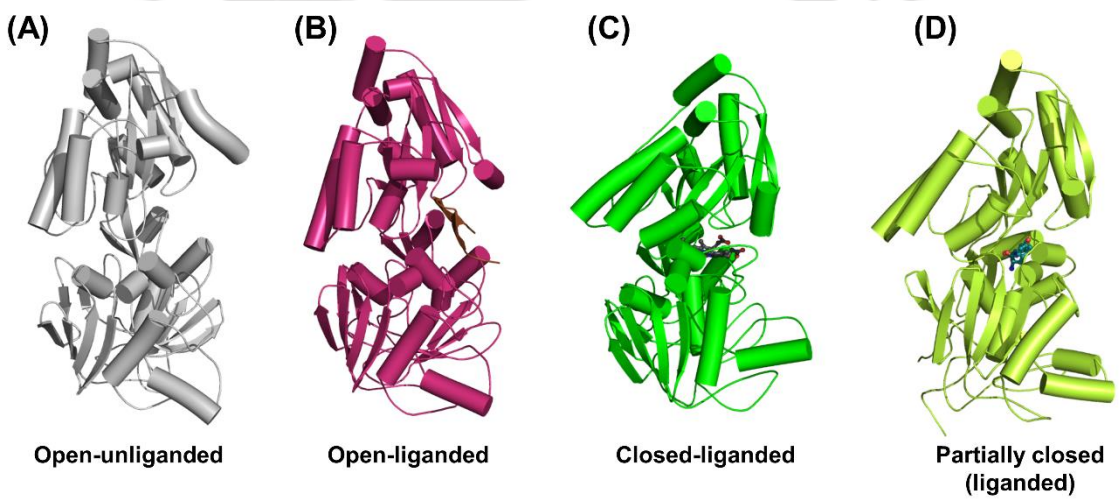


Figure 5.8. Ligand-based domain movement of *Ec*SapA. The various conformations of *Ec*SapA have been represented in the form of cartoons. (A)

Represents the open unliganded (grey) conformation. (B) Represents the open-liganded (dark pink) conformation with protegrin-1(brown) bound in the binding site. (C) Represents the closed-liganded (green) conformation with heme (black) bound in the binding site. (D) Represents the partially closed (liganded) (limon) conformation with Tyr-Ile bound in the binding site.

5.4. Discussion

The Sap transport system has been well established to be involved in the uptake of AMPs, followed by their proteolytic degradation within the bacterial cytosol (Shelton *et al.,* 2011). Apart from AMP uptake, the Sap transport system has also been reported to have multifunctional attributes that encompass the uptake of heme and oligopeptides (Mason *et al.,* 2011; Liang *et al.,* 2019). Akin to other Sap transporters, the results discussed in Chapter 4 also suggest multifunctional attributes of the *Ec*Sap transport system, wherein *Ec*SapA is suggested to be involved in the uptake of dipeptides, AMPs, and heme. Hence, this study utilizes MD simulations to validate the authenticity of the docked poses and further gain insights into the putative mechanisms enabling *Ec*SapA to bind such a wide variety of ligands.

The MD simulation studies of the apo and holo (unliganded) state of *Ec*SapA demonstrate both the conformations of the protein to be stable in a free-floating state, hence suggesting that only under the influence of a third factor (ligand or docking on *Ec*SapBC), the domain transition of *Ec*SapA may be observed. The open (apo) conformation of the SBPs is generally attributed to be unstable and is suggested to exist in a closed (holo) conformation in the free-floating state (Chai *et al.,* 2022). In the case of *Ec*SapA, even though the apo conformation demonstrated slightly higher fluctuations in comparison to the holo (unliganded) conformation, both forms were observed to be mostly stable throughout the trajectory, and none of the forms transitioned into the other. Despite the stable conformation, the apo state of *Ec*SapA was observed to open up further, wherein the α-helix 13 of domain III was observed to translate farther from domain II, hence suggesting further opening of *Ec*SapA. Thus, it could be comprehended

that the further opening of the *Ec*SapA apo conformation was most likely to attain its actual biologically relevant open conformation.

Further, all the *Ec*SapA ligand complexes were subjected to MD simulations to assess the validity of the obtained docked poses. Intriguingly, MD simulations of the *Ec*SapA-dipeptide complexes unraveled a selectivity amongst the dipeptides reported to bind the protein. The dipeptides Trp-Phe and Phe-Trp were rejected by *Ec*SapA post 300 ns while maintaining four dipeptides, Trp-Pro, Trp-Ile, Tyr-Ile, and Tyr-Phe, within its binding site. Amongst the four dipeptides, three (Trp-Pro, Trp-Ile, and Tyr-Ile) were observed to dock at the binding site B of *Ec*SapA. However, interestingly, during the simulation, all three dipeptides were observed to translate from binding site B to A. Hence, the results suggest the presence of an initial binding site and a final binding site in *Ec*SapA. Therefore, it may be proposed that the dipeptides initially bind at the binding site B (which sits on top of Domain I) in proximity to the hinge, hence inducing structural alterations, which ultimately lead them to move into the actual dipeptide binding site. This trend is in accordance with the well-established hinge-driven ligand-binding machinery of SBPs (Chandravanshi *et al.,* 2021).

*Ec*SapA also demonstrated an intriguing mechanism of binding protegrin-1 and protamine. Protegrin-1 was observed to bind *Ec*SapA in an open-liganded conformation. In contrast, protamine binds *Ec*Sap in a partially closed-liganded conformation, wherein part of the AMP is observed to be hanging out of the *Ec*SapA binding site. The mechanism of protamine-binding via *Ec*SapA is similar to the oligosaccharide-binding machinery reported in the case of maltose-binding proteins, wherein part of the oligosaccharide hangs outside the binding site allowing domain closure (Quiocho *et al.,* 1997).

The *Ec*SapA-heme complex was observed to be very stable, with no differences in the orientation of the *Ec*SapA domains. Akin to dipeptides, heme was also observed to bind at a specific interface between the domains I and III. The heme-binding site overlaps with the initial binding site of the dipeptides. Hence, most likely, despite having separate binding sites, *Ec*SapA may not be capable of binding both ligands simultaneously. The protegrin-1 binding site overlaps with

both the heme and dipeptide binding sites, suggesting competitive binding amongst these AMPs. The trend of competitive binding observed in *Ec*SapA is akin to the competitive ligand binding reported in *Hi*SapA (Mason *et al.,* 2011). Interestingly, the binding site of protamine, which is located in the interface of domains II and III, does not overlap as much with the other ligand binding sites. This demonstrates the preference of *Ec*SapA for taking up highly charged AMP molecules such as protamine.

The MD simulation data reveals varying degrees of movement in domain III of *Ec*SapA, leading to an increase or decrease in the binding-site volume as per the ligand. For instance, in the case of protamine, a partial closure of *Ec*SapA is observable. Further, the observations in *Ec*SapA are akin to that of *Hi*SapA, wherein the authors have suggested the protein demonstrate varying degrees of movement based on the size of the peptide bound by *Hi*SapA (Rivera *et al.,* 2024). Hence, similar to *Hi*SapA, *Ec*SapA is observed to demonstrate open, semi-open, and closed conformations.

5.5. Conclusions

The present study probes into the multifunctional aspects of *Ec*SapA, which is attributed to binding six dipeptides, two AMPs, and heme. With the aid of MD simulation-based studies, the dipeptide selectivity of *Ec*SapA was highlighted, wherein *Ec*SapA was observed to bind Trp-Pro, Trp-Ile, Tyr-Phe, and Tyr-Ile in a specific dipeptide-binding site. Further, the AMP and heme-binding propensity of *Ec*SapA were established. The results revealed a segregated binding site of *Ec*SapA, which enables *Ec*SapA to bind a wider range of ligands. Despite the segregated binding site, ligands compete for binding within *Ec*SapA, wherein domain III demonstrates varying degrees of movement based on the ligands, leading to varying binding-site volumes. Hence, the outcomes of this study establish the multifunctional aspects of *Ec*SapA and highlight the promiscuous and adaptable nature of the *Ec*SapA binding site, which could act as an idle platform for peptide-based drug delivery.

Appendix C. Supplementary data

Supplementary Figures: C1-C2

Supplementary Tables: C1

Chapter 6

In vitro characterization of *Ec*SapA

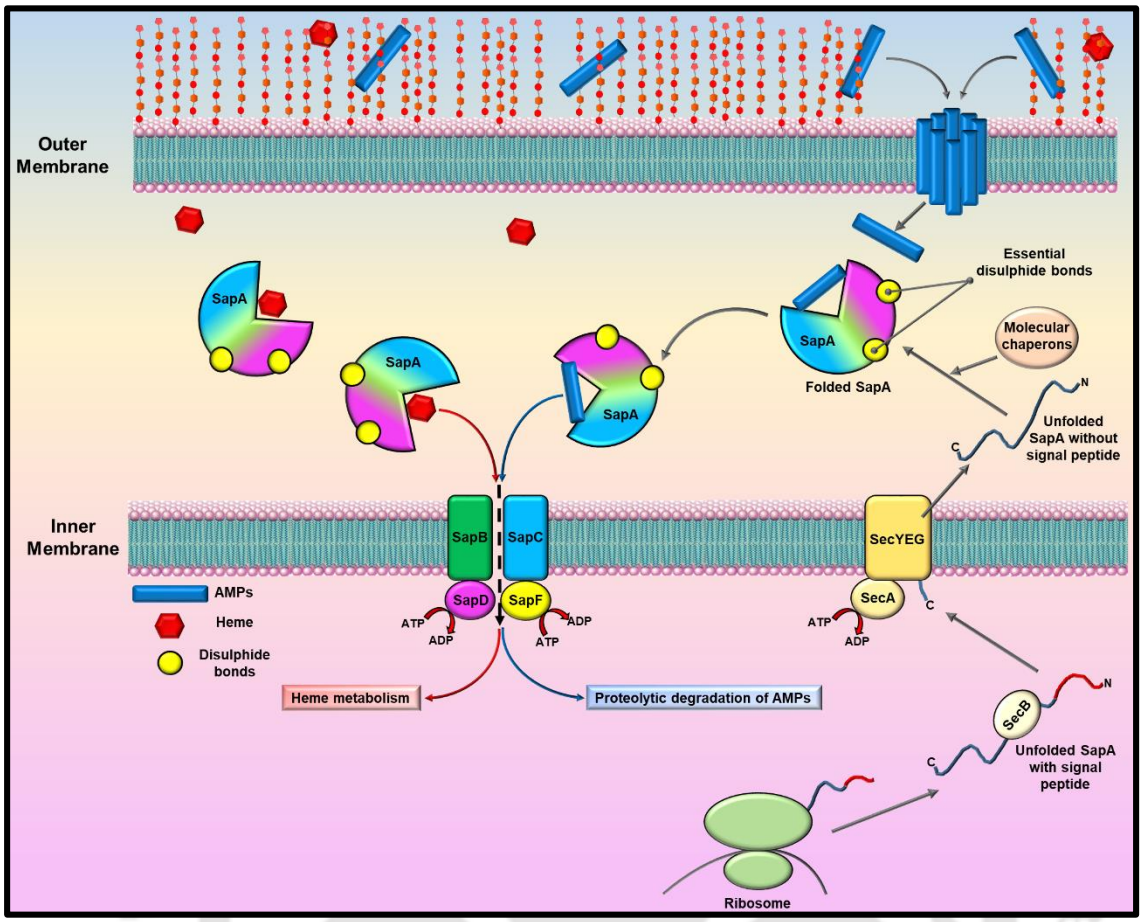
The manuscript for this chapter is under communication as:

Dasgupta P and Kanaujia SP. Biophysical characterization of a putative antimicrobial peptide-binding protein of *Escherichia coli* highlights its dual functionality.

Abstract

The human hosts are subjected to multiple resistance machinery against the bacterial systems. One such machinery encompasses the production of antimicrobial peptides (AMPs), which disrupts the bacterial membrane homeostasis, ultimately leading to the death of the bacteria. Further, the host also sequesters micronutrients (such as iron) essential for bacterial growth and inhibits them from establishing a niche. Hence, the bacterial system requires a system that is capable of sequestering these micronutrients, such as iron, from the hosts and combating such iron-based immunity issues. A particular transport system termed sensitivity to antimicrobial peptides (Sap) is suggested to be involved in AMP and heme uptake, wherein the AMPs are neutralized via the bacterial proteases, and the heme molecule is further metabolized. The present study targets the *Escherichia coli* Sap (*Ec*Sap) transport system and attempts to functionally characterize its substrate-binding protein *Ec*SapA utilizing biophysical strategies. Multiple strategies such as differential absorption spectroscopy (DAS), fluorescence, and electrophoretic mobility shift assay (EMSA) suggested the interaction of *Ec*SapA and heme with a K_d value of 4.8 µM. Further, differential scanning fluorimetry (DSF) also suggests the interaction of protamine with *Ec*SapA, which is indicative of the AMP-binding nature of *Ec*SapA. Thus, the current study establishes the bi-liganded attribute of *Ec*SapA and the functional significance of the *Ec*Sap transport system in the sustenance of *E. coli* within the host.

Graphical abstract



6.1. Introduction

The human host and bacterial pathogens are always intertwined in a tussle, wherein the host tries to hinder the bacterial pathogens from developing a niche (Mues and Chu, 2020). To facilitate the clearance of bacterial pathogens, the human hosts produce various antimicrobials (antimicrobial peptides) as well as sequester essential nutrients (iron) required by the bacterial pathogens to survive (Wang, 2014; Nairz and Weiss, 2020). Antimicrobial peptides (AMPs) are reported to target the bacterial membrane and disrupt their homeostasis, ultimately resulting in their death (Zhang *et al.,* 2021). Further, to dismantle the possibility of bacterial niche formation, the human hosts sequester iron from the environment of the bacterial system, thereby demonstrating nutritional immunity (Crawford and Wilson, 2015).

To combat both challenges, the bacterial system has developed multiple types of machinery to out-smart the human host and develop a suitable niche for pathogenesis. The Gram-negative pathogens have employed multiple strategies, such as manipulating the regulation of AMP production by the host, altering the charge of the outer membrane (OM), employing outer membrane proteases, forming a capsular layer over OM, and utilizing transporters for exporting/importing AMPs across the inner membrane (IM) (Gruenheid and Le Moual, 2012). Alongside, to combat the challenge of nutritional immunity posed by the host, the bacterial system employs various transporters and molecules that sequester iron-bound molecules essential for their sustenance (Murdoch and Skaar, 2022).

Both attacks deployed by human hosts are tackled by the bacterial transport systems. One such transport system is the Sap (sensitivity to antimicrobial peptides) transport system first reported in *Salmonella typhimurium*, wherein the presence of this system rendered the bacterium resistant to the AMP protamine. Further, the *S. typhimurium* Sap (*St*Sap) transport system was observed to belong to the ABC transporter superfamily, wherein it possesses a substrate-binding protein (SBP) SapA, two transmembrane domains (TMDs) SapBC, and two nucleotide-binding domains (NBDs) SapDF (Parra-Lopez *et al.,* 1993). The *Haemophilus influenzae* Sap (*Hi*Sap) transport system was extensively studied, wherein the system has been implicated for the uptake of AMPs (LL-37 Cathelicidin and β-Defensins) and heme (Mason *et al.,* 2011; Shelton *et al.,* 2011).

In the recent past, the membrane components (SapBCDF) of the *Escherichia coli* Sap (*Ec*Sap) transport system have been reported to function as a putrescine exporter, wherein the SBP component SapA was banished as an orphan SBP (Sugiyama *et al.*, 2016). Further, recently, the tertiary structure of *Hi*SapA was elucidated, wherein the protein was observed in a closed conformation. Surprisingly, the structure elucidated was observed to be bound to an RNA molecule non-specifically. The authors further state the constricted binding site of the protein may not be adequate for binding AMPs, and the results obtained might be an out-turn of the non-specific interaction of AMPs with the bound RNA

(Lukacik *et al.*, 2021). Both these reports threw ambiguity pertaining to the functionality of the Sap transport system.

Recently, another crystal structure of *Hi*SapA in open conformation was elucidated. Interestingly, the structure obtained was without any bound RNA, and further AMP binding to the protein was biophysically validated (Rivera *et al.*, 2024). Our results from Chapter 3 suggest that *Ec*SapA is a part of the *Ec*Sap transport system. Alongside the sequence and structure-based analysis elaborated in Chapter 4, it is suggested that *Ec*SapBC is similar to the ABC importer; hence, the exporter-like function of *Ec*SapBC appears bleak. Thus, the functional significance of the Sap transport system remains clouded due to the presence of contrasting reports.

In the current study, the functional prospects of *Ec*SapA were examined with the aid of experimental strategies. The outcomes of the study enlighten the importance of disulfide bonds and the sub-cellular localization of *Ec*SapA for its structural integrity. Further, based on biophysical strategies, the bi-functional attribute of *Ec*SapA that was earlier proposed in Chapter 4 was hence confirmed.

6.2. Materials and methods

6.2.1. Molecular cloning of *Ec*SapA

Molecular cloning strategies were employed to generate two different constructs of *EcsapA* cloned within the pET28a(+) vector. The first strategy was to clone a truncated construct of *EcsapA* (1581 bp), which was devoid of the 63 bp constituting the signal peptide of *EcsapA*. The signal peptide region of this protein was predicted with the aid of the software SignalP 5.0 (Almagro Armenteros *et al.*, 2019). To amplify this frame, the genome of *E. coli* BL21(DE3) was manually isolated. The truncated frame of *EcsapA* (tr*EcsapA*) was amplified via polymerase chain reaction (PCR) using the suitable forward (GTACATATGGCGCCTGAATCTCCCCCG) and reverse (GCGCCTCGAGTCATGGTTTTTTCACCTCATCCTG) primer sets with *Nde*I and *Xho*I (underlined) restriction sites (Figure 6.1A). The amplified gene was double digested with *Nde*I and *Xho*I and was ligated to the pET28a(+) vector digested by the same set of enzymes. The ligated frame was designed so that the protein

derived its 6X-His tag from the vector itself. The second strategy was to clone a full-length frame of *EcsapA* (1644 bp). The full-length frame of *Ec*SapA was amplified from the manually isolated *E. coli* BL21(DE3) genome via PCR utilizing suitable forward (GAAGGATCCATGCGCCAGGTATTATCGTCT) and reverse (GCTCTCGAGTCA**ATGATGATGATGATGATG**TGGTTTTTTCACCTCATC) primer sets with *Bam*HI and *Xho*I (underlined) restriction sites and 6X-His-tag (bold) (Figure 6.1B). The amplified gene was then double digested using *Bam*HI and *Xho*I and ligated within the pET28a(+) vector digested with the same set of enzymes. Both the obtained recombinant constructs were validated via double digestion and plasmid sequencing (Figure 6.1C-6.1D).

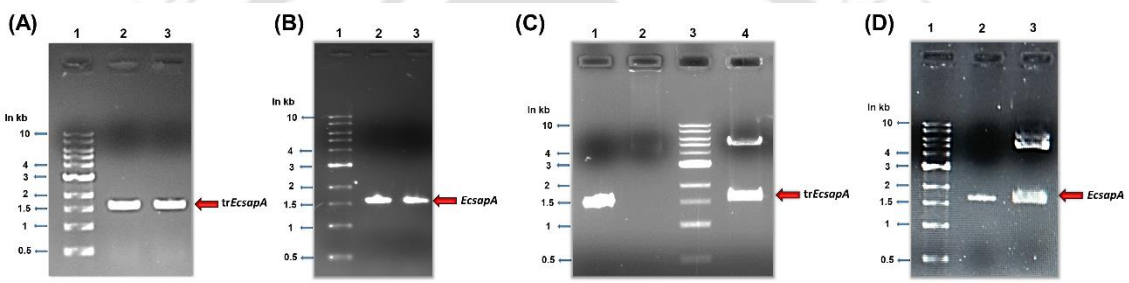


Figure 6.1. Cloning of various constructs of *EcsapA* (tr*EcsapA* and *EcsapA*). (A) Amplification of tr*EcsapA* at 66.9°C via static PCR (lane 1: DNA ladder, lane 2 and 3: amplified gene product). (B) Amplification of *EcsapA* at 69.7°C via static PCR (lane 1: DNA ladder, lane 2 and 3: amplified gene product). (C) Confirmation of tr*EcsapA* clone by double digestion of the vector pET28a(+) carrying the gene (tr*EcsapA*) with *Nde*I and *Xho*I restriction enzymes (lane 1: insert, lane 2: undigested vector pET28a(+), lane 3: DNA ladder, and lane 4: insert pop out post double digestion of the cloned vector pET28a(+)). (D) Confirmation of *EcsapA* clone by double digestion of the vector pET28a(+) carrying the gene (*EcsapA*) with *Bam*HI and *Xho*I restriction enzymes (lane 1: DNA ladder, lane 2: insert, lane 3: insert pop out post double digestion of the cloned vector pET28a(+)).

6.2.2. Overexpression and purification of *Ec*SapA

The expression system *E. coli* BL21(DE3) (Novagen, India) was transformed with the recombinant constructs of *Ec*SapA. The transformed cells were grown in Luria-Bertani (LB) broth at 37°C supplemented with 50 µg mL^{-1} kanamycin. The cells were allowed to grow up to an OD_{600} of 0.6, after which the cells were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated at 37°C for 4 hrs to attain maximum overexpressed protein (Figure 6.2A-6.2B). However, to our dismay, both the constructs were observed to be insoluble in this condition, and hence further optimizations were performed (Figure 6.2C-6.2D).

The full-length construct of *Ec*SapA was observed to solubilize when the culture was incubated at 25°C for 12 hrs post-induction with 1 mM IPTG (Figure 6.2E). However, we failed to solubilize tr*Ec*SapA despite rigorous optimizations. Hence, further methodologies would be pertaining to the full-length construct of *Ec*SapA. Post incubation at 25°C for 12 hrs, the cells were harvested by centrifuging at ~5000g for 10 mins. The harvested cells were then resuspended in chilled buffer A (20 mM Tris-Cl pH 8.0, 300 mM NaCl, 10 mM Imidazole, and 1 mM phenylmethylsulfonyl fluoride (PMSF)). The resuspended cells were then lysed by sonication at 25% amplitude using a 2 s on and 5 s off-cycle of 20 min. Following sonication, the insoluble fractions were removed by centrifuging at ~18500g for 40 min at 4°C. The clear supernatant fraction was loaded onto a Ni-nitrilotriacetic acid (Ni-NTA) affinity resin (Qiagen, Germany) packed with a pierce centrifuge column (Thermo Fisher Scientific, USA), which was pre-equilibrated with buffer A and pre-cooled to cold conditions. After passage of the soluble fraction, the column was subjected to a gradient wash with 10 column volumes of buffer A, followed by 10 column volumes of buffer B (20 mM Tris-Cl pH 8.0, 300 mM NaCl, 20 mM Imidazole, and 1 mM PMSF), and a final wash of 5 column volumes of buffer C (20 mM Tris-Cl pH 8.0, 300 mM NaCl, 50 mM Imidazole, and 1 mM PMSF). The recombinant protein construct bound to the Ni-NTA beads was eluted by increasing the imidazole concentration of buffer D to 250 mM (Figure 6.2F). The eluted proteins were pooled and subjected to dialysis, wherein imidazole was removed in a gradient manner. The final dialyzed protein was obtained in buffer E (20 mM Tris-Cl pH 8.0 and 150 mM NaCl). The dialyzed protein *Ec*SapA was concentrated utilizing the Vivaspin turbo 15 centricon

(30 kDa cut-off; Sartorius, Germany) up to a concentration of 17.5 mg mL^{-1} (Figure 6.2G). The protein concentration was estimated by measuring the absorbance of the protein at 280 nm and then calculating the concentration utilizing its theoretical molar extinction coefficient (ε_{280} = 89060 M^{-1} cm^{-1}). The protein quality and purity were analyzed and inspected via SDS-PAGE at each step of protein purification.

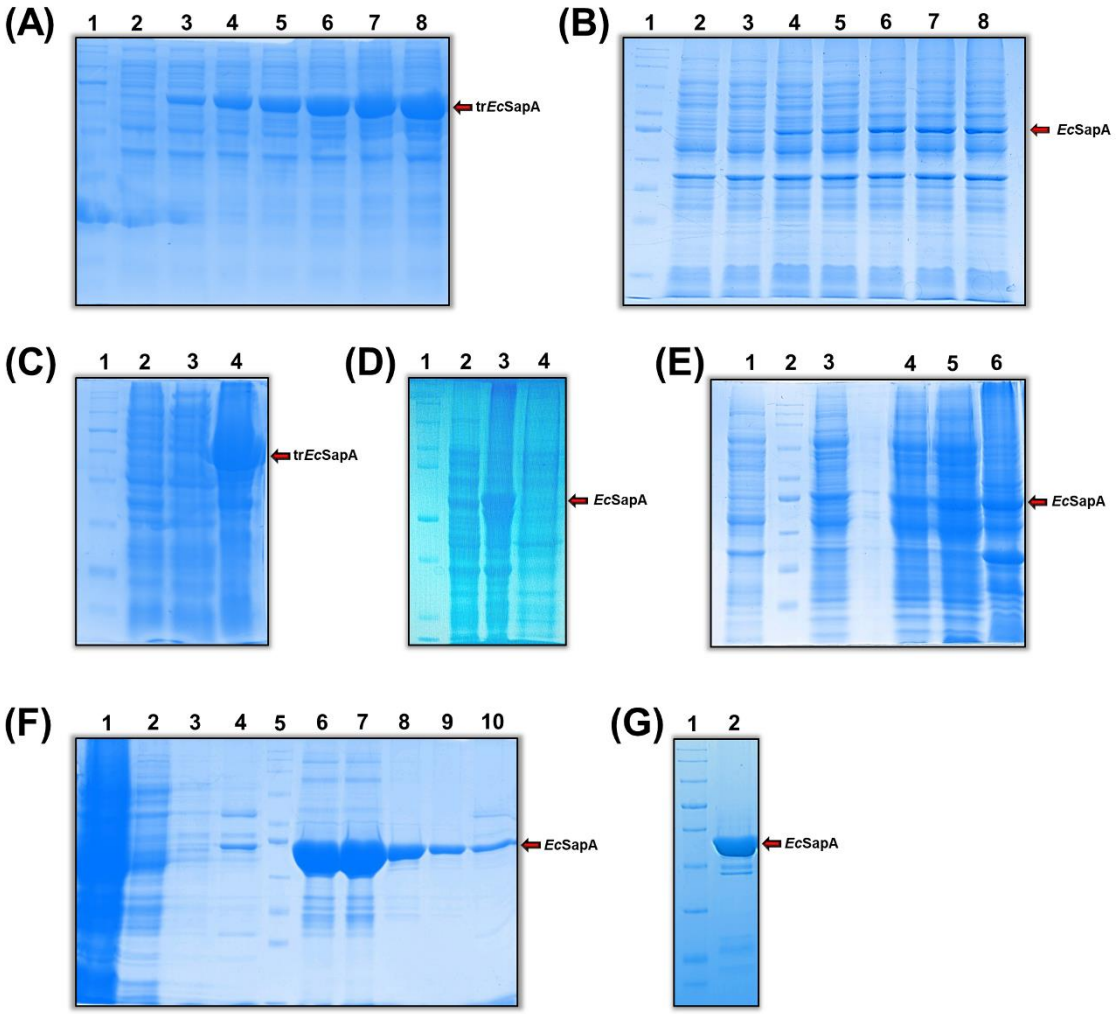


Figure 6.2. Overexpression, solubility, and purification of *Ec*SapA recombinant constructs. (A) Overexpression of tr*Ec*SapA in *E. coli* BL21(DE3) at 1 mM IPTG collected at different time intervals (lane 1: protein marker, lane 2: uninduced sample, lane 3: 1 hour, lane 4: 2 hours, lane 5: 3 hours, lane 6: 4 hours, lane 7: 5 hours, lane 8: 6 hours) post-induction. (B) Overexpression of *Ec*SapA in *E. coli* BL21(DE3) at 1 mM IPTG collected at different time intervals (lane 1: protein marker, lane 2: uninduced sample, lane 3: 1 hour, lane 4: 2 hours, lane 5:

3 hours, lane 6: 4 hours, lane 7: 5 hours, lane 8: 6 hours) post-induction. (C) SDS-PAGE analysis of tr*Ec*SapA solubility at 37°C for 4 hours (lane 1: protein marker, lane 2: uninduced sample, lane 3: soluble fraction, and lane 4: insoluble fractions). (D) SDS-PAGE analysis of *Ec*SapA solubility at 37°C for 4 hours (lane 1: protein marker, lane 2: soluble fraction, lane 3: insoluble fractions, and lane 4: uninduced sample). (E) SDS-PAGE analysis *Ec*SapA reveals the protein to solubilize at 25°C for 12 hours (lane 1: uninduced sample, lane 2: protein marker, lane 3: insoluble fractions, and lane 4-6: soluble fraction). (F) Purification profile of *Ec*SapA (lane 1: Unbound fraction, lanes 2-4: Flowthroughs, lane 5: protein marker, and lanes 6-10: elution fractions). (G) SDS-PAGE analysis of concentrated *Ec*SapA (lane 1: protein marker and lane 2: concentrated *Ec*SapA).

6.2.3. Crystallization trials of *Ec*SapA

In an attempt to elucidate the tertiary structure of *Ec*SapA, preliminary crystallization trials were performed. The purified *Ec*SapA (17.5 mg mL^{-1}) was screened with the multiple commercially available kits, wherein 2 µL each of protein and crystallization condition were mixed in a 1:1 ratio. The microbatch-under-oil method was utilized to screen the purified protein sample. Post-screening, certain tiny needle-shaped crystals were obtained as initial hits (Figure 6.3). The conditions in which the crystal hits were obtained have been rigorously optimized to obtain mountable crystals (Figure 6.3). The concentration of the components of the condition was varied, and multiple additive-based screening was also performed. Further, the condition was tweaked by the addition of varying percentages of various reagents such as glycerol, polyethylene glycol, low melting agarose, etc. Further, microseeding was also performed, wherein submicroscopic crystals were introduced into equilibrated solutions to provide a point of nucleation. All the optimizations were performed using the microbatch-under-oil and hanging-drop vapour-diffusion methods. All these optimizations yielded mountable crystals, which were then exposed to an X-ray beam to obtain their respective diffraction patterns. However, none of the diffraction patterns collected were suitable for further analysis.

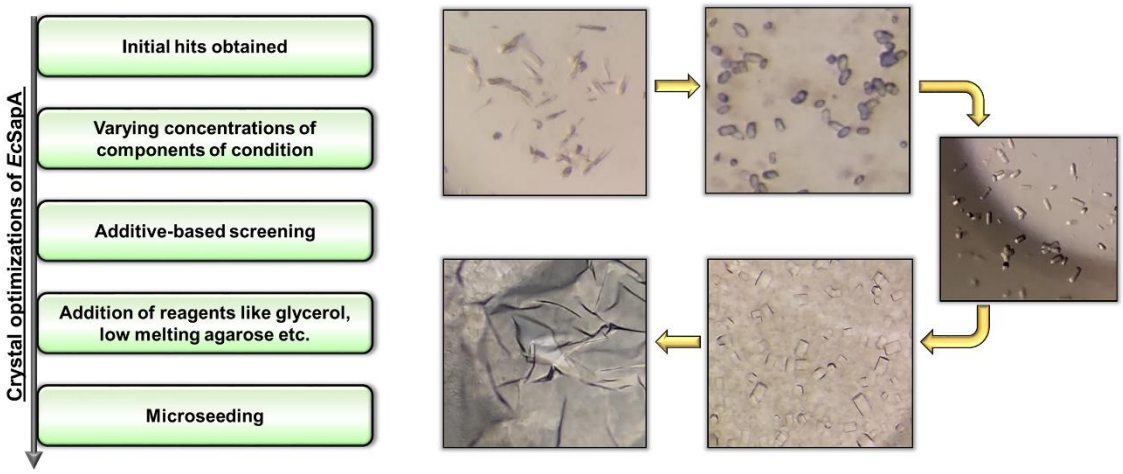


Figure 6.3. *Ec*SapA crystal optimization strategies. The various strategies employed for crystal optimization of *Ec*SapA have been represented in the left panel. The gradual improvement in the crystal morphologies of *Ec*SapA has been represented in the right panel.

6.2.4. Mass Spectrometric analysis

The accurate molecular weight of purified *Ec*SapA was estimated utilizing MALDI-ToF (matrix-assisted laser desorption/ionization-time of flight). An essential component in MALDI-ToF is the matrix being used. In this case, the sinapinic acid matrix was utilized, wherein sinapinic acid was dissolved in a 30:70 mixture of 30% acetonitrile and 0.1% trifluoroacetic acid to make a matrix concentration of 5 mg mL^{-1}. The mixture was sonicated for 30 mins and then subjected to centrifugation at ~20000g for 20 min to get rid of any excess crystals. The freshly prepared matrix is then mixed with 10 µM *Ec*SapA in a ratio of 3:1 (i.e., matrix:protein). Two microliters of the *Ec*SapA-sinapinic acid mix were spotted on the grooves of the ground steel MALDI plate and air-dried. The dried samples were ionized via the smart-beam-II solid-state laser of Bruker autoflex speed (Bruker Daltonics, Germany), and the raw data obtained were plotted using the in-built Flex control analysis software.

6.2.5. Size exclusion chromatography (SEC)

A HiLoad™ 26/600 Superdex™ 200 pg column (GE Healthcare, USA) has been utilized and was attached to an ÄKTA purifier (GE Healthcare, USA) for SEC-

based analysis. Prior to the run, the column was pre-equilibrated with buffer E. Post equilibration, standard protein samples such as Bovine serum albumin (BSA, Molecular weight (MW): 66 kDa), β-amylase (MW: 200 kDa), and carbonic anhydrase (CA, MW: 29 kDa) were analyzed at a flow rate of 1 mL min^{-1} at room temperature and the protein samples were detected by their absorbance at 280 nm wavelength. All the standard samples were prepared in buffer E to a final concentration of 10 mg mL^{-1}. Post analysis of the standard proteins, 10 mg mL^{-1} of *Ec*SapA was loaded onto the column and analyzed in similar conditions to that of the standard proteins. Post the run of all the samples, their elution volumes (V_e) were utilized to calculate the distribution coefficient (K_{av}) using the following formula:

$$K_{av} = \frac{\mathrm{V_e} - \mathrm{Vo}}{\mathrm{V_t} - \mathrm{Vo}}$$

Where V_o and V_t are the voids and the total volume of the column being utilized. The K_{av} of the standard samples was then plotted in a graph of K_{av} versus the log of the molecular weight of the protein (Log M_w) to obtain a standard graph. The molecular weight of our protein of interest was plotted and estimated on the standard graph.

6.2.6. Native-polyacrylamide gel electrophoresis (PAGE)-based analysis of proteins

To analyse via native PAGE, BSA (100 µM) was utilized as a reference protein to understand the adept run of the gel. Purified *Ec*SapA (80 µM) was prepared in buffer E. The samples (2 µL each) were mixed with 3x native loading dye (240 mM Tris-Cl pH 6.8, 30 % (v/v) glycerol, 0.03 % (w/v) bromophenol blue) and loaded onto 8% native gel. The gel was allowed to develop for 3 hours at 100 V. The resolved gel was then stained with Coomassie blue dye and then further developed to visualize the gel.

6.2.7. Dynamic light scattering (DLS)

The DLS-based experiments were performed on a LITESIZER 500 (Anton Paar, Austria) at 25°C. Purified *Ec*SapA was diluted to 1 µM in 50 mM Tris-Cl pH 8.0.

The prepared sample was then added to a polystyrene cuvette, and their light scattering was recorded with 10 autocorrelation functions (replicates). The recorded data was further analyzed to estimate the hydrodynamic diameter of *Ec*SapA. Further, to record the variation in the hydrodynamic diameter of *Ec*SapA with increasing temperatures (20°C to 90°C), the samples were incubated in the desired temperatures for 2 mins, following which their light scattering data were recorded. The recorded data were then visualized and analyzed using the program Origin v9.6.

6.2.8. Circular dichroism (CD)-based analysis

The JASCO J-1500 spectrometer (JASCO, Germany) was utilized to record the far-UV CD spectra of *Ec*SapA with the help of a 1 mm quartz cuvette. The samples were scanned from 190 nm to 260 nm wavelength. The purified protein *Ec*SapA is diluted using nuclease-free water (NFW) to a concentration of 3 µM. Prior to recording the spectra, the parameters of the spectrophotometer were adjusted, wherein response (2 s), sensitivity (100 mdeg), and scan speed (100 nm min^{-1}) were set. Each spectrum is a cumulative average of three independent scans post-subtraction of the blank sample (the same solvent that the protein is suspended in). Post the scan, the spectra were analyzed utilizing the Spectra Manager software integrated into the JASCO system, and the secondary structural elements were analyzed with the help of the web tool BeStSel (Micsonai *et al.*, 2018). Further, to observe the variation in the CD-spectra of *Ec*SapA with increasing temperature, the CD-spectra of *Ec*SapA was recorded from 20°C to 90°C with 20 s incubation at each temperature and with a heat elevation rate of 2°C min^{-1}. The denaturation profiles obtained were plotted using the software Origin v9.6. Further, the T_m of *Ec*SapA was estimated following a previously described method (Singh and Manoj, 2017). Briefly, the ellipticity (in mdeg) of *Ec*SapA at 222 nm was recorded at varying temperatures from 20°C to 90°C. The data obtained is utilized to calculate the fraction of molecules unfolded (F_U) utilizing the following formula:

$$F_U = \frac{Y_{obs} - Y}{Y_U - Y}$$

Where Y_{obs} is the CD signal at a given temperature, and Y_U and Y are the CD signals observed in the unfolded and folded forms of the protein. The data obtained is then plotted using the program Origin v9.6, wherein the data was fitted using the Boltzmann sigmoid equation to determine the T_m of *Ec*SapA.

6.2.9. Difference absorption spectroscopy (DAS)

A DAS-based analysis was employed for the preliminary analysis of hemin interaction with *Ec*SapA. Hemin for binding studies was prepared by dissolving in buffer E supplemented with 100 mM NaOH. In a microtiter plate, equimolar concentration (25 µM) of *Ec*SapA and hemin was mixed and incubated at room temperature (25°C) for 30 min. Post incubation, the samples were scanned from 300 nm to 600 nm using an infinite M200PRO microplate reader (TECAN, Switzerland). The obtained data was subtracted with an appropriate control sample (only hemin in buffer E). The primary spectra are obtained only post-subtraction of the free hemin spectra from the protein-hemin complex.

6.2.10. Electrophoretic mobility shift assay (EMSA) coupled with Western blot analysis

An EMSA-based study was devised to visualize the interaction of hemin with *Ec*SapA. The purified *Ec*SapA was diluted in buffer E to a concentration of 80 µM, mixed with increased concentrations of hemin (50 µM to 300 µM), and incubated at room temperature for 15 min. All the samples were prepared in the form of duplicates and were loaded onto two 8% native-polyacrylamide gel. Both the gels were allowed to develop for 2 hrs at 100V. Post-completion of the run, one gel was visualized with the help of Coomassie blue staining, and the other gel proceeded for Western blotting. The proteins from the second gel were transferred onto a nitrocellulose membrane, wherein the transfer buffer (24 mM tris, 192 mM glycine, and 20% v/v methanol) was supplemented with 0.04% w/v SDS (sodium dodecyl sulfate). The process of transfer was maintained at 4°C, and the proteins were transferred at 40V for 24 hrs. The transfer of native proteins onto the nitrocellulose membrane was performed based on previous reports with slight modifications (Preissler *et al.*, 2015). Post-transfer, the membrane was developed with the help of primary (mouse anti-His_6) and secondary (rabbit anti-mouse IgG) antibodies at a dilution of 1:3000 and 1:5000, respectively. The

secondary antibody used is conjugated with horseradish peroxidase, which enables band detection via chemiluminescence post-substrate addition (Bio-Rad, USA)**.** For densitometric analysis of the bands obtained in the immunoblot, the software ImageJ was utilized (Schneider *et al.*, 2012).

6.2.11. Intrinsic fluorescence-based analysis

The intrinsic fluorescence-based signals were recorded in a FluoroMax-4 spectrofluorometer (Horiba Scientific, Japan) at room temperature (25°C). Initially, the intrinsic fluorescence of varied concentrations of *Ec*SapA (1 µM to 20 µM) was recorded. The varying concentrations of *Ec*SapA were prepared by diluting in buffer E. The samples were added to a 10 mm quartz cuvette and excited at a wavelength of 285 nm. The fluorescence data was collected within the spectral range of 300 nm to 600 nm, employing a slit width of 5. For each sample, three independent scans were averaged and subtracted from adequate blank. Further, the same parameters were used when 5 µM of *Ec*SapA was titrated with increasing concentrations of urea (1 M – 8 M) to observe variation in the fluorescence spectra of *Ec*SapA. In the β-mercaptoethanol (β-Me)-based experiments, 5 µM *Ec*SapA was incubated with 5 mM β-Me with/without 8 M urea for 30 mins at room temperature (25°C). Post-incubation, the fluorescence spectra of each sample were recorded and analyzed.

Further, interaction experiments based on intrinsic fluorescence have been devised wherein 5 µM *Ec*SapA was incubated with increasing concentrations of hemin (2 µM-30 µM) and incubated at room temperature (25°C) for 30 mins. Following incubation, the fluorescence spectra of each sample were recorded. The spectra of each sample were subtracted with adequate blank (desired hemin concentration + buffer E) and then plotted in a fluorescence intensity (a.u) versus wavelength (nm) graph. The data obtained was utilized to determine the hemin binding affinity of *Ec*SapA based on the previously reported methodology (Mandal and Kanaujia, 2022). Briefly, a non-linear least square regression analysis was performed using the below-mentioned formula:

$$\frac{\Delta F}{\Delta F_{max}} = \frac{[ligand]_{tot}}{(K_d + [ligand]_{tot})}$$

Where ΔF is the difference in fluorescence intensity of the protein in the presence of a particular ligand concentration and in the absence of the ligand, ΔF_{max} difference at infinite [ligand], and $[ligand]_{tot}$ is the total ligand concentration (Bougie *et al.*, 2003). Further, the data obtained was utilized to estimate the stoichiometry using a linear form of Hill's equation as follows:

$$\log\left(\frac{\Delta F}{\Delta F_{max} - \Delta F}\right) = n \log[ion] - \log K'$$

Where n is the order of the reaction with respect to ligand concentration and *K´* is the ligand concentration, which results in 50% of ΔF_{max} (Bougie *et al.*, 2003). The fluorescence spectra and the binding study graphs were plotted using the software Origin.

6.2.12. Differential scanning fluorimetry (DSF)

DSF was performed in an Applied Biosystems 7500 real-time PCR system (Thermo Fischer Scientific, USA). The reaction was set up by mixing 5x SYPRO orange dye (Thermo Fischer Scientific, USA) with 10 µM of *Ec*SapA in buffer E. To the protein-dye mixture, 100 µM of Protamine sulfate (Sigma Aldrich, USA) was introduced and allowed to incubate for 15 mins prior to measurement. Alongside the test sample, a control experiment with only *Ec*SapA (10 µM) was also set up. The samples were added in 200 µL microcentrifuge tubes and were kept in the real-time PCR machine. The temperature range scanned was from 30°C to 95°C with a thermal scan rate set at 1°C/min. The data obtained were plotted on a relative fluorescence unit (a.u) versus temperature (°C) graph, and the shift in the T_m in each case was monitored.

6.2.13. Bioinformatic analysis

All the protein sequences utilized have been retrieved from the UniProtKB database (The UniProt Consortium, 2023). The homologs of *Ec*SapA were subjected to multiple sequence alignment (MSA) using the web tool Clustal Omega (Sievers and Higgins, 2014). Further, the alignments were rendered utilizing the web tool ESPript 3.0 (Gouet *et al.,* 2003). The tertiary structure, whenever required, was retrieved from the AlphaFold database (Varadi *et al.,* 2022). The stereo-chemical properties of the retrieved models were validated

using the PDBsum webserver (Laskowski *et al.,* 2018). The secondary structural elements of the predicted tertiary models were estimated using the DSSP program (Touw *et al.,* 2015). The electrostatic charged potential, whenever required, was calculated utilizing the Adaptive Poisson-Boltzmann Solver (APBS) tool available in PyMOL (Molecular Graphics System, Version 2.1.1 Schrödinger, LLC) (Baker *et al.,* 2001). All the structural figures have been visualized and generated utilizing PyMOL.

6.3 Results

6.3.1. The subcellular location of *Ec*SapA is essential for its structural integrity

*Ec*SapA was cloned in two different constructs (*Ec*SapA full-length and tr*Ec*SapA) and was subjected to over-expression (Figure 6.4A). The tr*Ec*SapA construct was devoid of signal peptide and hence was expected to be expressed within the cytoplasm, while the full-length construct is expected to be shuttled to the periplasmic space guided by the signal peptide following the Sec pathway as per predictions from web-tools. Both the constructs were observed to overexpress decently. Intriguingly, the tr*Ec*SapA construct, which was expected to demonstrate cytoplasmic expression, failed to solubilize, while the full-length construct expected to be shuttled into the periplasm was observed to be obtained in the soluble fraction. The soluble fraction was purified and concentrated to 17.5 mg mL^{-1} (Figure 6.2G). The above observation suggested inadequate folding of the tr*Ec*SapA construct within the cytoplasm. However, the full-length construct was observed to fold adequately and hence was obtained within the soluble fraction.

Hence, we speculated that tr*Ec*SapA failed to adequately fold within the cytoplasm, but *Ec*SapA full-length was capable of attaining an adequate fold when shuttled to the periplasmic space, wherein post-transport the signal peptide is cleaved off (Figure 6.4B). The protein sequence of *Ec*SapA (Uniprot id: Q47622) unraveled the presence of four cysteine residues (Cys39, Cys267, Cys454, and Cys467), which may plausibly form disulfide bonds crucial for attaining adept the tertiary fold of *Ec*SapA. The periplasmic space of Gram-

negative bacteria is oxidizing in nature and hence is favorable for disulfide bond formation, whereas the reducing environment of the bacterial cytoplasm inhibits the formation of disulfide bonds and hence dampening the adept folding of tr*Ec*SapA (Chautrand *et al.,* 2022). As we failed to solubilize tr*Ec*SapA, all the following experiments pertain to *Ec*SapA full-length, which shall henceforth be referred to as *Ec*SapA. A MALDI-ToF-based analysis was performed to estimate the size of *Ec*SapA. The molecular weight of purified *Ec*SapA was observed to be 60243.40 Da (~60.2 kDa) (Figure 6.4C). The observed molecular weight established the ferrying of *Ec*SapA to the periplasmic space wherein the signal peptide (1-21 amino acids) is cleaved, and the protein attains its adequate fold. Hence, the MALDI-ToF-based analysis confirms the essentiality of the periplasmic environment to attain adept structural attributes of *Ec*SapA.

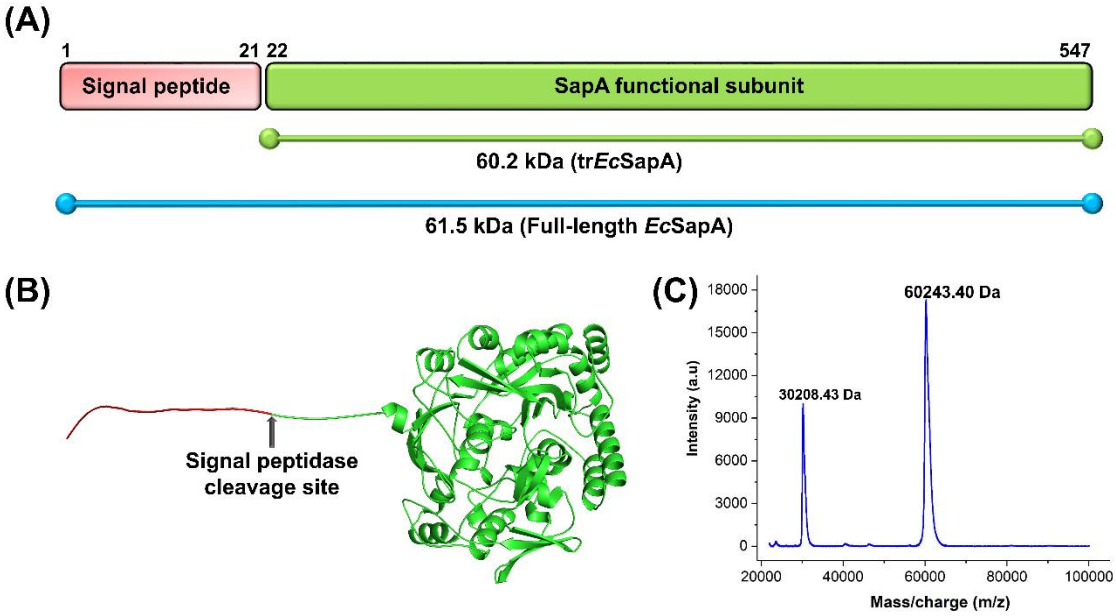


Figure 6.4. *Ec*SapA recombinant constructs and molecular weight estimation. (A) A schematic representation of *Ec*SapA protein frames is represented in the form of rounded rectangles, wherein the red and green boxes represent the signal peptide and functional subunit of *Ec*SapA, with the residue numbers representing the fragments mentioned above. The molecular weights and the frames mentioned within parenthesis are mentioned below the dumbbells. The green dumbbell represents the tr*Ec*SapA frame, and the blue dumbbell represents the full-length *Ec*SapA. (B) Cartoon representation of the predicted tertiary structure model of *Ec*SapA (UniProt id: Q47622), wherein the

signal peptide (red) and the functional subunit (green) have been represented. The predicted tertiary structure model was retrieved from AlphaFold database (C) MALDI-ToF based spectra of *Ec*SapA, wherein the experimental molecular weight was observed to be 60243.40 Da, which is observed to be equivalent to the truncated frame of *Ec*SapA (60279.18 Da). The m/2 peak is also observable at 30208.43 Da in the MALDI-ToF spectra of *Ec*SapA.

6.3.2. *Ec*SapA is monomeric and conformationally stable

Despite the prerequisite knowledge of *Ec*SapA to be an SBP, the potential oligomeric status of *Ec*SapA is uncertain. However, recently, the structure of its homolog *Hi*SapA (SI (sequence identity): 38%) has been reported to be monomeric (Lukacik *et al.,* 2021; Rivera *et al.,* 2024). The size exclusion chromatography (SEC) profile suggests *Ec*SapA to have a single homogeneous predominant population, which was observed to elute out at 219 mL. Based on the elution volume, the K_{av} of *Ec*SapA (0.53) was calculated. The K_{av} of standard proteins were estimated and plotted against a logarithmic (Log_{10}) scale of their molecular weight (M_w) (Figure D1). On plotting the K_{av} value of *Ec*SapA on this standard curve, the molecular weight of *Ec*SapA was estimated to be 57.4 kDa. Hence, the single peak observed can be suggested *Ec*SapA to be in monomeric form (Figure 6.5A). Intriguingly, the molecular weight of *Ec*SapA was observed to be underestimated in both FPLC and SDS-PAGE-based estimations. However, the adept molecular weight was earlier confirmed based on our MALDI-ToF-based experiments, and therefore, the discrepancies observed in the aforementioned might be anomalies observed in the case of these two strategies.

Further, the homogeneity of *Ec*SapA was analysed via native-PAGE and the hydrodynamic diameter of the protein was estimated via DLS (dynamic light scattering) experiments. The native PAGE represents a single crisp band in contrast to the multiple bands observable in the case of BSA (bovine serum albumin) representative of its multiple oligomeric forms (Figure 6.5A). Further, *Ec*SapA was observed to have a hydrodynamic diameter of 6.3 nm based on DLS-based experiments (Figure 6.5B). The observed hydrodynamic diameter was similar to the predicted value of 6.2 nm, estimated using the web tool HullRad

(Fleming and Fleming, 2018). Hence, the DLS data was in accordance with the FPLC and native-PAGE analysis, wherein *Ec*SapA was observed to be in a homogeneous and monomeric state.

Additionally, the secondary structural elements of *Ec*SapA were estimated using far-UV circular dichroism (CD) spectroscopy. The far-UV CD spectrum of *Ec*SapA demonstrated the signature spectra of an α/β-mixed protein, i.e., wherein the protein has substantial amounts of both α-helices and β-sheets (Janes, 2017). The spectra demonstrate four distinct peaks at 195 nm (positive peak), 208 nm, 215 nm, and 222 nm (negative peaks), respectively (Figure 6.5C). These distinct peaks observable are reminiscent of the fingerprint of α-helices (one positive peak at 193 nm and two negative peaks at 208 and 222 nm) and β-sheets (a positive peak at 196 nm and a negative peak at 218 nm) (Ranjbar and Gill, 2009). The secondary structural composition of *Ec*SapA was estimated utilizing the CD data to be 26% helices, 17.5% sheets, and 56.5% others (Figure 6.5C).

Moreover, the thermal stability of *Ec*SapA was estimated via thermal ramping by subjecting the protein to varying temperatures from 20°C to 90°C. The secondary structural elements of our protein of interest were observed to diminish from 50°C onwards (Figure 6.5D). Further, the gradual denaturation of the protein was mapped by observing the CD signal at 222 nm, following which the data was utilized to calculate the fraction of *Ec*SapA that unfolded. On plotting the unfolded fraction against temperature, the T_m of *Ec*SapA could be estimated to be 50.9°C (Figure 6.5E). The thermal denaturation of protein generally leads to aggregate formation (Devnani *et al.,* 2020). To inspect the same, *Ec*SapA was subjected to the same temperature range (20°C to 90°C) and analyzed via DLS. The DLS data revealed a similar trend wherein the hydrodynamic diameter was observed to drastically increase from 50°C onwards (6.3 nm at 20°C to 307 nm at 90°C) akin to the CD data (Figure 6.5F). The thermal stability of *Ec*SapA up to 40°C is indicative of its mesophilic nature and suggests the conformational stability of the protein.

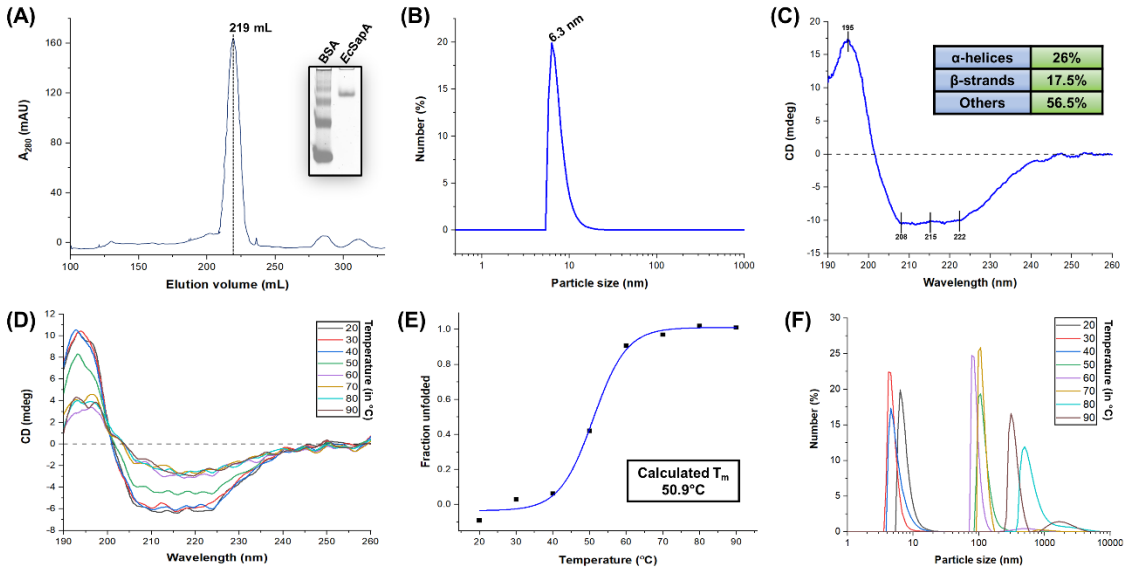


Figure 6.5. Conformation and stability of *Ec*SapA. (A) The SEC profile of *Ec*SapA on a Superdex 200 column has been represented as an absorbance of 280 nm (A_{280} in mAU) versus elution volume (mL). Alongside the SEC profile, the native-PAGE-based analysis of BSA and *Ec*SapA has been represented inset. (B) DLS-based hydrodynamic diameter estimation of *Ec*SapA, wherein a single peak corresponding to 6.3 nm hydrodynamic diameter is observable in a plot of the number of particles (in percentage) versus the particle size (nm). (C) A far-UV CD spectrum of *Ec*SapA along with the estimated secondary structural components (inset). The distinct peaks observed in the spectra have been mentioned in the spectra. (D) The far-UV CD spectra of *Ec*SapA on exposure to temperatures ranging from 20°C to 90°C. Each spectral line has been color coordinated with the temperature (mentioned within the box) it has been recorded. (E) The thermal unfolding curve of *Ec*SapA. The data has been plotted in the form of a fraction unfolded versus temperature (°C) graph. *Ec*SapA was observed to have a T_m of 50.9°C. (F) DLS-data of *Ec*SapA on exposure to temperatures ranging from 20°C to 90°C. The various DLS data have been color-coordinated with their respective temperature of exposure (mentioned within the box).

6.3.3. Disulfide bonds play a significant role in the structural integrity of *Ec*SapA

To gain structural insights, we initiated crystallization trials of *Ec*SapA. We were able to obtain crystals that were poorly diffracting. However, we failed to improve the crystal quality and hence were unable to elucidate the structure of *Ec*SapA (Figure D2). To estimate the structural features of *Ec*SapA, we utilized the predicted tertiary structure of *Ec*SapA from the AlphaFold 2.0 database. The results discussed in Chapter 4 suggest *Ec*SapA belongs to cluster C SBP and, in particular, belongs to cluster C-II SBPs. *Ec*SapA has been reported to be composed of three distinct domains (Domain I, II, and III) akin to the cluster C SBPs wherein domain I consists of hinge I and II (Figure 6.6A) (Chandravanshi *et al.,* 2021). Further, the tertiary structure model of *Ec*SapA was observed to consist of 17 α-helices, 5 η (3_{10})-helices, and 21 β-strands (Figure 6.6A). The model structure of *Ec*SapA is in accordance with the holo-structure of *Hi*SapA, which has been recently elucidated (Lukacik *et al.,* 2021). However, the sole difference in *Ec*SapA is the absence of the additional loop (residues 131-152) and α-helix (residues 153-163, helix 3) present solely in *Hi*SapA (Rivera *et al.,* 2024) (Figure D3).

Since the *Ec*SapA tertiary structure was unavailable, we employed biophysical strategies to gain insights into the structural features of *Ec*SapA. *Ec*SapA was composed of multiple tryptophan (11) residues. The spatial location of tryptophan residues was located within the modeled tertiary structure of *Ec*SapA, wherein three tryptophan residues (Trp251, Trp368, and Trp416) were observed to be buried (i.e., enclosed within a hydrophobic environment), while the remaining tryptophan residues were observed to be exposed (Figure D4A). Intrinsic fluorescence was employed, and varying concentrations of *Ec*SapA (1-15 µM) were excited with an excitation wavelength of 285 nm, and the emission spectra were collected from 300 nm to 500 nm (Figure D4B). The preliminary fluorescence experiment demonstrated a distinct bell-shaped curve with an emission maxima at 342 nm.

Further, the intrinsic fluorescent properties of *Ec*SapA were utilized to map the unfolding of *Ec*SapA in the presence of chaotropic agents. An experiment was

devised wherein the protein was exposed to an incremental concentration of urea from 1 M to 8 M. The samples were then subjected to intrinsic fluorescence-based analysis to understand the unfolding of the protein. The intensity of the fluorescence spectra of *Ec*SapA was observed to drop up to 3 M urea concentration (Figure 6.6B). However, intriguingly, from 4 M to 8 M urea concentrations, there was an increase in the fluorescence intensity combined with a redshift. As increasing amounts of urea were being titrated, it was hypothesized that certain factors hindered the complete denaturation of native *Ec*SapA. One of the key features observed in the case of *Ec*SapA is the presence of four cysteine residues (Cys60, Cys267, Cys475, and Cys488), which form two distinct disulfide bonds in between β1-β11 of domain I and α15-α16 of domain III (Figure 6.6C-6.6D). Hence, it was hypothesized that the presence of these disulfide bonds could be a plausible reason for the peculiar trend observed in the presence of urea.

Hence, to confirm the above-stated hypothesis, *Ec*SapA was exposed to a reducing agent β-Mercaptoethanol (β-Me) in the presence and absence of urea, to observe the effect β-Me has in the maintenance of *Ec*SapA tertiary structure. Intriguingly, as expected, the fluorescence intensity of *Ec*SapA was observed to demonstrate a distinct drop in the presence of β-Me alone and along with urea (Figure 6.6E). However, whenever *Ec*SapA is exposed to only urea, it tends to form aggregates, hence leading to a red shift of *Ec*SapA fluorescence spectra (Figure 6.6E). The aggregate forming tendency could be partly attributed to the presence of the two disulfide bonds. Thus, the presence of the disulfide bond in *Ec*SapA is observed to be an essential aspect of the maintenance and stability of the *Ec*SapA tertiary structure.

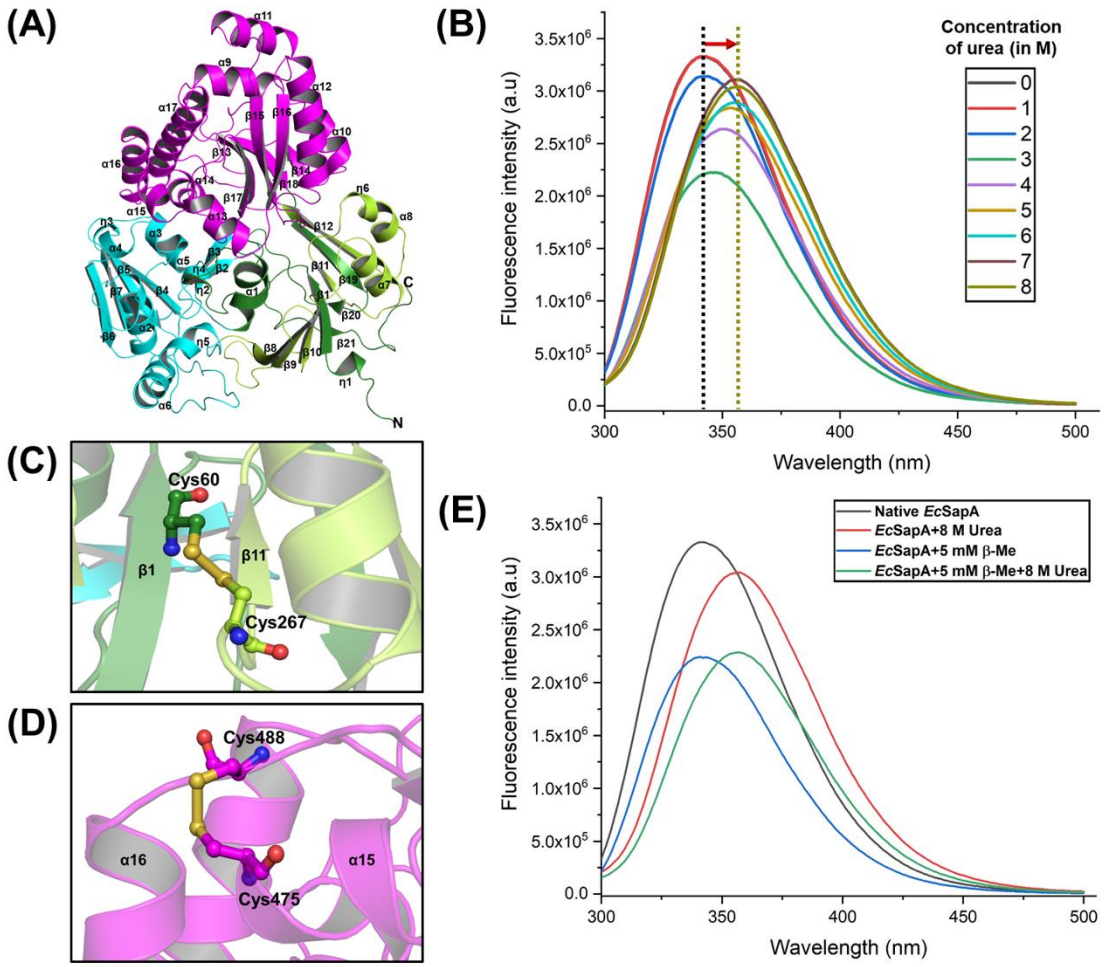


Figure 6.6. Significance of disulfide in the tertiary structure of *Ec*SapA. (A) A cartoon representation of the predicted tertiary structure of *Ec*SapA retrieved from the AlphaFold database. The structure is divided into three distinct domains: Domain I (green), Domain II (cyan), and Domain III (magenta). Further, Domain I is subdivided into hinge I (dark green) and hinge II (light green). The secondary structural elements of *Ec*SapA have been marked. (B) Urea-based denaturation of *Ec*SapA, wherein the increasing concentrations of urea (1M to 8M) have been demarcated by varying colors, which has been detailed in the box (inset). The dashed lines (native protein in black and 8M urea in olive green) and the red arrow represent the shift in λ_{max} of *Ec*SapA in the presence of urea. (C and D) The disulfide bonds (yellow ball and stick) of *Ec*SapA connect Cys60 (dark green ball and stick) with Cys267 (light green ball and stick) and Cys475 with Cys488 (magenta ball and stick). (E) Effect of 5 mM β-mercaptoethanol on the intrinsic fluorescence spectra of *Ec*SapA, wherein the spectral lines have been color coded based on the experiment detailed within the box inset.

6.3.4. Structure and sequence-based features negate dsRNA-binding by *Ec*SapA

Prior reports suggested *Hi*SapA binds RNA at a site in between domains I and II (Figure 6.7A) (Lukacik *et al.,* 2021). *Hi*SapA was reported to interact with RNA via two of its residues, Gln85 and Arg101, both located within domain II of *Hi*SapA (Figure 6.7B). Wherein Arg101 was observed to interact with 3'O and 2'OH of the ribose ring of the RNA nucleotide, and Gln85 interacted with 2'OH of the ribose ring of the RNA nucleotide. Along with the aforementioned residues, six lysine residues (Lys73, Lys258, Lys259, Lys262, Lys542, and Lys545) have also been reported to generate a favorable environment (positively-charged) for accommodating the RNA molecule (negatively-charged) (Lukacik *et al.,* 2021). To analyze the RNA-binding possibility of *Ec*SapA, a multiple sequence alignment (MSA) was performed with multiple SapA proteins to analyze the conservation of the above-stated residues. The MSA revealed that the residue Arg101 was majorly conserved in most cases, however, Gln85 was not conserved. Apart from Lys542, which was mostly conserved, all the other lysine residues were observed to be not conserved (Figure 6.7C). Intriguingly, various negatively charged residues are observable in place of Gln85 and Lys73 (Figure 6.7C). Further, the electrostatic potential charge-based analysis of the *Hi*SapA RNA-binding cleft revealed a positively-charged patch favorable to accommodating a negatively-charged RNA molecule (Figure 6.7D). On inspecting the similar region in *Ec*SapA, a negatively charged patch in accordance with the MSA results was observed (Figure 6.7E). Hence, the results suggest *Ec*SapA might not be capable of binding RNA akin to *Hi*SapA. To further ascertain the lack of RNA-binding affinity of *Ec*SapA, a docking study was devised wherein docking from multiple web tools (NPDock and HDOCK) failed to generate any *Ec*SapA-RNA complex structure (Tuszynska *et al.*, 2015; Yan *et al.*, 2020). Hence, the possibility of *Ec*SapA interacting with an RNA molecule appears to be vague. Furthermore, a recently elucidated structure of *Hi*SapA in open conformation was observed to be sans any RNA molecule to be bound with the protein of interest (Rivera *et al.,* 2024). Thus, the possibility of RNA-binding by *Hi*SapA remains unclear.

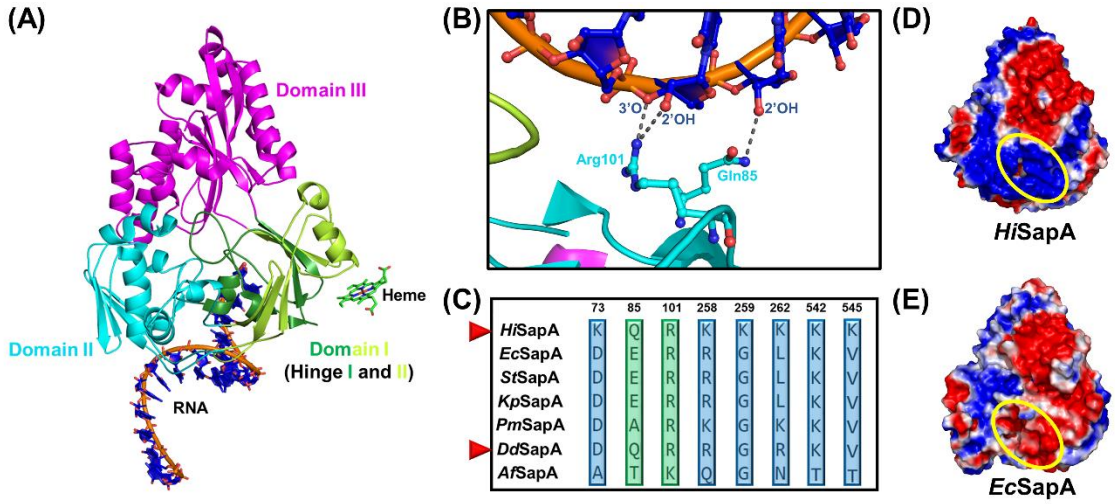


Figure 6.7. Interaction of dsRNA with SapA. (A) Cartoon representation of *Hi*SapA crystal structure bound to heme (green ball and stick) and RNA (blue ball and stick). The various domains of *Hi*SapA have been represented wherein Domain I (Hinge I, dark green; Hinge II, light green), Domain II (cyan), and Domain III (magenta) have been distinctly demarcated based on color. (B) The major residues (cyan ball and stick) of *Hi*SapA interact with the RNA backbone (blue ball and stick). (C) Multiple sequence alignment (MSA) of selected SapA proteins. The sequences of proteins used to generate the MSA are from *H. influenzae* (*Hi*SapA, P45285), *E. coli* (*Ec*SapA, Q47622), *S. typhimurium* (*St*SapA, P36634), *Klebsiella pneumoniae* (*Kp*SapA, W8UWE0), *Proteus mirabilis* (*Pm*SapA, B4EWK8), *Dickeya dadantii* (*Dd*SapA, E0SFZ5), and *Aliivibrio fischeri* (*Af*SapA, Q5E0R7). The conservation of the direct (green box) and the indirectly (blue box) interacting residues reported to interact with the dsRNA in *Hi*SapA is observed. The electrostatic charge potential of (D) *Hi*SapA crystal (PDB id: 7OFZ) and (E) *Ec*SapA tertiary model was retrieved from the AlphaFold database. The positively-charged and negatively-charged regions have been demarcated by blue surface (+1 kT/e) and red surface (-1 kT/e), respectively. The region of RNA interaction has been demarcated with a solid yellow circle in both proteins.

6.3.5. Difference absorption spectroscopy (DAS) provides hints of hemin-binding by *Ec*SapA

The *Hi*SapA has been suggested to bind heme and is suggested to be involved in helping the heme uptake by *H. influenzae* (Tanaka and Pinkett, 2019; Rodríguez-Arce *et al.,* 2019; Lukacik *et al.,* 2021). Further, a recent report from our group suggested that the heme-binding capability of *Ec*SapA is akin to *Hi*SapA. The molecular docking results discussed in Chapter 4 suggested the heme molecule dock between Domain I and Domain III of *Ec*SapA (Figure D5). Heme B is a porphyrin ring consisting of a molecule that is made up of four pyrrole rings connected by four methane bridges (Figure 6.8A). The porphyrin ring of Heme B produces a Soret peak when scanning the UV-Vis range (300 nm to 600 nm) (Figure 6.8B). The Soret peak is observable in the blue wavelength region of the visible spectrum, which results from an electron dipole moment that enables a π-π* transition within the porphyrin ring of heme.

Prior literature suggests that the interaction between hemoproteins and heme molecules results in a distinct Sorret peak, indicating protein-heme interaction (Mitra *et al.*, 2019; Mandal and Kanaujia, 2022). *Ec*SapA was subjected to DAS-based analysis to inspect the same. Upon mixing equimolar concentrations (25 µM) of *Ec*SapA and hemin, a red shift was observed in the Soret peak in comparison to the Soret peak of only hemin (Figure 6.8B). To further ascertain that the shift was a result of protein and hemin interaction, a similar scan within the same range was performed for only *Ec*SapA, and no such peak was observed (Figure 6.8C). Further, subtracting the spectra of free hemin from that of the complex resulted in a distinct Soret peak at 411 nm and a Q-peak at 534 nm, in accordance with prior reports (Figure 6.8D) (Mitra *et al.*, 2019; Mandal and Kanaujia, 2022). Hence, the above observed distinct peaks are indicative of heme-binding via *Ec*SapA, which is in accordance with prior computational observations discussed in Chapter 4.

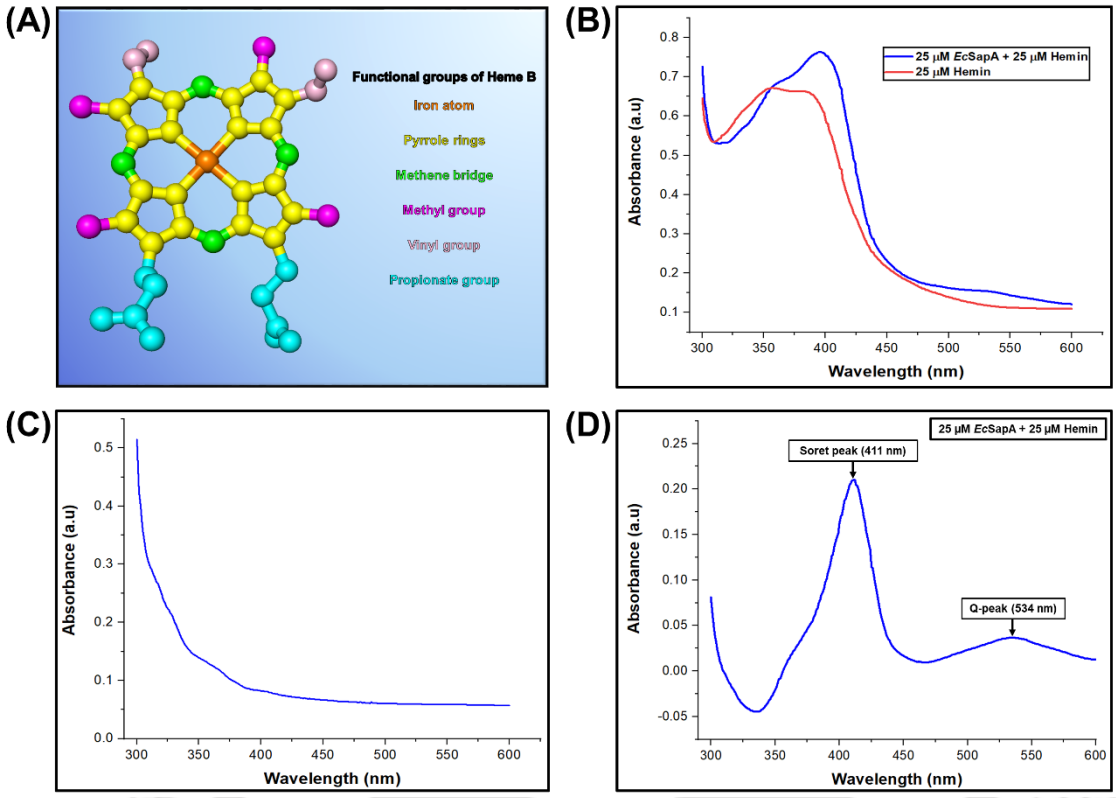


Figure 6.8. DAS-based analysis of *Ec*SapA. (A) Cartoon representation of heme B. The various functional groups of heme B, such as the iron atom (orange), pyrrole rings (yellow), methene bridge (green), methyl group (magenta), vinyl group (pink), and propionate group (cyan), have been represented in different colors mentioned within parenthesis. (B) DAS-based comparative analysis of *Ec*SapA-hemin complex (blue line) and only hemin (red line). A distinct shift was observed in the protein-hemin complex in comparison to the only hemin peak. (C) DAS analysis of heme-free *Ec*SapA (i.e., only the protein). (D) DAS analysis of *Ec*SapA-hemin complex after subtracting the spectra of only hemin. All the graphs are a plot of absorbance (a.u) versus wavelength (nm).

6.3.6. *Ec*SapA demonstrates a high affinity for heme-binding

Earlier reports of *Hi*SapA suggested enhanced migration of heme-protein complex on native-PAGE in comparison to only protein (Lukacik *et al.*, 2022). Hence, *Ec*SapA was subjected to incremental concentrations of hemin (50 to 300 µM) and analyzed via non-denaturing PAGE. As per expectations, upon coomassie-based staining, the *Ec*SapA-hemin complex was observed to migrate

faster in comparison to *Ec*SapA devoid of hemin (Figure 6.9A). Apart from protein bands, hemin bands are also observable on the gel. The hemin bands were visible even in unstained conditions, and the same regions retained the coomassie stain (Figure 6.9A).

Along with the shift, multiple additional bands were observable, hence raising doubt about the specific interaction of *Ec*SapA with hemin. Hence, to ascertain that the bands visualized belong to *Ec*SapA, a native-PAGE Western blot experiment was strategized. The proteins on the native gel were transferred onto a nitrocellulose membrane and then developed with the help of anti-His_6 antibodies. Post development, all the bands observed in the native gel were observed to be detected via the antibody, hence ascertaining those to be *Ec*SapA-hemin complex, providing direct evidence of *Ec*SapA-hemin interaction (Figure 6.9B). Akin to the coomassie stained gel, the bands corresponding to free hemin are also observable on the blot, as the brick red color of hemin was visible in the undeveloped blot (Figure 6.9B). Further, on performing densitometric analysis of the bands corresponding to the free hemin, a drop of 39% intensity was observable (Figure 6.9C). This drop in the band intensity is indicative of hemin binding to *Ec*SapA and the presence of excessive hemin. Further, post 50 µM hemin, three distinctive bands are visible, wherein the first corresponds to free SapA, the second corresponds to the specific interaction of *Ec*SapA with hemin, and the third (from 100 µM) corresponds to the non-specific interactions of hemin with *Ec*SapA (Figure 6.9B). It is noteworthy that the third non-specific band appears in the lanes consisting of the free hemin-band, thus suggesting the third band to be a non-specific interaction due to the presence of excess hemin (Figure 6.9B). Hence, from the native Western experiment, it could be deduced that *Ec*SapA was getting saturated at concentrations below 100 µM of hemin.

Hence, owing to the limitation in the gel-based detection, an intrinsic fluorescence-based study was performed. As discussed earlier, three buried tryptophan residues (Trp251, Trp368, and Trp416) remain aligned within the proposed binding site of *Ec*SapA; hence, the binding of hemin within the binding site cleft will result in changes in the tryptophan environment, indicating binding of the molecule. The results revealed a drop in the intrinsic fluorescence of

*Ec*SapA with increasing concentrations of hemin (2 µM to 30 µM) (Figure 6.9D). Further, the λ_{max} of *Ec*SapA was observed to demonstrate a blue shift of 9 nm from 340 nm to 331 nm, which is indicative of the change in the local environment of the tryptophan residues (Figure 6.9D). The fluorescence data obtained was utilized to calculate the emission intensity, for which the data was fitted using Hill's equation. The analysis revealed *Ec*SapA to bind hemin with a K_d of 4.8 µM and a stoichiometry (n) of 1.4, wherein an R-square value of 0.99 establishes adequate fitting (Figure 6.9E). The K_d obtained is in accordance with previous studies, wherein based on SPR-based studies, *Hi*SapA was suggested to bind heme with a K_d of 1.1 µM. Hence, our strategies suggest *Ec*SapA bind heme with high affinity.

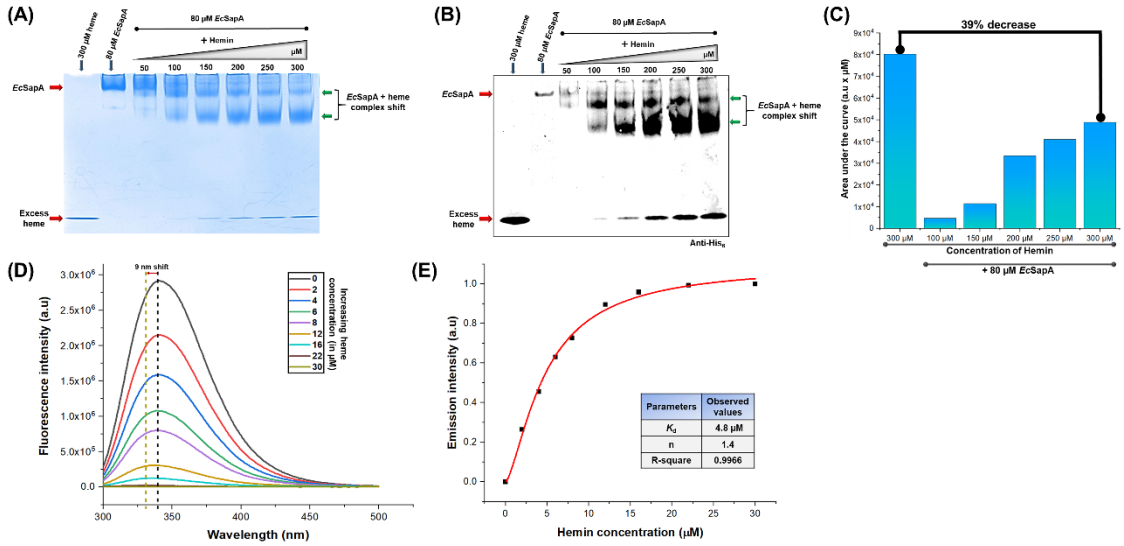


Figure 6.9. Biophysical strategies to demonstrate the heme-binding affinity of *Ec*SapA. (A) 8% Native-PAGE based EMSA analysis of *Ec*SapA-hemin complex, wherein 80 µM of *Ec*SapA is subjected to increasing concentrations of hemin (50 µM to 300 µM). (B) Anti-His_6-based immunoblot of *Ec*SapA-hemin complex, wherein only *Ec*SapA (80 µM) and only hemin (300 µM) were used as a negative control. (C) Densitometric analysis of free hemin bands observed in the immunoblot (B). A graph of the area under the curve versus the concentration of hemin is plotted in the form of histograms. The presence of hemin and *Ec*SapA in the samples is demarcated with the help of dumbbells below the X-axis of the graph. (D) Intrinsic fluorescence-based hemin-binding analysis of *Ec*SapA. The

increasing concentrations of hemin have been represented as varying colors, the details of which have been mentioned within the box (inset). The broken line represents the blue shift observed in the λ_{max} with the increasing heme concentrations, where the black lines represent only *Ec*SapA and olive-colored lines represent *Ec*SapA in the presence of 30 µM hemin. (E) The emission intensity calculated from the λ_{max} obtained from the fluorescence spectra is plotted onto a graph versus hemin concentration. The data points have been fitted utilizing Hill's equation, wherein the data is observed to saturate with increasing hemin concentrations. The estimated K_d and n values are mentioned in the box (inset).

6.3.7. *Ec*SapA can interact with protamine

Apart from heme, molecular docking studies discussed in Chapter 4 revealed the protamine sulfate-binding affinity of *Ec*SapA. Protamine is a bacteriolytic cationic AMP (antimicrobial peptide) wherein it is observed to interact with the Gram-negative inner cell membrane and hence disrupt the membrane homeostasis akin to other cationic AMPs (Pink *et al.*, 2014). Protamine was observed to dock at the interphase of Domain II and Domain III of *Ec*SapA and was observed to be held in position primarily via salt bridges formed between the positively-charged residues of protamine and the negatively-charged residues of *Ec*SapA (Figure 6.10A). To further inspect the interaction of protamine with *Ec*SapA, a differential scanning fluorimetry (DSF) was devised. *Ec*SapA (10 µM) was incubated with 10x concentration of protamine sulfate (100 µM) and then subjected to DSF. In DSF, the melting of the protein is mapped with the help of the fluorescent dye SYPRO orange. The native curve of *Ec*SapA was observed to have a T_m at 49.5°C in accordance with earlier observations (Figure 6.10B). Intriguingly, the melting temperature of *Ec*SapA incubated with protamine sulfate demonstrated a biphasic nature wherein two T_m's, 44°C and 60.5°C, were observed (Figure 6.10B). This biphasic phenomenon in literature has been attributed to instances in which there is domain-specific interaction between a ligand and a multi-domain protein, wherein certain domains are stabilized by the ligand (Gao *et al.,* 2020). Similar trends were observable in the *Ec*SapA-protamine sulfate docked pose, wherein protamine sulfate forms biased interactions with Domains II and III while

scarce interactions are observed with Domain I. In accordance with the docking data, the *Ec*SapA-protamine sulfate interaction study shows a biphasic spectrum, which in turn suggests Domain I to melt prior to Domain II and III (Figure 6.10B). Hence, our DSF-based analysis provided insights into the plausible role of *Ec*SapA in AMP uptake.

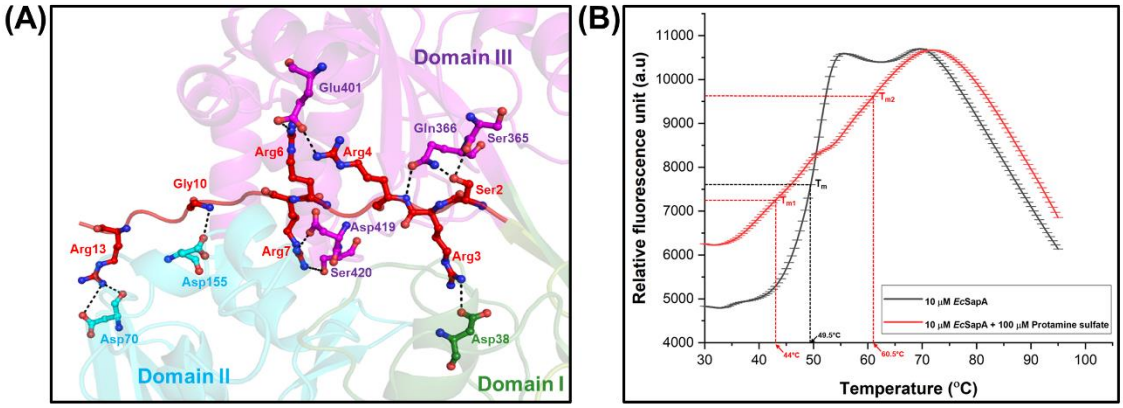


Figure 6.10. Interaction of *Ec*SapA with Protamine sulfate. (A) Docked-pose of protamine sulfate (red cartoon) with *Ec*SapA. The interacting residues have been represented in the form of ball and stick wherein the residues of protamine are red and the *Ec*SapA residues have been colored based on their domains: Domain I residue as green ball and stick, Domain II residues as cyan ball and stick, and Domain III as magenta ball and stick. (B) DSF-based melting temperature (T_m) estimation of *Ec*SapA (black) and *Ec*SapA-protamine sulfate (red) complex. The data is represented in a relative fluorescence (a.u) versus temperature (°C) graph.

6.4. Discussion

The Sap transport system is ubiquitous in several Gram-negative pathogens and hence is a common tool for resistance against the human immune system (Hsu *et al.*, 2019). However, the functional attributes of the Sap transport system, despite multiple reports, remain bleak, as a recent study elucidated the tertiary structure of *Hi*SapA, implicated the system to bind RNA in a non-specific manner, and further negated the AMP-binding possibility of *Hi*SapA (Lukacik *et al.*, 2021). Additionally, the functional attributes of the system have also been challenged in the case of the *Ec*Sap transport system. Wherein the membrane components

(*Ec*SapBCDF) have been suggested to function as a putrescine exporter, and *Ec*SapA has been suggested to be an orphan SBP unrelated to the *Ec*Sap transport system (Sugiyama *et al.*, 2016). However, a recently elucidated structure of the *Hi*SapA in open conformation sheds light on the proposed functionality of the system. In the recent structure of *Hi*SapA, no RNA molecule has been observed to be bound. Further, via biophysical strategies, interaction with human β-defensins was also established (Rivera *et al.,* 2024). In the case of the *Ec*Sap transport system, our results elaborated in Chapter 3 suggest *Ec*SapA to be a part of the *Ec*Sap transport system. Alongside, the sequence- and structure-based studies discussed in Chapter 4 suggest the architecture of *Ec*SapBC to be similar to that of type I ABC importers, and hence, the possibility of them functioning as an exporter remains obscure. In this study, we have explored the functional prospects of *Ec*SapA with the help of experimental strategies.

Our mass spectrometric analysis enlightened the significance of the subcellular location of *Ec*SapA in attaining adept fold. The tr*Ec*SapA construct was observed to get diverted in the insoluble fraction, while the full-length construct was observed to be in the soluble fraction. On inspection, the full-length construct attains the molecular weight corresponding to the truncated construct, hence suggesting truncation of the signal peptide from the full-length construct. Preliminary prediction tools predict *Ec*SapA to follow the Sec pathway, hence bringing the data to a full circle. The Sec pathway consists of a chaperon component SecB, an ATPase protein SecA, and a translocon complex constituted by SecYEG (Mori and Ito, 2001). In the Sec pathway, the SecB component binds the unfolded protein and helps stabilize the unfolded polypeptide. The SecB bound to the polypeptide then docks onto the translocon complex (SecYEG), where the unfolded protein is transferred onto the translocon complex. Post-transfer, the protein is translocated into the periplasm, where the transport is facilitated via multiple ATP-hydrolysis cycles (Tsirigotaki *et al.,* 2017). Finally, prior to the transfer, the signal peptide present in the protein is cleaved off from the polypeptide with the help of signal peptidases present alongside the SecYEG complex (Tsirigotaki *et al.,* 2017). The unfolded polypeptide attains its

adept fold with the aid of several molecular chaperons present in the periplasm (Rietsch and Beckwith, 1998).

Our fluorescence-based studies unraveled the essentiality of the disulfide bonds present in Domain I (Cys 60 and Cys267) and Domain III (Cys475 and Cys488). The absence of the disulfide bonds affects the structural attributes of *Ec*SapA and, hence, is crucial for their functioning. The bacterial cytosol has been widely attributed to be reducing in nature and hence suggested to house a lesser number of proteins consisting of disulfide bonds (Stewart *et al.,* 1998). Multiple periplasmic proteins are reported to possess disulfide bonds owing to the oxidizing environment and the presence of disulfide isomerases, which enable adept disulfide bond formations (Rietsch and Beckwith, 1998; Chautrand et al., 2022). Hence, it may be suggested that when tr*Ec*SapA is expressed within the cytosol, it fails to form disulfide bonds and, hence, fails to fold adeptly. Therefore, for adept folding and proper disulfide bond formation, *Ec*SapA is shuttled into the periplasm, wherein it is folded into its functional form.

The *Hi*Sap transport system has been suggested to be involved in the uptake of heme (Tanaka and Pinkett, 2019). In the case of *H. influenzae,* the organism is a heme auxotroph and largely depends on the *Hi*Sap transport system for heme uptake. Hence, the *Hi*SapA has multifunctional attributes wherein, apart from AMP-binding, it is also implicated in the moonlight as a heme and dipeptide-binding protein (Lukacik *et al.*, 2021). Akin to *Hi*SapA, our results elaborated in Chapter 4 also suggest the *Ec*Sap transport system to be multifunctional in nature, wherein it is implicated in binding AMPs, heme, and dipeptides. In our study, elaborate biophysical strategies enabled us to establish the heme-binding propensity of *Ec*SapA. The fluorescence-based studies suggested *Ec*SapA to bind heme with a K_d of 4.8 µM, which was similar to the K_d of *Hi*SapA (1.1 µM) (Tanaka and Pinkett, 2019).

Heme plays a crucial role in the pathogenesis of *E. coli*; hence, the bacterium possesses an elaborate system to acquire heme molecules from the host cells, enabling their sustenance (Figure 6.11). The host cells, such as red blood cells, macrophages, etc., are reported to harbor haemoglobins which house heme

molecules (Saha *et al.,* 2014). Reports suggest that with the aid of hemolysins, *E. coli* attacks the host cell membrane and ruptures it, releasing haemoglobins and hemoproteins (Skals *et al.,* 2009; Runyen-Janecky, 2013). To extract the heme molecules bound to the host proteins, *E. coli* is reported to produce haemoglobin proteases which cleave them and release heme molecules (Otto *et al.,* 2005). The free heme molecules are taken up in a TonB-dependent manner (via TonB-ExbB and ExbD complex) via the outer membrane receptor ChuA (*E. coli* heme utilization protein) (Torres and Payne, 1997). The free heme molecules are suggested to be transported into the cytosol via the ABC importer ChuTUV (Suits *et al.*, 2009). Certain reports suggest apart from ChuT, the dipeptide-binding protein *Ec*DppA is also observed to be involved in the uptake of heme from the periplasm (Létoffé *et al.,* 2006). Akin to its homolog, *Ec*SapA was also observed to dock heme, and further, the results of various biophysical strategies demonstrate the moonlighting role of *Ec*SapA in heme acquisition within *E. coli*. Hence, to fulfill the demand for heme in scarce situations, bacterial pathogens like *E. coli* most likely employ multiple transporters to meet their need.

The major ligand for the Sap transport system was suggested to be the AMPs. However, certain reports pertaining to the *Ec*Sap and *Hi*Sap transport systems made the AMP-binding possibility of the Sap transport system vague (Sugiyama *et al.*, 2016; Lukacik *et al.*, 2021). Recently, via biophysical strategies, Pinkett and coworkers suggested *Hi*SapA bind human β-defensins, which were in accordance with our results discussed in Chapter 4 (Rivera *et al.,* 2024). Further, our results discussed in Chapter 4 suggest a cationic AMP-binding possibility for *Ec*SapA. Our *in silico* predictions were experimentally validated in this study, hence providing the first evidence for *Ec*SapA-protamine interaction. Hence, both results align with prior reports of AMP uptake facilitated via the Sap transport system. Moreover, the biophysical strategies employed in this study establish the multifunctional attributes of *Ec*SapA and suggest its importance in sustaining within the host.

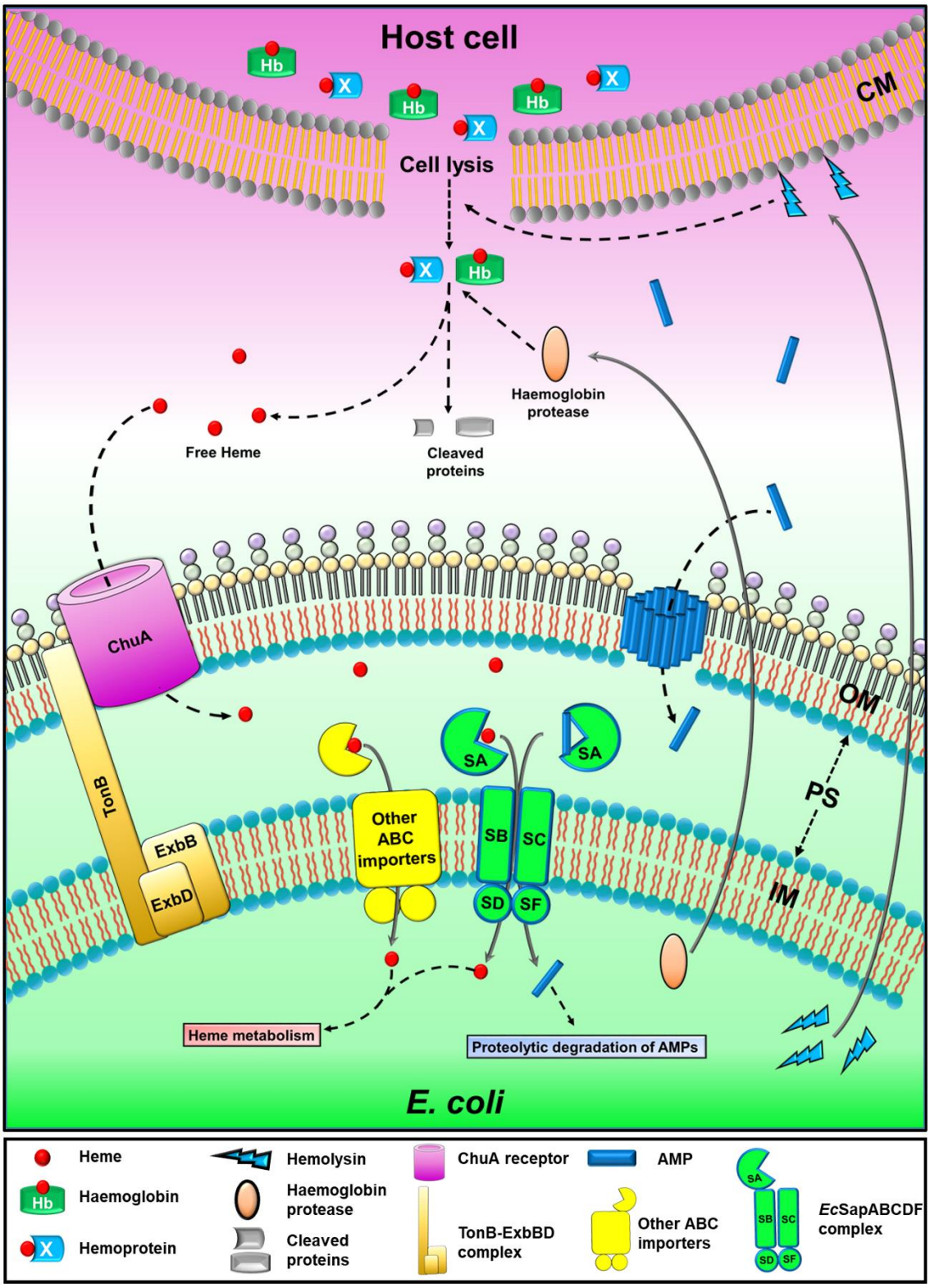


Figure 6.11. Schematic representation of competition between the host and *E. coli*. The hemolysins (cyan, distorted arrowheads) are produced and transported by *E. coli* to disrupt the cell, leading to the release of hemoproteins (cyan, curved square) and haemoglobins (green, rectangle). The haemoglobin

protease (red, oval) cleaves the proteins and releases free heme (red, circle) to be taken up by ChuA in a TonB-dependent manner. Further, the heme in the periplasm is taken up by the *Ec*Sap transport system and other ABC importers. The *Ec*Sap transport system also takes up cationic AMP molecules (blue, rectangle), ultimately leading to their proteolytic degradation. The details of the cartoon representation of each component in the figure have been elaborated in the box below the figure. Further, various abbreviations used are as follows: CM stands for cell membrane; OM is the outer membrane; PS is the periplasmic space; and IM is the inner membrane.

6.5. Conclusion

The outcomes of the study unravel the functional significance of *Ec*SapA. The study sheds light on the significance of the sub-cellular localization of *Ec*SapA for its structural stability. Further, the essentiality of adept disulfide bond formation and their maintenance in *Ec*SapA was observed to be a prerequisite for the structural integrity of the protein. Sequence- and structure-based analysis revealed no RNA-binding propensity in *Ec*SapA, unlike *Hi*SapA. The heme- and protamine sulfate-binding propensity of *Ec*SapA was also observed, highlighting the promiscuous binding site of *Ec*SapA, which enables its multifunctionality. Therefore, the outputs of this study may act as a pedestal to further peek into the multiple attributes of the *Ec*Sap transport system.

Appendix D. Supplementary data

Supplementary Figures: D1-D5

Summary

Summary

The results of this study make an effort to unravel the mysterious situation of the *Ec*Sap transport system. With the aid of computational and experimental strategies, the functional prospects of the *Ec*Sap transport system have been examined and the outputs of the research demonstrate certain crucial findings pertaining to the system in question.

The study was initiated by data mining all the available Sap components, which yielded 421 Sap transport systems belonging to unique bacterial species. Amongst the 421 entries, 352 were observed to follow the commonly reported genetic trends of the *sap* operon and hence were coined as consensus operons. The remaining 69 were observed to stray away from the conserved trends and hence were suggested to be non-consensus. Further, the Psp system and enoyl reductase were observed to be conserved upstream and downstream of the consensus *sap* operons. Both the Psp system and enoyl reductase are reported to be involved in bacterial inner membrane maintenance; hence their conservation in the immediate genetic neighborhood of *sap* operons is indicative of functional correlatedness. The *Ecsap* operon was observed to demonstrate all the trends of the consensus operon, hence suggesting *Ec*SapA to be a part of the *Ec*Sap transport system. Further, on inspection of the 69 non-consensus entries, three distinct variations were observed, which allowed us to classify them into three distinct groups as follows: Group I, operons consisted of fragmented or intermediate genes; Group II, operons consist of other components in place of *sap* components and Group III, operons have missing *sap* operon components. The trends observed in the non-consensus operons are reminiscent of stages of operon formation. Further, phylogenetic analysis with the consensus and non-consensus entries suggests a similar trend. Hence, this highlights the putative evolutionary trend of *sap* operon formation.

Having established the presence of *Ec*SapA in the *Ec*Sap transport system, the functional properties of *Ec*SapA were analyzed. Initial structure and sequence-based homology suggested the similarity of *Ec*SapA with dipeptide-binding proteins. Further, virtual screening studies enlightened six dipeptides (Trp-Phe,

Phe-Trp, Trp-Ile, Trp-Pro, Tyr-Phe, and Tyr-Ile) to be the putative ligands of *Ec*SapA. The docking analysis also unveiled two distinct dipeptide binding sites, A and B. Further, binding-site volume analysis revealed a wider binding site of *Ec*SapA, and the binding site charge-based analysis unveiled the AMP-binding possibility of *Ec*SapA. Moving ahead, protein-protein docking enlightened the protegrin-1 and protamine-binding capacity of *Ec*SapA. Further, molecular docking studies highlighted the heme-binding propensity of *Ec*SapA. Hence, the current study enlightened a multifunctional attribute of *Ec*SapA. The outputs of the study further reveal an importer-like trend in the case of *Ec*SapBC. Therefore, it suggests that *Ec*SapBC functions as an importer. Moreover, a charge-based assembly mechanism of *Ec*SapABCDF has also been proposed.

To establish the functional prospects of *Ec*SapA, all the docked poses were subjected to MD simulation-based studies. Wherein it was observed that four dipeptides (Trp-Ile, Trp-Pro, Tyr-Ile, and Try-Phe) were maintained within the binding site of *Ec*SapA, while the remaining two dipeptides (Trp-Phe and Phe-Trp) were gradually rejected. Further, the results demonstrate an initial and a final binding site for dipeptides. Hence, dipeptides are suggested to first bind in the initial binding site, which in turn would elicit conformational changes in *Ec*SapA, translating the dipeptides to the final binding site. Further, protegrin-1, protamine, and heme were also observed to be maintained within respective binding sites throughout the MD simulation runs of 1 µs each. Thus, a segregated binding site of *Ec*SapA was revealed to enable the accommodation of multiple ligands. Interestingly, the MD simulation data also unraveled the varied domain motions of *Ec*SapA in accordance to the ligand. As each ligand commands a varied domain movement, therefore *Ec*SapA may not be capable of binding more than one ligand at a time.

To gain further insights, *Ec*SapA was studied using biophysical strategies. The results indicate the essentiality of the sub-cellular localization of *Ec*SapA to attain adept fold. Further, the key role of the two disulfide bonds in the maintenance of *Ec*SapA structural integrity was also highlighted with the aid of a fluorescence spectroscopy-based study. In accordance with our prior data, *Ec*SapA was demonstrated to bind heme based on multiple biophysical strategies. Further,

differential scanning fluorimetry (DSF)-based analysis also highlighted the interaction of *Ec*SapA with protamine in accordance with our computational data.

Hence, succinctly, the results of this study establish the functional properties of the *Ec*Sap transport system with the aid of robust computational and biophysical strategies. The outcomes of this study establish *Ec*SapA to be a part of the *Ec*Sap transport system and further enlighten the probable evolutionary trail for the formation of the *sap* operon. Alongside this, the study throws light on the multifunctional prospects of *Ec*SapA. Hence, establishing the *Ec*Sap transport system as a key component in aiding the pathogenicity and sustenance of *E. coli,* therefore making it a lucrative target for future drug development studies.

Appendix A

Supplementary data to Chapter 3

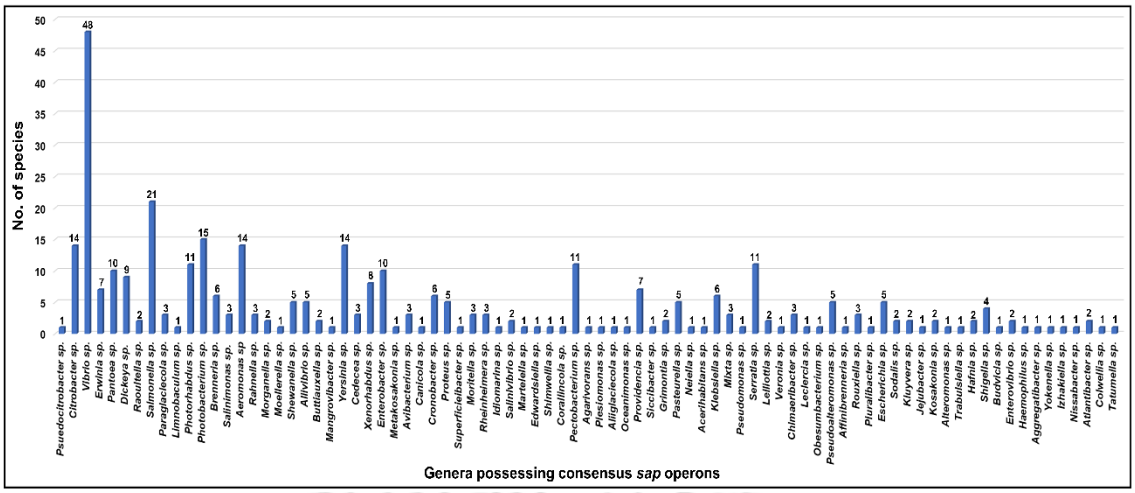


Figure A1. Distribution of consensus *sap* operons across 83 unique genera. The X-axis represents all the unique genera possessing consensus *sap* operon, while the Y-axis represents the number of species within each unique genera possessing *sap* operon.

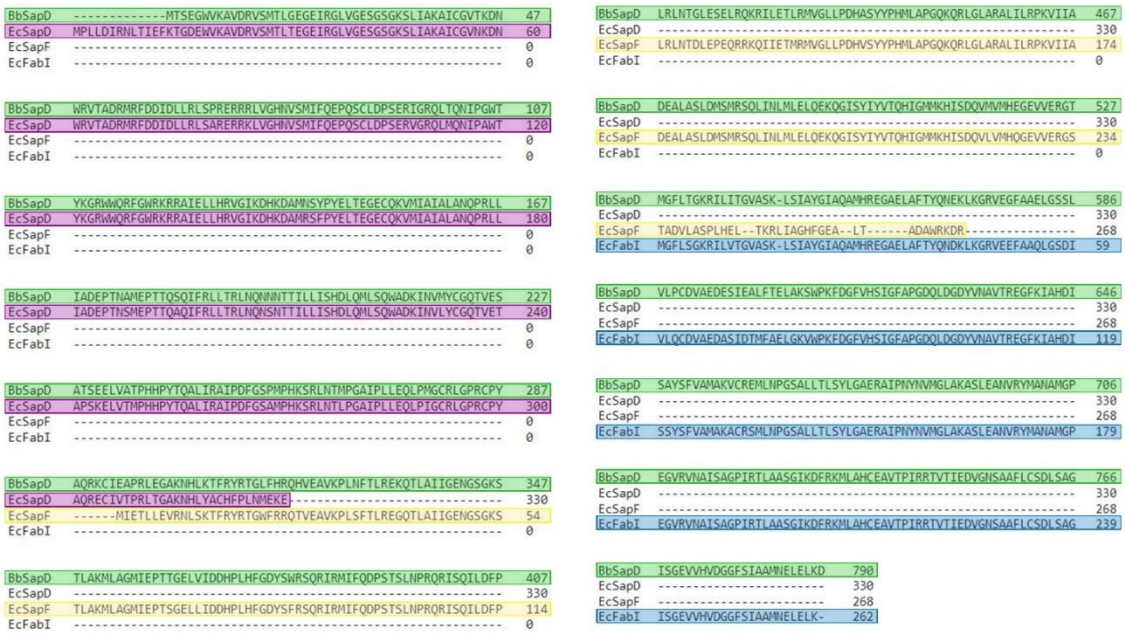


Figure A2. Multiple sequence alignment of *B. bassiana* SapD along with *E. coli* SapD, SapF, and FabI. The MSA was performed using the protein sequences of *Bb*SapD (*B. bassiana*, A0A0A2VES7), *Ec*SapD (*E. coli*, P0AAH4), *Ec*SapF (*E. coli*, P0AAH8), and *Ec*FabI (*E. coli*, P0AEK4); the organisms name and UniProt id of each protein are provided in the parenthesis. The sequence of *Bb*SapD, *Ec*SapD, *Ec*SapF, and *Ec*FabI has been highlighted with green, magenta, yellow, and blue, respectively.

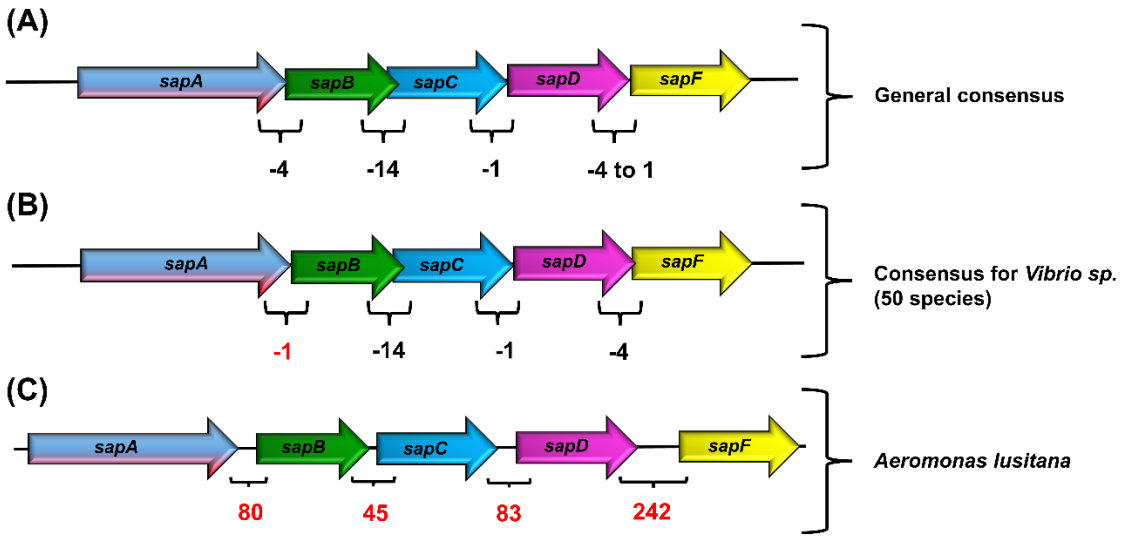
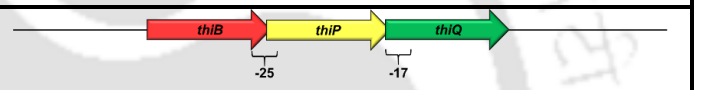
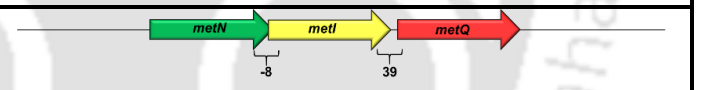
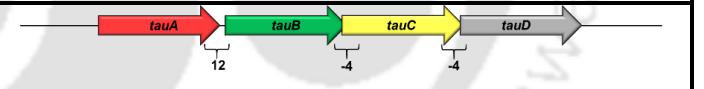
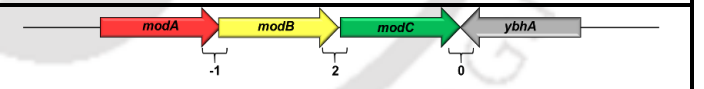
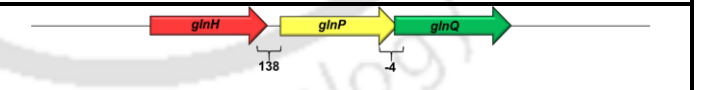
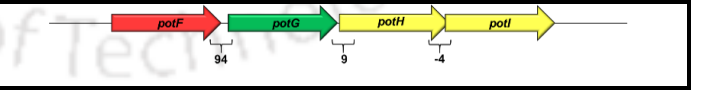
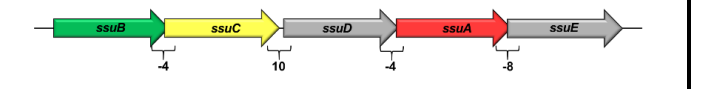
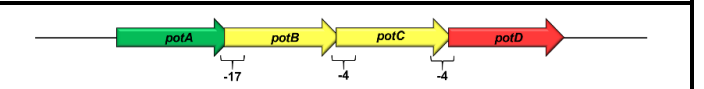
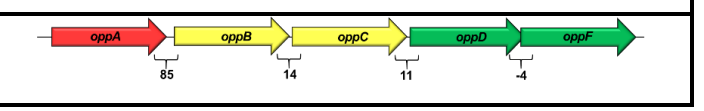
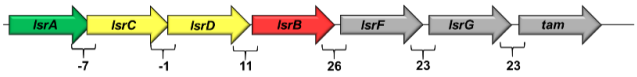
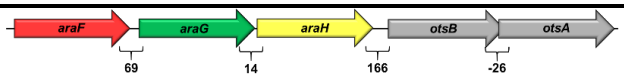
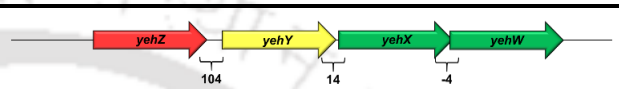
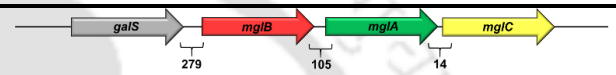
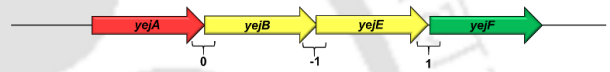
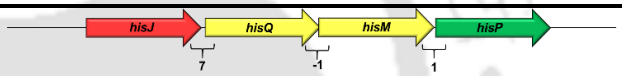
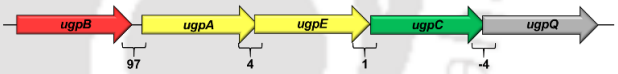
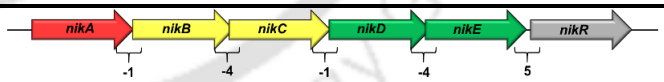
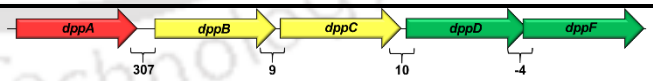
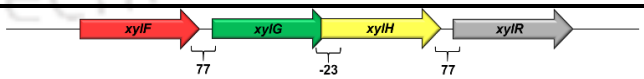
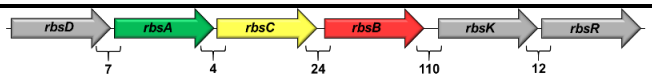
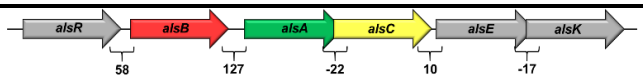
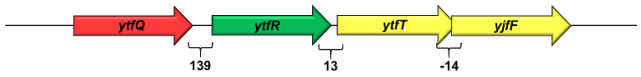
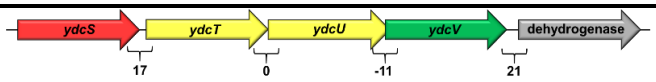
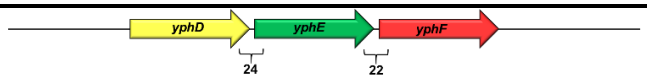
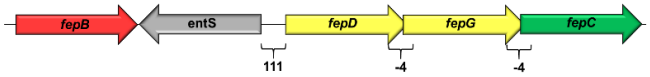
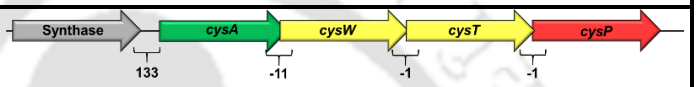
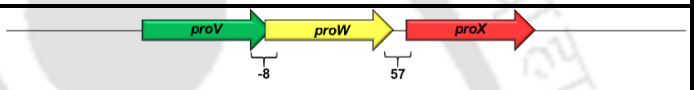
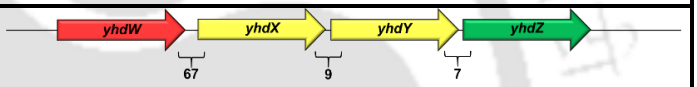
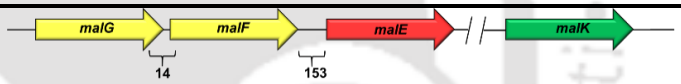
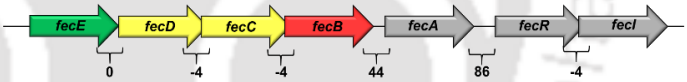


Figure A3. Varying intermediate distances of *sap* operons. (A) General consensus arrangement of *sap* operons. Varied intergenic distances of *sap* operons of (B) *Vibrio sp.* and (C) *Aeromonas lusitana.* The genetic arrangement of *sap* components is displayed as arrows in *sapA* (blue), *sapB* (green), *sapC* (light blue), *sapD* (magenta), and *sapF* (yellow). The gene names have been labeled inside the arrows. The numerical mentioned below the braced brackets indicate the intergenic distance between the adjacent gene pairs. The intergenic distances that stray away from the consensus value have been mentioned in red.

Table A1. Operonic arrangement of ABC importers of *E. coli* strain K-12 enlisted in the TransportDB database. The SBPs, TMDs, and NBDs have been represented as red, yellow, and green arrows, while other related genes of the operon have been represented as gray arrows. The gene name of each component has been mentioned inside the arrows.

S. No.	ABC importers in *E. coli* strain K-12	Operonic arrangement
1.	Glutamate/aspartate transporter	
2.	Glutathione transporter	
3.	Arginine /ornithine transporter	
4.	Thiamin transporter	
5.	Ferric hydroxamate transporter	
6.	Methionine transporter	
7.	Taurine transporter	
8.	Molybdenum transporter	
9.	Glutamine transporter	
10.	Putrescine transporter	
11.	Aliphatic sulphonate transporter	
12.	Spermidine/putrescine transporter	
13.	Oligopeptide transporter	

14.	D, D-dipeptide transporter	
15.	Autoinducer 2 transporter	
16.	Zinc transporter	
17.	L-Arabinose transporter	
18.	L-Cystine transporter	
19.	Glycine betaine transporter	
20.	Galactoside transporter	
21.	Antimicrobial peptide transporter	
22.	Histidine transporter	
23.	Phospholipid transporter	
24.	Glycerol-3-phosphate transporter	
25.	Branched chain amino acid transporter	
26.	Nickel transporter	
27.	Dipeptide transporter	
28.	D-xylose transporter	
29.	Phosphate transporter	
30.	Ribose transporter	
31.	D-allose transporter	

32.	Galactofuranose transporter	
33.	Ydc transporter	
34.	Yph transporter	
35.	Ferric enterobactin transporter	
36.	Ycj transporter	
37.	Ynj transporter	
38.	Sulphate transporter	
39.	L-Proline transporter	
40.	Amino acid transporter	
41.	Maltose transporter	
42.	Iron(III) dicitrate transporter	

Note: The curly braces demarcate the intergenic region with the distance (in bp) mentioned below the curly braces.

Table A2. A list of probable SapZ proteins was observed to be upstream of consensus *sap* operons following a similar genetic arrangement akin to the NTHI *sap* operon.

S. No.	Organism name (UniProt id)	Protein name	Intergenic distance
1.	*Vibrio cholerae* (A0A0H3AJU0)	conserved hypothetical protein	1
2.	*Vibrio gazogenes* (A0A1Z2SAU0)	hypothetical protein	2
3.	*Vibrio nigripulchritudo* (U4KEV0)	conserved hypothetical protein	9
4.	*Vibrio splendidus* (A0A2N7PGJ1)	hypothetical protein	37
5.	*Aliivibrio salmonicida* (B6ER97)	putative membrane protein	9
6.	*Paraglaciecola mesophila* (K6XY54)	hypothetical protein	89
7.	*Vibrio sinaloensis* (E8M9T5)	hypothetical protein	9
8.	*Yokenella regensburgei* (G9Z6Z1)	hypothetical protein	-3
9.	*Aliivibrio fischeri* (Q5E0S2)	hypothetical protein	10
10.	*Avibacterium avium (Pasteurella avium)* (A0A379AQ55)	Domain of uncharacterised function (DUF1083)	35
11.	*Pseudidiomarina homiensis* (A0A432Y493)	sodium/proline symporter PutP	-7
12.	*Vibrio orientalis* (C9QJY1)	hypothetical protein	8
13.	*Pseudoalteromonas piscicida* (A0A2A5JT36)	DNA-binding protein	19
14.	*Vibrio fluvialis* (S7I377)	hypothetical protein	0

15.	*Vibrio mimicus* (D2YJ16)	conserved hypothetical protein	34

Table A3. List of non-consensus *sap* operons and their respective operonic arrangements. The non-consensus entries are divided into three different groups: Group I, fragmented or intermediate genes (red); Group II, presence of other components (yellow); and Group III, missing genes (green).

Group	Organism name	UniProt id	Operonic construct	Remarks if any
Group I	*Rodentibacter pneumotropicus*	A0A3S4Y2K4	*sapA/sapB/sapC/sapD/sapF*_fragmented_operon	-
	Vibrio ponticus	A0A090PC72	*sapA/sapB/sapC/sapD/sapF_sapAC*_fragmented	-
	Kluyvera cryocrescens	A0A485A2H1	*sapA/sapB/sapC/sapD/sapF*_fragmented_operon	-
	Photobacterium aphoticum	A0A090QPK8	*sapA/sapB/sapC/sapD/sapF*_negative_direction_fragmented_operon	-
	Klebsiella variicola	A0A7H4SFB2	*sapA/sapB/sapC/sapD* is fragmented/*sapF*_negative_direction	-
	Haemophilus pittmaniae HK 85	F9QAX3	*sapA*_fragmented/*sapB/sapC/sapD/sapF*	-
	Vibrio variabilis	A0A090T2G9	*sapA/sapB/sapC/sapD/sapF*_negative_direction_*sapBCDF*_fragmented	-
	Serratia rubidaea	A0A4U9HFK1	*sapA/sapB/sapC/sapD/sapF*_negative_direction_*sapACD*_fragmented	-
	Klebsiella aerogenes	A0A3S4JJX6	*sapA/sapB/sapC/sapD*_fragmented/*sapF*	-

	Sodalis glossinidius	Q2NSU4	*sapA/sapB*_intermediate_genes_*sapC/sapD/sapF*	-
	Pantoea conspicua	A0A1X1C156	*sapA/sapB*_intermediate_genes_*sapC/sapD/sapF*_negative_direction	-
	Escherichia albertii	A0A1X3JVH2	*sapA/sapB*_intermediate_genes_*sapC/sapD/sapF*_negative_direction	-
	Pantoea coffeiphila	A0A2S9I6W3	*sapA/sapB*_intermediate_genes_*sapC/sapD/sapF*	-
	Salmonella schwarzengrund	A0A0N1QSB5	*sapA/sapB*_intermediate_genes_*sapC/sapD/sapF*	-
	Salmonella montevideo	A0A3Y0E377	*sapA/sapB*_intermediate_genes_*sapC/sapD/sapF*	-
	Enterobacter agglomerans (Erwinia herbicola) (Pantoea agglomerans)	A0A8I0HTD5	*sapA/sapB*_intermediate_genes_*sapC/sapD/sapF*_negative_direction	-
	Salmonella diarizonae	A0A5U3CWJ2	*sapA/sapB*_intermediate_genes_*sapC/sapD/sapF*	-
	Salmonella typhisuis	A0A735MXL3	*sapA/sapB*_intermediate_genes_*sapC/sapD/sapF*_negative_direction	-
	Klebsiella quasivariicola	A0A8B4TL28	*sapA/sapB*_intermediate_genes_*sapC/sapD/sapF*	-
	Kalamiella piersonii	A0A7Y8YB38	*sapA/sapB*_intermediate_genes_*sapC/sapD/sapF*	-
	Tatumella ptyseos ATCC 33301	A0A085JEH0	*sapA/sapB*_2_intermediate_genes_*sapC/sapD/sapF*_negative_direction	-
	Pantoea wallisii	A0A1X1DBD8	*sapA/sapB*_1_intermediate	-

			gene_*sapC/sapD/sapF*	
	Colwellia marinimaniae	A0A1Z5I7Y9	*sapA/sapB/sapC*_1_intermediate_gene_*sapD/sapF*	-
	Actinobacillus pleuropneumoniae	A0A3S4Z4N8	*sapA/sapB/sapC/sapD_sapF*(missing)_negative_direction	-
	Haemophilus paraphrohaemolyticus	I2NJ01	*sapA/sapB/sapC/sapD_sapF*(missing)_negative_direction	-
	Haemophilus parahaemolyticus	A0A377I0R7	*sapA/sapB/sapC/sapD_sapF*(missing)	-
	Actinobacillus minor	C5S2G4	*sapA/sapB/sapC/sapD_sapF*(missing)_negative_direction	-
	Bibersteinia trehalosi	W0R7L4	*sapA/sapB/sapC/sapD_sapF*(missing)_negative_direction	-
	Haemophilus ducreyi	Q7VM01	*sapA/sapB/sapC/sapD_sapF*(missing)	-
	Actinobacillus lignieresii	A0A380TUH1	*sapA/sapB/sapC/sapD_sapF*(missing)	-
Group II	*Pseudomonas sp.*	A0A2N8BEH0	*sapA/dppB/dppC/dppD/dppF*	Correct annotation of reported genes validated via sequence-based homology.
	Gordonia insulae	A0A3G8JU61	*sapA*/*tam_4*(methyltransferase)/*gsiCD*/*gsiA*	Correct annotation of reported genes validated via sequence-based homology.
	Providencia rustigianii	A0A379G359	*dppA/sapB/sapC/gsiA3/gsiA4*_negative_direction	Misannotated *sap* components validated via sequence-based homology.
	Halolamina pelagica	A0A0P7H0B4	*sapA/nikB/gsiD/malK18/malK17*_negative_direction	Correct annotation of reported genes validated via

				sequence-based homology.
	Vibrio thalassae	A0A240EG78	*dppa/sapB/sapC/ ddpD/gsiA*	Misannotated *sap* components validated via sequence-based homology.
	Photobacterium toruni	A0A1T4LKH3	*dppA/sapB/sapC/ ddpD/gsiA*_negative_direction	Misannotated *sap* components validated via sequence-based homology.
	Photorhabdus namnaonensis	A0A1B8YEP1	*dppA/sapB/sapC/ ddpD/gsiA*_negative_direction	Misannotated *sap* components validated via sequence-based homology.
	Photobacterium andalusiense	A0A1Y6M9E0	*dppA/sapB/sapC/ ddpD/gsiA*_negative_direction	Misannotated *sap* components validated via sequence-based homology.
	Shewanella khirikhana	A0A3Q9E4E6	*dppA/sapB/sapC/ ddpD/gsiA*_negative_direction	Misannotated *sap* components validated via sequence-based homology.
	Raoultella electrica	A0A514ESK2	*dppA/sapB*_intermediate_genes_*sapC/ddpD/gsiA*	Misannotated *sap* components validated via sequence-based homology.
	Rheinheimera nanhaiensis E407-8	I1DWW7	*dppA/dppB/sapC/ sapD/sapF*	Correct annotation of reported genes validated via sequence-based homology.
	Idiomarina loihiensis (strain ATCC BAA-735 / DSM 15497 / L2-TR)	Q5QUJ7	*dppA/dppB/sapC/ sapD/sapF*	Correct annotation of reported genes validated via sequence-based homology.
	Staphylococcus schweitzeri	A0A2K4AMP7	*sapA/gsiD/gsiC/gsiA(opposite direction)*	Correct annotation of reported genes validated via sequence-based homology.
	Solibacillus isronensis	K1L6X2	*sapA/gsiD/gsiC/gsiA(opposite*	Correct annotation of reported genes

			*direction)*_negative_direction	validated via sequence-based homology.
	Ruminiclostridium hungatei (Clostridium hungatei)	A0A1V4SM74	*nikA/sapF/ddpD/gsiD/gsiC*_negative_direction	Correct annotation of reported genes validated via sequence-based homology.
Group III	*Enterobacter hormaechei*	A0A2J0R4D8	*sapA/sapB*_incomplete_genome	-
	Erwinia billingiae	D8MSN3	*sapA/sapB*_other components are missing	-
	Pantoea latae	A0A1V9DLR6	*sapA/sapB*_remaining components not observed_negative_direction	-
	Ewingella americana	A0A502GK87	*sapA/sapB*_remaining components not observed	-
	Erwinia tracheiphila	A0A345CR28	*sapA/sapB*_remaining components not observed	-
	Rahnella woolbedingensis	A0A419NBH0	*sapA/sapB*_remaining components not observed	-
	Pantoea brenneri	A0A7Y6TQQ0	*sapA/sapB*_remaining components not observed	-
	Glaesserella parasuis ZJ0906	A0A806J7A0	*sapA/sapB*_remaining components not observed_negative_direction	-
	Candidatus *Symbiopectobacterium sp.* Dall1.0	A0A8J7KQ56	*sapA/sapB*_remaining components not observed	-
	Meiothermus luteus	A0A399F1M7	*sapA*_remaining components not observed	-
	Thermoplasmatales archaeon	A0A7U3GR96	*sapA*_remaining components, not observed_negative_direction	Probable nickel-binding protein based on sequence-based homology.

	Enhygromyxa salina	A0A0C1ZP53	*sapA*_remaining components not observed_negative_direction	-
	Salinivirga cyanobacteriivorans	A0A0S2I0E4	*sapA*_remaining components not observed	-
	Trinickia symbiotica	A0A2N7X9A4	*sapA*_remaining components not observed	-
	Arsenophonus endosymbiont of Bemisia tabaci	A0A4Q1B4T5	*sapA*_remaining components not observed	-
	Sphingopyxis sp.	A0A2Z5UQ54	*sapA*(only 75 amino acids long)_remaining components not observed	-
	Bacillus sp.	A0A553ZT36	only *sapA* (genomic data not available)	-
	Paraglaciecola polaris LMG 21857	K6ZSU4	*sapA/sapC/sapD/sapF_sapB*(missing)	-
	Beauveria bassiana	A0A0A2VIV3	*sapA/sapC/sapD*_mRNA based sequence_*sapBF* (missing)	-
	Hamiltonella defensa subsp. Acyrthosiphon pisum (strain 5AT)	C4K7X5	*sapB/sapC/sapD/sapF_sapA*(missing)_negative_direction	-
	Bacillus pacificus	A0A3P1B7C5	*sapB/ddpC*_remaining components not observed	-
	Gilliamella apicola	A0A556S8U2	*sapB/sapC/sapD/sapF_sapA*(missing)_negative_direction	-
	Streptococcus pseudopneumoniae (strain IS7493)	G0I6R5	*sapF*_remaining components, not observed_negative_direction	-
	Aggregatibacter actinomycetemcomitans	H0KG84	*sapA/sapB/sapC/sapD_sapF*(missing)_negative_direction	-

Table A4. Details of interactome analysis of non-consensus *sap* operon entries. The "organism name" column has been color-coded based on the group of non-consensus *sap* operon, wherein Group I is red, Group II is yellow, and Group III is green.

S. No.	Organism name	Protein interactome analysis output		
		Query protein (GenBank id)	Interacting proteins (GenBank id)	Score
1.	*Vibrio ponticus*	SapB (GAK87944.1)	SapD (GAK87941.1)	0.988
			SapC_1 (GAK87942.1)	0.975
			SapC_2 (GAK87943.1)	0.963
			SapA_1 (GAK87945.1)	0.935
			SapA_2 (GAK87946.1)	0.935
			SapA_3 (GAK87947.1)	0.935
			Dipeptide-binding ABC transporter (GAK85451.1)	0.848
			Oligopeptide ABC tranporter (GAK84486.1)	0.834
			SapF (GAK87940.1)	0.818
			OppF (GAK84481.1)	0.757
2.	*Photobacterium aphoticum*	SapA_2 (GAL04851.1)	SapA_1 (GAL04852.1)	0.995
			SapB_2 (GAL04849.1)	0.994
			SapB_1 (GAL04850.1)	0.994
			SapC_2 (GAL04847.1)	0.992
			SapC_1 (GAL04848.1)	0.991
			SapD (GAL04846.1)	0.975
			ABC permease (GAL07742.1)	0.931
			OppB (GAL05053.1)	0.918
			SapF_1 (GAL04845.1)	0.917
			ABC permease (GAL07741.1)	0.897

			SapF_2 (GAL04844.1)	0.864
			ABC permease (GAL06176.1)	0.817
3.	*Haemophilus pittmaniae* HK 85	SapD (EGV05302.1)	SapC (EGV05332.1)	0.995
			SapB (EGV05334.1)	0.993
			DppB (EGV04913.1)	0.889
			DppC (EGV04910.1)	0.885
			OppC (EGV05939.1)	0.86
			HbpA (EGV05700.1)	0.827
			SapA_1 (EGV05315.1) based on homology	0.808
			SapA_2 (EGV05314.1) based on homology	0.808
			SapF (EGV05320.1)	0.805
			OppB (EGV05914.1)	0.788
4.	*Klebsiella aerogenes*	SapA (VEC56604.1)	SapC (VEC56608.1)	0.999
			SapB (VEC56606.1)	0.999
			DppB (VEC48251.1)	0.86
			DdpB (AML38967.1)	0.796
			OppC (VEC53591.1)	0.778
			SapD_2 (VEC56613.1)	0.771
			DppC (VEC48253.1)	0.749
			Dipeptide ABC permease (AML38044.1)	0.736
			SapF (VEC56617.1)	0.71
			GsiC (AML34992.1)	0.697
5.	*Sodalis glossinidius*	SapA (BAE74781.1)	SapB (BAE74782.1)	0.997
			SapC (BAE74785.1)	0.94
			DppB (BAE73336.1)	0.844
			DppC (BAE73337.1)	0.815
			OppD_1 (BAE73338.1)	0.697

			OppF_1 (BAE73339.1)	0.695
			SapD (BAE74786.1) based on homology	0.663
			SapF (BAE74787.1) based on homology	0.562
			OppC (BAE74650.1)	0.503
			OppB_1 (BAE74649.1)	0.48
6.	*Pantoea conspicua*	SapA (ORM55238.1)	SapC (ORM55235.1)	0.987
			SapB (ORM55237.1)	0.983
			SapD (ORM55234.1) based on homology	0.881
			SapF (ORM55233.1) based on homology	0.848
			Unannotated protein (GCA_001743465_03470)	0.845
			DppC (ORM52486.1)	0.82
			OppC (ORM53237.1)	0.759
			GsiC (ORM55462.1)	0.729
			Peptide ABC permease (ORM54732.1)	0.708
			ZnuC (ORM54289.1)	0.664
7.	*Escherichia albertii*	SapA (EDS91296.1)	SapB (EDS91329.1)	0.999
			SapC (EDS91263.1)	0.986
			SapD (EDS91338.1)	0.945
			SapF (EDS91174.1)	0.924
			DppB (EDS93260.1)	0.897
			DppC (EDS93516.1)	0.858
			ABC permease (EDS90141.1)	0.828
			GsiC (EDS90932.1)	0.817

			Putative permease (EDS93472.1)	0.786
			YddQ (EDS90132.1)	0.775
8.	*Enterobacter agglomerans (Erwinia herbicola) (Pantoea agglomerans)*	SapB (SUB16630.1)	SapA (SUB16631.1) based on homology	0.997
			SapC (SUB16622.1)	0.991
			SapD (SUB16620.1)	0.96
			SapF (SUB16619.1)	0.801
			DppC (SUB19164.1)	0.78
			DdpC_2 (SUB15422.1)	0.777
			GsiD_2 (SUB17201.1)	0.775
			DdpC_3 (SUB16860.1)	0.774
			GsiD_3 (SUB17654.1)	0.762
			OppC (EZI32824.1)	0.748
9.	*Tatumella ptyseos* ATCC 33301	SapD (KFD18866.1)	SapC (KFD18867.1)	0.985
			SapB (KFD18870.1)	0.979
			SapF (KFD18865.1)	0.884
			GsiD_1 (KFD22720.1)	0.851
			GsiD (KFD22460.1)	0.823
			SapA (KFD18871.1) based on homology	0.82
			DdpB (KFD22502.1)	0.807
			DdpC (KFD22503.1)	0.797
			GsiC_2 (KFD22721.1)	0.794
			GsiD_6 (KFD21640.1)	0.783
10.	*Pantoea wallisii*	SapA (ORM73934.1)	SapB (ORM73935.1)	0.983
			SapD (ORM73937.1)	0.881
			SapF (ORM73938.1)	0.848
			DppB (ORM68999.1)	0.845
			DppC (ORM68998.1)	0.82
			OppC (ORM73981.1)	0.759

			GsiC (ORM69163.1)	0.729
			Peptide ABC permease (ORM71418.1)	0.708
			LptG (ORM73017.1)	0.684
			Acs (ORM75244.1)	0.658
11.	*Actinobacillus pleuropneumoniae*	SapA (ABN73892.1)	SapB (ABN73891.1)	0.998
			SapC (ABN73890.1)	0.99
			SapD (ABN73889.1)	0.968
			TyrR (ABN73893.1)	0.845
			DppB (ABN73173.1)	0.837
			DppC (ABN73174.1)	0.798
			SapF (ABN74335.1)	0.787
			ABC ATP-binding/permease (ABN73946.1)	0.636
			DppD (ABN73175.1)	0.609
			DNA binding protein (ABN73888.1)	0.6
12.	*Haemophilus paraphrohaemolyticus*	SapA (EIG25812.1)	SapB (EIG25812.1)	0.999
			SapC (EIG25859.1)	0.993
			SapD (EIG25896.1)	0.983
			ABC permease (EIG28059.1)	0.95
			Oligo/dipeptide transporter (EIG27349.1)	0.921
			SapF (EIG26673.1)	0.904
			DppC (EIG28055.1)	0.853
			ABC substrate binding protein (EIG25911.1)	0.843
			HbpA (EIG24526.1)	0.841

			ABC substrate binding protein (EIG27353.1)	0.838
13.	*Haemophilus parahaemolyticus*	SapD (RDF05747.1)	SapC (RDF05746.1)	0.995
			SapB (RDF05745.1)	0.993
			SapA (RDF05744.1)	0.963
			DppB (RDF00348.1)	0.92
			DppC (RDF00349.1)	0.898
			ABC permease (RDF05853.1)	0.866
			SapF (RDF02657.1)	0.802
			ABC substrate binding protein (RDF05852.1)	0.717
			ABC substrate binding protein (RDF05230.1)	0.715
			ABC ATP-binding protein (RDF05855.1)	0.692
14.	*Actinobacillus minor*	SapA (EER46932.1)	SapB (EER46931.1)	0.999
			SapC (EER46930.1)	0.993
			SapD (EER46929.1)	0.985
			DppB (EER45938.1)	0.946
			SapF (EER46477.1)	0.91
			LktB (EER46566.1) based on homology	0.904
			Peptide ABC substrate binding protein (EER46780.1)	0.872
			DppC (EER45939.1)	0.862
			Heme-binding protein A (EER47336.1)	0.835
			ABC substrate-binding protein (EER45936.1)	0.831
15.			SapB (AHG84027.1)	0.998

	Bibersteinia trehalosi	SapA (AHG84026.1)	SapC (AHG84028.1) based on homology	0.987
			SapD (AHG84029.1) based on homology	0.954
			SapF (AHG84468.1) based on homology	0.889
			DppB (AHG84601.1)	0.884
			DppC (AHG84602.1)	0.85
			TyrR (AHG84025.1) based on homology	0.845
			DppD (AHG84603.1) based on homology	0.741
			Peptide ABC substrate-binding protein (AHG84283.1)	0.72
			LktB (AHG84281.1) based on homology	0.7
16.	*Haemophilus ducreyi*	SapA (AAP96068.1)	SapB (AAP96069.1)	0.997
			SapC (AAP96070.1)	0.994
			SapD (AAP96071.1)	0.974
			SapF (AAP95751.1)	0.849
			DppB (AAP95290.1)	0.847
			TyrR (AAP96067.1)	0.845
			DppC (AAP95291.1)	0.814
			ABC permease/ATP binding protein (AAP96443.1)	0.779
			DppD (AAP95292.1)	0.697
			PolA (AAP96072.1)	0.687
17.	*Providencia rustigianii*	SapC (EFB73883.1)	SapA (EFB73881.1) based on homology	0.998

			SapB (SUC35377.1) based on homology	0.998
			SapD (EFB73884.1) based on homology	0.996
			SapF (EFB73885.1) based on homology	0.995
			DppB (EFB72901.1)	0.837
			DppA (EFB72902.1)	0.832
			ABC ATP binding protein (EFB72899.1)	0.816
			ABC ATP binding protein (EFB72898.1)	0.808
			ABC ATP binding protein (EFB73545.1)	0.795
			OppB (EFB73543.1)	0.785
18.	*Halolamina pelagica*	SapA (KPN31530.1)	DppB (KPN30552.1)	0.941
			GsiD (KPN31528.1)	0.903
			GsiC (KPN31556.1)	0.883
			GsiB (KPN30551.1)	0.855
			NikB (KPN31529.1)	0.791
			Dipeptide transporter (KPN31557.1)	0.787
			GsiD (KPN31558.1)	0.787
			MalK (KPN29090.1)	0.784
			NikB (KPN31613.1)	0.72
			Uncharacterized protein (KPN31606.1)	0.654
19.	*Rheinheimera nanhaiensis* E407-8	SapD (GAB58545.1)	DppA (GAB58542.1)	0.954
			DppB (GAB58543.1)	0.887
			SapC (GAB58544.1)	0.779
			SapF (GAB58546.1)	0.64

			Adk (GAB59355.1)	0.556
			NadD (GAB60389.1)	0.514
			Fibronectin type III domain protein (GAB59522.1)	0.458
			FhuC (GAB59022.1)	0.456
			Transcription initiation factor subunit 5 (GAB59997.1)	0.454
			HmuV (GAB58711.1)	0.427
20.	*Idiomarina loihiensis* (strain ATCC BAA-735 / DSM 15497 / L2-TR)	SapD (AAV81577.1)	SapC (AAV81576.1)	0.997
			DppB (AAV81575.1)	0.986
			DppA (AAV81574.1)	0.905
			SapF (AAV81578.1)	0.886
			PutP (AAV81579.1)	0.77
			PspF (AAV81573.1)	0.641
			Glutathione hydrolase proenzyme (AAV81580.1)	0.532
			HmuV (AAV80958.1)	0.524
			PspB (AAV81571.1)	0.498
			CysS (AAV81873.1)	0.479
21.	*Ruminiclostridium hungatei (Clostridium hungatei)*	SapF (OPX44924.1)	NikA (OPX44923.1)	0.922
			GsiC (OPX44927.1)	0.824
			GsiD (OPX44926.1)	0.808
			DdpD (OPX44925.1)	0.726
			OppD_1 (OPX45761.1)	0.723
			OppD_2 (OPX42617.1)	0.723
			DppB (OPX42619.1)	0.697
			OppC_2 (OPX42618.1)	0.672
			OppB (OPX45759.1)	0.653
			OppC_1 (OPX45760.1)	0.641

22.	*Erwinia billingiae*	SapA (CAX59840.1)	SapB (CAX59841.1)	0.997
			Oligopeptide ABC permease (CAX59872.1)	0.947
			DppB (CAX61939.1)	0.819
			DppC (CAX61938.1)	0.795
			Peptide ABC permease (CAX58463.1)	0.727
			Peptide ABC permease (CAX58287.1)	0.714
			DppB (CAX61880.1)	0.699
			Oligopeptide ABC ATPase protein (CAX59873.1)	0.631
			Cation transport regulator (CAX59842.1)	0.62
			YddQ (CAX58462.1)	0.61
23.	*Salinivirga cyanobacteriivorans*	SapA (ALO15827.1)	DppB (ALO13713.1)	0.742
			GsiD (ALO15346.1)	0.666
			GsiA (ALO16023.1)	0.621
			MntB (ALO14477.1)	0.538
			SMC (ALO14195.1)	0.507
			OpuAB (ALO14152.1)	0.479
			LuxQ (ALO15350.1)	0.47
			LTD domain-containing protein (ALO14753.1)	0.463
			MlaE (ALO16011.1)	0.454
			GbuA (ALO14151.1)	0.447
24.	*Trinickia symbiotica*	SapA (PMS38197.1)	DppC (PMS37713.1)	0.814
			DppB (PMS37712.1) based on homology	0.792

			ABC ATP binding protein (PMS37715.1)	0.629
			ABC ATP binding protein (PMS34802.1)	0.629
			ABC ATP binding protein (PMS34802.1)	0.629
			ABC permease (PMS34805.1)	0.614
			ABC permease (PMS38187.1)	0.6
			ABC permease (PMS36216.1)	0.595
			ABC substrate binding-protein (PMS35036.1)	0.592
			ABC permease (PMS34804.1)	0.577
25.	*Paraglaciecola polaris* LMG 21857	SapA (GAC33352.1)	SapC (GAC33353.1)	0.997
			SapD (GAC33354.1)	0.919
			SapF (GAC33355.1)	0.896
			LptF (GAC33869.1)	0.662
			CcmB (GAC33458.1)	0.652
			LptG (GAC33868.1)	0.649
			Eif2A (GAC31253.1)	0.643
			Agmatine deiminase (GAC32403.1)	0.641
			FbpC (GAC31472.1)	0.632
			PPE repeat containing protein (GAC31053.1)	0.63
26.	*Hamiltonella defensa subsp. Acyrthosiphon pisum* (strain 5AT)	SapD (ACQ68668.1)	SapB (ACQ68670.1)	0.994
			SapC (ACQ68669.1)	0.985
			SapF (ACQ68667.1)	0.885
			FabI (ACQ68666.1)	0.612

			dTDP-glucose 4,6-dehydratase (ACQ66958.1)	0.43
27.	*Streptococcus pseudopneumoniae* (strain IS7493)	SapF (AEL10001.1)	Amino acid ABC substrate-binding protein/Permease (AEL10002.1)	0.996
			Amino acid ABC permease (AEL10274.1)	0.849
			Amino acid ABC permease (AEL10898.1)	0.839
			Amino acid ABC permease (AEL10083.1)	0.832
			Amino acid ABC substrate-binding protein/Permease (AEL10083.1)	0.832
			Amino acid ABC permease (AEL09858.1)	0.83
			Amino acid ABC permease (AEL10409.1)	0.829
			Amino acid ABC permease (AEL09859.1)	0.822
			MetN (AEL09724.1)	0.819
			Lipoprotein (AEL09722.1)	0.802
28.	*Aggregatibacter actinomycetemcomitans*	SapA (EHK90108.1)	SapB (EHK90107.1)	0.997
			SapC (EHK90106.1) based on homology	0.993
			SapD (EHK90105.1)	0.986
			Oligopeptide-binding protein (EHK90133.1)	0.818

			Nickel-binding protein (EHK91585.1)	0.759
			OppC (EHK89465.1)	0.718
			OppB (EHK89466.1)	0.716
			ABC permease/ATP binding protein (EHK91205.1)	0.7
			ABC permease (EHK91206.1)	0.7
			Oligopeptide-binding protein (EHK89467.1)	0.682

Appendix B

Supplementary data to Chapter 4

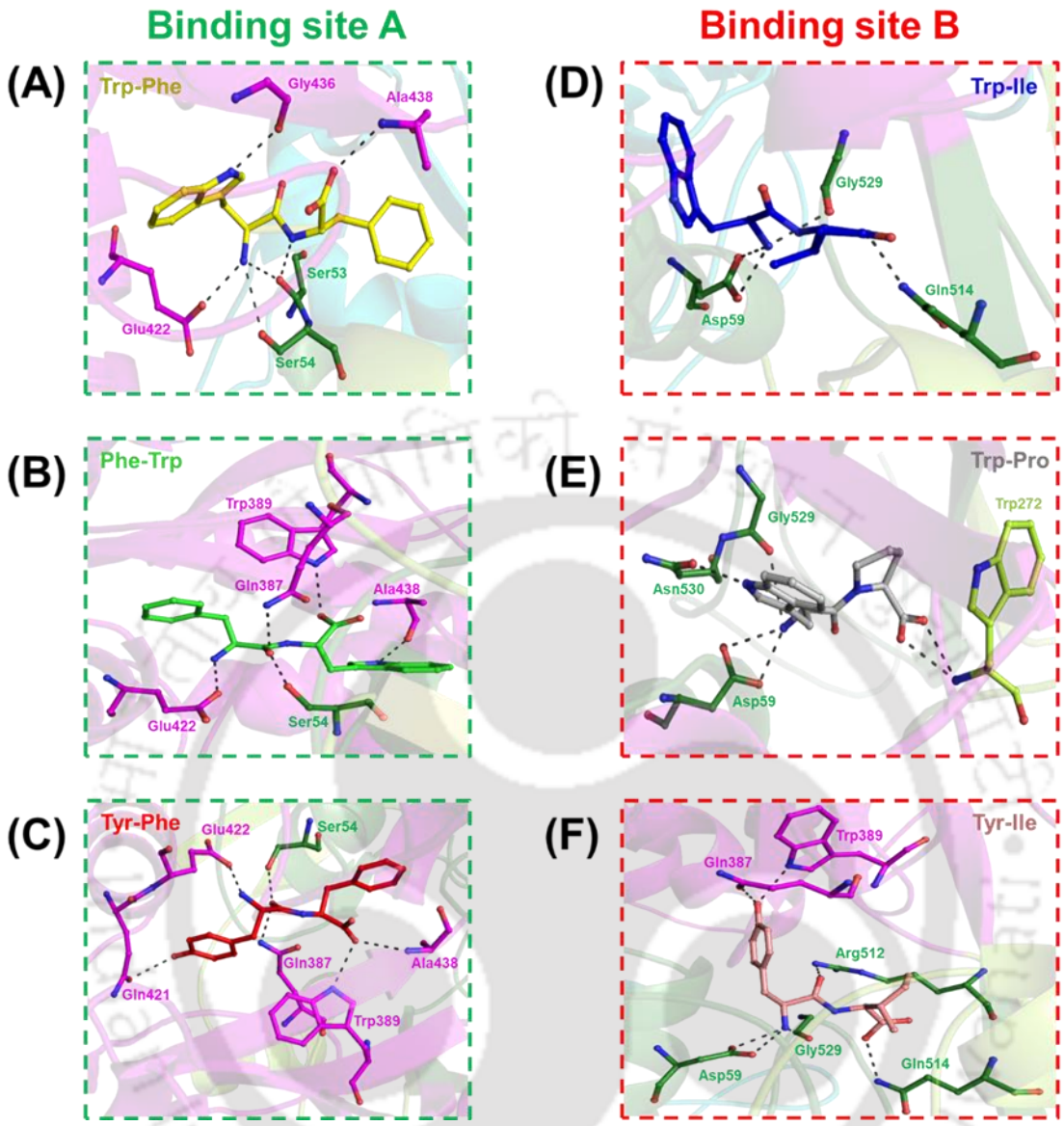


Figure B1. *Ec*SapA dipeptide docking studies. Docked poses of (A) Trp-Phe (yellow ball and stick), (B) Phe-Trp (green ball and stick), and (C) Tyr-Phe (red ball and stick) with *Ec*SapA bound in the binding site A. The docked poses of (D) Trp-Ile (blue ball and stick), (E) Trp-Pro (grey ball and stick), and (F) Tyr-Ile (pink ball and stick). The interacting residues of *Ec*SapA have been represented with a ball-and-stick model and colored based on their distribution in domains viz., Domain I (hinge I (dark green) and hinge II (light green)), Domain II (cyan), and Domain III (magenta).

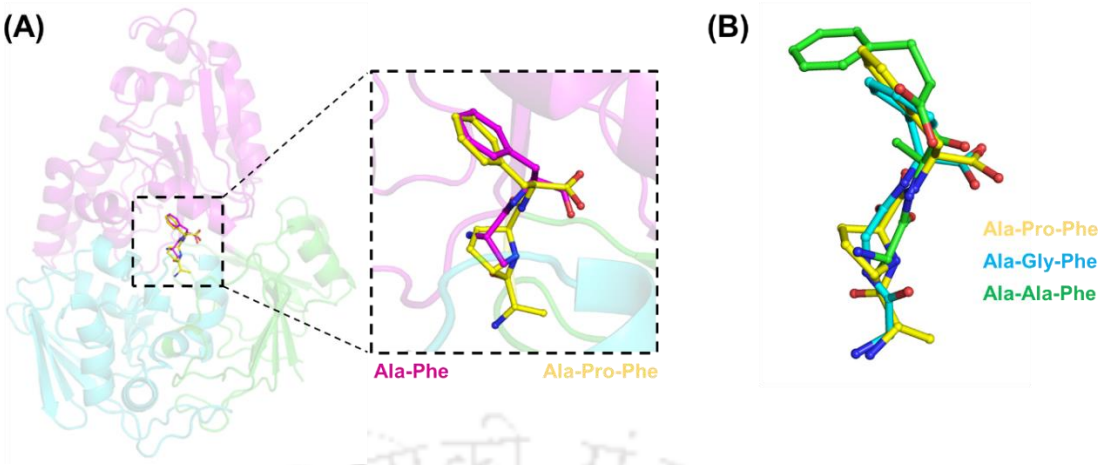


Figure B2. *Ps*DppA tripeptide-binding propensity. (A) Docked pose of tripeptide Ala-Pro-Phe (yellow ball-and-stick) with *Ps*DppA. The zoomed view of the binding-site core of *Ps*DppA consisting of Ala-Phe (magenta ball-and-stick) overlayed with Ala-Pro-Phe has been marked by dashed boxes. (B) Ala-Gly-Phe (cyan ball-and-stick) and Ala-Ala-Phe (green ball-and-stick) dock with a similar orientation to that of Ala-Pro-Phe (yellow ball-and-stick) within the *Ps*DppA binding site.

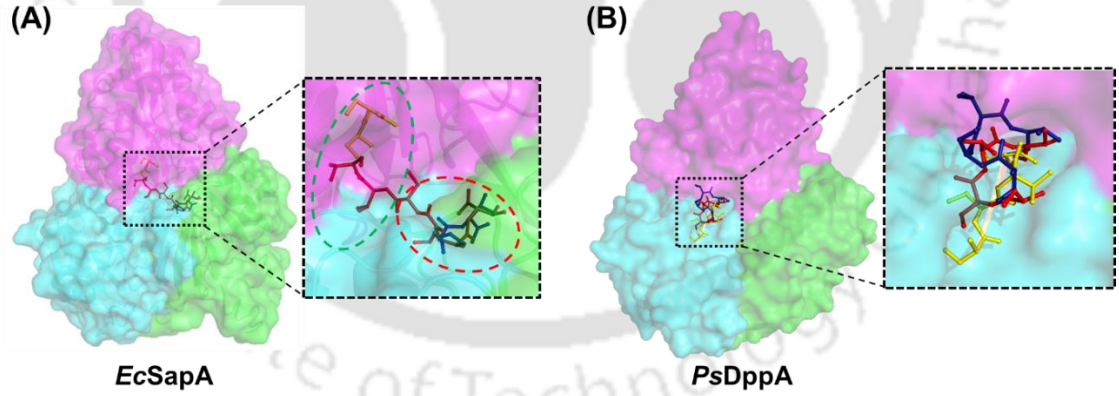


Figure B3. *Ec*SapA tetrapeptide docking studies. (A) The surface view of *Ec*SapA was observed to dock Ser-Ser-Val-Thr (red ball-and-stick model) at the interphase of binding site A (green dashed circle) and binding site B (red dashed circle). Further, Ser-Ser (yellow ball-and-stick model) docked in binding-site A, and Ser-Val (blue ball-and-stick model) and Val-Thr (brown ball-and-stick model) docked in binding site B. (B) In the surface view of *Ps*DppA, Ser-Ser-Val-Thr was observed to dock at the surface for all three top conformations (blue ball-and-stick model, conformation 1; red ball-and-stick model, conformation 2; yellow ball-

and-stick model, conformation 3). In both cases, the active sites were zoomed into and represented as dashed boxes.

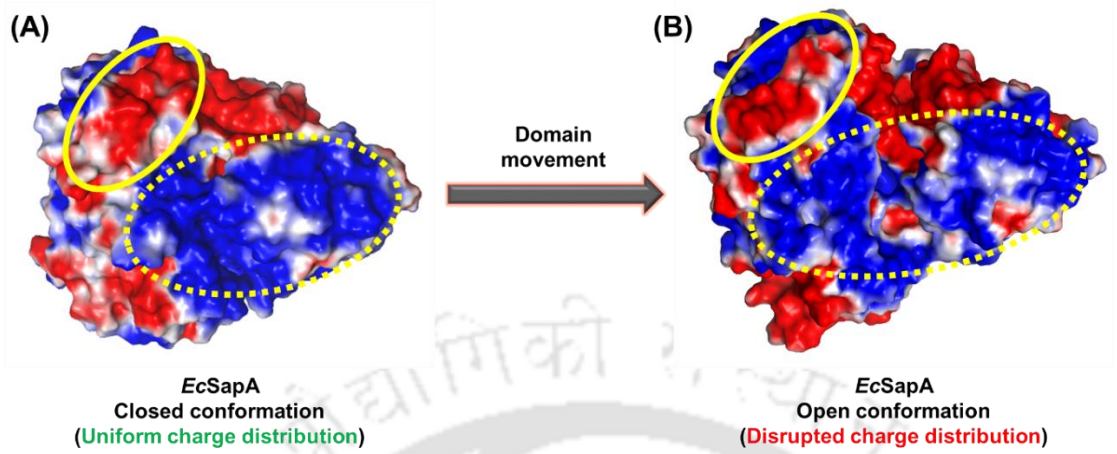


Figure B4. Domain movement of *Ec*SapA leads to disruption of uniform charge distribution. Electrostatic potential charge distribution of *Ec*SapA (A) open and (B) closed conformation. The matching circular patterns represent the region of contact for the complex formation. The dashed and solid yellow oval shape represent the site of the interaction of *Ec*SapA and *Ec*SapBC. The positively-charged and negatively-charged regions have been demarcated by blue surface (+1 kT/e) and red surface (-1 kT/e), respectively.

Table B1. Details of molecular docking results of various ligands bound with *Ps*DppA, *Ec*SapA, and *Hi*SapA. EFBE: estimated free energy of binding, NHBOND: number of hydrogen bonds.

Protein	Ligand	EFBE (kcal mol^{-1})	NHBOND	Interacting residues		
				Ligand		Protein atom
				Residue	atom	
*Ps*DppA	Ala-Phe	-10.26	9	Ala	N	Asp436 $O^{\delta1}$, Ala434 O
					O	Ala434 N
				Phe	N	Thr48 O
					O	Gly50 N, Arg383 $N^{\eta2}$
					OXT	Arg383 N^{ε}, Arg383 $N^{\eta2}$, Tyr385O^{η}
	Gly-Phe	-9.76	8	Gly	N	Asp436 $O^{\delta1}$, Ala434 O
					O	Ala434 N
				Phe	N	Thr48 O
					O	Gly50 N, Arg383 $N^{\eta2}$
					OXT	Arg383 N^{ε}, Tyr385O^{η}
	Ala-Glu	-9.74	12	Ala	N	Asp436 $O^{\delta1}$, Ala434 O
					O	Ala434 N
				Glu	N	Thr48 O
					O	Arg383 N^{ε}, Tyr385O^{η}
					OXT	Gly50 N and Arg383 $N^{\eta2}$
					$O^{\varepsilon2}$	Thr48 O, Thr48 $O^{\gamma1}$, Thr49 N, Thr49 $O^{\gamma1}$
	Met-Leu	-9.61	8	Met	N	Asp436 $O^{\delta1}$, Ala434 O
					O	Ala434 N
				Leu	N	Thr48 O
					O	Gly50 N, Arg383 $N^{\eta2}$
					OXT	Arg383 N^{ε}, Tyr385O^{η}
	Val-Thr	-9.32	8	Val	N	Asp436 $O^{\delta1}$, Ala434 O

					O	Ala434 N
				Thr	N	Thr48 O
					O	Gly50 N, Arg383 $N^{\eta 2}$
					OXT	Arg383 N^{ε}, Tyr385O^{η}
	Ala-Gln	-9.09	8	Ala	N	Asp436 $O^{\delta 1}$, Ala434 O
					O	Ala434 N
				Gln	N	Thr48 O
					O	Gly50 N
					OXT	Arg383 N^{ε}, Tyr385 O^{η}
					$N^{\varepsilon 2}$	Gly432 O
	Gly-Glu	-8.9	11	Gly	N	Asp436 $O^{\delta 1}$, Ala434 O
					O	Ala434 N
				Glu	N	Thr48 O
					O	Gly50 N, Arg383 $N^{\eta 2}$
					OXT	Tyr385 O^{η}
					$O^{\varepsilon 1}$	Thr48 O, Thr48 $O^{\gamma 1}$, Thr49 N, Thr49 $O^{\gamma 1}$
	Gly-Leu	-8.71	8	Gly	N	Asp436 $O^{\delta 1}$, Ala434 O
					O	Ala434 N
				Leu	N	Thr48 O
					O	Gly50 N, Arg383 $N^{\eta 2}$
					OXT	Arg383 N^{ε}, Tyr385 O^{η}
	Ala-Pro-Phe	-9.91	7	Ala	N	Asp181 $O^{\delta 2}$, Asp436 $O^{\delta 1}$
				Pro	O	Ala434 N
				Phe	N	Thr48 O
					O	Arg383 N^{ε}, Tyr385 O^{η}
					OXT	Gly50 N
	Ala-Gly-Phe	-9.5	9	Ala	N	Asp181 $O^{\delta 2}$, Asp436 $O^{\delta 1}$
					O	Gly50 N
				Gly	N	Asp436 $O^{\delta 1}$

					O	Ala434 N
				Phe	N	Thr48 O
					O	Arg383 N^{ε}, Tyr385 O^{η}
					OXT	Gly50 N
	Ala-Ala-Phe	-8.9	9	Ala	N	Thr48 O, Asp436 $O^{\delta1}$
				Ala	N	Thr48 O
				Phe	O	Thr48 $O^{\gamma1}$, Lys457 N^{ζ}
					OXT	Thr48 O, Thr48 $O^{\gamma1}$, Thr49 N, Thr49 $O^{\gamma1}$
*Ec*SapA	Trp-Phe	-8.59	6	Trp	N	Ser53 O, Ser54 O^{γ}, Glu422 $O^{\varepsilon1}$
					$N^{\varepsilon1}$	Gly436 O
				Phe	N	Ser53 O
					OXT	Ala438 N
	Trp-Pro	-8.54	6	Trp	N	Asp59 $O^{\delta1}$, Asp59 $O^{\delta2}$, Gly529 O
					$N^{\varepsilon1}$	Asn530 $O^{\delta1}$
				Pro	O, OXT	Trp272 N
	Phe-Trp	-8.41	5	Phe	N	Glu422 $O^{\varepsilon1}$
					O	Ser54 O^{γ}, Gln387 $N^{\varepsilon2}$
				Trp	O	Trp389 $N^{\varepsilon1}$
					$N^{\varepsilon1}$	Ala438 O
	Tyr-Phe	-8.26	6	Tyr	O	Ser54 O^{γ}, Gln387 $N^{\varepsilon2}$
					N	Glu422 $O^{\varepsilon1}$
					O^{η}	Gln421 $O^{\varepsilon1}$
				Phe	OXT	Ala438 N, Trp389 $N^{\varepsilon1}$
	Trp-Ile	-8.2	4	Trp	N	Asp59 $O^{\delta1}$, Asp59 $O^{\delta2}$, Gly529 O
				Ile	O	Gln514 $N^{\varepsilon2}$
	Tyr-Ile	-8.01	7	Tyr	O	Arg512 $N^{\eta2}$
					N	Asp59 $O^{\delta1}$, Asp59 $O^{\delta2}$, Gly529 O

					O^{η}	Gln387 $O^{\varepsilon 1}$, Trp389 $N^{\varepsilon 1}$
				Ile	O	Gln514 $N^{\varepsilon 2}$
	Ser-Ser-Val-Thr	-4.95	9	Ser	O	Arg512 $N^{\eta 2}$
					O^{γ}	Asn530 $O^{\delta 1}$, Asp59 $O^{\delta 1}$
				Ser	O^{γ}	Gln387 $O^{\varepsilon 1}$, Trp389 $N^{\varepsilon 1}$
				Val	O	Ala438 N
				Thr	O	Ser54 O^{γ}, Gln387 $O^{\varepsilon 1}$
					$O^{\gamma 1}$	Ser53 O
	Heme B	-8.35	8	Not applicable	Fe	Thr384 O
					NA	Ser41 O
					O1A	Asn45 $O^{\delta 1}$, Gln43 O, Arg385 $N^{\eta 1}$
					O2A	Arg385 $N^{\eta 1}$
					O1D	Gln43 N
					O2D	Gln43 $N^{\varepsilon 2}$
*Hi*SapA	Heme B	-10.98	4	Not applicable	O2A	Arg438 $N^{\eta 1}$
					O1D	Arg534 $N^{\eta 1}$
					O2D	Arg534 $N^{\eta 1}$, Arg534 $N^{\eta 2}$

Table B2. Probable dipeptide ligands of *Ps*DppA apart from the eight cognate ligands of *Ps*DppA. EFBE: estimated free energy of binding.

Rank	Dipeptide	EFBE (kcal mol^{-1})
1	Pro-Phe	-10.05
2	Ile-Ile	-9.73
3	Leu-Ile	-9.68
4	Phe-Ala	-9.68
5	Ala-Ile	-9.66
6	Val-Leu	-9.64
7	Leu-Val	-9.54
8	Val-Val	-9.51
9	Val-Ile	-9.49
10	Ile-Val	-9.48
11	Phe-Pro	-9.37
12	Val-Asp	-9.3
13	Val-Ser	-9.27
14	Leu-Leu	-9.05
15	Leu-Pro	-9.01
16	Ser-Ile	-9.01
17	Leu-Ala	-8.96
18	Pro-Leu	-8.96
19	Pro-Ile	-8.92
20	Cys-Val	-8.89
21	Ile-Ala	-8.83
22	Ile-Pro	-8.77
23	Ala-Leu	-8.74
24	Ala-Val	-8.72
25	Val-Ala	-8.71

Table B3. Degree of Domain III movement of tertiary structure models of selected SapA proteins.

Protein	Degree of movement (°)
*Hi*SapA	32.2
*Kp*SapA	33.0
*Af*SapA	40.4
*Dd*SapA	48.8
*Pm*SapA	42.2
*St*SapA	42.3
*Ec*SapA	33.2

Table B4. Details of protein-protein docking results of *Ec*SapA with protamine and protegrin-1. EFBE: estimated free energy of binding, NHBOND: number of hydrogen bonds.

Ligand	EFBE (kcal mol^{-1})	NHBOND	Interacting residues		
			Ligand		Protein atom
			Residue no.	atom	
Protamine	-1009.1	36	Ser2	O	Ala388 N
				O^{γ}	Ser386 O, Gln387 $N^{\varepsilon 2}$
			Arg3	N	Gln387 $O^{\varepsilon 1}$
				$N^{\eta 1}$	Asp59 $O^{\delta 1}$, Gly529 O
				$N^{\eta 2}$	Asp59 $O^{\delta 1}$, Asp59 $O^{\delta 2}$
			Arg4	N	Gln387 $O^{\varepsilon 1}$

				$N^{\eta1}$	Glu422 $O^{\varepsilon1}$, Glu422 $O^{\varepsilon2}$
				$N^{\eta2}$	Glu422 $O^{\varepsilon1}$
			Arg5	N^{ε}	Asp440 $O^{\delta1}$, Asp440 $O^{\delta2}$
				$N^{\eta1}$	Asp440 $O^{\delta1}$
				$N^{\eta2}$	His186 $N^{\varepsilon2}$
			Arg6	N	Asp440 $O^{\delta1}$
				O	Trp437 $N^{\varepsilon1}$
				$N^{\eta1}$	Thr461 $O^{\gamma1}$
				$N^{\eta2}$	Glu422 $O^{\varepsilon2}$
			Arg7	N^{ε}	Ser441 O^{γ}
				$N^{\eta1}$	Tyr147 O^{η}, Asp440 O, Ser441 O^{γ}
				$N^{\eta2}$	Tyr147 O^{η}
			Arg8	N^{ε}	Ser459 O^{γ}
				$N^{\eta1}$	Gln152 $O^{\varepsilon1}$
				$N^{\eta2}$	Gln152 $O^{\varepsilon1}$
			Gly10	N	Asp176 $O^{\delta1}$
			Arg11	$N^{\eta1}$	Tyr96 O^{η}
				$N^{\eta2}$	Tyr96 O^{η}
			Arg12	$N^{\eta1}$	Gln174 $O^{\varepsilon1}$
				$N^{\eta2}$	Gln174 $O^{\varepsilon1}$
			Arg13	$N^{\eta1}$	Asp91 $O^{\delta2}$
				$N^{\eta2}$	Asp91 O, Asp91 $O^{\delta2}$
Protegrin-1	-1121.3	10	Arg4	$N^{\eta1}$	Asp59 $O^{\delta1}$
			Leu5	N	Val37 O
				O	Val37 N, Gln387 $N^{\varepsilon2}$
			Tyr7	O^{η}	Arg385 O
			Arg9	$N^{\eta1}$	Ser435 O^{γ}

			Arg10	N	Glu422 $O^{\varepsilon 1}$
				$N^{\eta 1}$	Ser459 O
			Arg11	$N^{\eta 2}$	Asp440 $O^{\delta 1}$
			Cys15	N	Gln387 $O^{\varepsilon 1}$
				O	Ala388 N

Appendix C

Supplementary data to Chapter 5

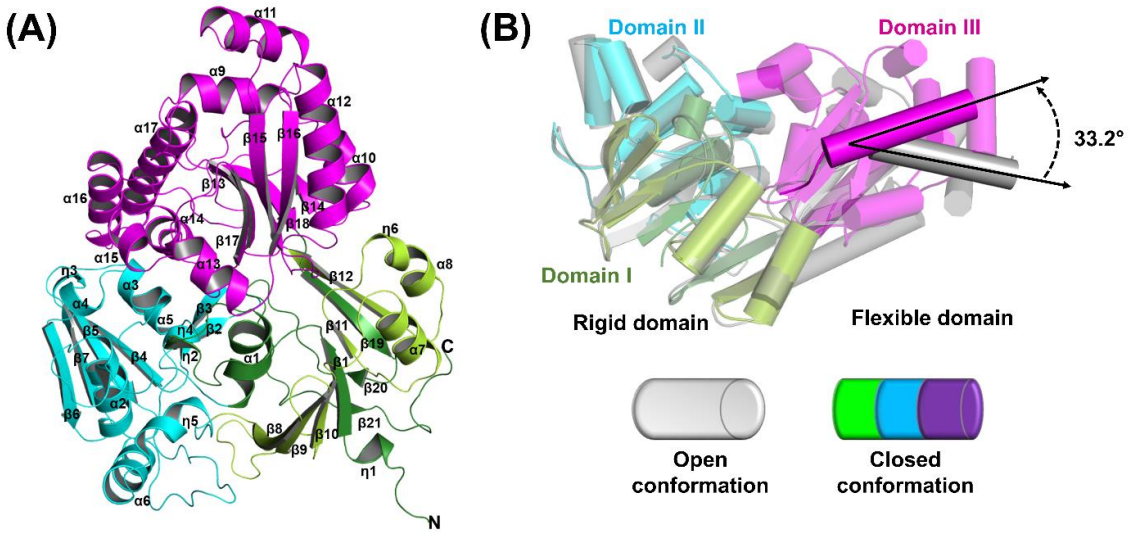


Figure C1. Tertiary model and domain movement of *Ec*SapA. (A) A cartoon representation of the predicted tertiary structure of *Ec*SapA (closed conformation). The structure is divided into three distinct domains: Domain I (green), Domain II (cyan), and Domain III (magenta). Further, Domain I is subdivided into hinges I (dark green) and II (light green). The secondary structural elements of *Ec*SapA have been marked. (B) The domain movement of *Ec*SapA was mapped by superposing the open (grey) and closed conformations. A significant movement in domain III (magenta) of the closed conformation is observable in the case of *Ec*SapA. The domains I and II of *Ec*SapA were color-coded in green and cyan. The angle of movement of domain III observed in *Ec*SapA has been denoted by a dashed arrow.

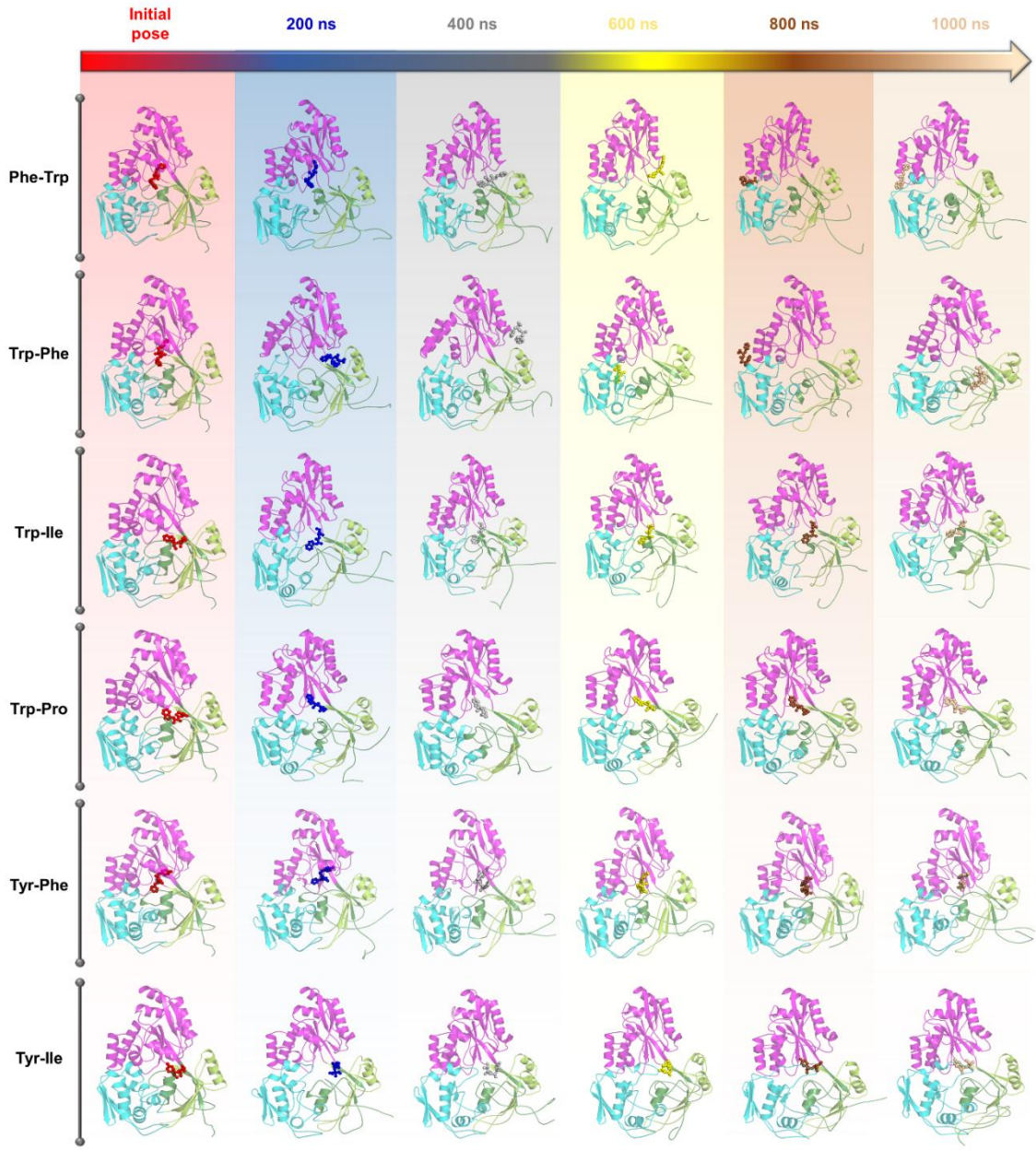


Figure C2. Frames of *Ec*SapA-dipeptide complexes extracted from the MD simulation trajectory. The various frames extracted from the MD simulation trajectories across the 1000 ns time period. Various frames extracted are as follows: initial equilibrated structure (red), 200 ns (blue), 400 ns (grey), 600 ns (yellow), 800 ns (brown), and 1000 ns (wheat) time points. In the extreme left panel, the dipeptides bound to *Ec*SapA across the entire trajectory of the MD simulations have been mentioned. The dipeptides are represented in the form of a ball-and-stick model, while *Ec*SapA is represented in the form of cartoons. *Ec*SapA has been color-coded on the basis of its distinct domains, which

comprise domain I, which includes hinge I (dark green) and hinge II (light green), alongside domain II (cyan) and domain III (magenta).

Table C1. Details of the number of hydrogen bonds throughout the MD simulations trajectory of the *Ec*SapA-ligand complexes. H-bond: Hydrogen bond.

S. No.	Protein	Ligand	No. of H-bond in equilibrated structure	Average No. of H-bond across the trajectory
1.	*Ec*SapA	Phe-Trp	4	1.4
2.		Trp-Phe	4	1.3
3.		Tyr-Phe	7	2.1
4.		Heme	4	1.9
5.		Tyr-Ile	3	3.2
6.		Trp-Pro	3	2.7
7.		Trp-Ile	5	2.7
8.		Protegrin-1	6	9.5
9.		Protamine	18	10.4

Appendix D

Supplementary data to Chapter 6

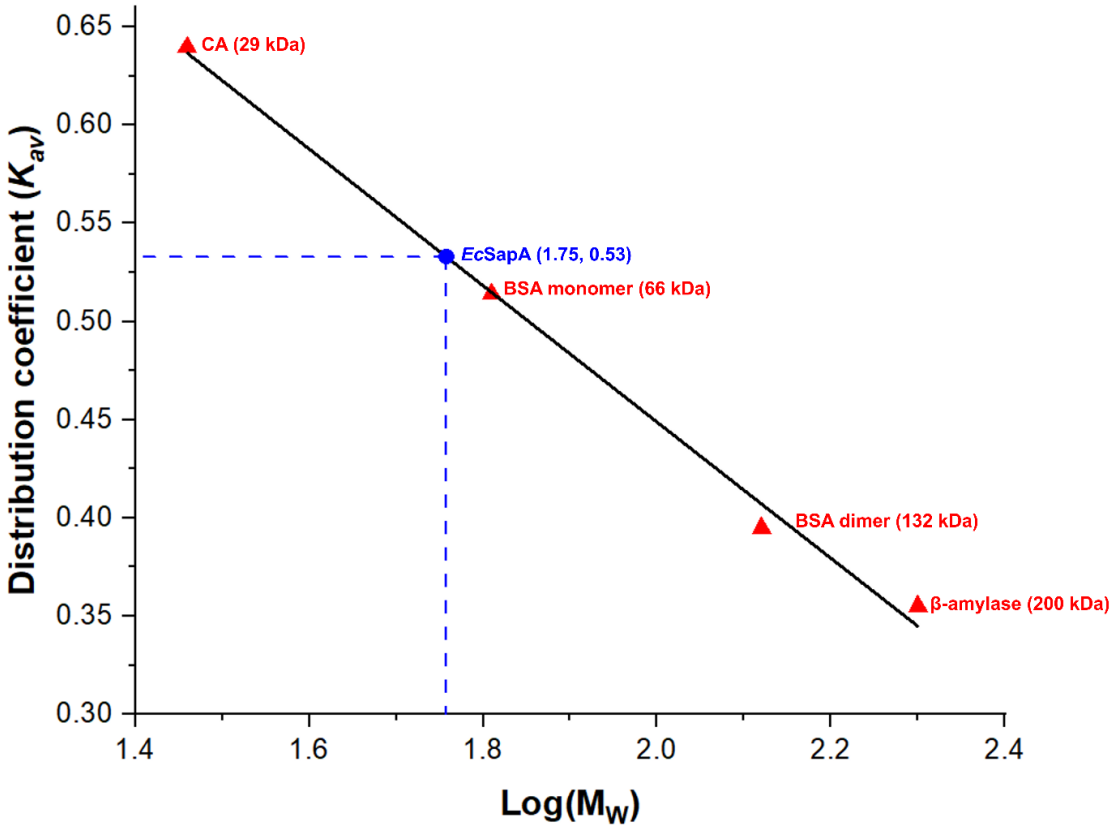


Figure D1. FPLC-based size estimation of *Ec*SapA. A standard graph was plotted in the form of K_{av} versus Log(M_w) of standard proteins (red triangle). The standard proteins used are carbonic anhydrase (CA), bovine serum albumin (BSA), and β-amylase. The molecular weight of *Ec*SapA was estimated from the standard graph by plotting its estimated K_{av} value (blue circle). The protein names and their reported molecular weight (in parenthesis) have been mentioned adjacent to each triangle. The estimated K_{av} and Log M_w values (broken droplines) of *Ec*SapA have been mentioned within the parenthesis adjacent to the circle.

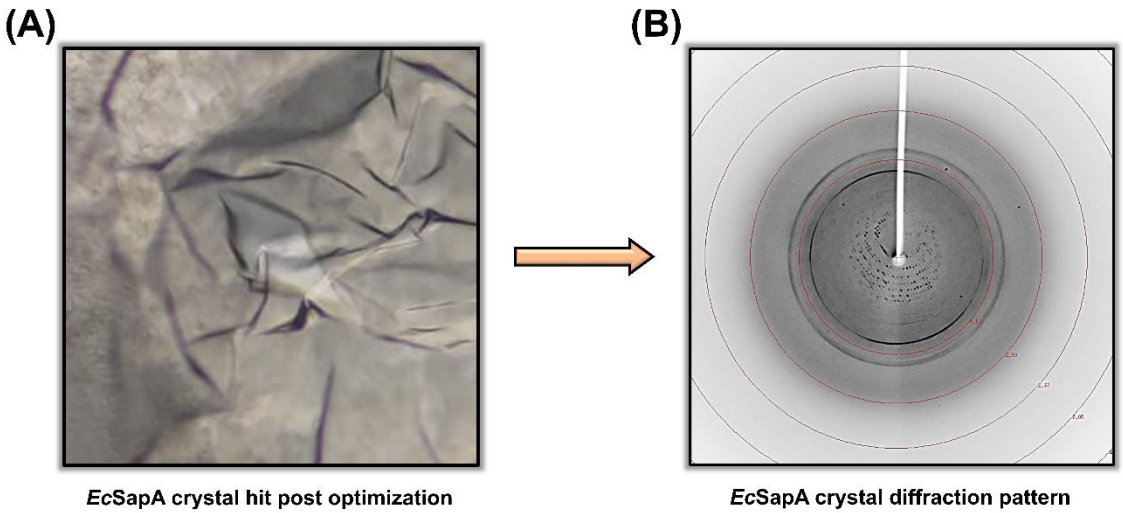


Figure D2. *Ec*SapA crystal diffraction pattern. (A) Crystal of *Ec*SapA obtained after multiple rounds of optimizations. (B) X-ray diffraction pattern of *Ec*SapA crystal.

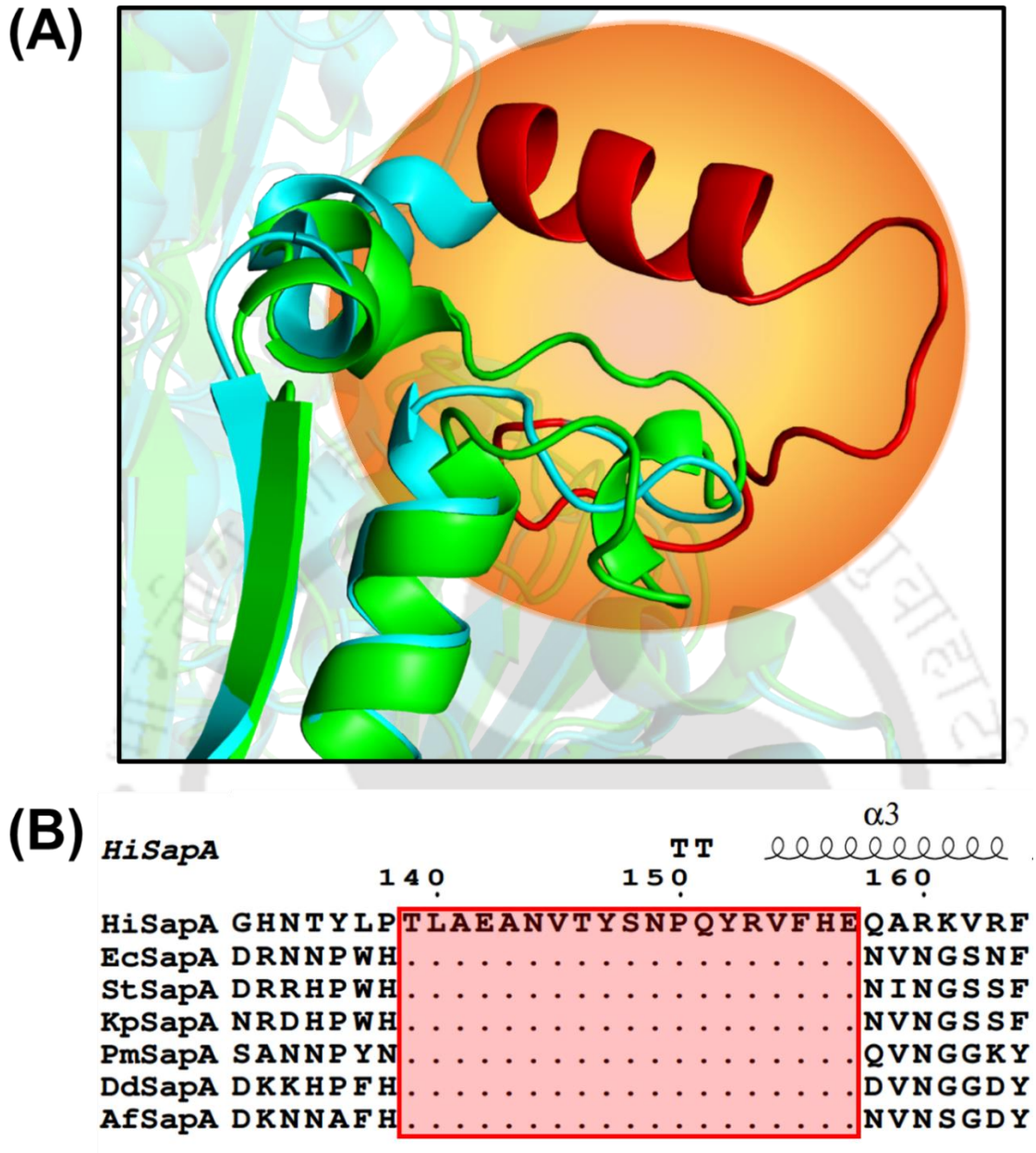


Figure D3. Structural variation between *Hi*SapA and *Ec*SapA. (A) Cartoon representation of a superimposed tertiary model of *Ec*SapA (green) with the crystal structure of *Hi*SapA (cyan). The additional loop and helix present in *Hi*SapA have been represented in red cartoon. (B) Multiple sequence alignment of selected SapA proteins. The protein sequences utilized for the MSA are from *H. influenzae* (*Hi*SapA, P45285), *E. coli* (*Ec*SapA, Q47622), *S. typhimurium* (*St*SapA, P36634), *Klebsiella pneumoniae* (*Kp*SapA, W8UWE0), *Proteus mirabilis* (*Pm*SapA, B4EWK8), *Dickeya dadantii* (*Dd*SapA, E0SFZ5), and *Aliivibrio fischeri* (*Af*SapA, Q5E0R7). The additional region present in *Hi*SapA has been shaded red.

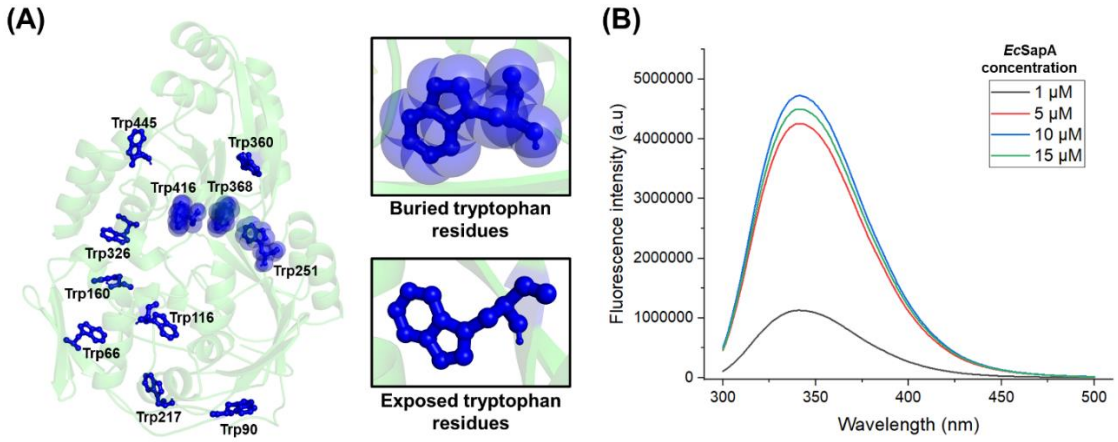


Figure D4. Intrinsic fluorescence properties of *Ec*SapA. (A) Cartoon representation of the predicted tertiary structure model of *Ec*SapA (green). The tryptophan residues of *Ec*SapA have been represented in the form of a blue ball-and-stick model. The buried residues are shaded with translucent circles, as represented in the box (inset). (B) The intrinsic fluorescence spectra of varying concentrations (1µM to 15 µM) of *Ec*SapA. All the fluorescence spectra of the varying concentrations have been represented in different colors mentioned within the box (inset).

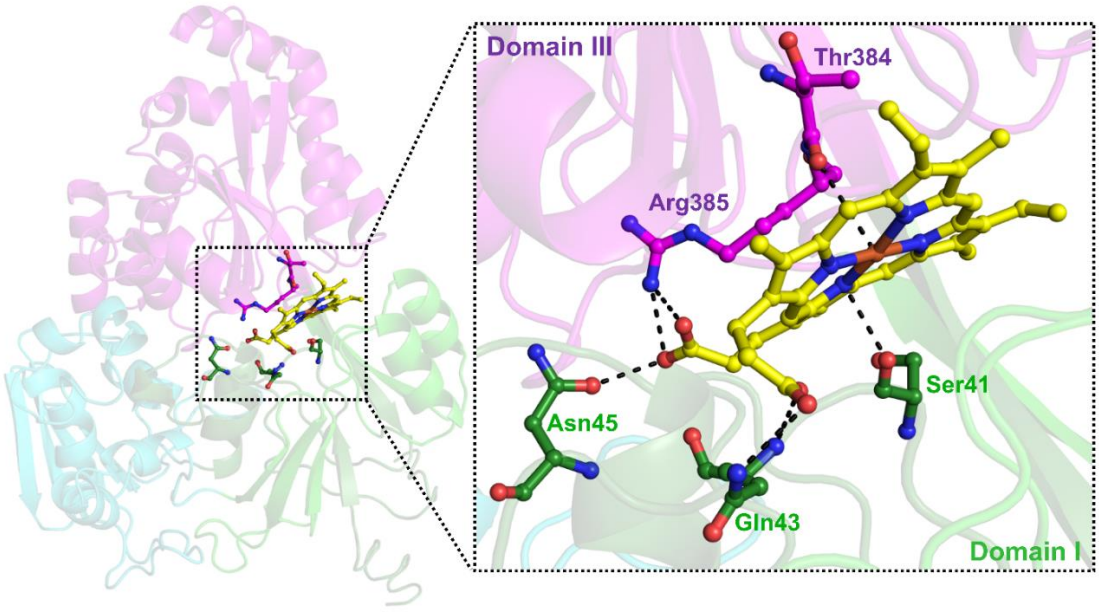


Figure D5. **Heme-docked pose of *Ec*SapA.** Cartoon-based representation of the hemin (yellow ball and stick) docked pose in *Ec*SapA. The dotted line represents the zoomed view of the residues interacting with the docked heme molecule. The interacting residues of *Ec*SapA have been represented in a ball-and-stick model and colored based on their distribution in domains viz., Domain I (green) and Domain III (magenta).

Bibliography

1. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, and Lindahl E (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1**, 19-25.
2. Albright LM, Huala E, and Ausubel FM (1989). Prokaryotic signal transduction mediated by sensor and regulator protein pairs. *Annu. Rev. Genet.*, **23**, 311-336.
3. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, and Nielsen H (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.*, **37(4)**, 420-423.
4. Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol.*, **215(3)**, 403-410.
5. Ambudkar SV, Kim IW, Xia D, and Sauna ZE (2006). The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding. *FEBS Lett.*, **580(4)**, 1049-1055.
6. Arnold FM, Weber MS, Gonda I, Gallenito MJ, Adenau S, Egloff P, Zimmermann I, Hutter CA, Hürlimann LM, Peters EE, and Piel J (2020). The ABC exporter IrtAB imports and reduces mycobacterial siderophores. *Nature*, **580(7803)**, 413-417.
7. Assaf R, Xia F, and Stevens R (2021). Detecting operons in bacterial genomes via visual representation learning. *Sci. Rep.*, **11(1)**, 2124.
8. Baker NA, Sept D, Holst MJ, and McCammon JA (2001). The adaptive multilevel finite element solution of the Poisson-Boltzmann equation on massively parallel computers. *IBM J. Res. Dev.*, **45(3.4)**, 427-438.
9. Belas R, Manos J, and Suvanasuthi R (2004). *Proteus mirabilis* ZapA metalloprotease degrades a broad spectrum of substrates, including antimicrobial peptides. *Infect. Immun.*, **72**, 5159-5167.
10. Benfield AH and Henriques ST (2020). Mode-of-action of antimicrobial peptides: membrane disruption vs. intracellular mechanisms. *Front. Med. Tech.*, **2**, 610997.

11. Bengoechea JA and Skurnik M (2000). Temperature-regulated efflux pump/potassium antiporter system mediates resistance to cationic antimicrobial peptides in *Yersinia*. *Mol. Microbiol.*, **37**, 67-80.
12. Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, and Westbrook J (2000). The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.*, **7(11)**, 957-959.
13. Berntsson RPA, Doeven MK, Fusetti F, Duurkens RH, Sengupta D, Marrink SJ, Thunnissen AMW, Poolman B, and Slotboom DJ (2009). The structural basis for peptide selection by the transport receptor OppA. *EMBO J.*, **28(9)**, 1332-1340.
14. Berntsson RPA, Smits SH, Schmitt L, Slotboom DJ, and Poolman B (2010). A structural classification of substrate-binding proteins. *FEBS Lett.*, **584(12)**, 2606-2617.
15. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, and Schwede T (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, **42(W1)**, W252-W258.
16. Bina XR, Provenzano D, and Nguyen N (2008). *Vibrio cholerae* RND family efflux systems are required for antimicrobial resistance, optimal virulence factor production, and colonization of the infant mouse small intestine. *Infection*, **76**, 3595–3605.
17. Blair JM, Richmond GE, and Piddock LJ (2014). Multidrug efflux pumps in Gram-negative bacteria and their role in antibiotic resistance. *Future Microbiol.*, **9(10)**,1165-1177.
18. Blair JM, Zeth K, Bavro VN, and Sancho-Vaello E (2022). The role of bacterial transport systems in the removal of host antimicrobial peptides in Gram-negative bacteria. *FEMS Microbiol. Rev.*, **46(6)**, 32.
19. Bolintineanu DS, Vivcharuk V, and Kaznessis YN (2012). Multiscale models of the antimicrobial peptide protegrin-1 on gram-negative bacteria membranes. *Int. J. Mol. Sci.*, **13(9)**, 11000-11011.
20. Bougie I, Charpentier S, and Bisaillon M (2003). Characterization of the metal ion binding properties of the hepatitis C virus RNA polymerase. *J. Biol. Chem.*, **278(6)**, 3868-3875.

21. Brandis G (2021). Reconstructing the evolutionary history of a highly conserved operon cluster in Gammaproteobacteria and Bacilli. *Genome Biol. Evol.*, **13(4)**, evab041.
22. Brogden KA (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.*, **3(3)**, 238-250.
23. Bussi G, Donadio D, and Parrinello M (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys*., **126(1)**, 014101.
24. Butler MS, Blaskovich MA, and Cooper MA (2013). Antibiotics in the clinical pipeline in 2013. *J. Antibiot. (Tokyo)*, **66(10)**, 571-591.
25. Cabiaux V, Agerberths B, Johansson J, Homblé F, Goormaghtigh E, and Ruysschaert JM (1994). Secondary structure and membrane interaction of PR-39, a Pro+ Arg-rich antibacterial peptide. *Eur. J. Biochem.*, **224(3)**, 1019-1027.
26. Campos MA, Vargas MA, Regueiro V, Llompart CM, Albertí S, and Bengoechea JA (2004). Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. *Infect. Immun.*, **72(12)**, 7107-7114.
27. Capone R, Mustata M, Jang H, Arce FT, Nussinov R, and Lal R (2010). Antimicrobial protegrin-1 forms ion channels: molecular dynamic simulation, atomic force microscopy, and electrical conductance studies. *Biophys J.*, **98(11)**, 2644-2652.
28. Casteels P, Ampe C, Riviere L, Van Damme J, Elicone C, Fleming M, Jacobs F, and Tempst P (1990). Isolation and characterization of abaecin, a major antibacterial response peptide in the honeybee (*Apis mellifera*). *Eur. J. Biochem.*, **187(2)**, 381-386.
29. Chai HH, Ham JS, Kim TH, and Lim D (2022). Identifying ligand-binding specificity of the oligopeptide receptor OppA from *Bifidobacterium longum* KACC91563 by structure-based molecular modeling. *Arab. J. Chem.*, **15(11)**, 104198.
30. Chakraborty K, Ghosh S, Koley H, Mukhopadhyay AK, Ramamurthy T, Saha DR, Mukhopadhyay D, Roychowdhury S, Hamabata T, Takeda Y, and Das S (2008). Bacterial exotoxins downregulate cathelicidin (hCAP-18/LL-37) and human β-defensin 1 (HBD-1) expression in the intestinal epithelial cells. *Cell. Microbiol.*, **10(12)**, 2520-2537.

31. Chan C, Burrows LL, and Deber CM (2004). Helix induction in antimicrobial peptides by alginate in biofilms. *J. Biol. Chem.*, 279, 38749–38754.
32. Chandravanshi M, Gogoi P, and Kanaujia SP (2020a). Structural and thermodynamic correlation illuminates the selective transport mechanism of disaccharide α-glycosides through ABC transporter. *FEBS J.*, **287(8)**, 1576-1597.
33. Chandravanshi M, Samanta R, and Kanaujia SP (2020b). Conformational trapping of a β-glucosides-binding protein unveils the selective two-step ligand-binding mechanism of ABC importers. *J. Mol. Biol.*, **432(20)**, 5711-5734.
34. Chandravanshi M, Sharma A, Dasgupta P, Mandal SK, and Kanaujia SP (2019). Identification and characterization of ABC transporters for carbohydrate uptake in *Thermus thermophilus* HB8. *Gene*, **696**, 135-148.
35. Chandravanshi M, Tripathi SK, and Kanaujia SP (2021). An updated classification and mechanistic insights into ligand binding of the substrate-binding proteins. *FEBS Lett.*, **595(18)**, 2395-2409.
36. Chaplin DD (2010). Overview of the immune response. *J. Allergy Clin. Immunol.*, **125(2)**, 3-23.
37. Chautrand T, Souak D, Chevalier S, and Duclairoir-Poc C (2022). Gram-negative bacterial envelope homeostasis under oxidative and nitrosative stress. *Microorganisms*, **10(5)**, 924.
38. Chen HY, Weng SF, and Lin JW (2000). Identification and analysis of the *sap* genes from *Vibrio fischeri* belonging to the ATP-binding cassette gene family required for peptide transport and resistance to antimicrobial peptides. *Biochem. Biophys. Res. Commun.*, **269(3)**, 743-748.
39. Chen J, Lu G, Lin J, Davidson AL, and Quiocho FA (2003). A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol. Cell*, **12(3)**, 651-661.
40. Chen YC, Chuang YC, Chang CC, Jeang CL, and Chang MC (2004). A K^+ uptake protein, TrkA, is required for serum, protamine, and polymyxin B resistance in *Vibrio vulnificus*. *Infect. Immun.*, **72(2)**, 629-636.

41. Cheng JT, Hale JD, Elliot M, Hancock RE, and Straus SK (2009). Effect of membrane composition on antimicrobial peptides aurein 2.2 and 2.3 from Australian southern bell frogs. *Biophys. J.*, **96(2)**, 552-565.
42. Chitsaz M, Booth L, Blyth MT, O'Mara ML, and Brown MH (2019). Multidrug resistance in Neisseria gonorrhoeae: identification of functionally important residues in the MtrD efflux protein. *MBio*, **10(6)**, 2277.
43. Ciulla MG and Gelain F (2023). Structure–activity relationships of antibacterial peptides. *Microb. Biotechnol.*, **16(4)**, 757-777.
44. Conibear AC, Rosengren KJ, Harvey PJ, and Craik DJ (2012). Structural characterization of the cyclic cystine ladder motif of θ-defensins. *Biochemistry*, **51(48)**, 9718-9726.
45. Crawford A and Wilson D (2015). Essential metals at the host–pathogen interface: nutritional immunity and micronutrient assimilation by human fungal pathogens. *FEMS Yeast Res.*, **15(7)**, 071.
46. Crow A, Greene NP, Kaplan E, and Koronakis V (2017). Structure and mechanotransmission mechanism of the MacB ABC transporter superfamily. *Proc. Natl. Acad. Sci. U. S. A.*, **114(47)**, 12572-12577.
47. Darden T, York D, and Pedersen L (1993). Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.,* **98(12)**, 10089-10092.
48. Darwin AJ (2013). Stress relief during host infection: the phage shock protein response supports bacterial virulence in various ways. *PLoS Pathog.*, **9(7)**, e1003388.
49. Dawson RJ and Locher KP (2006). Structure of a bacterial multidrug ABC transporter. *Nature*, **443**, 180–185.
50. de Boer M, Gouridis G, Vietrov R, Begg SL, Schuurman-Wolters GK, Husada F, Eleftheriadis N, Poolman B, McDevitt CA, and Cordes T (2019). Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers. *Elife*, **8**, 44652.
51. de Leeuw E, Li C, Zeng P, Li C, Diepeveen-de Buin M, Lu WY, Breukink E, and Lu W (2010). Functional interaction of human neutrophil peptide-1 with the cell wall precursor lipid II. *FEBS Lett.*, **584(8)**, 1543-1548.

52. De Santis M, Mantovani A, and Selmi C (2020). The other side of the innate immune system: humoral arms favoring cancer. *Cell. Mol. Immunol.*, **17(10)**, 1024-1025.
53. Desta IT, Porter KA, Xia B, Kozakov D, and Vajda S (2020). Performance and its limits in rigid body protein-protein docking. *Structure*, **28(9)**, 1071-1081.
54. Devnani B, Ong L, Kentish S, and Gras S (2020). Heat induced denaturation, aggregation and gelation of almond proteins in skim and full fat almond milk. *Food Chem.*, **325**, 126901.
55. Dhople V, Krukemeyer A, and Ramamoorthy A (2006). The human beta-defensin-3, an antibacterial peptide with multiple biological functions. *Biochim. Biophys. Acta Biomembr.*, **1758(9)**, 1499-1512.
56. Doeven MK, Van den Bogaart G, Krasnikov V, and Poolman B (2008). Probing receptor-translocator interactions in the oligopeptide ABC transporter by fluorescence correlation spectroscopy. *Biophysical J.*, **94(10)**, 3956-3965.
57. Dong N, Wang C, Zhang T, Zhang L, Xue C, Feng X, Bi C, and Shan A (2019). Bioactivity and bactericidal mechanism of histidine-rich β-hairpin peptide against Gram-negative bacteria. *Int. J. Mol. Sci.*, **20(16)**, 3954.
58. Duarte-Mata DI and Salinas-Carmona MC (2023). Antimicrobial peptides immune modulation role in intracellular bacterial infection. *Front. Immunol.*, **14**, 1119574.
59. Dupont A, Heinbockel L, Brandenburg K, and Hornef MW (2014). Antimicrobial peptides and the enteric mucus layer act in concert to protect the intestinal mucosa. *Gut Microbes*, **5(6)**, 761-765.
60. Ekiert DC, Coudray N, and Bhabha G (2022). Structure and mechanism of the bacterial lipid ABC transporter, MlaFEDB. *Curr. Opin. Struct. Biol.*, **76**, 102429.
61. Elbourne LD, Tetu SG, Hassan KA, and Paulsen IT (2017). TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.*, **45(D1)**, 320-324.
62. Emsley P, Lohkamp B, Scott WG, and Cowtan K (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66(4)**, 486-501.

63. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, and Pedersen LG (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.,* **103(19)**, 8577-8593.
64. Eswarappa SM, Panguluri KK, Hensel M, and Chakravortty D (2008). The *yejABEF* operon of *Salmonella* confers resistance to antimicrobial peptides and contributes to its virulence. *Microbiology*, **154(2)**, 666-678.
65. Farber DL, Yudanin NA, and Restifo NP (2014). Human memory T cells: generation, compartmentalization and homeostasis. *Nat. Rev. Immunol.*, **14(1)**, 24-35.
66. Feng SH, Zhang WX, Yang J, Yang Y, and Shen HB (2020). Topology prediction improvement of α-helical transmembrane proteins through helix-tail modeling and multiscale deep learning fusion. *J. Mol. Biol.*, **432(4)**, 1279-1296.
67. Fernando DM, Gee CT, Griffith EC, Meyer CJ, Wilt LA, Tangallapally R, Wallace MJ, Miller DJ, and Lee RE (2022). Biophysical analysis of the *Mycobacteria tuberculosis* peptide binding protein DppA reveals a stringent peptide binding pocket. *Tuberculosis*, **132**, 102157.
68. Fleming PJ and Fleming KG (2018). HullRad: Fast calculations of folded and disordered protein and nucleic acid hydrodynamic properties. *Biophys. J.*, **114(4)**, 856-869.
69. Flores-Kim J and Darwin AJ (2016). The phage shock protein response. *Annu. Rev. Microbiol.*, **70**, 83-101.
70. Fondi M, Emiliani G, and Fani R (2009). Origin and evolution of operons and metabolic pathways. *Res. Microbiol.*, **160(7)**, 502-512.
71. Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, and Olson AJ (2016). Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.*, **11(5)**, 905-919.
72. Foschiatti M, Cescutti P, Tossi A, and Rizzo R (2009). Inhibition of cathelicidin activity by bacterial exopolysaccharides. *Mol. Microbial.*, **72(5)**, 1137-1146.
73. Fukami-Kobayashi K, Tateno Y, and Nishikawa K (1999). Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J. Mol. Biol.,* **286(1)**, 279-290.

74. Fukamizo T, Kitaoku Y, and Suginta W (2019). Periplasmic solute-binding proteins: Structure classification and chitooligosaccharide recognition. *Int. J. Biol. Macromol.*, **128**, 985-993.
75. Galvan EM, Lasaro MA, and Schifferli DM (2008). Capsular antigen fraction 1 and Pla modulate the susceptibility of *Yersinia pestis* to pulmonary antimicrobial peptides such as cathelicidin. *Infect. Immun.*, **76**, 1456-1464.
76. Gao K, Oerlemans R, and Groves MR (2020). Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys. Rev.*, **12(1)**, 85-104.
77. Garcia-Vallvé S, Palau J, and Romeu A (1999). Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli* and *Bacillus subtilis*. *Mol. Biol. Evol.*, **16(9)**, 1125-1134.
78. Gasteiger J and Marsili M (1980). Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*, **36(22)**, 3219-3228.
79. Getahun YA, Ali DA, Taye BW, and Alemayehu YA (2022). Multidrug-Resistant Microbial Therapy Using Antimicrobial Peptides and the CRISPR/Cas9 System. *Vet Med (Auckl)*, **11**, 173-190.
80. Gouet P, Robert X, and Courcelle E (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.*, **31(13)**, 3320-3323.
81. Graf M, Mardirossian M, Nguyen F, Seefeldt AC, Guichard G, Scocchi M, Innis CA, and Wilson DN (2017). Proline-rich antimicrobial peptides targeting protein synthesis. *Nat. Prod. Rep.*, **34(7)**, 702-711.
82. Greco I, Molchanova N, Holmedal E, Jenssen H, Hummel BD, Watts JL, Håkansson J, Hansen PR, and Svenson J (2020). Correlation between hemolytic activity, cytotoxicity and systemic in vivo toxicity of synthetic antimicrobial peptides. *Sci. Rep.*, **10(1)**, 13206.
83. Groisman EA, Parra-Lopez C, Salcedo M, Lipps CJ, and Heffron F (1992). Resistance to host antimicrobial peptides is necessary for Salmonella virulence. *Proc. Natl. Acad. Sci. U. S. A.*, **89(24)**, 11939-11943.
84. Gruenheid S and Le Moual H (2012). Resistance to antimicrobial peptides in Gram-negative bacteria. *FEMS Microbiol. Lett.*, **330(2)**, 81-89.

85. Guina T, Yi EC, Wang H, Hackett M, and Miller SI (2000). A PhoP-regulated outer membrane protease of *Salmonella enterica* serovar Typhimurium promotes resistance to alphahelical antimicrobial peptides. *J. Bacteriol.*, **182**, 4077-4086.
86. Gunn JS (2008). The Salmonella PmrAB regulon: lipopolysaccharide modifications, antimicrobial peptide resistance and more. *Trends Microbiol.*, **16,** 284-290.
87. Guo L, Lim KB, Poduje CM, Daniel M, Gunn JS, Hackett M, and Miller SI (1998). Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell*, **95**, 189-198.
88. Gutner M, Chaushu S, Balter D, and Bachrach G (2009). Saliva enables the antimicrobial activity of LL-37 in the presence of proteases of *Porphyromonas gingivalis*. *Infect. Immun.,* **77**, 5558–5563.
89. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, and Hutchison GR (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics*, **4(1)**, 1-17.
90. Harms C, Domoto Y, Celik C, Rahe E, Stumpe S, Schmid R, Nakamura T, and Bakker EP (2001). Identification of the ABC protein SapD as the subunit that confers ATP dependence to the K^+-uptake systems TrkH and TrkG from *Escherichia coli* K-12. *Microbiology*, **147(11)**, 2991-3003.
91. He K, Ludtke SJ, Worcester DL, and Huang HW (1996). Neutron scattering in the plane of membranes: structure of alamethicin pores. *Biophys. J.*, **70(6)**, 2659-2666.
92. He SW, Zhang J, Li NQ, Zhou S, Yue B, and Zhang M (2017). A TFPI-1 peptide that induces degradation of bacterial nucleic acids, and inhibits bacterial and viral infection in half-smooth tongue sole, *Cynoglossus semilaevis*. *Fish Shellf. Immunol.*, **60**, 466-473.
93. Hess B (2008). P-LINCS: Aparallel linear constraint solver for molecular simulation. *J. Chem. Theory Comp*., **4(1)**, 116-122.
94. Higgins CF (1992). ABC transporters: from microorganisms to man. *Annu. Rev. Cell. Biol.*, **8(1)**, 67-113.
95. Higgins CF, Haag PD, Nikaido K, Ardeshir F, Garcia G, and Ames GFL (1982). Complete nucleotide sequence and identification of membrane

components of the histidine transport operon of *S. typhimurium*. *Nature*, **298(5876)**, 723-727.

96. Hollenstein K, Frei DC, and Locher KP (2007). Structure of an ABC transporter in complex with its binding protein. *Nature*, **446(7132)**, 213-216.
97. Hollingsworth SA and Dror RO (2018). Molecular dynamics simulation for all. *Neuron*, **99(6)**, 1129-1143.
98. Hollmann A, Martinez M, Maturana P, Semorile LC, and Maffia PC (2018). Antimicrobial peptides: interaction with model and biological membranes and synergism with chemical antibiotics. *Front. Chem.*, **6**, 204.
99. Honeycutt JD, Wenner N, Li Y, Brewer SM, Massis LM, Brubaker SW, Chairatana P, Owen SV, Canals R, Hinton JCD, and Monack DM (2020). Genetic variation in the MacAB-TolC efflux pump influences pathogenesis of invasive Salmonella isolates from Africa. *PLoS Pathog.*, **16(8)**, 1008763.
100. Hsu CR, Chang IW, Hsieh PF, Lin TL, Liu PY, Huang CH, Li KT, and Wang JT (2019). A novel role for the *Klebsiella pneumoniae* Sap (Sensitivity to Antimicrobial Peptides) transporter in intestinal cell interactions, innate immune responses, liver abscess, and virulence. *J. Infect. Dis.*, **219(8)**, 1294-1306.
101. Hsu STD, Breukink E, Tischenko E, Lutters MA, De Kruijff B, Kaptein R, Bonvin AM, and Van Nuland NA (2004). The nisin–lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics. *Nat. Struct. Mol. Biol.*, **11(10)**, 963-967.
102. Huan Y, Kong Q, Mou H, and Yi H (2020). Antimicrobial peptides: classification, design, application and research progress in multiple fields. *Front. Microbiol.*, **11**, 2559.
103. Huang J and MacKerell Jr AD (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.*, **34(25)**, 2135-2145.
104. Huang YH, Colgrave ML, Daly NL, Keleshian A, Martinac B, and Craik DJ (2009). The biological activity of the prototypic cyclotide kalata b1 is modulated by the formation of multimeric pores. *J. Biol. Chem.*, **284(31)**, 20699-20707.

105. Imura Y, Nishida M, Ogawa Y, Takakura Y, and Matsuzaki K (2007). Action mechanism of tachyplesin I and effects of PEGylation. *Biochim. Biophys. Acta Biomembr.*, **1768(5)**, 1160-1169.
106. Inoue T, Moran I, Shinnakasu R, Phan TG, and Kurosaki T (2018). Generation of memory B cells and their reactivation. *Immunol. Rev.*, **283(1)**, 138-149.
107. Islam D, Bandholtz L, Nilsson J, Wigzell H, Christensson B, Agerberth B, and Gudmundsson G (2001). Downregulation of bactericidal peptides in enteric infections: a novel immune escape mechanism with bacterial DNA as a potential regulator. *Nat. Med.*, **7**, 180-185.
108. Janes RW (2017). PDB2CD visualises dynamics within protein structures. *Eur. Biophys. J.*, **46(7)**, 607-616.
109. Jean-François F, Elezgaray J, Berson P, Vacher P, and Dufourc EJ (2008). Pore formation induced by an antimicrobial peptide: electrostatic effects. *Biophys. J.*, **95(12)**, 5748-5756.
110. Joo HS, Fu CI, and Otto M (2016). Bacterial strategies of resistance to antimicrobial peptides. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **371(1695)**, 20150292.
111. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, and Klein ML (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79(2)**, 926-935.
112. Joskova M, Mokry J, and Franova S (2020). Respiratory cilia as a therapeutic target of phosphodiesterase inhibitors. *Front. Pharmacol.*, **11**, 609.
113. Kai-Larsen Y, Lüthje P, Chromek M, Peters V, Wang X, Holm Å, Kádas L, Hedlund KO, Johansson J, Chapman MR, and Jacobson SH (2010). Uropathogenic *Escherichia coli* modulates immune responses and its curli fimbriae interact with the antimicrobial peptide LL-37. *PLoS Pathog.*, **6(7)**, 1001010.
114. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, and Xu J (2012). Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.*, **7(8)**, 1511-1522.
115. Karpowich NK, Huang HH, Smith PC, and Hunt JF (2003). Crystal structures of the BtuF periplasmic-binding protein for vitamin B12 suggest

a functionally important reduction in protein mobility upon ligand binding. *J. Biol. Chem.*, **278(10)**, 8429-8434.

116. Katiki M, Neetu N, Pratap S, and Kumar P (2022). Biochemical and structural basis for *Moraxella catarrhalis* enoyl-acyl carrier protein reductase (FabI) inhibition by triclosan and estradiol. *Biochim.*, **198**, 8-22.
117. Kawulka KE, Sprules T, Diaper CM, Whittal RM, McKay RT, Mercier P, Zuber P, and Vederas JC (2004). Structure of subtilosin A, a cyclic antimicrobial peptide from Bacillus subtilis with unusual sulfur to α-carbon cross-links: formation and reduction of α-thio-α-amino acid derivatives. *Biochemistry*, **43(12)**, 3385-3395.
118. Kelley LA, Mezulis S, Yates CM, Wass MN, and Sternberg MJ (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10(6)**, 845-858.
119. Killick J, Morisse G, Sieger D, and Astier AL (2018). Complement as a regulator of adaptive immunity. *Semin. Immunopathol.*, **40**, 37-48.
120. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, and Zaslavsky L (2023). PubChem 2023 update. *Nucleic Acids Res.*, **51(D1)**, 1373-1380.
121. Ko J, Park H, Heo L, and Seok C (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.*, **40(W1)**, W294-W297.
122. Koehbach J (2017). Structure-activity relationships of insect defensins. *Front. Chem.*, **5**, 45.
123. Koehbach J and Craik DJ (2019). The vast structural diversity of antimicrobial peptides. *Trends Pharmacol. Sci.*, **40(7)**, 517-528.
124. Kokryakov VN, Harwig SS, Panyutich EA, Shevchenko AA, Aleshina GM, Shamova OV, Korneva HA, and Lehrer RI (1993). Protegrins: leukocyte antimicrobial peptides that combine features of corticostatic defensins and tachyplesins. *FEBS Lett.*, **327(2)**, 231-236.
125. Kooi C and Sokol PA (2009). *Burkholderia cenocepacia* zinc metalloproteases influence resistance to antimicrobial peptides. *Microbiology*, **155**, 2818-2825.
126. Krizsan A, Prahl C, Goldbach T, Knappe D, and Hoffmann R (2015). Short proline-rich antimicrobial peptides inhibit either the bacterial 70S ribosome

or the assembly of its large 50S subunit. *Chembiochem*, **16(16)**, 2304-2308.

127. Kumar P, Kizhakkedathu JN, and Straus SK (2018). Antimicrobial peptides: diversity, mechanism of action and strategies to improve the activity and biocompatibility in vivo. *Biomolecules*, **8(1)**, 4.
128. Kumar S, Stecher G, Li M, Knyaz C, and Tamura K (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, **35(6)**, 1547.
129. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, and Thornton JM (2018). PDBsum: Structural summaries of PDB entries. *Protein Sci.*, **27(1)**, 129-134.
130. Le CF, Fang CM, and Sekaran SD (2017). Intracellular targeting mechanisms by antimicrobial peptides. *Antimicrob Agents Chemother.*, **61(4)**, 10-128.
131. Le Sage V, Zhu L, Lepage C, Portt A, Viau C, Daigle F, Gruenheid S, and Le Moual H (2009). An outer membrane protease of the omptin family prevents activation of the *Citrobacter rodentium* PhoPQ two-component system by antimicrobial peptides. *Mol. Microbiol.*, **74**, 98-111.
132. Lee YH, Deka RK, Norgard MV, Radolf JD, and Hasemann CA (1999). *Treponema pallidum* TroA is a periplasmic zinc-binding protein with a helical backbone. *Nat. Struct. Biol.*, **6**, 628–633.
133. Lei J, Sun L, Huang S, Zhu C, Li P, He J, Mackey V, Coy DH, and He Q (2019). The antimicrobial peptides and their potential clinical applications. *Am. J. Transl. Res.*, **11(7)**, 3919.
134. Létoffé S, Delepelaire P, and Wandersman C (2006). The housekeeping dipeptide permease is the Escherichia coli heme transporter and functions with two optional peptide binding proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **103(34)**, 12891-12896.
135. Letunic I and Bork P (2021). Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.*, **49(W1)**, 293-296.
136. Li CY, Chen XL, Qin QL, Wang P, Zhang WX, Xie BB, Su HN, Zhang XY, Zhou BC, and Zhang YZ (2015). Structural insights into the multispecific

recognition of dipeptides of deep-sea gram-negative bacterium *Pseudoalteromonas sp.* strain SM9913. *J. Bacteriol.*, **197(6)**, 1125-1134.

137. Liang H, Mao Y, Sun Y, and Gao H (2019). Transcriptional regulator ArcA mediates expression of oligopeptide transport systems both directly and indirectly in *Shewanella oneidensis*. *Sci. Rep.*, **9(1)**, 13839.
138. Linton KJ (2007). Structure and function of ABC transporters. *Physiology*, **22(2)**, 122-130.
139. Lipkin RB and Lazaridis T (2015). Implicit membrane investigation of the stability of antimicrobial peptide β-barrels and arcs. *J. Membr. Biol.*, **248**, 469-486.
140. Llanos P, Henriquez M, Minic J, Elmorjani K, Marion D, Riquelme G, Molgó J, and Benoit E (2006). Puroindoline-a and α1-purothionin form ion channels in giant liposomes but exert different toxic actions on murine cells. *FEBS J.*, **273(8)**, 1710-1722.
141. Lohner K and Prossnigg F (2009). Biological activity and structural aspects of PGLa interaction with membrane mimetic systems. *Biochim. Biophys. Acta Biomembr.*, **1788(8)**, 1656-1666.
142. Lomize AL, Todd SC, and Pogozheva ID (2022). Spatial arrangement of proteins in planar and curved membranes by PPM 3.0. *Protein Sci.*, **31(1)**, 209-220.
143. López-Solanilla E, García-Olmedo F, and Rodríguez-Palenzuela P (1998). Inactivation of the sapA to sapF locus of *Erwinia chrysanthemi* reveals common features in plant and animal bacterial pathogenesis. *Plant Cell,* **10**, 917–924.
144. Low WY, Thong S, and Chng SS (2021). ATP disrupts lipid-binding equilibrium to drive retrograde transport critical for bacterial outer membrane asymmetry. *Proc. Natl. Acad. Sci. U. S. A.*, **118(50)**, e2110055118.
145. Lukacik P, Owen CD, Harris G, Bolla JR, Picaud S, Alibay I, Nettleship JE, Bird LE, Owens RJ, Biggin PC, and Filippakopoulos P (2021). The structure of nontypeable *Haemophilus influenzae* SapA in a closed conformation reveals a constricted ligand-binding cavity and a novel RNA binding motif. *Plos one*, **16(10)**, 0256070.

146. Lyu Y, Fitriyanti M, and Narsimhan G (2019). Nucleation and growth of pores in 1, 2-Dimyristoyl-sn-glycero-3-phosphocholine (DMPC)/cholesterol bilayer by antimicrobial peptides melittin, its mutants and cecropin P1. *Colloids. Surf. B Biointerfaces*, **173**, 121-127.
147. Mabrouk DM (2022). Antimicrobial peptides: features, applications and the potential use against covid-19. *Mol. Biol. Rep.*, ***49*(10)**, 10039-10050.
148. Mackler BF and Kreil G (1977). Honey bee venom melittin. Correlation of nonspecific inflammatory activities with amino acid sequences. *Inflammation*, **2**, 55-65.
149. Maisetta G, Brancatisano FL, Esin S, Campa M, and Batoni G (2011). Gingipains produced by *Porphyromonas gingivalis* ATCC49417 degrade human-beta-defensin 3 and affect peptide's antibacterial activity *in vitro*. *Peptides*, **32**, 1073-1077.
150. Majewska M, Zamlynny V, Pieta IS, Nowakowski R, and Pieta P (2021). Interaction of LL-37 human cathelicidin peptide with a model microbial-like lipid membrane. *Bioelectrochemistry*, **141**, 107842.
151. Mak TW, Saunders ME, and Jett BD (2013). Introduction to the Immune Response. Primer to the immune response, Academic cell press, 3-20.
152. Mandal SK and Kanaujia SP (2022). Role of an orphan substrate-binding protein MhuP in transient heme transfer in *Mycobacterium tuberculosis*. *Int. J. Biol. Macromol.*, **211**, 342-356.
153. Mandal SK, Adhikari R, Sharma A, Chandravanshi M, Gogoi P, and Kanaujia SP (2019). Designating ligand specificities to metal uptake ABC transporters in *Thermus thermophilus* HB8. *Metallomics*, **11**, 597-612.
154. Mandard N, Bulet P, Caille A, Daffre S, and Vovelle F (2002). The solution structure of gomesin, an antimicrobial cysteine-rich peptide from the spider. *Eur. J. Biochem.*, **269(4)**, 1190-1198.
155. Mandard N, Sodano P, Labbe H, Bonmatin JM, Bulet P, Hetru C, Ptak M, and Vovelle F (1998). Solution structure of thanatin, a potent bactericidal and fungicidal insect peptide, determined from proton two-dimensional nuclear magnetic resonance data. *Eur. J. Biochem.*, **256(2)**, 404-410.
156. Mao B, Pear MR, McCammon JA, and Quiocho FA (1982). Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *J. Biol. Chem.*, **257(3)**, 1131-1133.

157. Maqbool A, Horler RS, Muller A, Wilkinson AJ, Wilson KS, and Thomas GH (2015). The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochem. Soc. Trans.*, **43(5)**, 1011-1017.
158. Mardirossian M, Grzela R, Giglione C, Meinnel T, Gennaro R, Mergaert P, and Scocchi M (2014). The host antimicrobial peptide Bac71-35 binds to bacterial ribosomal proteins and inhibits protein synthesis. *Chem. Biol.*, **21(12)**, 1639-1647.
159. Mardirossian M, Pérébaskine N, Benincasa M, Gambato S, Hofmann S, Huter P, Müller C, Hilpert K, Innis CA, Tossi A, and Wilson DN (2018). The dolphin proline-rich antimicrobial peptide Tur1A inhibits protein synthesis by targeting the bacterial ribosome. *Cell Chem. Biol.*, **25(5)**, 530-539.
160. Marshall JS, Warrington R, Watson W, and Kim HL (2018). An introduction to immunology and immunopathology. *Allergy Asthma Clin. Immunol.*, **14(2)**, 1-10.
161. Martins A, Spengler G, Rodrigues L, Viveiros M, Ramos J, Martins M, Couto I, Fanning S, Pagès JM, Bolla JM, and Molnar J (2009). pH modulation of efflux pump activity of multi-drug resistant *Escherichia coli*: protection during its passage and eventual colonization of the colon. *PloS One*, **4(8)**, 6656.
162. Mason KM, Bruggeman ME, Munson RS, and Bakaletz LO (2006). The non-typeable *Haemophilus influenzae* Sap transporter provides a mechanism of antimicrobial peptide resistance and SapD-dependent potassium acquisition. *Mol. Microbiol.*, **62(5)**, 1357-1372.
163. Mason KM, Munson Jr RS, and Bakaletz LO (2005). A mutation in the sap operon attenuates survival of nontypeable *Haemophilus influenzae* in a chinchilla model of otitis media. *Infect. Immun.*, **73(1)**, 599-608.
164. Mason KM, Raffel FK, Ray WC, and Bakaletz LO (2011). Heme utilization by nontypeable *Haemophilus influenzae* is essential and dependent on Sap transporter function. *J. Bacteriol.*, **193(10)**, 2527-2535.
165. Matsuzaki K, Murase O, Fujii N, and Miyajima K (1996). An antimicrobial peptide, magainin 2, induced rapid flip-flop of phospholipids coupled with pore formation and peptide translocation. *Biochemistry*, **35(35)**, 11361-11368.

166. Matsuzaki K, Sugishita KI, Ishibe N, Ueha M, Nakata S, Miyajima K, and Epand RM (1998). Relationship of membrane curvature to the formation of pores by magainin 2. *Biochemistry*, **37(34)**, 11856-11863.
167. Matthyssen T, Li W, Holden JA, Lenzo JC, Hadjigol S, and O'Brien-Simpson NM (2022). The potential of modified and multimeric antimicrobial peptide materials as superbug killers. *Front. Chem.*, **9**, 795433.
168. Mattiuzzo M, Bandiera A, Gennaro R, Benincasa M, Pacor S, Antcheva N, and Scocchi M (2007). Role of the *Escherichia coli* SbmA in the antimicrobial activity of proline-rich peptides. *Mol. Microbiol.*, **66(1)**, 151-163.
169. McCoy AJ, Liu H, Falla TJ, and Gunn JS (2001). Identification of *Proteus mirabilis* mutants with increased sensitivity to antimicrobial peptides. *Antimicrob. Agents Chemother.*, **45(7)**, 2030-2037.
170. McDermott AM (2013). Antimicrobial compounds in tears. *Exp. Eye Res.*, **117**, 53-61.
171. McDonald C, Jovanovic G, Ces O, and Buck M (2015). Membrane stored curvature elastic stress modulates recruitment of maintenance proteins PspA and Vipp1. *MBio*, **6(5)**, 10-1128.
172. Micsonai A, Wien F, Bulyáki É, Kun J, Moussong É, Lee YH, Goto Y, Réfrégiers M, and Kardos J (2018). BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.*, **46(W1)**, 315-322.
173. Miller SI, Ernst RK, and Bader MW (2005). LPS, TLR4 and infectious disease diversity. *Nature Rev. Microbiol.*, **3(1)**, 36-46.
174. Mirghani R, Saba T, Khaliq H, Mitchell J, Do L, Chambi L, Diaz K, Kennedy T, Alkassab K, Huynh T, and Elmi M (2022). Biofilms: Formation, drug resistance and alternatives to conventional approaches. *AIMS microbiol.*, **8(3)**, 239.
175. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, and Gough J (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.*, **47(D1)**, 351-360.

176. Mitra A, Ko YH, Cingolani G, and Niederweis M (2019). Heme and hemoglobin utilization by Mycobacterium tuberculosis. *Nat. Commun.*, **10(1)**, 4260.
177. Miyata T, Tokunaga F, Yoneya T, Yoshikawa K, Iwanaga S, Niwa M, Takao T, and Shimonishi Y (1989). Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin II, and polyphemusins I and II: chemical structures and biological activity. *J. Biochem.*, **106(4)**, 663-668.
178. Mookherjee N, Anderson MA, Haagsman HP, and Davidson DJ (2020). Antimicrobial host defence peptides: functions and clinical potential. *Nat. Rev. Drug Discov.*, **19(5)**, 311-332.
179. Mori H and Ito K (2001). The Sec protein-translocation pathway. *Trends Microbial.*, **9(10)**, 494-500.
180. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, and Olson AJ (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, **30(16)**, 2785-2791.
181. Mount KL, Townsend CA, Rinker SD, Gu X, Fortney KR, Zwickl BW, Janowicz DM, Spinola SM, Katz BP, and Bauer ME (2010). *Haemophilus ducreyi* SapA contributes to cathelicidin resistance and virulence in humans. *Infect. Immun.*, **78(3)**, 1176-1184.
182. Mues N and Chu HW (2020). Out-Smarting the Host: Bacteria maneuvering the immune response to favor their survival. *Front. Immunol.*, **11**, 819.
183. Müller L, Di Benedetto S, and Pawelec G (2019). The immune system and its dysregulation with aging. *Biochemistry and cell biology of ageing: Part II clinical science*, 21-43.
184. Muniz LR, Knosp C, and Yeretssian G (2012). Intestinal antimicrobial peptides during homeostasis, infection, and disease. *Front Immunol.*, **3**, 310.
185. Murdoch CC and Skaar EP (2022). Nutritional immunity: the battle for nutrient metals at the host–pathogen interface. *Nat. Rev. Microbiol.*, **20(11)**, 657-670.
186. Nairz M and Weiss G (2020). Iron in infection and immunity. *Mol. Aspects Med.*, **75**, 100864.

187. Nayab S, Aslam MA, Rahman SU, Sindhu ZUD, Sajid S, Zafar N, Razaq M, Kanwar R, and Amanullah (2022). A review of antimicrobial peptides: its function, mode of action and therapeutic potential. *Int. J. Pept. Res. Ther.*, **28(1)**, 46.
188. Nguyen HN, Jain A, Eulenstein O, and Friedberg I (2019). Tracing the ancestry of operons in bacteria. *Bioinformatics*, **35(17)**, 2998-3004.
189. Omardien S, Drijfhout JW, Vaz FM, Wenzel M, Hamoen LW, Zaat SA, and Brul S (2018). Bactericidal activity of amphipathic cationic antimicrobial peptides involves altering the membrane fluidity when interacting with the phospholipid bilayer. *Biochim. Biophys. Acta Biomembr.*, **1860(11)**, 2404-2415.
190. Oppenheim FG, Xu T, McMillian FM, Levitz SM, Diamond RD, Offner GD, and Troxler RF (1988). Histatins, a novel family of histidine-rich proteins in human parotid secretion. Isolation, characterization, primary structure, and fungistatic effects on *Candida albicans*. *J. Biol. Chem.*, **263(16)**, 7472-7477.
191. Oppenheim JJ, Biragyn A, Kwak LW, and Yang D (2003). Roles of antimicrobial peptides such as defensins in innate and adaptive immunity. *Ann Rheum Dis*., **62(Suppl 2)**, 17-21.
192. Oren Z and Shai Y (1998). Mode of action of linear amphipathic α-helical antimicrobial peptides. *Biopolymers.*, **47(6)**, 451-463.
193. Oren Z, Lerman JC, Gudmundsson GH, Agerberth B, and Shai Y (1999). Structure and organization of the human antimicrobial peptide LL-37 in phospholipid membranes: relevance to the molecular basis for its non-cell-selective activity. *Biochem. J.*, **341(3)**, 501-513.
194. Oswald C, Holland IB, and Schmitt L (2006). The motor domains of ABC-transporters: what can structures tell us? *Naunyn Schmiedebergs Arch. Pharmacol.*, **372**, 385-399.
195. Otto BR, Sijbrandi R, Luirink J, Oudega B, Heddle JG, Mizutani K, Park SY, and Tame JR (2005). Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. *J. Biol. Chem.*, **280(17)**, 17339-17345.
196. Padilla E, Llobet E, Doménech-Sánchez A, Martínez-Martínez L, Bengoechea JA, and Albertí S (2010). *Klebsiella pneumoniae* AcrAB efflux

pump contributes to antimicrobial resistance and virulence. *Antimicrob. Agents Chemother.*, **54(1)**, 177-183.

197. Pandey S, Modak A, Phale PS, and Bhaumik P (2016). High resolution structures of periplasmic glucose-binding protein of *Pseudomonas putida* CSV86 reveal structural basis of its substrate specificity. *J. Biol. Chem.*, **291(15)**, 7844-7857.
198. Pandidan S and Mechler A (2019). Nano-viscosimetry analysis of the membrane disrupting action of the bee venom peptide melittin. *Sci. Rep.*, **9(1)**, 10841.
199. Pandidan S and Mechler A (2021). Latest developments on the mechanism of action of membrane disrupting peptides. *Biophys. Rev.*, **7(3)**, 173.
200. Parra-Lopez C, Baer MT, and Groisman EA (1993). Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in *Salmonella typhimurium*. *EMBO J.*, **12(11)**, 4053-4062.
201. Parrinello M and Rahman A (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. App. Phys*., **52(12)**, 7182-7190.
202. Pence HE and Williams A (2010). ChemSpider: an online chemical information resource. *J. Chem. Educ.*, **87(11)**, 1123–1124.
203. Pink DA, Hasan FM, Quinn BE, Winterhalter M, Mohan M, and Gill TA (2014). Interaction of protamine with gram-negative bacteria membranes: possible alternative mechanisms of internalization in *Escherichia coli*, *Salmonella typhimurium* and *Pseudomonas aeruginosa*. *J. Pept. Sci.*, **20(4)**, 240-250.
204. Pollard AJ and Bijker EM (2021). A guide to vaccinology: from basic principles to new developments. *Nat. Rev. Immunol.*, **21(2)**, 83-100.
205. Porcelli F, Ramamoorthy A, Barany G, and Veglia G (2013). On the role of NMR spectroscopy for characterization of antimicrobial peptides. *Membrane Proteins: Folding, Association, and Design*, 159-180.
206. Powers JPS and Hancock RE (2003). The relationship between peptide structure and antibacterial activity. *Peptides*, ***24*(11)**, 1681-1691.

207. Preissler S, Rato C, Chen R, Antrobus R, Ding S, Fearnley IM, and Ron D (2015). AMPylation matches BiP activity to client protein load in the endoplasmic reticulum. *Elife*, **4**, 12621.
208. Price MN, Arkin AP, and Alm EJ (2006). The life-cycle of operons. *PLoS Genet.*, **2(6)**, e96.
209. Quiocho FA, Spurlino JC, and Rodseth LE (1997). Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure*, **5(8)**, 997-1015.
210. Raghava S, Giorda KM, Romano FB, Heuck AP, and Hebert DN (2011). The SV40 late protein VP4 is a viroporin that forms pores to disrupt membranes for viral release. *PLoS Pathog.*, **7(6)**, 1002116.
211. Rahman MM, Machuca MA, Khan MF, Barlow CK, Schittenhelm RB, and Roujeinikova A (2019). Molecular basis of unexpected specificity of ABC transporter-associated substrate-binding protein DppA from *Helicobacter pylori*. *J. Bacteriol.*, **201(20)**, 10-1128.
212. Ranjbar B and Gill P (2009). Circular dichroism techniques: biomolecular and nanostructural analyses-a review. *Chem. Biol. Drug Des.*, **74(2)**, 101-120.
213. Rees DC, Johnson E, and Lewinson O (2009). ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.*, **10(3)**, 218.
214. Rice AJ, Park A, and Pinkett HW (2014). Diversity in ABC transporters: type I, II and III importers. *Crit. Rev. Biochem. Mol. Biol.*, **49(5)**, 426-437.
215. Rice P, Longden I, and Bleasby A (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.*, **16(6)**, 276-277.
216. Richards SM, Strandberg KL, and Gunn JS (2010). *Salmonella*-regulated lipopolysaccharide modifications. *Subcell. Biochem.,* **53**, 101-122.
217. Rietsch A and Beckwith J (1998). The genetics of disulfide bond metabolism. *Annu. Rev. Genet.*, **32(1)**, 163-184.
218. Rinker SD, Trombley MP, Gu X, Fortney KR, and Bauer ME (2011). Deletion of *mtrC* in *Haemophilus ducreyi* increases sensitivity to human antimicrobial peptides and activates the *CpxRA* regulon. *Infect. Immun.*, **79**, 2324-2334.
219. Rivera KG, Tanaka KJ, Buechel ER, Origel Jr O, Harrison A, Mason KM, and Pinkett HW (2024). Antimicrobial Peptide Recognition Motif of the

Substrate Binding Protein SapA from Nontypeable *Haemophilus influenzae*. *Biochemistry*, **63(3)**, 294-311.

220. Rodríguez-Arce I, Al-Jubair T, Euba B, Fernández-Calvet A, Gil-Campillo C, Martí S, Törnroth-Horsefield S, Riesbeck K, and Garmendia J (2019). Moonlighting of Haemophilus influenzae heme acquisition systems contributes to the host airway-pathogen interplay in a coordinated manner. *Virulence*, **10(1)**, 315-333.
221. Rončević T, Puizina J, and Tossi A (2019). Antimicrobial peptides as anti-infective agents in pre-post-antibiotic era? *Int. J. Mol. Sci.*, **20(22)**, 5713.
222. Rosengren KJ, Clark RJ, Daly NL, Göransson U, Jones A, and Craik DJ (2003). Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. *J. Am. Chem. Soc.*, **125(41)**, 12464-12474.
223. Runyen-Janecky LJ (2013). Role and regulation of heme iron acquisition in gram-negative pathogens. *Front. Cell. Infect. Microbiol.,* **3**, 55.
224. Saether O, Craik DJ, Campbell ID, Sletten K, Juul J, and Norman DG (1995). Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1. *Biochemistry*, **34(13)**, 4147-4158.
225. Saha D, Patgaonkar M, Shroff A, Ayyar K, Bashir T, and Reddy KVR (2014). Hemoglobin Expression in Nonerythroid Cells: Novel or Ubiquitous? *Int. J. Inflamm*. **2014**, 1–8.
226. Salamov VSA and Solovyevand A (2011). Automatic annotation of microbial genomes and metagenomic sequences. *Metagenomics and its applications in agriculture, biomedicine and environmental studies*, Nova Science Publishers, New York, 61-78.
227. Salzman NH, Chou MM, de Jong H, Liu L, Porter EM, and Paterson Y (2003). Enteric salmonella infection inhibits Paneth cell antimicrobial peptide expression. *Infect. Immun.*, **71(3)**, 1109-1115.
228. Saurin W, Hofnung M, and Dassa E (1999). Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J. Mol. Evol.*, **48(1)**, 22-41.
229. Scheepers GH, Lycklam a Nijeholt JA, and Poolman B (2016). An updated structural classification of substrate-binding proteins. *FEBS Lett.*, **590(23)**, 4393-4401.

230. Schmidtchen A, Frick IM, and Björck L (2001). Dermatan sulphate is released by proteinases of common pathogenic bacteria and inactivates antibacterial α-defensin. *Mol. Microbiol.*, **39(3)**, 708-713.
231. Schneider CA, Rasband WS, and Eliceiri KW (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9(7)**, 671–675.
232. Seefeldt AC, Nguyen F, Antunes S, Pérébaskine N, Graf M, Arenz S, Inampudi KK, Douat C, Guichard G, Wilson DN, and Innis CA (2015). The proline-rich antimicrobial peptide Onc112 inhibits translation by blocking and destabilizing the initiation complex. *Nat. Struct. Mol. Biol.*, **22(6)**, 470-475.
233. Selsted ME, Novotny MJ, Morris WL, Tang YQ, Smith W, and Cullor JS (1992). Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils. *J. Biol. Chem.*, **267(7)**, 4292-4295.
234. Seyfi R, Kahaki FA, Ebrahimi T, Montazersaheb S, Eyvazi S, Babaeipour V, and Tarhriz V (2020). Antimicrobial peptides (AMPs): roles, functions and mechanism of action. *Int. J. Pept. Res. Ther.*, **26**, 1451-1463.
235. Shadman H, Gallops CE, Ziebarth JD, DeRouchey JE, and Wang Y (2022). Exploring Structures and Dynamics of Protamine Molecules through Molecular Dynamics Simulations. *ACS Omega*, ***7*(46)**, 42083-42095.
236. Shafee TM, Lay FT, Phan TK, Anderson MA, and Hulett MD (2017). Convergent evolution of defensin sequence, structure and function. *Cell. Mol. Life Sci.*, **74**, 663-682.
237. Shafer WM, Qu X, Waring AJ, and Lehrer RI (1998). Modulation of *Neisseria gonorrhoeae* susceptibility to vertebrate antibacterial peptides due to a member of the resistance/nodulation/division efflux pump family. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 1829-1833.
238. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13(11)**, 2498-2504.
239. Shelton CL, Raffel FK, Beatty WL, Johnson SM, and Mason KM (2011). Sap transporter mediated import and subsequent degradation of antimicrobial peptides in *Haemophilus*. *PLOS Pathog.*, **7**, 1002360.

240. Shenkarev ZO, Balandin SV, Trunov KI, Paramonov AS, Sukhanov SV, Barsukov LI, Arseniev AS, and Ovchinnikova TV (2011). Molecular mechanism of action of β-hairpin antimicrobial peptide arenicin: oligomeric structure in dodecylphosphocholine micelles and pore formation in planar lipid bilayers. *Biochemistry*, **50(28)**, 6255-6265.
241. Shishido SN, Varahan S, Yuan K, Li X, and Fleming SD (2012). Humoral innate immune response and disease. *Clin. Immunol.*, **144(2)**, 142-158.
242. Shitan N, Bazin I, Dan K, Obata K, Kigawa K, Ueda K, Sato F, Forestier C, and Yazaki K (2003). Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in *Coptis japonica*. *Proc. Natl. Acad. Sci. U. S. A.*, **100(2)**, 751-756.
243. Sievers F and Higgins DG (2014). Clustal omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.*, **1079**, 105-116.
244. Singh A, Duche RT, Wandhare AG, Sian JK, Singh BP, Sihag MK, Singh KS, Sangwan V, Talan S, and Panwar H (2023). Milk-derived antimicrobial peptides: overview, applications, and future perspectives. *Probiotics Antimicrob.*, **15(1)**, 44-62.
245. Singh MK and Manoj N (2017). Structural role of a conserved active site cis proline in the *Thermotoga maritima* acetyl esterase from the carbohydrate esterase family 7. *Proteins.*, **85(4)**, 694-708.
246. Skals M, Jorgensen NR, Leipziger J, and Praetorius HA (2009). α-Hemolysin from *Escherichia coli* uses endogenous amplification through P2X receptor activation to induce hemolysis. *Proc. Natl. Acad. Sci. U. S. A.*, **106(10)**, 4030-4035.
247. Smith MW, Tyreman DR, Payne GM, Marshall NJ, and Payne JW (1999). Substrate specificity of the periplasmic dipeptide-binding protein from *Escherichia coli*: experimental basis for the design of peptide prodrugs. *Microbiology*, **145(10)**, 2891-2901.
248. Sohlenkamp C and Geiger O (2016). Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiol. Rev.*, **40(1)**, 133-159.
249. Sollome J and Fry RC (2015). Environmental Contaminants and the Immune System: A Systems Perspective. *Syst. Biol. Toxicol. Environ. Health*, 171-186.

250. Sperandio B, Regnault B, Guo J, Zhang Z, Stanley SL Jr, Sansonetti PJ, and Pedron T (2008). Virulent *Shigella flexneri* subverts the host innate immune response through manipulation of antimicrobial peptide gene expression. *J. Exp. Med.*, **205**, 1121-1132.
251. Sproston NR and Ashworth JJ (2018). Role of C-reactive protein at sites of inflammation and infection. *Front. Immunol.*, **9**, 754.
252. Steiner H, Hultmark D, Engström Å, Bennich H, and Boman HG (1981). Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature*, **292(5820)**, 246-248.
253. Stewart EJ, Åslund F, and Beckwith J (1998). Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for the thioredoxins. *EMBO J.*, **17(19)**, 5543-5550.
254. Stumpe S, Schmid R, Stephens DL, Georgiou G, and Bakker EP (1998). Identification of OmpT as the protease that hydrolyzes the antimicrobial peptide protamine before it enters growing cells of *Escherichia coli*. *J. Bacteriol.*, **180**, 4002-4006.
255. Su T, Cheng MT, and Lin SH (2014). Investigating the effectiveness of an educational card game for learning how human immunology is regulated. *CBE Life Sci. Educ.*, **13(3)**, 504-515.
256. Subbalakshmi C and Sitaram N (1998). Mechanism of antimicrobial action of indolicidin. *FEMS Microbiol. Lett.*, **160(1)**, 91-96.
257. Sugiyama Y, Nakamura A, Matsumoto M, Kanbe A, Sakanaka M, Higashi K, Igarashi K, Katayama T, Suzuki H, and Kurihara S (2016). A novel putrescine exporter SapBCDF of *Escherichia coli*. *J. Biol. Chem.*, **291**, 26343-26351.
258. Suits MD, Lang J, Pal GP, Couture M, and Jia Z (2009). Structure and heme binding properties of *Escherichia coli* O157: H7 ChuX. *Protein Sci.*, **18(4)**, 825-838.
259. Sumathi K, Ananthalakshmi P, Roshan MM, and Sekar K (2006). 3dSS: 3D structural superposition. *Nucleic Acids Res.*, **34(2)**, 128-132.
260. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, and Bork P (2023). The STRING database in 2023: protein–protein association networks and

functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.*, **51(D1)**, 638-646.

261. Taboada B, Estrada K, Ciria R, and Merino E (2018). Operon-mapper: A web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*, **34(23)**, 4118-4120.
262. Tal N and Schuldiner S (2009). A coordinated network of transporters with overlapping specificities provides a robust survival strategy. *Proc. Natl. Acad. Sci. U. S. A.*, **106(22)**, 9051-9056.
263. Tanaka KJ and Pinkett HW (2019). Oligopeptide-binding protein from nontypeable *Haemophilus influenzae* has ligand-specific sites to accommodate peptides and heme in the binding pocket. *J. Biol. Chem.*, **294(3)**, 1070-1082.
264. Tang C, Schwieters CD, and Clore GM (2007). Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature*, **449(7165)**, 1078-1082.
265. Taylor D, Cawley G, and Hayward S (2014). Quantitative method for the assignment of hinge and shear mechanism in protein domain movements. *Bioinformatics*, **30(22)**, 3189-3196.
266. Terasaka K, Blakeslee JJ, Titapiwatanakun B, Peer WA, Bandyopadhyay A, Makam SN, Lee OR, Richards EL, Murphy AS, Sato F, and Yazaki K (2005). PGP4, an ATP binding cassette P-glycoprotein, catalyzes auxin transport in *Arabidopsis thaliana* roots. *Plant Cell*, **17(11)**, 2922-2939.
267. The UniProt Consortium (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.*, **51(D1)**, 523–531.
268. Thomas C, Aller SG, Beis K, Carpenter EP, Chang G, Chen L, Dassa E, Dean M, Duong Van Hoa F, Ekiert D, and Ford R (2020). Structural and functional diversity calls for a new classification of ABC transporters. *FEBS Lett.*, **594(23)**, 3767-3775.
269. Thomassin J, Brannon J, Gibbs BF, Gruenheid S, and Le Moual H (2012). OmpT outer membrane proteases of enterohemorrhagic and enteropathogenic *Escherichia coli* contribute differently to the degradation of human LL-37. *Infect. Immun.*, **80**, 483-492.

270. Tian W, Chen C, Lei X, Zhao J, and Liang J (2018). CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.*, **46(W1)**, W363-W367.
271. Tomar TS, Dasgupta P, and Kanaujia SP (2023). Operon Finder: a deep learning-based web server for accurate prediction of prokaryotic operons. *J. Mol. Biol.*, **435(14)**, 167921.
272. Torres AG and Payne SM (1997). Haem iron-transport system in enterohaemorrhagic *Escherichia coli* O157: H7. *Mol. Microbiol.*, **23(4)**, 825-833.
273. Touw WG, Baakman C, Black J, Te Beek TA, Krieger E, Joosten RP, and Vriend G (2015). A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.*, **43(D1)**, 364-368.
274. Tsirigotaki A, De Geyter J, Economou A, and Karamanou S (2017). Protein export through the bacterial Sec pathway. *Nat. Rev. Microbiol.*, **15(1)**, 21-36.
275. Turner PJ (2005). XMGRACE, Version 5.1.19. Beaverton, OR: Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology.
276. Turvey SE and Broide DH (2010). Innate immunity. *J. Allergy Clin. Immunol.*, **125(2)**, 24-32.
277. Tuszynska I, Magnus M, Jonak K, Dawson W, and Bujnicki JM (2015). NPDock: a web server for protein–nucleic acid docking. *Nucleic Acids Res.*, **43(W1)**, 425-430.
278. Uberoi A, Bartow-McKenney C, Zheng Q, Flowers L, Campbell A, Knight SA, Chan N, Wei M, Lovins V, Bugayev J, and Horwinski J (2021). Commensal microbiota regulates skin barrier function and repair via signaling through the aryl hydrocarbon receptor. *Cell Host Microbe.*, **29(8)**, 1235-1248.
279. Vanommeslaeghe K and MacKerell Jr AD (2012). Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model.*, **52(12)**, 3144-3154.
280. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, and Žídek A (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage

of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50(D1)**, 439-444.

281. Von Heijne G (1985). Signal sequences: the limits of variation. *J. Mol. Biol.*, **184(1)**, 99-105.
282. Walrant A, Bauzá A, Girardet C, Alves ID, Lecomte S, Illien F, Cardon S, Chaianantakul N, Pallerla M, Burlina F, and Frontera A (2020). Ionpair-π interactions favor cell penetration of arginine/tryptophan-rich cell-penetrating peptides. *Biochim. Biophys. Acta Biomembr.*, **1862(2)**, 183098.
283. Wang E, Qiang X, Li J, Zhu S, and Wang P (2016a). The *in vitro* immune-modulating properties of a sweat gland-derived antimicrobial peptide dermcidin. *Shock*, **45(1)**, 28.
284. Wang G (2014). Human antimicrobial peptides and proteins. *Pharmaceuticals*, **7(5)**, 545-594.
285. Wang H, Peng H, Li W, Cheng P, and Gong M (2021). The toxins of *Beauveria bassiana* and the strategies to improve their virulence to insects. *Front. Microbiol.*, 2375.
286. Wang X, Preston III JF, and Romeo T (2004). The pgaABCD locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation. *J. Bacteriol.*, **186(9)**, 2724-2734.
287. Wang Z, Bie P, Cheng J, Lu L, Cui B, and Wu Q (2016b). The ABC transporter YejABEF is required for resistance to antimicrobial peptides and the virulence of *Brucella melitensis*. *Sci. Rep.*, **6(1)**, 31876.
288. Weatherspoon-Griffin N, Yang D, Kong W, Hua Z, and Shi Y (2014). The CpxR/CpxA two-component regulatory system up-regulates the multidrug resistance cascade to facilitate *Escherichia coli* resistance to a model antimicrobial peptide. *J. Biol. Chem.*, **289(47)**, 32571-32582.
289. White SH, Wimley WC, and Selsted ME (1995). Structure, function, and membrane integration of defensins. *Curr. Opin. Struct. Biol.*, **5(4)**, 521-527.
290. Wiesner J and Vilcinskas A (2010). Antimicrobial peptides: the ancient arm of the human immune system. *Virulence*, **1(5)**, 440-464.
291. Wilkens S (2015). Structure and mechanism of ABC transporters. *F1000prime rep.*, **7**, 14.

292. Xhindoli D, Pacor S, Benincasa M, Scocchi M, Gennaro R, and Tossi A (2016). The human cathelicidin LL-37 - A pore-forming antibacterial peptide and host-cell modulator. *Biochim. Biophys. Acta Biomembr.*, **1858(3)**, 546-566.
293. Xie F, Langford PR, and Wang C (2017). The SapA protein is involved in resistance to antimicrobial peptide PR-39 and virulence of *Actinobacillus pleuropneumoniae*. *Front. Microbiol.*, ***8***, 259698.
294. Yamamura R, Inoue KY, Nishino K, and Yamasaki S (2023). Intestinal and fecal pH in human health. *Front. Microbiomes*, **2**, 1192316.
295. Yan Y, Tao H, He J, and Huang SY (2020). The HDOCK server for integrated protein–protein docking. *Nat. Protoc.*, **15(5)**, 1829-1852.
296. Yang J, Yan R, Roy A, Xu D, Poisson J, and Zhang Y (2015). The I-TASSER Suite: protein structure and function prediction. *Nat. Methods*, **12(1)**, 7-8.
297. Yang L, Harroun TA, Weiss TM, Ding L, and Huang HW (2001). Barrel-stave model or toroidal model? A case study on melittin pores. *Biophys. J.*, **81(3)**, 1475-1485.
298. Yu J, Zhou Y, Tanaka I, and Yao M (2010). Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics*, **26(1)**, 46-52.
299. Zaitseva J, Jenewein S, Jumpertz T, Holland IB, and Schmitt L (2005). H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *EMBO J.,* 24, 1901–1910.
300. Zasloff M (1987). Magainins, a class of antimicrobial peptides from Xenopus skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc. Natl. Acad. Sci. U. S. A.*, **84(15)**, 5449-5453.
301. Zhang LJ and Gallo RL (2016). Antimicrobial peptides. *Curr. Biol.*, **26(1)**, 14–19.
302. Zhang QY, Yan ZB, Meng YM, Hong XY, Shao G, Ma JJ, Cheng XR, Liu J, Kang J, and Fu CY (2021). Antimicrobial peptides: mechanism of action, activity and clinical potential. *Mil. Med. Res.*, **8**, 1-25.
303. Zhang YM and Rock CO (2008). Membrane lipid homeostasis in bacteria. *Nat. Rev. Microbiol.*, **6(3)**, 222-233.

304. Zheng D, Liwinski T, and Elinav E (2020). Interaction between microbiota and immunity in health and disease. *Cell Res.*, **30(6)**, 492-506.
305. Zhu S, Peigneur S, Gao B, Umetsu Y, Ohki S, and Tytgat J (2014). Experimental conversion of a defensin into a neurotoxin: implications for origin of toxic function. *Mol. Biol. Evol.*, **31(3)**, 546-559.
306. Ziaja M, Dziedzic A, Szafraniec K, and Piastowska-Ciesielska A (2020). Cecropins in cancer therapies-where we have been? *Eur. J. Pharmacol.*, **882**, 173317.

List of Publications and Conferences

Publications from thesis

1. **Dasgupta P**, Vinil K, and Kanaujia SP (2024). Evolutionary trends indicate a coherent organization of *sap* operons. *Res. Microbiol.*, 175(8), 104228.
2. **Dasgupta P** and Kanaujia SP. Enlightening the multifarious attributes of the *Escherichia coli* Sap transport system: A computational perspective. *(Under review)*
3. **Dasgupta P** and Kanaujia SP. Unlocking conformational dynamics of *Ec*SapA: Insights into the emphasis of ligand dependency using molecular dynamics simulations. *(Under preparation)*
4. **Dasgupta P** and Kanaujia SP. Biophysical characterization of a putative antimicrobial peptide-binding protein of *Escherichia coli* highlights its dual functionality. *(Under communication)*

Publications from other work

1. Tomar TS, **Dasgupta P**, and Kanaujia SP (2022). Operon Finder: A deep learning-based web server for accurate prediction of prokaryotic operons. *J. Mol. Biol.*, 435(14), 167921**.**
2. Chandravanshi M, Sharma A, **Dasgupta P**, Mandal SK and Kanaujia SP (2019). Identification and characterization of ABC transporters for carbohydrate uptake in *Thermus thermophilus* HB8. *Gene*, 696, 135-148.

Conferences attended

1. **Pratik Dasgupta** and Shankar Prasad Kanaujia. Evolutionary insights into the conservation of a class of antimicrobial peptide transporter operons. Research and Industrial Conclave – Integration'2024. August 09-11, 2024, IIT Guwahati, Assam, India. (*Oral Presentation*).
2. **Pratik Dasgupta** and Shankar Prasad Kanaujia. Unravelling the multifarious behaviour of a putative antimicrobial peptide-binding protein via computational strategies. National Seminar on Crystallography (NSC-50). November 22-24, 2023, CSIR-IMTech Chandigarh, Chandigarh, India. (*Poster Presentation*)

3. **Pratik Dasgupta** and Shankar Prasad Kanaujia. *In silico* analysis reveals the evolutionary conservation of a class of antimicrobial peptide (AMP) transporter operons. National Seminar on Crystallography (NSC-49). November 28-30, 2022, University of Jammu, Jammu, India. *(Oral Presentation)*
4. **Pratik Dasgupta** and Shankar Prasad Kanaujia. *In silico* analysis of a putative antimicrobial peptide-binding protein highlights its promiscuous nature. National Seminar on Crystallography (NSC-48). November 25-27, 2021, IIT Roorkee, Uttarakhand, India. (*Poster Presentation*)
5. **Pratik Dasgupta** and Shankar Prasad Kanaujia. *In silico* studies enlighten the salient features of a putative antimicrobial peptide-binding protein. PDB 50th Anniversary Symposium in Asia. November 24, 2021, PDBj and Institute for Protein Research, Osaka University, Japan. (*Poster Presentation*)
6. **Pratik Dasgupta** and Shankar Prasad Kanaujia. Insights into a putative antimicrobial peptide binding protein: An *in silico* approach. Bringing Molecular Structure to Life: 50 Years of the PDB. October 20-22, 2021, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. *(Poster Presentation)*