Structural Characterization of Small basic protein (Sbp) and Accumulation associated protein (Aap) – two Proteins involved in Biofilm Formation in *Staphylococcus epidermidis*

DISSERTATION

zur Erlangung des akademischen Grades eines Doktors
der Naturwissenschaften (Dr. rer. nat.)

Fachbereich Chemie
der Universität Hamburg

Vorgelegt von

Madiha Fayyaz
aus Lahore, Pakistan
Hamburg, 2017

Die vorliegende Arbeit wurde im Zeitraum von April 2013 bis Dezember 2017 in der Arbeitsgruppe von Prof. Dr. Ch. Betzel im Laboratorium für Strukturbiologie von Infektion und Entzündung am Institut für Biochemie und Molekularbiologie des Fachbereichs Chemie sowie in der Arbeitsgruppe von Prof. Dr. Holger Rohde im Institut für Mikrobiologie am Universitätsklinikum Hamburg Eppendorf der Universität Hamburg durchgeführt.

Gutachter: Prof. Ch. Betzel
Gutachter: Prof. R. Bredehorst

Tag der Disputation: 31.03.2017

To my parents (Nusrat Saba & Muhammad Fayyaz), siblings and homeland, Pakistan

Contents of Thesis

ABBREVIATIONS

2D	two dimensional
3D	three dimensional
Å	Angstrom (10^{-10} m)
M	micro (10^{-6}m)
Aa	amino acid(s)
Aap	Accumulation associated protein
Amp	Ampicillin
approx.	Approximately
APS	ammonium per-sulfate
AtlE	Autolysin E
Bp	base pair
°C	degree Celsius
CD	circular dichroism
cDNA	*complementary* DNA
DESY	Deutsches Elektronen Synchrotron (German electron synchrotron)
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
dNTPs	2'-deoxynucleoside-5'-triphosphate
DTT	Dithiothreitol
E.	*Escherichia*
EDTA	ethylene diamine tetraacetic acid
EMBL	European Molecular Biology Laboratory
Embp	Extracellular matrix binding protein

ESI	Electrospray ionization
et al.	*et alii*
G	gram (unit)
HF	High Fidelity
Hr	Hour
I	Intensity
IMD	Implanted medical devices
IPTG	isopropyl-β-D-thiogalactopyranoside
Kd	dissociation constant
kDa	kilo Dalton
L	litre (unit)
LB	Luria Bertani
LPS	Lipopolysaccharides
M	Molar
M	milli (10^{-3}m)
Mda	Mega Dalton
mM	milli Molar
Mg	milli gram
Min	Minutes
MRSA	methicillin resistant *S. Aureus*
MS	mass spectrometry
MW	molecular weight (g/mol or Da)
MWCO	molecular weight cut off
Ni-NTA	Nickel-nitrilotriacetic acid
Nm	nanometer
NSD	normalized spatial discrepancy
OD	Optical Density

ORFs	open reading frames
P.	*Pseudomonas*
PIC	Protease inhibitor cocktail
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PCT	pre-crystallization test
PEG	polyethylene glycol
PIA	polysaccharide intercellualar adhesion
Rg	radius of gyration
R_H	hydrodynamic radius
Rpm	revolutions per minute
S.	*Staphylococcus*
S	scattering vector
S	second(s)
SAXS	small angle X-ray scattering
Sbp	small basic protein
SDS	sodium dodecyl sulfate
SEC	size-exclusion chromatography
sec.	Secondary
SFX	serial femtosecond crystallography
SOC	super optimal broth with catabolic repression
T	temperature [K]
T	time [s]
TAE	Tris-acetate-EDTA
TEMED	*N,N,N',N'*-tetramethylethylenediamine
TEV	(protease) tobacco etch virus (protease)

TMV	Tobacco mosaic virus
TOF	time of flight
UV	ultra violet
UV-vis	ultraviolet-visible light
v/v	volumeper volume
w/v	weightper volume

Abbreviations of Nucleotides

A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uridine

Abbreviations of Amino Acids

A	Ala	Alanine
R	Arg	Arginine
N	Asn	Asparagine
D	Asp	Aspartate
C	Cys	Cysteine
E	Glu	Glutamate
Q	Gln	Glutamine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
L	Leu	Leucine
K	Lys	Lysine
M	Met	Methionine

F	Phe	Phenylalanine
P	Pro	Proline
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
V	Val	Valine

1. Introduction

1.1. Bacterial biofilm formation as a significant virulence principle

A large number of biotic and abiotic surfaces can be infected by a single or mixed microbial species forming biofilms, a meshwork of unicellular organisms enclosed in an extracellular bacterial derived matrix composed of proteins, polysaccharides and nucleic acid (Fey and Olson, 2010; O'Toole *et al*., 2000). Bacteria reside in biofilms not only in nature but they also get attached to the household surfaces such as curtains, shower-heads, drinking water systems and bath sinks (Mullis and Falkinham, 2013; Rożej *et al*., 2015; Xu *et al*., 2014). Although mostly biofilms are formed by mixed multiple microbial species, they can also consist of a single microbial species (Adal and Farr, 1996). *Pseudomonas aeruginosa* has emerged as the most studied single-species, biofilm-forming gram-negative bacterium, although, as detailed in this review, among the gram-negative bacteria, *Pseudomonas aeruginosa*, *Escherichia coli* and *Vibrio cholerae* have also been studied in detail. Gram-positive biofilm-forming bacteria that have been studied include *Staphylococcus epidermidis*, *Staphylococcus aureus* and enterococci (O'Toole *et al*., 2000).

Staphylococcus epidermidis and different *Pseudomonas species* assemble biofilm communities resembling mushroom like structures, while most of the biofilms are arranged in multiple cell layers. Flat, dried and wrinkled colonies on agar plates are formed by *Vibrio cholerae* and *Bacillus subtilis* (Bester *et al*., 2011; López *et al*., 2010; Seper *et al*., 2014). Biofilms formed in silica are a major threat to food processing and are formed by *Anoxybacillus flavithermus* (Saw *et al*., 2008). Small yellow air balloon shaped biofilms have been observed to be organized by *Myxococcus xanthus* species (Kim *et al*., 2009; Zhang *et al*., 2005). A matrix which is a slimy substance surrounds around and on the surface of each bacterial cell is a general feature of all biofilms formed by bacteria. It helps in the survival of bacteria by protecting them from external environment and providing nutritious substances and water to the cells (Bester *et al*., 2011; López *et al*., 2010). This matrix is composed of carbohydrates, proteins, extracellular DNA and phospholipids (Becker *et al*., 2014; Christner *et al*., 2012; Linnes *et al*., 2013; Reichhardt *et al*., 2015). A phenomenon referred to as quorum sensing has been observed, in which passive diffusion of nutrients and other

substances through the porous matrix helps in sharing them between the cells and acting as cell communicating components (Banat *et al*., 2014). The survival and growth of cells depend on each other, as the cells get benefits from each other (Bester *et al*., 2011; Xu *et al*., 2014).The most remarkable function of the matrix is to protect the cells against external environment and mechanical shattering. The matrix also deals with the chemical effects on cells, such as disinfectants, antimicrobials and antibiotics. In summary, the biofilm safeguards cells against mechanical, chemical and physical stresses, limited nutrient availability and tearing and shearing forces (Banat *et al*., 2014; Taylor *et al*., 2014).

1.2. *Staphylococcus epidermidis,* its prevalence and pathogenicity

Staphylococcus epidermidis belongs to the group of coagulase negative staphylococci (CoNS). It is the most frequently found species in the human epithelia and the most abundant species responsible for infection (Kloos and Musselwhite, 1975). An average of 10-24 different strains are carried by a person (Grice *et al*., 2009). It is the third most common cause of nosocomial infections and is frequently found in infections of implanted medical devices. The two major features predominate the cause of infection - the prevalence of *S. epidermidis* on human skin, which is due to its exceptional ability to stick to skin and biomaterials, such as indwelling medical devices and forming multilayered biofilms (Cheung *et al*., 2010; Otto, 2009). The genome analysis of *S. epidermidis* illustrates that the species is equipped with the genes, which provide protection against harsh environmental conditions e.g. extreme salt concentrations and osmotic pressure in its natural habitat (Gill *et al*., 2005; Zhang *et al*., 2003). The infection normally starts when the bacterium finds its way to the patient from skin during implantation of medical device. The use of such implanted medical devices (IMDs) has been increased due to certain medical health conditions, which has ultimately resulted in increased rate of infection (O'Grady *et al*., 2002).

Antibiotic susceptible *S. epidermidis,* that are part of normal skin flora, are replaced by antibiotic resistant nosocomial strains of *S. epidermidis* upon admission of a patient to a hospital (Ahlstrand *et al*., 2011; Rohde *et al*., 2004).These infection causing strains can be transmitted from patient to patient. Health care providers, dealing with patients, might be a

cause of spread from one unit to the other, one hospital to another and even from one country to another country (Kozitskaya *et al*., 2005; Widerström *et al*., 2012).

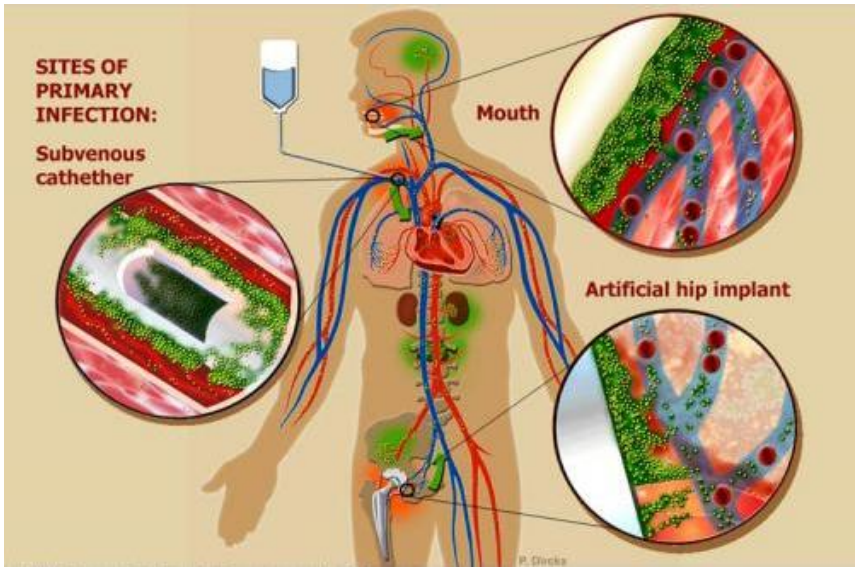

Figure 1: Sites of infections in which biofilm formation is of relevance to pathogenesis.

[https://www.biofilm.montana.edu/files/CBE/images/CBE03_1n2infect(1).preview.jpg]

1.3. Biofilm Formation

Every bacterial species forms biofilm by passing through four main steps- primary attachment (adhesion), dispersion (accumulation), maturation and liberation or detachment (McCann *et al*., 2008). These stages are regulated by different proteins, polysaccharides and lipids. The availability of nutrients, surrounding medium and various other factors like pH, oxygen level and temperature play key roles in biofilm formation (Fey and Olson, 2010). When a medical device is implanted into the body of a patient, surface conditioning of the implant takes place by the macromolecular proteinaceous and polysaccharide containing components in the body fluids such as blood, urine, saliva, mucous etc. and a conditioning film is formed by adsorption of these elements onto the device (Choong and Whitfield, 2000).

The chemistry of the abiotic surface of the device also plays a major role in adhesion of the substances, because each attachment protein has its own specificity towards a specific surface, for instance, the small basic protein (Sbp) expressed by *S. epidermidis* 1457 tends to bind to glass and the surfaces made by polystyrene, while fibronectin is an important adherent for extracellular matrix binding protein (Embp) produced by *S. epidermidis* 1585 (Patel *et al.*, 2012). *S. epidermidis* also attaches to the unmodified native surfaces and a number of unspecific chemo-physical factors, such as van der Waals forces, steric hindrances, surface tension, electrostatic and hydrophobic forces are involved in attaching the bacteria to native surfaces (Dunne, 2002). A staphylococcal protein autolysin E, AtlE, plays a vital role in adhering bacteria to both conditioned and unconditioned polymer surfaces (Rohde *et al*., 2006)and attaches to vitronectin and polystyrene (Heilmann *et al*., 1996). A reduced virulence has been observed by the strains lacking AtlE (Patel *et al*., 2012).

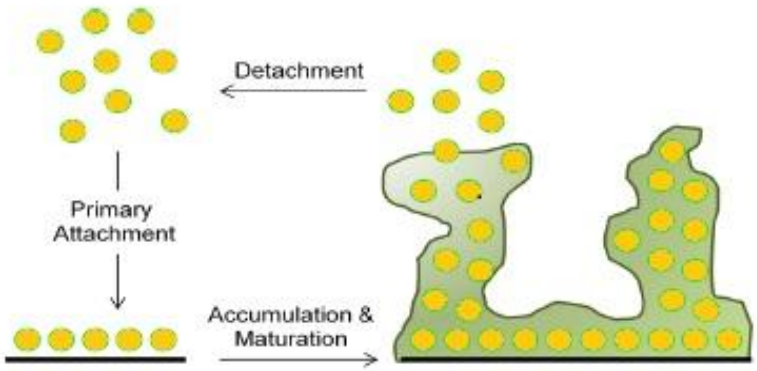

Figure 2: Schematic organization of *S. epidermidis* biofilm formation. Multiple factors induce primary attachment of bacterial cells to the biotic or abiotic surface and then the expression of polysaccharides and proteins take place resulting accumulation and maturation phases of biofilm formation. Some of the cells detach from biofilms naturally and start the process again after colonizing new surfaces (Rohde *et al*., 2006).

A polysaccharide responsible for primary attachment of bacterial cells to the surface in *S. epidermidis* is polysaccharide intercellular adhesin (PIA), which is an N- succinylated β- 1,6 linked polyglucosamine (Mack D, 1994). The biosynthesis machinery necessary for PIA production is encoded by the *icaADBC* operon (Gerke *et al*., 1998; Heilmann *et al*., 1996). PIA mediated intercellular adhesion and its production ultimately leads to aggregation of

bacteria during maturation phase (Patel *et al*., 2012). Additionally, the other non-protein molecule, teichoic acid, which is found in cell wall interacts with fibronectin associated surfaces and hence participates in the attachment process of bacteria to the surface to start biofilm formation (Hussain *et al*., 2001). After adherence of bacterial cells to the surface, the accumulation phase of the biofilm formation takes place, in which bacteria proliferate, inter-bacterial adhesion of cells takes place and multilayered cell clusters are formed resulting in a complicated meshwork of aggregated bacteria on the surface (McCann *et al*., 2008). At this step, the expression levels of accumulation associated protein (Aap) and Bap homologue protein (Bhp) raises dramatically in addition to the expression of PIA (Patel *et al*., 2012). It results in a step towards the formation of a robust mature biofilm in which the importance of adhesive proteins in inevitable (Nakano *et al*., 1998; Patel *et al*., 2007). The bacterial strain which lack PIA and Aap, Embp plays its role in intercellular adhesion. However, a broad distribution of *icaADBC,* PIA and Aap in the clinically important population of *S. epidermidis* suggests, that these factors work cooperatively during accumulation phase of biofilm formation (Christner *et al*., 2012; Rohde *et al*., 2004; Rohde *et al*., 2007).

The maturation phase of biofilm assembly is characterized by the differentiation of cell layers and the construction of river like channels. A highly organized mushroom-like biofilm structure by the generation of slime glycocalyx encased with surface bound bacterial cells in a gelatinous matrix has also been observed (Dunne, 2002; Rohde *et al*., 2006). The river-like channels participate in bathing the cells by transporting fresh nutrients and medium to the cells through their river-like channels just like a mushroom does (Rohde *et al*., 2006). Finally, the mature cells detach the bacterial consortium to colonize a new surface point (Otto, 2014). The detachment of cells and formation of more infecting sites are of key importance for the prevalence of biofilm forming infections (Kong *et al*., 2006). Toxaemia linked to acute staphylococcal infections is also contributed by dissemination of bacterial cells from the mature consortium (Yarwood and Schlievert, 2003). The quorum sensing accessory gene regulatory system (Agr system) plays a vital role in the detachment process (Vuong *et al*., 2004; Yarwood and Schlievert, 2003). The Agr system additionally plays a role in the production of recognition of peptide based pheromones leading to the formation of surfactant-like δ-hemolysin, which enhance biofilm detachment. At the stationary phase,

accumulation associated protein supports in biofilm detachment process, while being down regulated. Additionally, AtlE is upregulated by the Agr system during cell detachment (Wang *et al*., 2011).

1.4. Factors involved in biofilm formation

The mechanisms involved in biofilm formation are complex. The biofilm entities in staphylococcal species are either protein or polysaccharide based, depending upon the genetic makeup of the organism (Götz, 2002; Rohde *et al*., 2007; Büttner *et al*., 2015). The extracellular matrix also contains extracellaular DNA (eDNA), which is supposed to be involved in early accumulation stage of biofilm formation (Büttner *et al*., 2015; Nakano *et al*., 1998).There is a need of a balanced expression of all the components involved in a mature biofilm formation. A small basic protein (Sbp), accumulation associated protein (Aap), polysaccharide intercellular adhesion (PIA) and the extracellular matrix binding protein (Embp) play inevitable roles in biofilm formation (Götz, 2002; Līduma *et al*., 2012; Macintosh *et al*., 2009; Patel *et al*., 2012; Reiter *et al*., 2014). An interesting finding is that biofilm formation seems to be a survival strategy of the cells, because the factors involved in biofilm formation are usually induced by stress including high levels of zinc, glucose, glucosamine, abrupt variability in temperature, pH change, high osmolarity and the presence of ethanol (Cerca *et al*., 2011; Conrady *et al*., 2008; Götz, 2002; Mack *et al*., 1992).

1.4.1. Polysaccharide intercellular adhesin (PIA)

Polysaccharide intercellular adhesin (PIA) and its structural homologs such as poly N-acetyl glucosamine (PNAG) in *Staphylococcus aureus* are of primary importance. Such carbohydrate based adhesins are not only found in staphylococcus but also in other bacterial organisms. The structures are encoded by orthologous *icaADBC* operons (Kaplan *et al*., 2004; Wang *et al*., 2004). This operon is carried by most of the food processing and clinical samples of *S. epidermidis* (Frebourg *et al*., 2000; Rohde *et al*., 2004). Each gene in the operon has its own specific function, which helps in biofilm formation (Vuong *et al*., 2004). IcaA and IcaD are found in the cytoplasmic membrane. IcaA acts as a catalytic enzyme representing N-acetylglucosaminyltransferase activity, while IcaD has been found to enhance

the activity of IcaA. IcaAD makes N-acetylglucosamine oligomers, which are essential for the synthesis of PIA. There is another gene *ica*C present in between *ica*A and *ica*D having the features of an integral membrane. The long chain PIA synthesis requires the i*ca*C gene as an essential element (Gerke *et al.*, 1998).

PIA constitutes the matrix in biofilm formation. It is a homoglycan (molecular weight >230 kDa) mainly consisting of approximately130 β-1, 6-linked-2-acetamido-2-deoxy-D-glucopyranosyl residues. Positive charges, which are significant for the adhesive properties of the molecule, are produced by the deacetylation of 15% residues of PIA. The rest of the molecule remains N-acetylated (Mack *et al*., 1996; Rohde *et al*., 2010). Ester linked succinates are also present in PIA, that introduce negative charges into the molecule additionally. The simultaneous presence of negative and positive charges makes the intercellular adhesive properties of PIA (Gerke *et al*., 1998; Vuong *et al*., 2004).

Figure 3: Synthesis of PIA depends on *icaADBC* operon. *IcaR* is a regulator at the upstream of the operon (Vuong *et al*., 2004). The ability of formation of biofilms depends on this operon by synthesizing PIA in PIA dependent biofilm formation strains, such as *S. epidermidis* M10 (Mullis and Falkinham, 2013).

In animal models of foreign material infection, PIA negative mutants were significantly less virulent as compared to isogenic wild type strains (Rupp *et al*., 1999; Mack *et al*., 2000) demonstrating that PIA is of significant importance for the establishment of a *S. epidermidis* biofilm infection.

1.4.2. Extracellular matrix binding protein (Embp)

Different bacterial species express extracellular matrix binding protein with slight variations, e.g. Emb in *Streptococcus defectivus* and Ebh in *S. aureus* (Christner *et al*., 2012; Clarke *et al*., 2002; Manganelli and van de Rijn, 1999). Embp in *S. epidermidis* is encoded by the

Embp gene and is 1.1 MDa in size. Embp expression is co-related to high osmotic stress (Linnes *et al*., 2013).

1.4.3. Accumulation associated protein (Aap)

Accumulation associated protein (Aap) is a thin cell wall anchored protein and extends 120nm away from the cell wall in the form of fibrils (Becker *et al*., 2014). It is expressed in protein as well as polysaccharide based biofilms (Conrady *et al*., 2008). Aap not only contributes to the first attachment of bacterial cells to the surfaces, but also plays its role in the accumulation phase of biofilm formation (Becker *et al*., 2014; Līduma *et al*., 2012; Schaeffer *et al*., 2015). The protein is encoded by the *aap* gene. The molecular size ranges from 140-220 kDa, which depends upon the number of domain B repeats (Rohde *et al*., 2007). It contains four regions named as domain A, L-type lectin domain, domain B and a LPXTG-motif containing C-terminal cell wall anchor. The N-terminal domain A repeat region consisting of 11 degenerate 16 aa-repeats, and is involved in primary attachment to abiotic surfaces (Becker *et al*., 2014; Schaeffer *et al*., 2015). Domain B is composed of a variable number of 5 to 17 nearly identical 128-aa repeats and terminates into a "half-repeat" and a "collagen-like" repeat (Rohde *et al*., 2007). Each repeat is composed of a 78 aa G5- and a 50 aa E domain Repeats are referred to as G5 domains (Conrady *et al*., 2013; Macintosh *et al*., 2009). The G5 domain consists of 6 β-strands and has 5 glycine residues conserved in each G5 domain. A 50 residue space region E is found in between two consecutive G5 domains. The region is also of similar sequence as G5 domain (Conrady *et al*., 2013). SasG is an *S. aureus* homologue of Aap in *S. epidermidis*. Both have similar structures and are supposed to have similar functions as adhesive components in biofilm formation (Schaeffer *et al*., 2015). G5.E repeats of Aap and SasG oligomerize and form a twisted rope-like structure in the presence of Zn^{2+} (Conrady *et al*., 2008; Conrady *et al*., 2013).

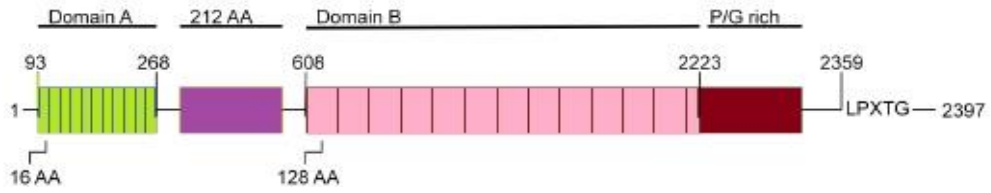

Figure 4: Schematic representation of Aap. Aap domain A, a site of 212 amino acids for proteolytic processing and domain B consisting of G repeats. The cell wall anchor motif consists of a G rich region and a LPXTG motif at the C-terminus (Conrady *et al*., 2008, modified).

An L-type lectin domain consisting of 212 amino acids is present between domain A and domain B, which serves as a cleavage site for proteolytic processes (Rohde *et al*., 2015; Paharik *et al*., 2016). Aap induces biofilm formation after removal of its domain A by proteolytic cleavage by metalloprotease SepA (Rohde *et al*., 2005). An LPXTG motif at the C-terminal behind domain B is associated with the covalent attachment of Aap to the surface of bacteria (Becker *et al*., 2014; Conrady *et al*., 2008; Schaeffer *et al*., 2015).

1.4.4. Small Basic Protein (Sbp)

An 18 kDa protein with its basic isoelectric point (pI=9.8) is referred as small basic protein (Sbp). It is encoded by 513 nucleotides. It is an extracellular protein that plays a pivotal role in forming a highly ordered architecture of the biofilm matrix (Decker *et al*., 2015). Sbp makes a biofilm scaffold, which provides support for cell to cell adhesion and aggregation for PIA and Aap dependent biofilm formation. The direct induction of cell aggregation in biofilm by Sbp has not been observed keeping the structural and functional organization of Sbp onto biofilm in consideration. Hence, Sbp has significant relevance with the protein (Aap) as well as polysaccharide (PIA) and is one of the significant co-factors required for fostering intercellular adhesins (Decker *et al*., 2015). The strains in which Aap, PIA or Embp have been studies as intercellular adhesins show, that the spatial distribution of these components is different in biofilms (Schommer *et al*., 2011). The predominantly cell surface localized proteinaceous intercellular adhesins (Aap and Embp) are also found in the intercellular matrix in rather small amounts (Decker *et al*., 2015).

A distinct spatial distribution pattern of Sbp localization within the extracellular matrix and accumulation at biofim-surface interface has been observed (Decker *et al*., 2015). It has been speculated that Sbp primes the surfaces of implants during colonization of cells during biofilm formation by *S. epidermidis* and promotes the stabilization of tethering of mature biofilms on surfaces e.g. polystyrene (Decker *et al*., 2015). Sbp and domain B of Aap co-localize on living biofilms. Sbp is necessary for the biofilm formation which is mediated by Aap domain B. Besides the direct involvement of Sbp in Aap mediated biofilm formation, it could also impart an impact through indirect modalities e.g. by recruiting additional and yet unknown factors to the cell surface, which could act as ligands for Aap.

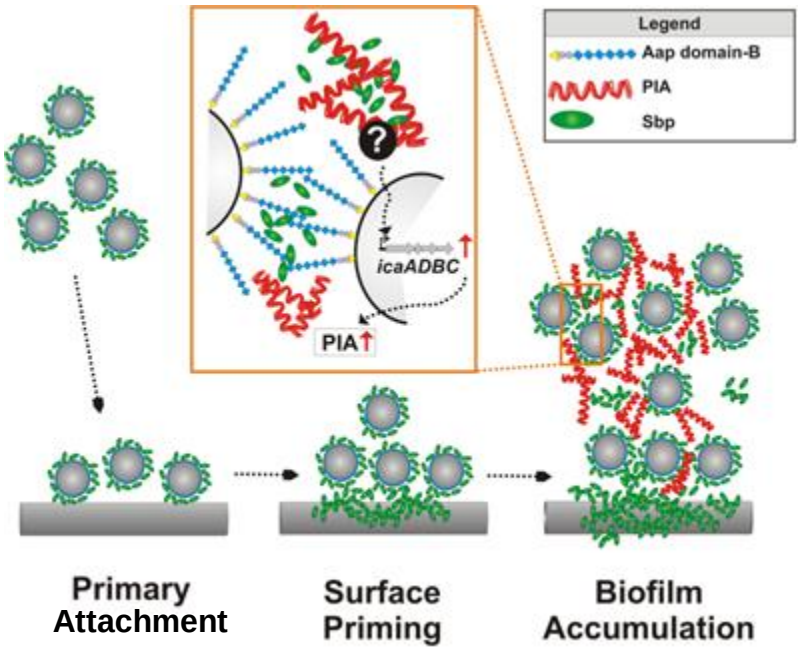

Figure 5: Model of the functioning of Sbp in *S. epidermidis* based biofilm formation. *S. epidermidis* cells that are free-floating having Sbp bound on their surfaces attached to artificial surfaces resulting in the localization of Sbp at the bacterial-substrate interface. The stability of cells along the foreign material interactions and adherence in the accumulation step require the deposition of Sbp as surface priming process. Surface priming and accumulation steps take place more or less simultaneously. The formed biofilm extracellular matrix has Sbp co-localized with PIA. The biofilm mediated by PIA depends upon the presence of Sbp, which stimulates (unknown mechanism so far) *icaADBC* transcription and subsequent release of PIA. Additionally, Aap domain-B dependent

bacterial aggregation requires Sbp as necessary factor, most probably through molecular interactions. Some other factors, fibronectin and lipids, involved in the biofilm formation cannot be ignored here (Decker *et al.*, 2015).

Sbp plays a vital role in PIA mediated biofilm formation as well (Decker *et al*., 2015). The integration of cells into the biofilm consortium by PIA takes place as a result of its ionic interactions with charged cell surfaces (Vergara-Irigaray *et al*., 2008). A homologue of PIA in *E. coli* called PNAG interacts with lipopolysaccharides (LPS) (Amini *et al*., 2009). Sbp provides potential charged character to the surface to which PIA binds, as a result, Sbp is the possible receptor for PIA (Decker *et al*., 2015). It is most likely that in the strain with the mutant Sbp, the PIA production becomes reduced due to the down regulation of *icaADBC* expression (Decker *et al*., 2015). To understand the involvement of Sbp, along with other factors, in signaling the stimulation of biofilm formation will be of key importance in the future.

1.5. Therapeutic options, antibiotic treatment and resistance

It is highly challenging to disrupt a once matured biofilm on an implanted medical device (IMD) and in most of the cases, the removal of the implant is the only choice to get rid of infection, which usually needs a second surgery and patients are again prone to risks of other infections (Rohde *et al*., 2006). The disruption of bacterial cells through antimicrobial agents, phagocytosis or antibiotics provides protection against biofilm formation. Normally, macrophages and neutrophils in the human immune system disrupt the planktonic bacterial cells through phagocytosis (Schell *et al.*, 2014). The antimicrobial characteristic of neutrophils is defined by defensins, lysozymes, cathelicidins and reactive oxygen species whereas,nitrogen species and pro-inflammatory cytokines are produced by macrophages (Gruszka*et al*., 2015; Scherr *et al*., 2014). Lytic enzymes are released by dying phagocytes to kill the planktonic cells (Scherr *et al*., 2014). However, the organization of bacterial cells into biofilm results in the ineffectiveness of all the above strategies. Moreover, the serum proteins of the patient e.g. fibrinogen, fibronectin, collagen, vitronectin, thrombospondin, bone sialoprotein and elastin provide considerable support to bind to the extracellular binding proteins of the micro-organisms e.g. Embp in *S. epidermidis*, Emb in *Streptococcus*

defectivus and Ebh in *S. aureus* (Christner *et al*., 2010; Clarke *et al*., 2002; Linnes *et al*., 2013; Manganelli and van de Rijn, 1999). Collagen is one of the substances that support the attachment of biofilm. Bacterial compounds activate macrophages, which lead to the production of urea and ornithine in higher amounts that ultimately results in tissue remodeling and collagen formation instead of phagocytosis (Scherr *et al*., 2014). Biofilms support the survival of cells within the consortium even in the presence of antimicrobial peptides. The effector mechanisms of the host immune system are also not very effective against such kind of organized cells (Foster, 2005; Vuong *et al*., 2004). Antibiotics, antimicrobials and disinfectants also face resistance due to the production of Aap, PIA and Embp (Ganeshnarayan *et al*., 2009; Knobloch *et al*., 2002). The disruption of a biofilm matrix is therefore important to expose the cells for phagocytic killing, disinfectants and antimicrobials (Banat *et al*., 2014; Ganeshnarayan *et al*., 2009; Schaeffer *et al*., 2015).

1.6. Advanced methods in structural biology

With the discovery of X-rays, the method of crystallography was established to obtain and observe the diffraction patterns of X-rays, which have been diffracted at the crystal lattice of different materials. X-ray crystallography provides insights, which have revolutionized the understanding of the structures of different biomolecules ranging from viruses, DNA, RNA and pharmaceutical materials to proteins, but it requires the growth of a good quality crystals. Recently, a number of new approaches to further revolutionize crystallography have been established e.g. in vivo crystallization and serial femtosecond crystallography (SFX). The later approach is used to record high resolution diffraction data sets from many small crystals in low μm size range by merging the data to three dimensional dataset afterwards (Boutet *et al*., 2012).

X-ray crystallography is now routinely applied to analyze the binding of pharmaceutical drugs to potential target proteins. This can also provide insights for introducing promising changes that possibly can improve the functions of potential drugs. The 3D-structure of a protein also answers the question of the particular function exhibited by that protein. In recent years, much advancement has been achieved in solving structures of proteins, which are involved in disease development, or contribute in performing a vital function in an

organism. In this way, small molecules, which can potentially be targets against defined parts of proteins, can be screened. These molecules can inhibit proteins and, thereby, potentially help to stop a disease progression. Consequently, these molecules can be used in clinical trials, to evaluate their significance and can be considered as potential drugs.

In order to study a protein in terms of its shape, size, variability in different conditions, such as different temperature, pH and flexibility properties in the presence and absence of a potential ligand, solution scattering techniques e.g. small angle X-ray scattering (SAXS) have been developed at 3rd generation synchrotron sources of PETRA III, DESY, Hamburg. The intensities of scattered X-rays from the protein molecules in solution are recorded at small scattered angles. In this context, for example, the conformational changes induced by sugar in the structural organization of botulinum toxin have already been studied by SAXS (Sagane *et al*., 2013). Moreover, the oligomerization of *E. coli* DnaB/C "helicase loader" complex has also been studied by this method (Arias-Palomo *et al*., 2013).

Another method that has caught more attention recently to study biomolecules and their oligomeric states is native mass spectrometry. The basic principle of native mass spectrometry is electrospray ionization (ESI) and its discovery was awarded the Nobel Prize in 2002. The mass of a huge tobacco mosaic virus (TMV) by electrospray ionization with its subsequent conversion into gaseous phase, without affecting its infectivity, has been determined with high accuracy (Siuzdak *et al*., 1996).

Cryo-electron microscopy (cryo-EM) is another advanced technology to study the cell architecture and proteins at molecular resolution (Milne *et al*., 2013). Imaging of biological structures using electrons was first demonstrated with the bacteriorhodopsin structure determination at ~ 7 Å resolution in 1975 (Henderson and Unwin, 1975). This study paved the way to determine the atomic resolution models of icosahedral viruses (Zhang *et al*., 2010). Moreover, cryo-EM generated density maps are routinely combined with other technologies, such as nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography to obtain high atomic resolution models for complex molecular assemblies.

2. AIM OF THE WORK

Hospital acquired infections are becoming life-threatening in high risk populations, including immuno-compromised and old patients. Device associated infections contribute to a major part in such kinds of infections.80 % of the device related infections are caused by skin commensals (the most notable *Staphylococcus epidermidis*). Along with other molecules, Sbp and Aap are two important proteins involved in *S. epidermidis* based biofilm formation.

Aim 1

- The first part of this work was focused on the structural elucidations of Sbp including its stabilization, analysis of its behavior under different conditions, secondary structure and shape determination by applying biochemical methods and complementary techniques such as expression, purification, DLS, CD spectrometry, SAXS and native mass spectrometry.

Aim 2

- The second part of the work was emphasized to investigate the structural basis of G5 subdomain of Aap domain B to analyze the interaction between Aap and Sbp. As it had been found out that Sbp helps in Aap domain B mediated biofilm formation, it was assumed that there are some molecular interactions between these two proteins. It was worth to gain information and structural insights into the interactions. For this, three G5 constructs of Aap domain B were cloned, expressed and characterized.

Besides this, data to be obtained should contribute to the goal of discovering new therapeutic agent against nosocomial infections.

3. MATERIALS AND METHODS

3.1. Materials

3.1.1. Devices

Table 1: List of the devices used in this work.

Acrylamide gel chamber	BIO-Radmini protean-Tetra System
Agarose gel chamber	MWG. Biotech. Electrophoresis constant power supply ECPS
Agarose gel imaging system	BIO-RAD Quantity one-4.5.2 (Basic) 3000/150
ÄKTA FPLC purification system	ÄKTA Purifier P-901; GE System No. 1282332 Made in Sweden
Balance	Kern & Sohn GmbH Germany, Kern PCB
CD spectrometer	J-815 (Jasco, UK)
Centrifuges	Multifuge IS-R, Sorval ^{R} RC 26 PLUS, SIGMA ^{R} 3-18K
Crystal imaging system/device	CrystalScore (Diversified Scientific Inc., USA), microscope SZX12 with camera DP10 (both Olympus, Japan)
Crystal plate incubator	RUMED 3001 (Rubarth, Germany) incubators
DLS instrument	Xtal concepts Spectrosize ^{TM} 300
Electrophoresis power supply	BIO-RAD power PAC 1000
Freezer (-20 °C and -80 °C)	Liebherr Profiline, Herafreeze BASIC Thermoscientific
Hot-plate magnetic stirrer	Grhardt, IK MAG ^{R} RCT
Incubators	CERTOMAT ^{R} BS-1 Sertorius stedim Biotech. INFORS HT Multitorn
Micropipettes	Gilson Made in France
Microwave	Bosch
Nanodrop	PeqLab
Orbital shaker	Edmund Buhler Labortechnik Materialtechnik Johanna Otto GmbH Germany
PCR machines	Peq Lab Biotechnologie GmbH
Pipetting robots	Honeybee 961 (Zinsser Analytic Gmbh, Germany), Oryx 4 (Douglas, UK)

pH meter	HANNA instrument, HI pH/ORP meter
Sonifier	Branson digital sonifier
Spectrophotometer	Bio-Rad Smartspec ™ 3000

3.1.2. Consumables

All the plastic consumables including reaction tubes, syringes and pipette tips were obtained from Sarstedt (Germany).

3.1.3. Expression vectors

Two different vectors pDEST17 and pET 302 NT-His (Invitrogen) were used in this study.

(1) pET 302 NT-His vector

The pET 302 NT-His contains T7*lac* promoter to promote high level expression of the gene of interest in *E. coli*. An N-terminal 6 x His-tag is present to detect and purify protein. It also has an ampicillin resistance marker gene for selection in *E. coli*.

(2) pDEST17 vector

The pDEST17 is an N-terminal fusion vector, which contains an initiation codon ATG upstream of 6 x His-tag. In order to ensure a proper initiation of translation in *E. coli*, a Shine-Dalgarno RBS sequence is included upstream of ATG.

3.1.4. Amino acid composition of proteins

Table 2: List of amino acid sequences of the proteins.

Amino Acid sequences of proteins (N-C terminal)	
Sbp	N N V E A A T G N S M K T V Q Q L N K G D K S L E N V K I G E S M K S V L K K Y S H P I Y S Y N P N S N E K Y Y E F R T D K G V L L V T A N G K K E R G N V T R V S M T Y N N A N G P S Y K A V K Q Q L G H K A I S R V H Y N N V T G N F G Y I Q K G Q A S Y Q F S S N S P K D K N V K L Y R I D L N K
Aap $G5_1EG5_2EG5_3$	V D G D P I I S T K E I P F N K K R E F D P N L A P G T E K V V Q K G E P G I E T T T T P T Y V N P N T G E K V G E G E P T E K I T K Q P V D E I V H Y G G E E I K P G H K D E F D P N A P K G S Q T T Q P G K P G V K N P D T G E V V T P P V D D V T K Y G P V D G D P I T S T E E I P F D K K R E F N P D L K P G E E R V K Q K G E P G T K T I T T P T T K N P L T G E K V G E G E P

	T E K I T K Q P V D E I T E Y G G E E I K P G H K D E F D P N A P K G S Q E D V P G K P G V K N P D T G E V V T P P V D D V T K Y G P V D G D P I T S T E E I P F D K K R E F N P D L K P G E E R V K Q K G E P G T K T I T T P T T K N P L T G E K V G E G E P T E K V T
Aap $G5_1EG5_2$	V D G D P I I S T K E I P F N K K R E F D P N L A P G T E K V V Q K G E P G I E T T T T P T Y V N P N T G E K V G E G E P T E K I T K Q P V D E I V H Y G G E E I K P G H K D E F D P N A P K G S Q T T Q P G K P G V K N P D T G E V V T P P V D D V T K Y G P V D G D P I T S T E E I P F D K K R E F N P D L K P G E E R V K Q K G E P G T K T I T T P T T K N P L T G E K V G E G E P T E K I T
Aap $G5_1E$	V D G D P I I S T K E I P F N K K R E F D P N L A P G T E K V V Q K G E P G I E T T T T P T Y V N P N T G E K V G E G E P T E K I T K Q P V D E I V H Y G G E E I K P G H K D E F D P N A P K G S Q T T Q P G K P G V K N P D T G E V V T P P V D D V T K Y G P

3.1.5. Primers for cloning

Table 3: Primer sequences for cloning of respective genes.

Target Protein	**Type**	**Sequence (5'-3')**
Sbp	Forward	CACGTGAATTCGGAAAACCTGTATTTTCAGGGCAACAACGTTGAAGCGGC
	Reverse	AGCCGGATCCGATTATTTATTTAAGTCTATACGATA
aap G5 constructs	Forward	ATGCATCATCATCATCATCACGTGAATTCGGAAAACTGTATTTTCAGGGCATGGTGGATGGTGACCCG
aap $G5_1EG5_2EG5_3$	Reverse	CATCGATCTCGAGCGAATTCTTATTAGGTAACTTTCTCCGTCGG
aap $G5_1EG5_2$	Reverse	CATCGATCTCGAGCGAATTCTTAGGTGATCTTTTCAGTCGGTTC
aap $G5_1E$	Reverse	CATCGATCTCGAGCGAATTCTTAAGGCCCATATTTCGTCACG

3.1.6. Restriction enzymes

Table 4: Restriction enzymes used

Enzyme	Buffer	Supplier
EcoRI-HF	Cutsmart	New England BioLabs
BamHI-HF	Cutsmart	New England BioLabs

3.1.7. Bacterial strains and competent cells

Table 5: List of bacterial strains used.

Strain	Properties	Resistance/selection marker	Suppliers
E. coli One shot BL21Star™(DE_3)	Transformation efficiency: 1–5 x 107 cfu/µg pUC19 DNA	Ampicillin	Invitrogen Life Technologies
One shot Top10 cells		Kanamycin	Invitrogen Life Technologies
E. coli BL21AI	High yield of protein	Ampicillin	Invitrogen Life Technologies
Staphylococcus epidermidis 1457	Wild type,*aap* positive, *sbp* positive	_	AG Rohde, UKE, Hamburg

3.1.8. Antibodies

Table 6: List of antibodies.

Antibody	Detection	Dilution	Supplier
Monoclonal Anti-polyhistidine-Alkaline Phosphatase, mouse	His-tag	1:10,000	Sigma-Aldrich (St. Louis, USA)
Anti-mouse IgG Horseradish peroxidase (HRP) labeled (in sheep)	His-tag	1:10,000	GE Healthcare UK
Rabbit anti-rSbp serum (91435	Sbp	1:10,000	AG Rohde, UKE, Hamburg

Rabbit anti-rDomB serum	Aap Domain B	1:10,000	AG Rohde, UKE, Hamburg
Anti-rabbit IgG HRP labeled (in goat)	Sbp and Aap	1:10,000	Sigma-Aldrich

3.1.9. Buffers and solution

All buffers and solutions were prepared in double distilled H_2O.

Table 7: Protein purification buffers.

For Sbp	
Lysis, column and wash buffer	20 mM $NaPO_4$, 500 mM NaCl, 20 mM Imidazole, pH 7.4 protease inhibitor cocktail (PIC) 1 tablet/ 100 ml
Elution buffer	20 mM $NaPO_4$, 500 mM NaCl, 250 mM Imidazole, pH 7.4
Buffer for cleavage of His-tag	20 mM $NaPO_4$, 500 mM NaCl, pH 7.4
Gel filtration buffer	50 mM $NaPO_4$,.150 mM NaF, pH 7.4
CD buffer	20 mM $NaPO_4$,.100 mM NaF, pH 7.4
DLS and SAXS	50 mM $NaPO_4$,.150 mM NaF, pH 7.4
For three G5 protein constructs of Aap	
Lysis, column and washing buffer	20 mM $NaPO_4$, 500 mM NaCl, 20 mM Imidazole, pH 7.4 PIC 1 tablet/ 100 ml
Elution buffer	20 mM $NaPO_4$,500 mM NaCl, 250 mM Imidazole, pH 7.4
Buffer for cleavage of Tag	20 mM $NaPO_4$, 500 mM NaCl
Gel filtration buffer	20 mM Tris, pH 7.0, 100 mM NaCl
CD, DLS and SAXS	20 mM Tris, pH 7.0, 100 mM NaCl

Table 8: Bacterial cell culture media.

Solution	**Composition**
LB-medium/ agar	1.0 % Bacto-Trypton, 0.5 % Bacto-Yeast-Extract, 1.0 % NaCl in ddH_2O, autoclaved.

	For LB-agar: 1.5 % agar in LB-medium
Selective antibiotic	100 mg/ml ampicillin in 10 ml ddH_2O; stock solution
IPTG	1 M in dH_2O, autoclaved; stock solution
Arabinose	20 g in 100ml dH_2O
SOC media	20 g trypton, 0.6 g NaCl, 5.0 g yeast extract, 0.2 g KCl, 10 mM $MgCl_2.6H_2O$, 10 mM $MgCl_2.7H_2O$, 20 mM glucose, ad 1000ml ddH_2O, pH 7.0

Both media were autoclaved for 20 min. at 121 °C.

Table 9: Agarose gel electrophoresis preparation.

Solution	**Composition / Suppliers**
1.2 % Agarose	1.2 g in 100ml in ddH_2O
DNA Loading dye	0.05 % Bromophenol Blue, 0.25 % Xylene cyanol, 1 mM EDTA, 50 % glycerol
Nucleic acid staining solution	Red safe ™ iNtRON Biotechnology
1×TAE, agarose gel electrode buffer	80 mM Tris, 40 mM Acetic Acid, 2 mM EDTA
DNA marker	GeneRuler ™Thermofischer scientific

Table 10: SDS PAGE preparation

Solution	**Composition / Supplier**
Sample loading buffer	4 x LDS NuPAGE novex life technologies
Electrode buffer diluted to 1 x for final use	10 x NuPAGE MES buffer, novex life technologies
Resolving gel buffer	1.5 M Tris-HCl, pH 8.8
Stacking gel buffer	0.5 M Tris-HCl, pH 6.8
SDS buffer	10 % (w/v) in ddH_2O
Acrylamide	1.6 ml in stacking gel, 5 ml in resolving gel for 15 % gel
APS	50 µl of 10 % (w/v) in ddH_2O

TEMED	5 µl in resolving gel and 10 µl in stacking gel
Staining solution	1.0 % (w/v) coomassie blue R 250, 50 % ddH_2O, 40 % (v/v) methanol and 10 % (v/v) acetic acid
Destaining solution	50 % ddH_2O, 40 % (v/v) methanol and 10 % (v/v) acetic acid
Protein marker (SDS PAGE)	PageRuler Plus Prestained protein ladder Thermoscientific(10 to 250 kDa)
Protein marker (SDS PAGE)	PageRuler Prestained protein ladder (10 to 180 kDa)

Table 11: Ni Sepharose HisTrap FF column by GE Healthcare Life Sciences regeneration buffers.

Solution	**Composition**
Stripping buffer	20 mM $NaPO_4$, 0.5 mM NaCl, 50mM EDTA, pH 7.4
Packing solution	0.1 M $ZnCl_2$ and 0.1 M $NiSO_4$ in case of regeneration with Ni^{2+}

Table 12: Native PAGE preparation.

Native gel	**Suppliers**
Native PAGE marker	NativeMark unstained protein ladder life technologies
Native cathode, anode and running buffers	Novex Life technologies
LDS sample buffer (4 x)	Novex Life technologies

Table 13: Western blot preparation.

Solutions	**Compositions / Suppliers**
20 xTransfer buffer	28.8 g glycine , 6.04 g tris base, 200 ml methanol, 1.6 l ddH_2O, pH 7.2
1 x Transfer buffer	50 ml methanol, 25 ml 20 x transfer buffer, 425 ml ddH_2O,
PBS (Phosphate buffered saline)	140 mM NaCl , 25 mM KCl, 0.5 mM $MgCl_2$, 1 mM $CaCl_2$, 10 mM Na_2HPO_4 (pH 7.5)

PBST	1 x PBS + 0.05 % (v/v) Tween 20
Blocking Solution	4 % (w/v) BSA (Bovine serum albumin) in PBS
Transfer membrane	PVDF Immobilien-P Transfer membrane, pore size 0.45 µm
ECL western blot detection reagents	GE Healthcare (Buckinghamshire, UK)
Super RX medical X-ray films	Fujifilm, Tokyo, Japan

3.2. Molecular biology methods

3.2.1. Cloning

3.2.1.1. Polymerase Chain Reaction (PCR)

A widely applied method to amplify a gene sequence from a single copy into a large number of copies is termed as polymerase chain reaction (PCR). The specific sense and antisense primers having an N-terminus 6 x His-tag site and a TEV protease recognition cleavage site were synthesized by eurofins MWG operon (Hamburg, Germany). In order to check the annealing temperature of primers to bind with the gene of interest, gradient PCR was performed over a range of annealing temperatures from 68 °C to 74 °C. After confirming the annealing temperature, DNA sequence of the required gene was amplified by using gene specific primers and high fidelity Phusion polymerase (Thermofischer). The primers were diluted to a final concentration of 10 pM from the original stock solution to carry out the reaction. For 0.5 µl (approx. 75 ng) of DNA template, 0.75 µl of each primer was used and reaction was carried out in a thermocycler (Peq Lab Biotechnologie GmBH). The amplification reaction in the thermocycler was set as the initial denaturation step for 45 sec at 96 °C followed by denaturation for 35 cycles of 45 sec at 96 °C each. The annealing temperature was set at 71 °C (based on the result of gradient PCR for specific oligonucleotides of primers), which was followed by final elongation for 45 sec at 72 °C. The components of the reaction for the final volume of 25 µl are listed below.

Table 14: The PCR reaction mixture.

Component	Amount
DNA template up to 100 ng	0.5 µl
5x HF-polymerase ($MgCl_2$) buffer	5.0 µl
dNTPs	0.5 µl
Forward primer	0.75 µl
Reverse primer	0.75 µl
HF- Phusion polymerase	0.25 µl
$H_2O_{\ inj}$	*ad.*25µl

3.2.1.2. Agarose gel electrophoresis

The separation of DNA fragments based on their respective sizes can be performed by agarose gel electrophoresis. The size of the fragment of negatively charged DNA, the electric power applied and the concentration of agarose gel determine the movement of a PCR product towards the anode. 1.2 % agarose gel was prepared by dissolving 1.2 g of agarose in 1 x TAE buffer supplemented with 5 µl red safe-dye (iNtRON Biotechnology) to visualize DNA fragments under UV-light. The samples (5 µl each) were applied to a gel after mixing them with 6 x loading dye (1 µl). To estimate the size of DNA fragments, a suitable marker was used as size reference. A constant voltage of 100 V (Electrophoresis constant power supply ECPS) was applied through the buffer in an agarose gel chamber (MWG. Biotech.) and the gel tray was placed in appropriate direction in the chamber. After running the gel, the PCR products were visualized under a UV lamp in an agarose gel imaging system (BIO-RAD Quantity one-4.5.2 (Basic) 3000/150).

3.2.1.3. DNA purification

The cleaning of the PCR product from salts, DNA fragments and gel was done using a PCR clean-up NucleoSpin Extract II Kit (Macherey Nagel, Düren, Germany), according to manufacturer's protocol. The portion of the gel carrying the required DNA fragment was cut with a sharp scalpel and cleaning up was done. In order to avoid the damage of DNA caused by UV exposure, only a single well of the gel was exposed to UV light to determine the exact position of amplified product on the gel. In the last step of cleaning DNA was eluted in 50 µl of H_2O*inj*.

3.2.1.4. Restriction digestion of plasmid and template

Restriction digestion of plasmid pET 302 NT-His and a DNA template was performed using BamHI-HF and EcoRI-HF (New England BioLabs) enzymes for 2 hrs at 37 °C. The following reaction mixture was prepared.

Table 15: The reaction composition of restriction digestion.

EcoRI-HF	1.5 µl

BamHI-HF	1.5 µl
Cutsmart buffer	5.0 µl
DNA (PCR product and plasmid separately)	5.0 µl

3.2.1.5. Ligation

The molar ratio of vector and insert DNA fragment was calculated by using the online Promega BioMath Calculator [https://www.promega.de/resources/tools/biomath-calculators] according to the size of insert and vector. The digested vector and DNA fragment were ligated with the help of T4 DNA ligase. The following reaction mixture of total 20 µl volume was incubated overnight at 16 °C for this purpose.

Table 16: Ligation reaction components and their amounts

pET 302-NT His	5.0 µl
Insert	1.0 µl
10 x T4 DNA ligase	2.0 µl
T4 DNA ligase	1.0 µl
Nuclease free H_2O	*ad.* 20µl

3.2.1.6. Transformation into Top 10 cells

The ligated reaction mixture was transformed into chemically competent one shot Top10 cells (Invitrogen). An aliquot of commercial competent cells was placed on ice and mixed with 1.5 µl of ligated mixture by pipetting up and down very gently. The mixture was placed on ice for 30 min. A heating block was preheated to 42 °C and cell-DNA mixture was incubated for 1 min at 42 °C. The sample was cooled down on ice for 1 min subsequently and 500 µl of sterile SOC-medium was added into it. The suspension was then incubated for 30 min at 37 °C with mild shaking. Cells (50 µl and 100 µl with SOC medium) were plated onto LB agar plates in a sterile environment. Plates, containing ampicillin antibiotic for selection, were placed at 37 °C for incubating overnight. Next day, colonies were picked and inoculated to grow in 5 ml LB medium at 37 °C with shaking at 180 rpm for further analysis.

3.2.1.7. Isolation and purification of plasmids

The potential Top10 cells, carrying the expected plasmid, were collected from the overnight 5 ml grown culture. Plasmids were isolated using Qiagen Plasmid Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's protocol. Nucleic acid concentrations of the isolated plasmids were determined using a nanodrop spectrophotometer (peqLab) measuring the absorption at 260 nm.

3.2.1.8. Colony PCR, restriction digestion and sequencing of cloned gene

In order to see cloning of the expected gene into plasmid, colony PCR and digestion reaction were performed using primers and restriction enzymes respectively (see section 3.2.1.1 and 3.2.1.3). To further investigate the success of cloning of the exact gene of interest without any mutations or insertions, plasmid DNA was sequenced based on Sanger method by MWG eurofins sequencing lab (Hamburg, Germany). Samples were prepared by mixing 1 µl of forward and reverse primers in 5 µl of DNA. A confirmed plasmid with successful cloned gene was further transformed into *E. coli* expression BL21 Star cells, according to heat shock transformation protocol, described in section 3.2.1.6.

3.2.2. Gibson cloning

The method has been developed by Dr. Daniel Gibson and his colleagues. This method assembles the multiple overlapping DNA fragments in a single tube reaction.A single buffer with three different enzymatic activities- the exonuclease activity creating single stranded 3' overhang to anneal the fragments, the polymerase activity and ligase activity to seal the nicks in assembled DNA- comprises the Gibson Assembly Master Mix (New England BioLabs). A double stranded fully sealed DNA molecule, capable of direct transformation, is the end result of this assembly. The primers (one forward and three reverse) were designed keeping a 6 x His tag site and a TEV protease cleavage site in consideration and got them synthesized from eurofins MWG operon (Hamburg, Germany). The annealing temperature was determined by setting gradient PCR as described in section 3.2.1.1. All the three genes were

amplified by PCR, as described in section 3.2.1.1 but with the exception of annealing temperature, which was set at 71 °C according to primer sequences and gradient PCR results.

3.2.2.1. Linearization of pET 302 NT-His Vector

The pET 302 NT-His vector was digested with high fidelity restriction enzyme EcoRI-HF (New England BioLabs). 2 µl of enzyme were mixed with 50 µl of plasmid in the presence of 5 µl of cutsmart buffer. The reaction was allowed to stay at 37 °C for 2 hrs to complete digestion. The concentrations of a digested vector and amplified PCR products were determined using Nanodrop spectrophotometer (peqLab).

3.2.2.2. Assembly Protocol

The amount of each gene to vector ratio, which will probably be suitable for a ligation reaction, was calculated by applying the online molar ratio calculator tool in three different ratios (1:1, 2:1 and 3:1). The assembly of the gene fragments into a vector was performed according to the Gibson Assembly Cloning Kit's (New England BioLabs) protocol.

Table 17: Gibson assembly reaction mixtures.

Amplified $G5_1E$ PCR product	2 µl (200 ng)	_	_
$G5_1EG5_2$	_	3 µl (150 ng)	_
$G5_1EG5_2$ $EG5_3$	_	_	2 µl (180 ng)
Linearized vector	1 µl (400 ng)	1 µl (400 ng)	1 µl (400 ng)
Gibson assembly master mix (2X)	10 µl	10 µl	10 µl
Deionized H_2O	7 µl	6 µl	7 µl
Total volume	20 µl	20 µl	20 µl

Each of the three reaction mixtures was incubated at 50 °C in a thermocycler for 15 min. Samples were stored on ice for subsequent transformation.

3.2.2.3. Heat shock transformation, plasmid isolation

Transformation of plasmids carrying the respective gene in the form of an above explained assembly mixture was done in one shot top10 cells according to the protocol described in section 3.2.1.6. The clones were grown in LB medium to isolate plasmid by Qiagen Plasmid Mini Kit. DNA sequencing was done by MWG eurofins lab (Hamburg, Germany) using the respective primers. The plasmids with successful clones were transformed into *E.coli* BL21 Star expression system for further purpose.

3.3. Biochemical methods

3.3.1. Test expression of recombinant proteins

A culture of 5 ml LB-medium containing the appropriate antibiotic (ampicillin 1:1000) was inoculated with a single colony from the transformed *E. coli* BL21 Star cells and incubated over-night at 37 °C under continuous shaking at 180 rpm. A negative control was also prepared at the same time. Next day, three flasks with 50 ml LB-medium containing the right proportion (1:1000) of ampicillin were inoculated with 500 µl of overnight grown culture each and incubated at 37 °C with shaking at 180 rpm until the OD_{600} value reached 0.4 determined by a spectrophotometer. 1 ml of growth culture was collected in an Eppendorf tube from one of the flasks for electrophoretic analysis and divided the grown culture of that particular flask in two equal parts. In one part, recombinant protein expression was induced by adding 1 mM IPTG and 1 ml culture sample was taken after 2 hrs, 3 hrs, 3.5 hrs and 4 hrs incubation in four eppendorf tubes respectively. From the other 25 ml non-induced part of the culture, samples were taken in exactly the same way. 1 ml culture from each sample, which was already taken, was centrifuged at 13,000 rpm for 1 min. and saved the cell pellet at -80°C for protein expression analysis. The grown culture in two remaining flasks was also divided into two equal parts when OD_{600} value reached 0.7 and 1.0 respectively and induced those with 1 mM IPTG in the same way described above.

Note: Test expression for all recombinant proteins - Sbp, three constructs of Aap and a formerly cloned Sbp in pDEST 17 vector was performed the same way, as described above

except the fact that inducer for recombinant Sbp expression in pDEST17 vector in expression system BL21AI cells was 20 % L-Arabinose instead of 1 mM IPTG.

3.3.2. Large scale expression of proteins

The cells for large scale expression of the recombinant proteins were harvested after inducing the cells, which were grown in 6 L LB-media culture (in 1 L flask each), with 1 mM IPTG or 20 % L-Arabinose, depending upon the complete construct (Sbp/pET302 NT-His and Sbp/pDEST17). The induction was done, when OD_{600} value reached 0.7. The cells were harvested by centrifugation at 4 °C at 6000 x g for 15 min after 3.5 hrs post induction expression of the proteins in cells at 37 °C. The supernatant was discarded and the cell pellet was saved at -20 °C for further studies.

G5 construct expression cells were induced by 1 mM IPTG, when OD_{600} value reached 0.9. After induction, the growth temperature of cells was reduced to 20 °C and expression of protein was allowed to continue overnight with shaking at 180 rpm. The expressed cells were harvested by centrifugation at 4 °C at 6000 x g for 15 min. and stored at -20 °C for further research. 4 L LB-media culture was used for expression of each *aap* construct.

3.3.3. Cell lysis and affinity purification

The stored *E. coli* BL21 Star cells expressing the target protein were resuspended in lysis buffer (containing 20 mM $NaPO_4$, 500 mM NaCl & 40 mM imidazol, pH=7.4). To guarantee the maximum solubilization of the expressed protein, suspended cells were sonicated 14 times for 10 sec alongwith 20 sec resting incubation on ice between each step (Branson digital sonifier). The supernatant containing proteins along with the target protein was collected after centrifuging the cell lysate at 14,000 rpm for 25 min. In order to minimize the cell debris in the lysate, the supernatant was filtered through 0.1 µm filter membrane unit.

The full protein purification has been done at 4 °C consisting of three steps a) affinity chromatography, b) dialysis and c) size exclusion chromatography. Purification of the recombinant Sbp containing the polyhistidine (6 x) tag sequence was carried out using Zn^{2+} NTA affinity chromatography. An ÄKTA FPLC purification system (ÄKTA Purifier P-901;

GE Healthcare, UK) connected to a UV detector and a fraction collector was applied. An accompanying UNICORN software was supporting to create a method for a better purification plan by separating the steps of binding of protein to the column, washing and subsequent elution steps as well as the elution parameters. The purifier was also supported by the chromatogram evaluation in UNICORN software. A Zn^{2+} nitrilotriacetic acid agarose column (5 ml), HisTrap™ HP (GE Healthcare) was used. As the column was pre-packed with Ni^{2+}, it had to be regenerated with Zn^{2+} ions. Prior to incubating the cell lysate in the column for purification, the regenerated column with Zn^{2+} was equilibrated with binding buffer, elution buffer (to remove any already attached protein especially when the column was not freshly regenerated) and then again with binding buffer using 5 column volumes (CVs) each buffer. The stored supernatant (10 ml) was added to the column, the matrix resin was washed with 8-10 CVs of wash buffer to get rid of the unbounded proteins followed by an elution step, where the desired protein was eluted by increasing the amount of imidazole in fractions of 1 ml in falcon tubes. All the steps for purification were done using the Äkta purifier (P-901; GE Healthcare, UK) in a cold room at 4 °C to maintain the stability of protein.

3.3.4. Exchange of buffer and concentration of proteins

The affinity purified protein was dialyzed to change the buffer composition in between the purification steps with the storage buffer 50 mM $NaPO_4$, 200 mM at NaF pH=7.4 in case of Sbp using dialysis Slide-A-Lyzer™ G2 Dialysis Cassettes, 3.5 K MWCO, 70 ml. The purified protein was concentrated by centrifugation using Amicon Ultracentrifugal filter of 3,000 MCWO size at 3500 rpm. For Aap G5 constructs, the storage buffer was 20 mM Tris, pH 7.0, 100 mM NaCl.

3.3.5. Size exclusion chromatography

In order to further purify the target protein and separate the protein molecules on the basis of molecular size, gel exclusion chromatography was performed using a Superdex^{TM} 200 10/300GL column (GE Healthcare, Chalfont St. Giles, Great Britain) of 24 ml volume capacity. The column was first equilibrated with ddH_2O and then with the respective buffer

before running the target protein through it. 500 µl of concentrated protein suspension was applied onto the column. The purest form of protein was obtained in fractions of 400 µl by elution at the respective elution volume corresponding to the molecular weight of protein.

Size exclusion chromatography was also performed for interaction analysis of Sbp and Aap constructs. Purified tag free (see section 3.3.6.) Sbp and each Aap G5 construct (one construct in one experiment) were dialyzed against the same buffer 50 mM $NaPO_4$, 200 mM, 1 mM $ZnCl_2$, NaF pH=7.4 and mixed together as 1:1. The mixture was left to stand for 1-2 hrs at 4 °C and allowed it to pass through gel filtration Superdex^{TM} 200 10/300GL column to elute the complex, if formed. The respective proteins were collected as fractions of 400 µl in each eppendorf tube using ÄKTA Purifier (P-901; GE Healthcare, UK).

3.3.6. TEV protease digestion and separation of TEV cleaved proteins

The 6 x His-tag on each protein, which was used to bind with Ni^{2+} and Zn^{2+} resins during purification, was necessary to be removed from protein. The cleavage of the His-tag was achieved by incubating the protein sample with TEV protease at a molar ratio of 1:10 and 1:50. The cleavage reaction was performed at RT with mild shaking overnight. Samples were analyzed by SDS PAGE to ensure the progress of digestion of His-tag. The final purified tag free protein was collected as flow through, after passing through Ni-NTA matrix. His-tag was bound to the column matrix, while protein passed through the column and collected in the flow through.

3.3.7. Protein quantification

The concentrations of the purified proteins in their respective buffer solutions were determined by measuring the specific absorption of UV-light at a wavelength of 280 nm according to Lambert and Beer law using a Nanodrop spectrophotometer (peqLab). The formula is given below:

$$A_{280} = \varepsilon * c * d$$

Whereas, ε refers to a molar extinction coefficient (M^{-1} cm^{-1}), A_{280} represents specific absorption at 280 nm, d stands for diameter of the cuvette/ path length (cm) and c is the concentration of protein (mg/ml).

Elution buffer was used as a blank reference by applying 1.5 µl to the sensor arm of nanodrop. Afterwards, the protein sample was measured by adjusting the molecular weight and extinction co-efficient of each protein in the settings. The concentration determination was repeated and an average reading was calculated for each protein. Molar extinction co-efficient of protein is theoretically calculated by online ProtParam tool in Expasy (Gasteiger *et al*., 2005).

3.3.8. SDS Polyacrylamide Gel Electrophoresis (SDS PAGE)

SDS PAGE is a widely applied method in protein biochemistry to determine the purity and size of proteins under denaturing conditions. A discontinuous polyacrylamide gel (stacking and resolving) provides a support medium to separate proteins in an electric field, hence, named as electrophoresis. An anionic detergent sodium dodecyl sulfate (SDS) is used to denature proteins by maintaining a net negative charge within a wide pH range. The overall negative charge, imparted by SDS on polypeptides of protein in proportion to its relative mass, destroys most of the structure of protein and makes it linear. The negatively charged proteins are strongly attracted to anode terminal in an electrophoresis chamber. Polyacrylamide behaves as a sieve for protein molecules. The smaller molecules pass faster through the pores of gel, while bigger molecules migrate slowly. As polypeptides denatured by SDS have same charge-to-mass ratio hence, the separation of protein molecules is based on molecular weight.

After preparation of SDS gel, it was placed in a gel chamber vertically. Protein samples (18 µl each) were mixed with 6 µl of sample loading buffer (4 x LDS NuPAGE novex life technologies) and incubated at 100 °C for 10 min to be denatured. The samples were loaded onto the wells of a gel placed in 1 x MES buffer (NuPAGE MES buffer, novex life technologies) in the gel chamber. The gel chamber was connected to electrophoresis power

supply (BIO-RAD power PAC 1000). A power supply of 120 V was selected for the migration of protein samples through gel for a time period of 1.5 hrs.

For SDS PAGE analysis of Sbp, a 15 % SDS gel was prepared, while a 12 % SDS gel was used for the G5 constructs.

3.3.9. Native gel electrophoresis

The electrophoresis of protein molecules at their native state, keeping the folding states and net charges intact, is termed as native gel electrophoresis. The separation of proteins on the basis of their isoelectric points and hydrodynamic radii can be achieved by native gel electrophoresis.

Pre-cast NativePAGE ^{TM} Novex Bis-Tris Gels (4-16 % gradient gels) were obtained from novex life technologies for blue or clear native PAGE. The protein samples were mixed with 4 x native PAGE sample buffer (novex life technologies) according to manufacturer's protocol. The samples were loaded onto the wells of a gel, which was placed in native PAGE cathode buffer (novex life technologies). The anode buffer was poured into the side chambers of electrophoresis apparatus. In order to avoid the denaturation of protein samples, electrophoresis was carried out in cold room. The power for the migration of protein samples was increased from 50 V to 150 V, while passing the samples from stacking to resolving portion of gel. Moreover, a standard native PAGE marker (NativeMark unstained protein ladder life technologies) was used to approximate molecular weight sizes of protein samples.

3.3.10. Coomassie staining and destaining of SDS and native PAGE

To visualize protein bands on SDS and native gels, the resolving part of the gels were stained in the coomassie solution for 30 min. The destaining of the gels was done to make the bands clear and prominent in the destaining solution for 1 hr. The gels were stored in tap water afterwards. Documentation of the results was done by scanning the gels.

3.3.11. Western blot analysis

Western blotting was applied for immunological detection of purified Aap constructs. Proteins were separated on 12 % SDS PAGE gels and blotted over polyvinylidene difluoride (PVDF) membrane by the following method. For this, 4 sponge filters soaked in 1 x transfer buffer were place in a blotting chamber (Invitrogen). The following series of placement of gel along with blotting membrane was done on the top of sponge filters: one layer of filter paper, the SDS gel, one activated (activation for 15 sec in 100 % methanol and 15 sec washing in 1 x transfer buffer) PVDF membrane, another layer of filter paper and 4 soaked sponge filters. The set up was firmly packed in the blotting chamber. The chamber was filled with 1 x transfer buffer and blotting was performed for 1 hr at 30 V. The blocking of the blotted membrane was performed in 3 % BSA in 1 x PBS-buffer at 4 °C and shaking it overnight slowly. Next day, the membrane was washed in 1 x PBST solution for three times for 15 min (each wash). The primary antibody Rabbit anti-rDomB was diluted 1:10,000 in 1 x PBS-T to make final 10 ml solution. The membrane was incubated in this dilution for 1 hr at RT. Washing of the membrane was repeated for three times and incubation in the diluted (1:10,000 in 1 x PBS-T) secondary anti-rabbit peroxidase antibody was performed for 1 hr at RT with shaking. The washing step was repeated three times, as done previously. The membrane was placed on a thin plastic sheet and 1 ml of ECL-TM western blotting detection reagents were poured on it. The plastic sheet was carefully wrapped over the membrane to spread the reagent evenly without touching the membrane directly. The signal was detected by developing it on Super RX medical X-ray films after illuminating from 20 sec to 1 min.

3.4. Biophysical methods

3.4.1. Mass spectrometry based protein identification

Trypsin is commonly used to cleave a protein into peptides, which are used to analyze by liquid chromatography-tandem mass spectrometry (LC-MS/MS) for identification of proteins. The digestion of each protein construct by trypsin and mass spectrometry data collection were performed in co-operation with the group of Prof. Dr. Buck (UKE, Hamburg). The samples in the form of coomassie stained bands were excised from SDS

PAGE gels and destaining was done. The standard protocol was followed by which the excised samples were treated with 100 % acetonitrile (ACN), NH_4HCO_3 and dithiothreitol DTT to reduce disulfide bonds, if any. The samples were also modified by iodine acetic amide solution. Digestion of protein samples by trypsin in trypsin resuspension buffer (sequencing grade; Promega) was conducted overnight. The digested protein in the form of peptides was extracted from the gel by adding 100 % (v/v) ACN. Peptide samples were desalted by reverse phase chromatography. An ESI ion trap (LC/MSD Trap XCT Ultra II) instrument was used to detect peptides that were identified and listed using the mascot search engine (Matrix Science).

3.4.2. Native Mass Spectrometry

Native mass spectrometry (MS) is an analytical method used to receive structural information from an analyte. Particular of interest are protein and protein complexes up to mega Dalton size, often assembled of different subunits, whose heterogeneity can be observed by native MS. Mass spectrometers, used for this purpose, are based on the principle of electrospray ionization (ESI). The protein sample is converted from solution into the gas phase by droplet evaporation and fission upon spraying from a micrometer sized capillary. ESI is considered to be a gentle ionization method, which allows the soluble protein to retain its tertiary and quaternary fold upon transfer into the gas phase (Metwally*et al*., 2015), compared to Matrix-assisted laser desorption ionization (MALDI), which requires dried sample. In order to prevent signal loss by salt adducts, which would be inevitable upon ESI, the buffer substances have to be volatile. Ammonium salts are commonly used for this purpose, because of their volatility at low pressure, wide pH range and adjustable ionic strength. Multiple charged ions with a Gaussian distribution of a single protein can be generated, depending on the accessible surface area in solution. Unfolded protein contains a relatively higher surface area and therefore a higher charge compared to its folded form. The readout from common mass analyzers is usually a mass to charge ratio (*m/z*) where specific mass and charge can be determined from two adjacent peaks of the same analyte species with one charge increment by the following formula.

Peak1 $(m/z)_1$ $$\left[\frac{m}{z}\right]^1 = \frac{m+nH}{n}$$

Peak2 $(m/z)_2$ $$\left[\frac{m}{z}\right]^2 = \frac{m+(n-1)H}{(n-1)}$$

It is assumed that m>>nH and m >> (n-1)H

Peak 1: $$\left[\frac{m}{z}\right]^1 = \frac{m+nH}{n} = \frac{m}{n}$$

Peak 2: $$\left[\frac{m}{z}\right]^2 = \frac{m+(n-1)H}{(n-1)} = \frac{m}{(n-1)}$$

Where n and n-1 represent the number of charges on consecutive peaks.

$$n = \frac{\left[\frac{m}{z}\right]^2}{\left[\frac{m}{z}\right]^2 - \left[\frac{m}{z}\right]^1}$$

$$m = n * (m/z)_1$$

m represents the mass of a macromolecule and n is the number of charges.

The mass spectrometers used for the described experiments were an LCT Premier and Q-ToF2 mass spectrometer (Waters/Micromass, UK) modified for high masses (MS Vision, NL). A nano-ESI source was used for ionization and a time-of-flight (ToF) for mass analysis before multi-channel plate (MCP) detection. MS control and sample analysis were carried out by MassLynx 4.1 (Waters, UK).

The buffer exchange was performed using centrifugal filter units at 15,600 g (Vivaspin 500, MWCO 5000, Sartorius) or dialysis devices (Slide-A-Lyzer 100 µL, 3500 MWCO, ThermoScientific) at 4°C. The storage buffer was exchanged against 200-400 mM AmAc (ammonium acetate) (99.99% purity, Sigma Aldrich), pH 6.6 and pH 7.4 depending on the type of protein. In case of interaction analysis between proteins, 0.1 mM Zn (II) acetate was added into ammonium acetate solution, after buffer exchange. For ESI 5-10 µL of the sample to be analyzed were inserted into a gold-coated glass capillary. The capillary voltage was between 1000 V and 1500 V. The voltage at the sample cone applied was varied between 100 V and 150 V. In LCT, the gas pressures were 10 mbar inthe source region and $1.0 * 10^2$ mbar in the hexapole region. For MS/MS analysis the Q-ToF mass spectrometer was used with a pressure of 10 mbar in the source region. The collision cell was pressure elevated by argon to

1.3 x 10^{-2} mbar and collision energy was set between 10 V and 200 V. Measurements were performed in positive ion mode.

3.4.3. Microscale thermophoresis (MST)

MST is a powerful technique to analyze biomolecular interactions. The basic principle that gave the name to the technique is thermophoresis in which the directed movement of molecules is studied over a range of temperatures. The movement of molecules depends upon the size, hydration shell and charge of molecules. Any change in molecular properties is quantified by this highly sensitive technique. During experiment, a temperature gradient is established by an infra-red (IR) laser and the directed movement of molecules is detected and recorded by the signal of an intrinsically attached fluorophore.

MST experiments were performed using Monolith NT™ Protein Labeling Kit (Nano Temper Technologies). Purified tag free Aap constructs were labelled with NT-647-NHS fluorescent dye according to manufacturer's protocol. N-hydroxysuccinimide (NHS)-ester chemistry of the fluorescent dye reacts with primary amines of protein to establish a dye-protein conjugate. The concentration of proteins was adjusted to 15 µM using a labeling buffer. Solid fluorescent dye was dissolved in 100% DMSO (approx. 435 µM) and mixed thoroughly just before performing the reaction. The concentration of the dye was maintained at 2 folds concentration of protein using labeling buffer. Protein to be labeled (three G5 constructs separately) and dye were mixed in 1:1 ratio in a total volume of 200 µl and let them incubate in dark for 30 min. at RT. The excessive amount of dye was removed by passing the above labeled reaction through a supplied column having a resin to bind with the excessive dye and eluting the labeled protein in a number of fractions. Sbp was allowed to mix with the above labeled G5 in the presence of 1 mM $ZnCl_2$. Different ratios of Sbp to G5 constructs (1:1, 1:2 and 2:1) were made and allowed them to stand for at least 30 min. Monolith NT™ Standard Treated Capillaries were filled with the above set reaction mixtures and the amount of bound and unbound proteins were detected as fluorescent signals using Monolith NT.115 instrument. The obtained spectra were analyzed using affinity analysis software.

3.4.4. Circular dichroism (CD) spectrometry

A spectroscopic technique used to estimate the secondary structure of a protein and polypeptides in solution providing insights into the tertiary structure is called CD spectroscopy. It is also useful to investigate the stability of a protein by measuring its melting temperature. According to the supplier's instructions, the CD spectrometer (J-815, Jasco, UK) equipped with a peltier element was calibrated. Precise investigations related to thermal stability of a protein fold were supported by the peltier element. CD measurements were performed applying a Jasco J-810 CD spectrometer (JASCO) and using a spectra manager program.

The sample solutions were prepared in concentrations of 0.1-0.5 mg/ml and measured in a quartz cuvette with a layer thickness of 0.1 cm at room temperature. Within a range of wavelength between 190-260nm *(far-UV spectrum*), the ellipticity of the sample protein (100 µl) was measured after centrifuging the protein solution at 13,000 rpm at 4°C for 30 min. 10 measurements were taken in total, which were combined later by arithmetic averaging. The ellipticity θ is the result of difference in absorbance of clockwise and anti-clockwise circular polarized light and is calculated by the following formula.

$$\theta = \frac{180.\ln 10}{4\pi} * (A_L - A_R)$$

The determined ellipticity was converted into MRE (Mean Residue Ellipticity) according to following equation by CD software.

$$\mathbf{MRE} = \frac{\mathbf{MRW}.\boldsymbol{\theta}}{\mathbf{10}.\mathbf{cd}}$$

Whereas MRE refers to the mean amino acid residue weight, d: path length through the sample and c: protein concentration.

A typical curve showing minima at 208 nm and 220 nm and a maximum at 192 nm depicts the presence of α- helices, as shown as a green curve in figure 6. Proteins appearing at 215 nm minimum and a maximum at 195 nm have mainly β-sheet fold (blue curve in figure 6).

Random coiled structures of a protein appear at a minimum between 190 and 200 nm and a maximum at 220 nm in a CD spectrum (described as red curve in figure 6). The prediction of secondary structural elements in a protein was basically determined from this standard curve, while all values are not absolute.

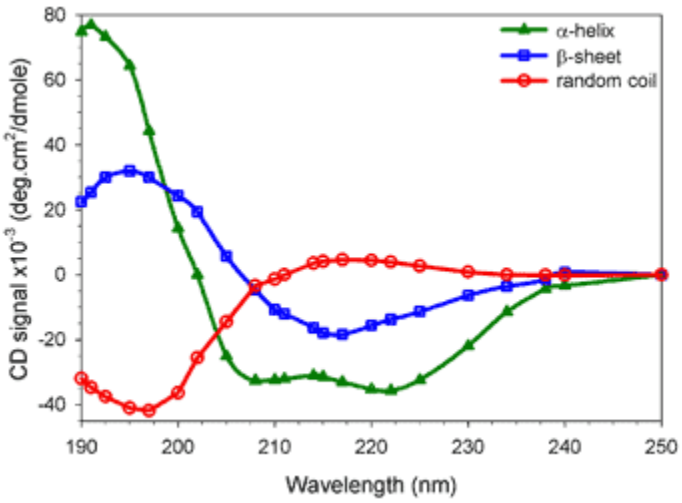

Figure 6: Standard CD spectrum of a protein. CD spectrum representing α-helices is shown in green. β-sheets can be seen in a blue curve and random coil in red.
http://www.fbs.leeds.ac.uk/facilities/cd.

To investigate the folding stability of Sbp, its melting temperature (T_m) was also determined at a wavelength of 220 nm, using CD spectrometry.

3.4.5. Dynamic light scattering (DLS)

Dynamic light scattering is used to investigate the particle distribution in a solution by determination of the hydrodynamic radii (*Rh*) of the particles, which allows the calculation of the corresponding molecular weights of particles. Thus, DLS can be applied to determine whether a protein solution is monodisperse or strongly aggregated. The behavior of complex fluids, such as polymers can also be studied by applying this technique. The Spectroscatter device 201 (Molecular Dimensions, UK) was used for DLS measurements. This allows using at least 15 µl of protein sample in a quartz cuvette. The samples were irradiated by a red light class 3b laser (λ = 690 nm; laser power 10-50 mW) to detect isotropic scattering at an angle of 90°. The samples from lower to higher concentration were measured by DLS to see the

change in the hydrodynamic radius and aggregation behavior of Sbp. Finally, the sample with a concentration of 10 mg/ml was centrifuged at 4 °C for 30 min at 13,000 rpm to remove high molecular weight aggregates and then analyzed in a quartz cuvette with a path length of 1 cm at room temperature. Measurements were accumulated per sample using an autopilot function. The protein samples of Aap constructs were also studied by this method.

Stokes-Einstein equation is used to determine *Rh* by its dependence on diffusion co-efficient D_T and viscosity η.

$$Rh = \frac{K_B T}{6\pi\eta D_T}$$

K_B is the Boltzmann's constant and T is the absolute temperature. Since particles in solution have Brownian motion, the back scattered light has a frequency shift, which results in the variation of intensity of light and recorded by a highly sensitive detector within the instrument.

3.4.6. Small-angle X-ray scattering (SAXS)

In order to obtain a low resolution structural data (including size and shape) of randomly oriented protein molecules in solution, a purified, filtered, homogenous and monodisperse protein sample was applied to SAXS at the EMBL BioSAXS beamline P12 at PETRA III 3rd generation synchrotron source DESY Hamburg, Germany. The elastically scattered X-rays (wavelength 0.1-0.2 nm) by the samples were recorded at low angles (typically 0.1 - 10°) and analyzed.

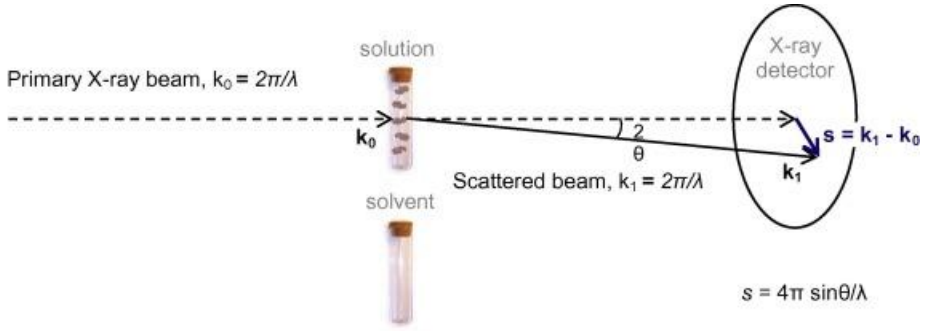

Figure 7: Schematic representation of a SAXS experiment (Kikhney and Svergun, 2015).

The shape and size of macromolecules, pore size, characteristic distances between partially ordered materials and some other data can be investigated by having information from the scattered pattern at the above mentioned small angular range. Structure information of macromolecules between 5 and 25 nm and of repeat distances in partially ordered systems of up to 150 nm can be delivered by SAXS (Walenta, 1985). A number of different concentrations of Sbp (1 mg/ml, 2 mg/ml, 4 mg/ml, 6 mg/ml, 8 mg/ml and 16 mg/ml) were used to perform the SAXS experiment. Likewise, a number of different concentrations were used to determine the ab-initio models of Aap constructs as well. The monodispersity of each sample was ensured by DLS prior to X-ray exposure, which was later on confirmed by Linear Guinier plots from experimental data using Primus software. The final scattering curves were obtained by extrapolating low angle data. The pair-distribution function was calculated by GNOM (Svergun *et al*., 2013). Ab initio models of the respective proteins were calculated using DAMMIF (Frank *et al*., 2009) and the automated mode of DAMAVER (Volkov *et al*., 2003).

3.4.7. Pre-crystallization test (PCT)

Crystallography is one of the most advanced and widely used technologies for the determination of a high resolution 3D-structure of a protein. Crystals of the protein of interest need to be grown and can subsequently be analyzed using X-ray radiation. To optimize the concentration of proteins for screening different crystallization conditions, a Pre-Crystallization Test (PCT) was used. The four reagents of the PCT kit (Hampton Research USA), which were used to evaluate ideal protein concentration for crystallization screening, are as following:

1) Reagent A1: 0.1 M Tris hydrochloride pH 8.5, 2.0 M $(NH_4)_2SO_4$

2) Reagent B1: 0.1 M Tris hydrochloride pH 8.5, 1.0 M $(NH_4)_2SO_4$

3) Reagent A2: 0.1 M Tris hydrochloride pH 8.5, 0.2 M $MgCl_2$ 6(H_20), 30% w/v PEG 4,000

4) Reagent B2: 0.1 M Tris hydrochloride pH 8.5, 0.2 M $MgCl_2$ 6(H_20), 15% w/v PEG 4,000

Different concentrations of proteins were used to avoid an immediate amorphous precipitate and a clear drop, to reach the best point at which the crystallization screening could be performed.

3.4.8. Robotic crystallization screening

Major advancements in structural proteomics have been observed with the increasing use of robotics, which play a vital role in automation of crystallization experiments by reducing the amount of protein samples by an order of magnitude. The time period to perform a large number of conditions has also been reduced, hence improved the reproducibility of experiments. A large number of set ups in a comparably shorter time has also become possible. After purification, the protein samples were concentrated, centrifuged at 14,000 rpm for 1 hr and filtered to remove the precipitated part. The disparity of the samples was monitored by DLS. The honeybee 961 pipetting robot (Zinsser Analytic GmBh, Germany) was used to carry out different sets of crystallization conditions in which protein and precipitant were transferred to 96 well Nextal Qia1 plates (Qiagen, Germany). The commercially available screens Morpheus, Stura, and PACT (all Molecular Dimensions, UK) and JCSG+, Classic, Cryos, ComPAS, and AmSO4 Suite (all Qiagen, Germany), were used to identify potential crystallization conditions. 400 nl of protein solution were mixed with 400 nl of the respective precipitant solution with the help of honeybee robot in each well. The reservoir was filled with 45 µl of precipitant solution to set up sitting drop vapour diffusion method. The plates were sealed and stored at RT. Plates were also stored at 4 °C for another parallel set up.

3.4.9. Optimization of initial crystallization conditions

After obtaining initial crystallization conditions for crystals, their growing conditions were optimized by varying the precipitant condition just around the condition found in the initial screen. The concentration of polyethylene glycols (PEGs), if found in original condition, was increased and decreased by 10 %. The slight changes in each component of the initial screen were made e.g., pH, salt, buffer, PEGs and temperature. Conditions were optimized in 24-well Linbro plates (Hampton Research, USA), using the hanging drop vapor diffusion method. 1.5 µl of a protein solution and 1.5 µl precipitant were added together in the droplet and the reservoir was filled with 1 ml precipitant solution. Conditions were also optimized in the 48-well MRC sitting drop plates (Molecular Dimensions, UK). Different ratios of precipitant to protein sample were made in a total volume of 4 µl and the reservoir was filled with 50 µl of precipitant solution. The set ups in the plates were made as duplicates to place

them at 4 °C and 16 °C. All the droplets were manually pipetted. Crystallization under oil method has also been considered. For crystallization under oil, a paraffin oil (Applichem Germany) treated Terazaki plate (Nunc, Denmark) was used to fill all wells with oil. Then, 2 µl of protein with 2 µl of precipitant were pipetted into each well under the microscope, whereas the oil covered the droplet.

3.4.10. Data Collection

3.4.10.1. Native Diffraction Data

$G5_1EG5_2EG5_3$ (one of the Aap G5 constructs) crystal was fished on a nylon loop under a microscope. Subsequent flash cooling of the mounted crystal was done in a liquid nitrogen stream at 100 K. The cryo- protected crystal was mounted on the goniometer at the beamline P13 at PETRA III, DESY, Hamburg. A native diffraction data-set was collected and initial processing of data was done, using the software XDS (Kabsch, 2010) and iMOSFLM (Leslie, 1992). For scaling, the SCALA (Evans, 2006) program in the CCP4 package was used.

3.4.10.2. Matthews Coefficient (V_M)

The solvent content in a crystal is determined by the Matthew's co-efficient (Matthews, 1968). It is one of the most important parameters to determine the number of molecules in an asymmetric unit.

$$V_M = \frac{\text{Volume of unit cell}}{\text{Molecular weight of molecule} * Z * X}$$

Z in the above equation represents the number of asymmetric units in the unit cell (the number of symmetry operators in space group). The unknown variable, X, is the number of molecules in an asymmetric unit.

3.5. Bioinformatics tools

Different bioinformatics tools were used for sequence and structure analyses and predictions, which are described below.

3.5.1. Basic Local Alignment Search Tool (BLAST)

BLAST is an algorithm used to compare biological sequence information in proteins and nucleic acids (Altschul *et al*., 1990). There are different tools available at the BLAST server for comparing protein and nucleotide sequences against protein and nucleotide databases (blastp and blastn).

3.5.2. ProtParam

ProtParam is a bioinformatics tool, which allows the computation of various chemical and physical parameters of a given protein, such as molecular weight, extinction co-efficient, theoretical pI, amino acid and atomic composition, estimated half-life and instability index (Gasteiger *et al*., 2005)

3.5.3. PDBsum

It is a web-based database providing a graphical summary of the key information on the macromolecular structure and includes images of the structure, annotated plots of each protein chain's secondary structure and detailed structural analysis (Laskowski, R. A. *et al*., 1997).

3.5.4. Easy Sequencing in PostScript (ESPript)

The ESPript program was used for rendering secondary structure information and sequence similarities from aligned sequences (Robert and Gouet, 2014).

3.6. Homology model building servers

5.6.1. Iterative Threading ASSEmbly Refinement (I-Tasser): was used to develop homology models by identifying templates from PDB by multiple threading approaches (http://zhanglab.ccmb.med.umich.edu/I-TASSER) (Yang *et al*., 2014).

3.6.2. Swiss modeling server: was used to obtain a homology model with reference to the amino acid sequence of the protein of interest. https://swissmodel.expasy.org/ (Biasini *et al*., 2014).

3.6.3. Raptor X: Determination of estimated structures of sub domains in the proteins of interest was also performed by Raptor X(Ma *et al*., 2013).

3.6.4. PRALINE: was used for multiple sequence alignment in combination with the prediction of secondary structure elements by generating homology-extended multiple alignment (Simossis and Heringa, 2005).

Apart from the above mentioned homology model servers, Phyre2, IntFOLD2, HHpred and M4T servers were also used to predict and compare homology models.

4. RESULTS

4.1. Structural characterization of Sbp

In order to achieve the first aim of the work, the following strategy was applied.

4.1.1. Expression and purification of Sbp/His by Zn^{2+}NTA

The *sbp* gene with 6 x His tag (yellow box) at the N-terminus and a linker of 18 amino acids (red box) at the upstream of gene was already cloned into pDEST17 expression vector in *E. coli* BL21AI cells. The amino acid sequence of Sbp/His construct is given below.

M S Y Y H H H H H H L E S T S L Y K K A G S A A A P F T N N V E A A T G N
S M K T V Q Q L N K G D K S L E N V K I G E S M K S V L K K Y S H P I Y S
Y N P N S N E K Y Y E F R T D K G V L L V T A N G K K E R G N V T R V S M
T Y N N A N G P S Y K A V K Q Q L G H K A I S R V H Y N N V T G N F G Y I
Q K G Q A S Y Q F S S N S P K D K N V K L Y R I D L N K

After performing test expression, recombinant *sbp*/His was expressed in *E. coli* BL21AI cells on large scale. For purification of the soluble protein, affinity chromatography using a zinc matrix column was applied, as the 6 x His-tag in protein binds with high affinity to Zn^{2+} ions. After equilibration of the column, using the appropriate buffers (described in section 3.1.9), the solubilized protein, obtained from the centrifuged supernatant fraction of the lysed cells was applied onto the column.

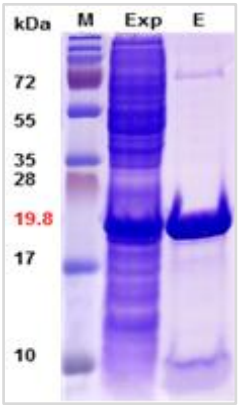

Figure 8: SDS PAGE analysis of purification of Sbp.M represents the marker. Lane Exp shows expression of *sbp* and lane E represents the elution of ~80 % purified Sbp after Zn^{2+} affinity chromatography.

The high affinity of the protein towards the matrix led to a highly specific binding of the protein tag to the matrix. After several washing steps, using the appropriate buffer (20 mM

$NaPO_4$, 500 mM NaCl, 20 mM Imidazole at pH 7.4) to remove non-specifically bound proteins, Sbp/His was eluted in elution buffer described in section 3.1.9 and the fractions were analyzed by 15 % SDS PAGE.

In order to remove the remaining impurities in the eluted Sbp from affinity purified chromatography, gel filtration was performed. A calibration curve was prepared using the standard proteins ovalbumin (48 kDa), ribonuclease A (13.5 kDa), aprotinin (150 kDa), conalbumin (6.6 kDa) and carbonic anhydrase (72 kDa). Subsequently, the relative molecular weight of the eluted protein was interpolated from an exponential calibration plot of an elution volume versus molecular weight. The calculated molecular weight indicates that Sbp forms a stable monomer in solution [see figure 9 (B)].

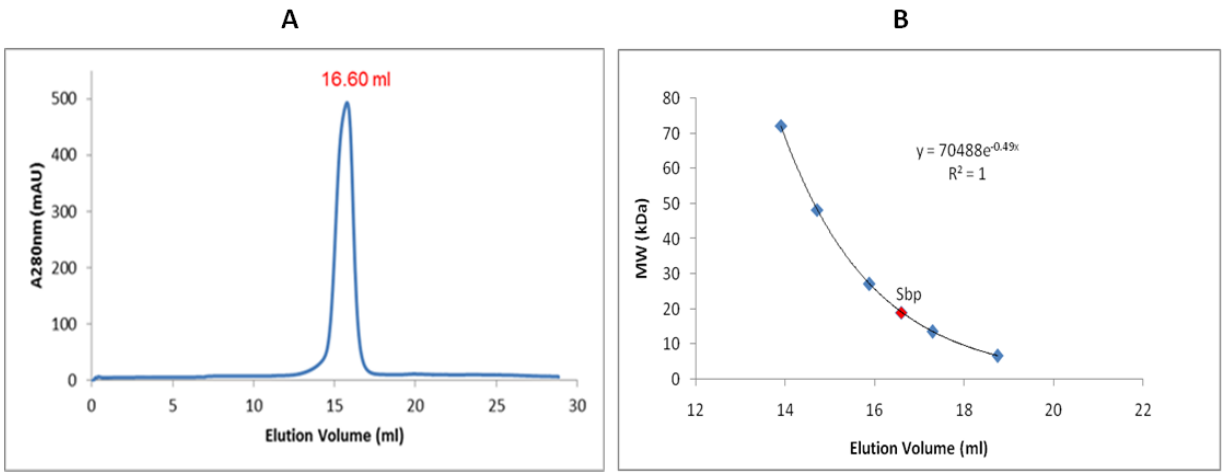

Figure 9: (A) Superdex 200 SEC profile of Sbp. The elution volume of Sbp at 16.6 ml corresponds to the calculated molecular weight of Sbp, which is ~19.8 kDa. **(B)** Calibration plot representing the molecular weight of Sbp along with standard proteins.

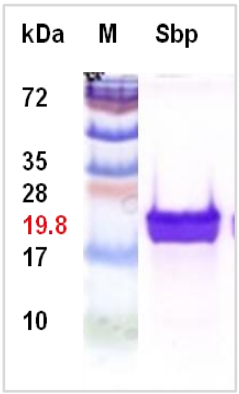

Figure 10: SDS PAGE analysis. Lanes M represents marker, while lane Sbp is purified Sbp after SEC.

4.1.2. Stability assays and biochemical characterization

To assess the thermal stability of Sbp at a range of temperatures from 5 °C to 95 °C in different compositions of solvent buffers, a thermofluor assay was performed. The results showed that the highest melting temperature (T_m) of Sbp is 40 °C as shown in figure 11.

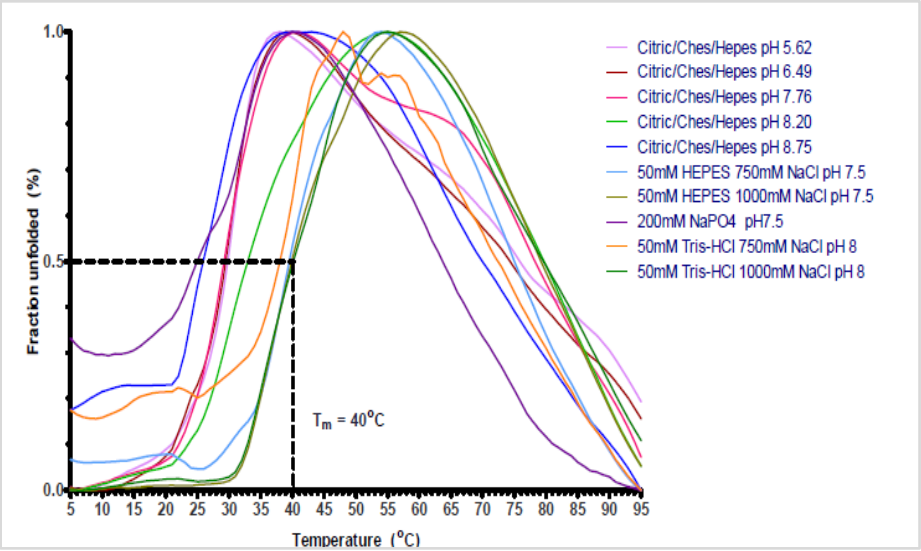

Figure 11: Thermofluor assay of Sbp. The denaturation of Sbp starts at 40 °C in most of the buffers mentioned in the right side of the figure.

Unfortunately, the protein was not stable in any of the buffers used in this experiment, shown in figure 11. Over a time period of three hour and longer, precipitation of Sbp was observed during exchange of buffer.

To further characterize the effects of chemically different buffer conditions on the stability and aggregation behavior of Sbp, dynamic light scattering (DLS) was applied. Different buffer conditions were used to optimize the stability of protein, based on selecting the most monodisperse protein solution. Before using the sample for DLS measurements, the impurities, consisting of degraded and precipitated protein, were removed. For this, the protein sample was centrifuged at 14,000 rpm for 1 hr and filtered afterwards. After trying a wide range of different buffers with different molarity and different pH settings, a buffer

solution containing 50 mM $NaPO_4$, 150 mM NaF at pH=7.4 was found to be a *suitable* solvent to stabilize Sbp over a long period of time of approx. 10 days.

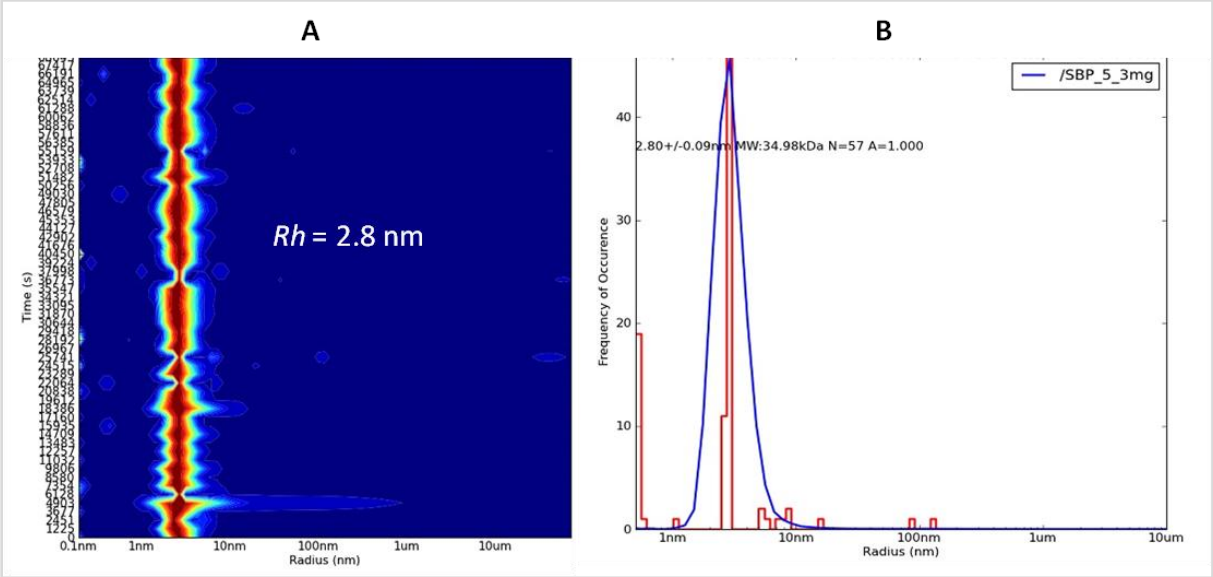

Figure 12: Results of DLS measurements with Sbp. (A) represents the radius distribution of Sbp particles over a range of 54 measurements. **(B)** The histogram (blue) and statistical distribution (red) resulted in a calculated hydrodynamic radius (*Rh*) of 2.8 nm ± 0.1 nm**.**

The DLS measurement to calculate the hydrodynamic radius (*Rh*) of Sbp particles was done at a protein concentration of 5.3 mg/ml. In order to see the changes in *Rh* with increasing time, 54 measurements were conducted sequentially. Each measurement was performed for 30 sec and the gap between each measurement was 2 min. *Rh* of Sbp was calculated as 2.8 nm ± 0.1 nm and remained constant over time, as shown in figure 12.

The effect of increasing temperature on Sbp in the above stated buffer (50 mM $NaPO_4$, 150 mM NaF at pH=7.4) was analyzed by CD melting analysis of Sbp. The melting curve was obtained for a range of temperature from 20 °C to 40 °C at a wavelength of 220 nm. A decrease in the absorption of Sbp, as the temperature was raised above 40 °C is the indication that Sbp has started degrading. The change in ellipticity towards the baseline also reflects the loss of secondary structures in Sbp. T_m of Sbp estimated from CD melting curve was ~40 °C, as shown in figure 13.

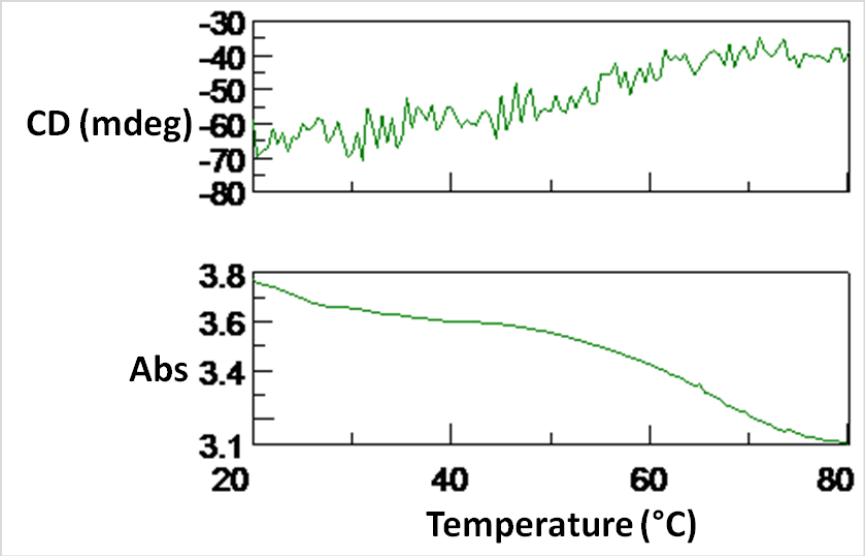

Figure 13: CD spectrum of Sbp as a function of temperature.

4.1.3. Sequence alignment, homology model predictions and analysis

The secondary structure of Sbp was predicted using different bioinformatics tools. All the used bioinformatics tools, including Swiss modeling tool, RaptorX, I-Tasser and IntFOLD server, took more or less the same homology models as templates to predict secondary structure of Sbp. PDB IDs: 4H0A from *S. aureus* (Joint centre for structural genomics, to be published), 2QZB from *E. coli* (Bonanno *et al*., to be published) and 4YGT from *Bacillus subtilis* (Joint centre for structural genomics, to be published) were the most common homology models used as templates by the above mentioned tools.

The multiple sequence alignment, using the amino acid sequences of templates taken by the before mentioned bioinformatics tools, revealed that there is 24 % sequence homology with 82 % coverage between Sbp and a cysteine-rich secretory protein (PDB ID: 4H0A) from *S. aureus*. The amino acid sequence alignment of Sbp and 4H0A by PRALINE (Simossis and Heringa, 2005) is shown in figure 14.

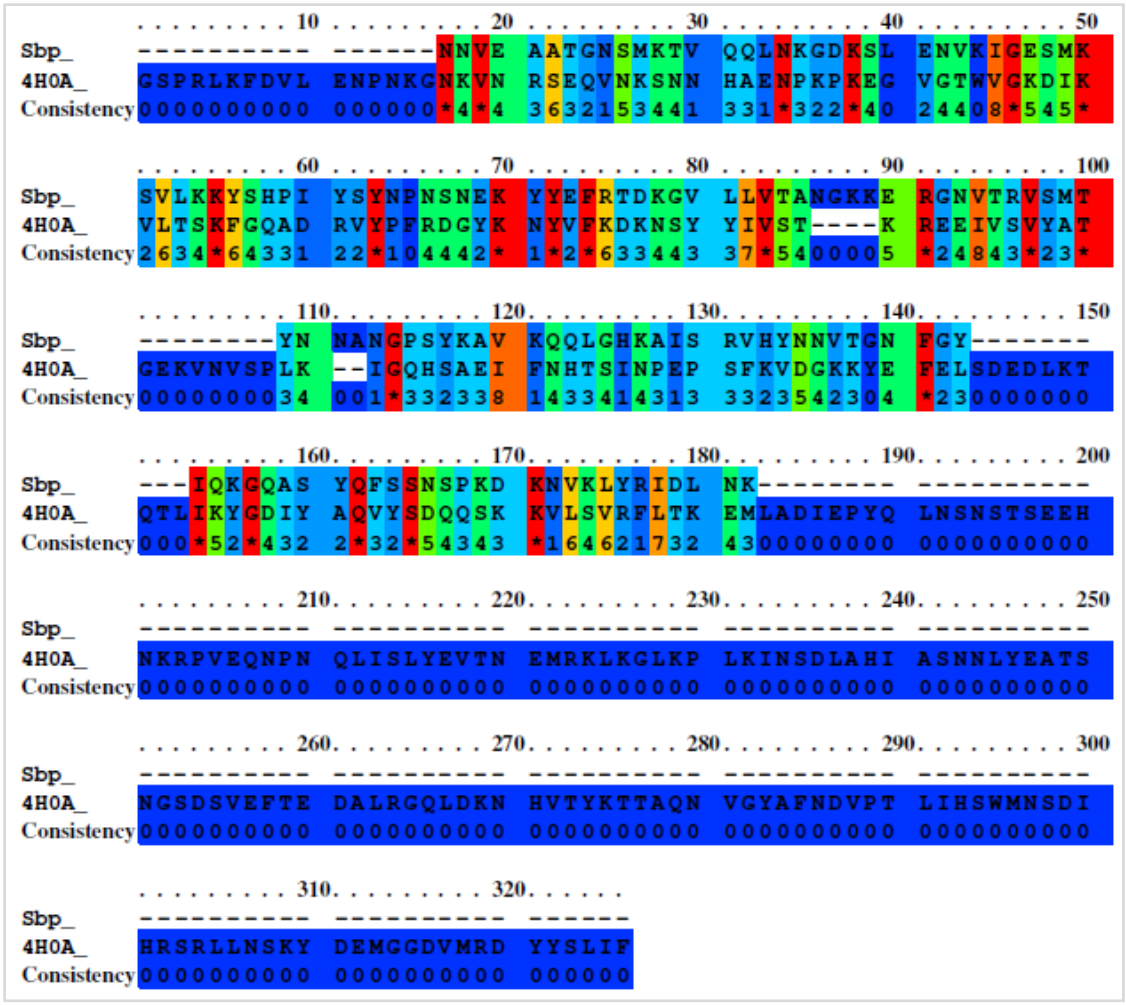

Figure 14: The amino acid sequence alignment of Sbp with a homologue (PDB ID: 4H0A). Amino acids shown in red boxes are conserved in both protein sequences. The colour codes refer to going from highly non-conserved residues to conserved amino acids. 0 1 2 3 4 5 6 7 8 9 10

Figure 15 (A) shows the calculated model of Sbp predicted by Swiss modeling, while (B) and (C) are the Sbp structure prediction models by RaptorX and I-Tasser. The model (D) is the representation of the predicted model by the IntFOLD server. All the models showed the dominance of β-sheets as compared to α-helices. Moreover, the coil structure was also clearly visible in all the predicted models, as shown in figure 15.

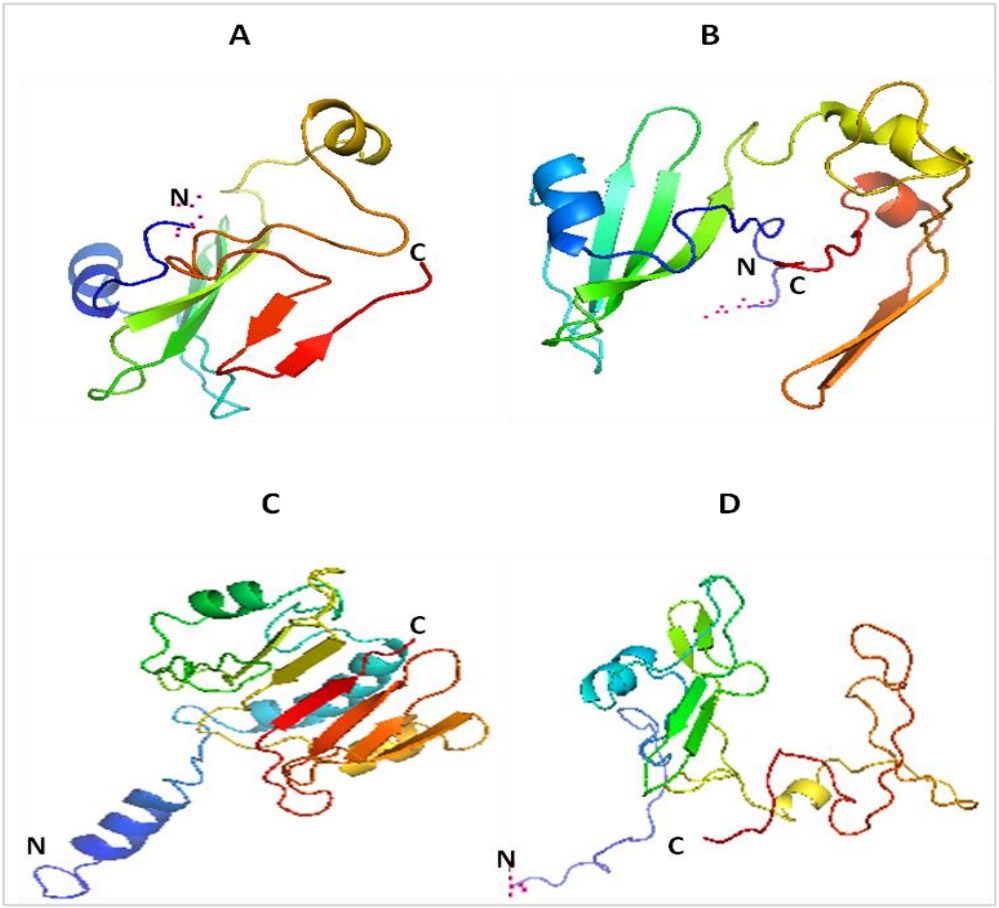

Figure 15: (A) The predicted structure model of Sbp using the swiss homology-modelling server, showing the presence of β-sheets, relatively lower number of α-helices and coils in the structure of Sbp. **(B)** The prediction of Sbp calculated by RaptorX. **(C)** and **(D)** The structure of Sbp predicted by I-Tasser and IntFold respectively, representing the presence of a high ratio of β-sheets as compared to α-helices. The coiled parts are also shown in the models. The N and C termini of each model are also shown n the figures.

The secondary structure predicted by aligning sequences of Sbp and 4H0A by PRALINE is shown in figure 16.

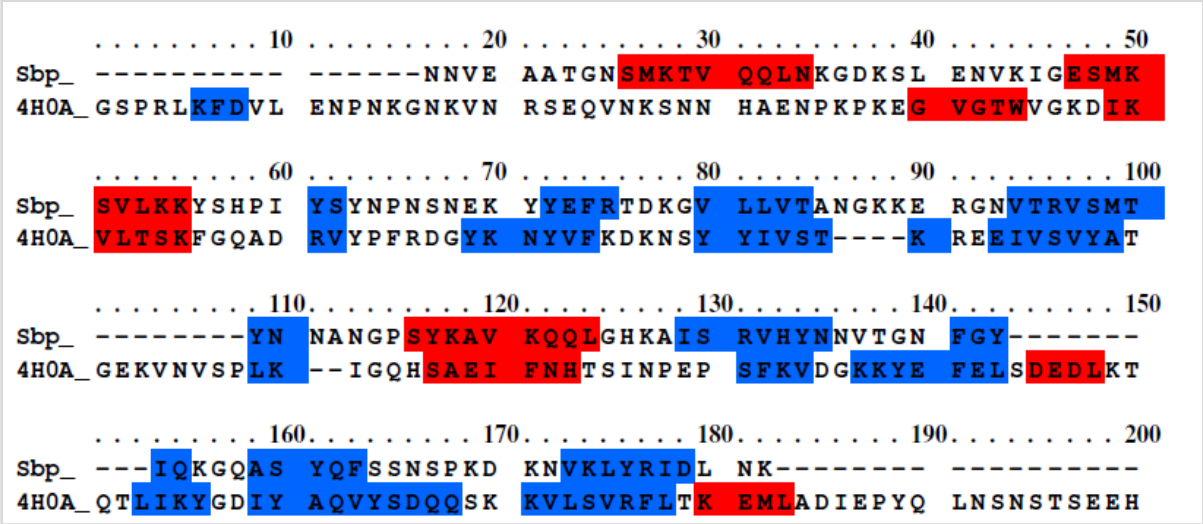

Figure 16: The amino acid sequence of Sbp aligned with 200 amino acids of 4H0A, using the PRALINE, online analysis tool (Simossis and Heringa, 2005). The sequence shown in red boxes refers to α-helices, while the blue boxes correspond to β-sheets predicted by PRALINE.

The verification of secondary structure composition and folding state of Sbp in solution was done by circular dichroism spectrometry. The obtained spectrum is the characteristic for a typical β-sheet rich protein. One minimum at 206 nm was observed, which suggests that β-sheets are predominant in Sbp along with regions of low structural complexity. The positive ellipticity at lower wavelengths, below 192 nm, is the indication that Sbp is well folded and has defined and partly compact secondary structure.

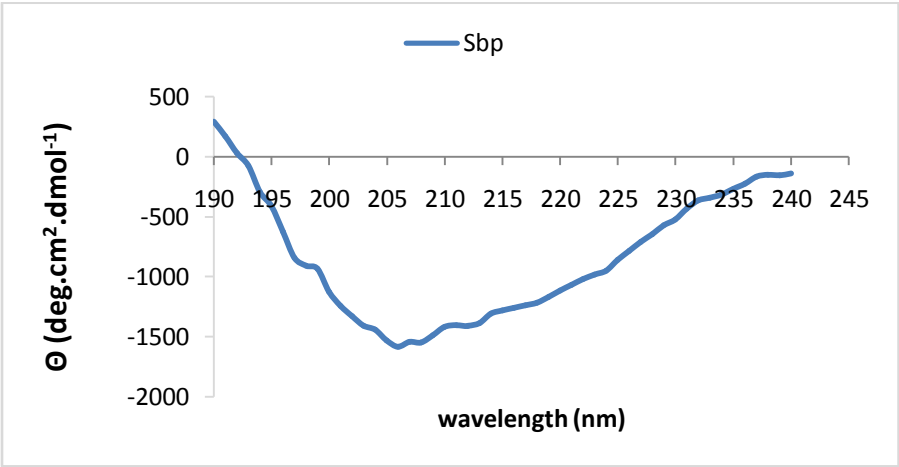

Figure 17: CD spectrum of Sbp.

CD spectrometry verified the overall folding of Sbp applying Yang's algorithm (Yang *et al*., 2015). The approximation for the fold of Sbp is 2 % α-helices, 45 % β-sheets and 51 % random coil (figure 17). In parallel, an approximation according to Reed's algorithm (Norma *et al*., 2006) is 6 % α-helices, 46 % β-sheets and 33 % random coil. Both approximations are overall in good agreement with each other.

The secondary structure estimation of Sbp obtained by CD spectrometry correlated well to the estimation of secondary structure by bioinformatics tools analysis.

4.1.4. Crystallization experiments

Once, the homogeneity and monodispersity of protein was confirmed, concentrated Sbp was applied to a number of crystallization set ups to screen a potential condition suitable for Sbp crystal growth. The protein was concentrated up to 10 mg/ml. A variety of available crystallization screens (Morpheus, PACT, Stura and ComPAS suites, each with 96 variations of conditions) were used to set up the initial crystallization conditions using the Honeybee 961 dispensing robot (Genomic solutions, UK). The principle of sitting drop vapour diffusion method was applied, while setting the experiment in 2 well MRC plates (NUNC) at 293K. 400 nl of Sbp were mixed with the same amount of precipitant solution, whereas 45 µl of precipitant was used as reservoir for establishing equilibration during vapour diffusion.

The plates were checked to identify a potential condition for crystallization on regular basis, but unfortunately no crystals were observed and no condition was found that gave a promising starting point for optimization of the crystallization. Repeating the experiment with the crystallization screens (Cryos, Classic and JCSG+), by using a new purification batch of Sbp, also did not result in successful crystal formation.

4.1.5. Small angle X-ray scattering measurements

High resolution X-ray crystallography provides atomic-resolution details of the structure of protein molecules, however, several intrinsic structural properties of a diverse range of samples, for example oligomerization, concentration-dependent aggregation and flexibility in solution may not be captured by crystallization techniques. Therefore, *in solution* methods,

such as small angle X-ray scattering (SAXS) proves beneficial in determining the low-resolution shapes and global conformation/structure(s) of proteins in solution. Sbp scattering data were collected at the advanced 3rd generation synchrotron source PETRA III at DESY, Hamburg, on the EMBL-P12 bioSAXS beamline (Blanchet *et al*., 2015). In terms of monodispersity, the samples were first verified by DLS. Monodisperse solutions of purified Sbp with 4 different concentrations (2 mg/ml, 3 mg/ml, 4 mg/ml and 5 mg/ml) were applied to SAXS measurements. Twenty successive measurements with an exposure time of 50 milliseconds were conducted.

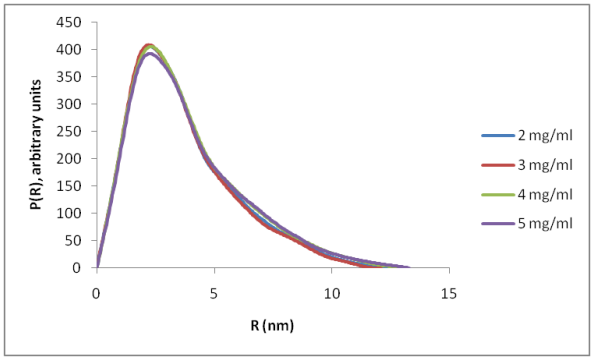

Figure 18: Pair-wise distance distribution function *P(R)* of Sbp representing the maximal dimension (D_{max}) increasing from 11.5 nm to 14 nm for the samples with increasing concentration of 2-5 mg/ml.

A low-resolution structure of Sbp monomer was generated *ab initio* using a bead modeling approach. The scattering data obtained from 2 mg/ml Sbp was used to generate *ab initio* model.

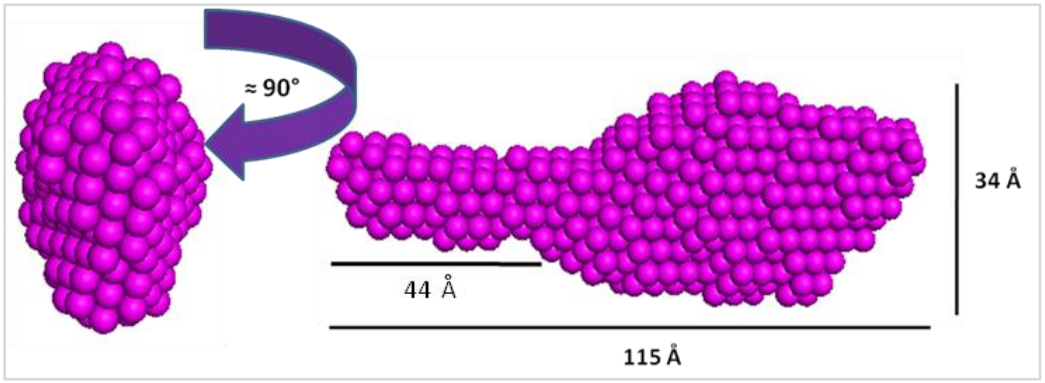

Figure 19: The spatially aligned average and volume-corrected *ab initio* model of an Sbp monomer generated by *DAMMFILT* derived from ten individual *DAMMIF* dummy atom models.

The *ab initio* model has an elongated ellipsoidal shape with a considerably large globular head and a small tail, as shown in figure 19. The χ^2 (chi square) and *p* (probability) values of

the experimental data fitting to the model were 0.9 and 1.0 respectively, determined by data comparison in PRIMUSqt.

The spatial superposition of a predicted I-Tasser model of Sbp onto the low resolution *ab initio* model was performed by *SUPCOMB* (Kozin *et al*., 2001). The predicted I-Tasser structure model fitted well to the globular head region of *ab initio* model of Sbp, as shown in figure 20. However, an α-helix structure at the N-terminus of the predicted model was observed to fit in a part of extended tail of the *ab initio* model of Sbp.

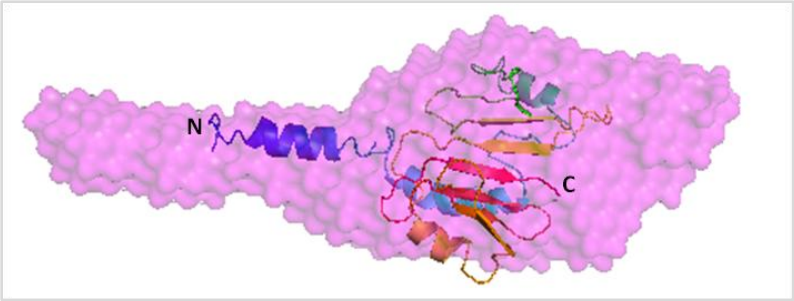

Figure 20: The spatial alignment of the calculated (predicted) I-Tasser homology model of Sbp onto the *ab initio* model of Sbp from SAXS using *SUPCOMB* (Kozin *et al*., 2001).

There could be a number of reasons that Sbp crystals were not obtained to unravel its atomic structure in detail. One of the reasons could be that Sbp has a 6 x His tag and a linker region, consisting of 18 amino acids. The linker region is present between His tag and start of Sbp sequence. Unfortunately, there was no restriction site available to cleave the tag off from protein, so a new construct of Sbp was designed and cloned, as described below.

4.1.6. Cloning of Sbp into pET302NT-His vector

Sbp gene was amplified from genomic DNA using gene specific primers. The N-terminus was extended by a 6 x His tag and a TEV protease cleavage recognition site during the synthesis of primers. The restriction sites for EcoRI (5' end) and BamHI (3'end) were also introduced into the primers. The PCR amplification of the cloned *sbp*, using gene specific primers, showed that the obtained band corresponded well to the expected sizes. The amplified gene *sbp* appeared at 513 base pairs, when compared with a DNA marker, as

shown in figure 21. The PCR amplified product of Sbp was digested using EcoRI and BamHI high fidelity restriction enzymes. pET 302 NT-His vector was also linearized by the same enzymes. Ligation of the gene into the vector was done using T4 DNA ligase, as described in chapter 3.2.1. The obtained plasmids were isolated by Qiagen Plasmid Mini Kit (Qiagen, Hilden, Germany) and transformed into BL21 Star *E. coli* cells.

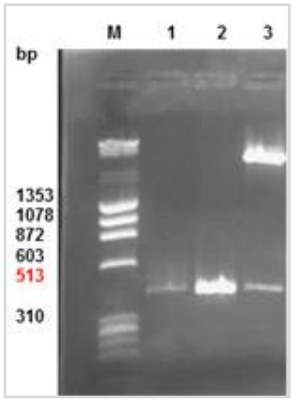

Figure 21: Agarose gel electrophoretic analysis of the amplification of Sbp gene by PCR. Lane M represents the marker. Lane 2 shows the amplification of*sbp* gene from genomic DNA and lane 3 shows the amplification of *sbp* from a selected clone after transformation with the ligated pET 302 NT-His vector containing *sbp*.

The PCR amplified product of Sbp was digested using EcoRI and BamHI high fidelity restriction enzymes. The pET 302 NT-His vector was also linearized by the same enzymes. Ligation of the gene into a vector was done using T4 DNA ligase as described in section 3.2.1. The obtained plasmids were isolated by using the Qiagen Plasmid Mini Kit (Qiagen, Hilden, Germany) and transformed into BL21 Star *E. coli* cells. The amino acid sequence of Sbp with 6 x His tag (yellow box) and TEV protease cleavage recognition (red box) is given below. The arrow head indicates the cleavage site, where TEV protease cleaves off the His tag.

M S Y Y H H H H H H V N S E N L Y F Q↑G N N V E A A T G N S M K T V Q
Q L N K G D K S L E N V K I G E S M K S V L K K Y S H P I Y S Y N P N
S N E K Y Y E F R T D K G V L L V T A N G K K E R G N V T R V S M T Y
N N A N G P S Y K A V K Q Q L G H K A I S R V H Y N N V T G N F G Y I
Q K G Q A S Y Q F S S N S P K D K N V K L Y R I D L N K

4.1.7. Test expression of *sbp*

Test expression of *sbp in E. coli* BL21 Star cells carrying the plasmid pET 302 NT-His with cloned *sbp* gene was performed using 1 mM IPTG as inducer, as described in section 3.3.1.

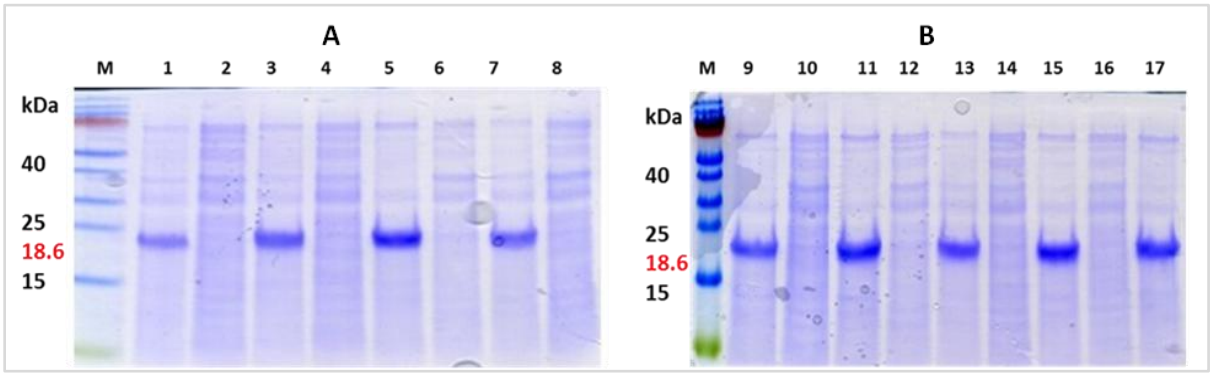

Figure 22: SDS-PAGE analysis of test expression of *sbp* in BL21 Star cells at RT. (A) Lane M shows marker band, lanes 1, 3 and 5 represent the expression of *sbp* when induced with IPTG at OD_{600} 0.4 and using collected cells after 2, 3 and 4 hrs respectively. Lane 7 in **(A)** and 9 & 11 in **(B)** are *sbp* expression, when induced with IPTG at OD_{600} 0.7 and collected cells after 2, 3 and 4 hrs respectively. Lane 13, 15 and 17 are *sbp* expression when induced with IPTG at OD_{600} 1.0 and collected cells after 2, 3 and 4 hrs respectively. Lanes 2, 4, 6, 8, 10, 12, 14 and 16 represent the respective non induced (NI) samples, showing no *sbp* expression.

The obtained results indicated that the expression of *sbp* was better when it was induced at OD_{600} 0.7 and 1.0, as compared to lower cell count expression. I chose the expression time point when OD_{600} value reached 1.0 after 3 hrs of induction, as shown in figure 22 (B) lane 15.

4.1.8. Purification and cleavage of 6 x His tag of Sbp by TEV protease

Purification of recombinant Sbp (rSbp) was performed following the steps, as described in sections 4.1.1. In order to obtain pure and 6 x His tag free Sbp, the affinity purified protein was dialyzed against the initial buffer without imidazole overnight, using Slide-A-Lyzer

Dialysis Cassette G2 with 3500 Da MWCO (Amicon Ultra). Sbp was concentrated up to 1mg/ml. The cleavage reaction with TEV protease (27 kDa) was performed at RT with mild shaking for overnight. A ratio of the TEV protease to Sbp (1:10) was used for complete cleavage. As cleavage reaction takes place in a reduced environment, 1 mM dithiothreitol (DTT) was added to the solution. The cleavage was verified by SDS PAGE. In order to remove the cleaved tag, remaining tagged Sbp and TEV protease enzyme from the protein solution, affinity chromatography was performed. The pure tag free Sbp was collected in the flow through. The final yield of the protein was ~10 mg out of 6 liters of *E. coli* cell culture and the purity of Sbp was approximately 95 %, as shown in figure 23.

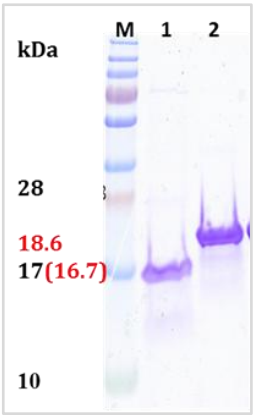

Figure 23: SDS PAGE analysis of Sbp. M represents marker**.** Lane 1 shows the His tag free purified Sbp and lane 2 represents the uncleaved Sbp as a reference.

4.1.9. DLS measurements and crystallization experiments of tag free Sbp

After removing His tag, size distribution profile of pure Sbp was determined by DLS. *Rh* measurements of Sbp were done over a range of concentrations from 0.2 mg/ml to 10 mg/ml at an interval of 10 days to check the stability of the protein. The histograms obtained from DLS measurements represented in figure 24 indicated that *Rh* gradually increased from 2.25 nm to 2.72 nm with an increasing concentration of 0.2 mg/ml to 10 mg/ml. Furthermore, aggregation of Sbp was observed at a concentration of 10 mg/ml.

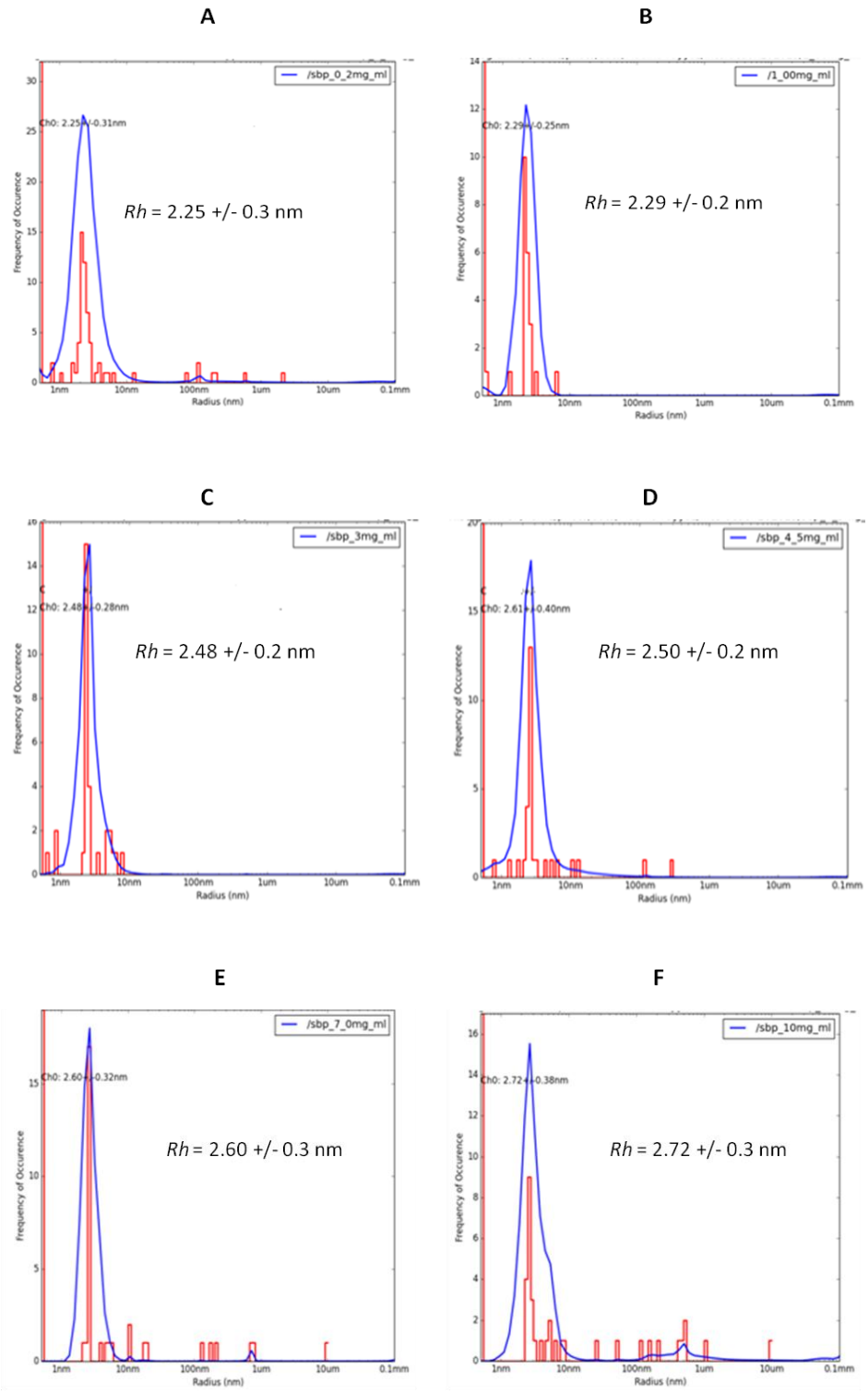

Figure 24: Histograms of *Rh* of Sbp calculated from DLS measurements. (A) *Rh* of 2.25 nm +/- 0.3 nm was obtained for a concentration of 0.2 mg/ml. **(B)** *Rh* of 2.29 nm +/- 0.1 nm at a concentration of 1.0 mg/ml. **(C)** *Rh* of 2.48 nm +/- 0.2 nm at a concentration of 3.0 mg/ml. **(D)** and

(E) represent the likewise increase in *Rh*. **(F)** An increase in the *Rh* has been observed at a concentration of 10.0 mg/ml resulting in an *Rh* of 2.72 nm +/- 0.3 nm.

To determine a suitable precipitant was the next step for Sbp to crystallize. A number of crystallization set ups were prepared with a variety of available crystallization screens, as described in sections 3.4.8. and 4.1.4. Once, the homogeneity and monodispersity of the protein was confirmed, concentrated Sbp was applied to a number of crystallization set ups to screen a potential condition suitable for Sbp crystal growth. The protein was concentrated up to 10 mg/ml. A variety of available crystallization screens (Morpheus, PACT, Stura and ComPAS suites, each with 96 variations of conditions) were used to set up the initial crystallization conditions using Honeybee 961 dispensing robot (Genomic solutions, UK). The principle of sitting drop vapour diffusion method was also applied, while setting the experiment in 2 well MRC plates (NUNC) at 293K. 400 nl of Sbp were mixed with the same amount of precipitant solution, whereas 45 µl of precipitant was used as reservoir for establishing equilibration during vapour diffusion.

The plates were checked to identify a potential condition for crystallization on regular basis, but unfortunately, no crystals were observed and no condition was found that gave a promising starting point for optimization of the crystallization. Repeating the experiment with the crystallization screens (Cryos, Classic and JCSG+), by using a new purification batch of Sbp, also did not result in successful crystal formation.

4.1.10. SAXS measurements of tag free Sbp

Monodisperse solutions of purified Sbp with 6 different concentrations (1 mg/ml, 2 mg/ml, 4 mg/ml, 6 mg/ml, 8 mg/ml and 16 mg/ml) were applied to SAXS measurements, as described in section 4.1.5. The data were radially averaged to produce 1D-scattering profiles (Franke *et al*., 2012) and compared to each other to assess potential radiation damage (Franke *et al*., 2015). Only those scattering profiles without measurable damage were averaged. Scattering from the matched solvent blank was subtracted from the sample scattering to generate the final SAXS profiles of the protein in solution (figure 25). The scattering intensity [$I(s)$] of

each sample was investigated up to a scattering vector, or momentum transfer, of $s = 3.5\ nm^{-1}$ ($s = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle and λ, the X-ray wavelength was 0.124 nm.

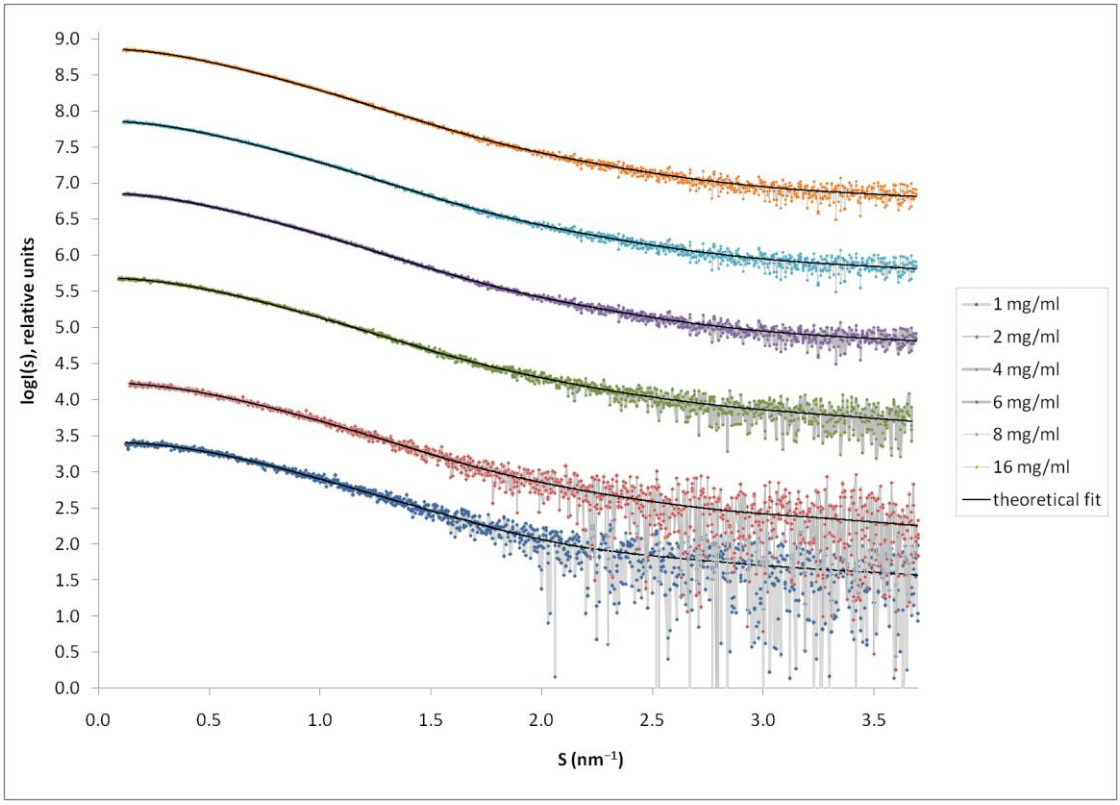

Figure 25: Small angle X-ray scattering data of 6 different concentrations (1mg/ml, 2 mg/ml, 4 mg/ml, 6 mg/ml, 8 mg/ml and 16 mg/ml) of Sbp are represented in the form of a plot of scattering intensity $I(s)$ versus scattering vectors (s), showing angle dependent changes in scattering intensity. The calculated fits, along the data points of each concentration of Sbp, are generated by *GNOM* in the *PRIMUSqt* software package and are represented as black curves. Note: For display purposes only, the up-scaling of data has been performed by adding a factor of 1 along scattering intensity data of each concentration to show all fits on the same plot.

The radii of gyration (R_g) of the particles were determined by Guinier analysis using *AutoRg* implemented in *PRIMUSqt* (software package). The linear section of the Guinier plot [$\ln I(s)$ versus s^2] at very low angles ($sR_g < 1.3$) is useful, as the slope of the plot relates to R_g, while the extrapolated intensity to zero angle, $I(0)$, relates to the particle volume and hence the molecular weight (MW). In addition, *GNOM* (Svergun. 1992) was used to calculate the probable real-space atom pair distribution profile [$P(R)$ vs R], from which the real-space R_g

and maximum dimensions (D_{max}) were determined, by computing the inverse indirect fourier transform of the SAXS data [see figure 26 (B)].

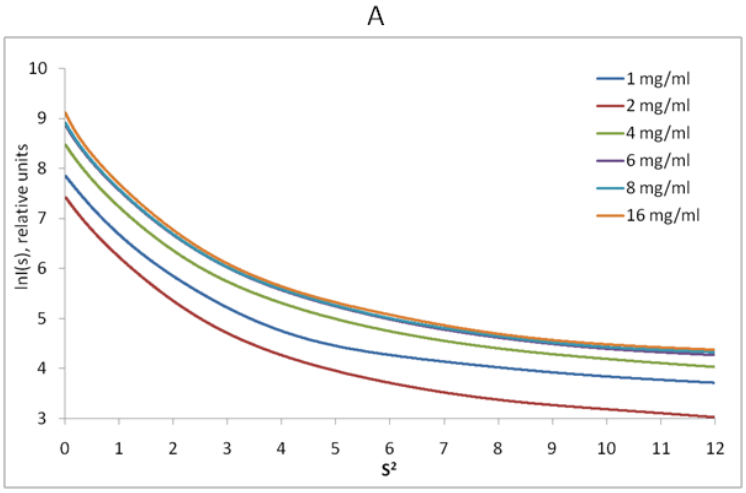

Figure 26: (A) Guinier plot analysis [ln$I(s)$ vs s^2] of Sbp with increasing concentrations from 1 mg/ml to 16 mg/ml**.**

The distance distribution profile represents the frequency of distances between scattering centres within the protein (electrons) which are, in effect, the probable frequency of distances between all atom pairs and this provides information about the shape of the protein in solution.

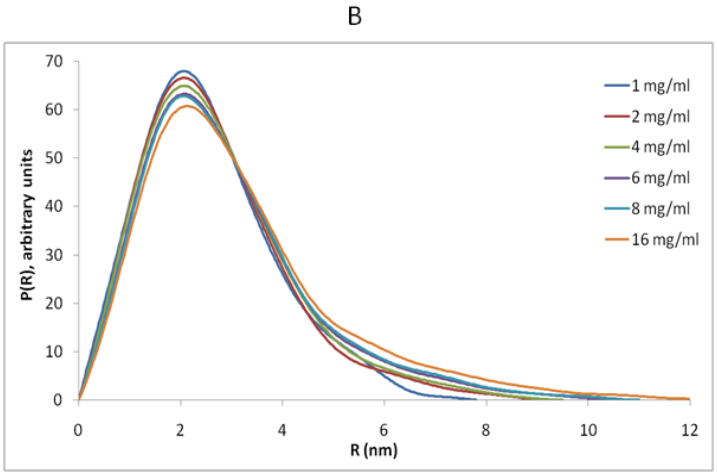

Figure 26 (B) Pair-wise distance real-space distribution function, $P(R)$, of Sbp.

Concentration dependent oligomerization /aggregation is observed in the scattering data. The R_g of Sbp calculated from Guinier plot and $P(R)$ show a gradual increase with increasing sample concentration [see figure 26 (A)]. The $P(R)$ profile shown in figure 26 (B) also represent an increase in D_{max} of samples from lower to higher concentrations. The approximate molecular weight of each sample was calculated from the concentration and scattering intensity of respective samples after standardizing against bovine serum albumin (BSA). The MW of 15 kDa was estimated from forward scattering of the sample with 1 mg/ml concentration. The rest of the concentrations (2–16 mg/ml) showed an increase in MW from 17–36 kDa, as shown in table 18. These observations suggest that Sbp has a propensity to self-assemble into high molecular weight oligomers.

Table 18: Statistics obtained from SAXS measurements of Sbp

Concentration used (mg/ml)	***Rg* (nm) Guinier Plot**	***Rg*(nm) *P(R)***	**MW (kDa) from *I(0)***	***D_{max}* from *P(R)***
1	1.98	2.04	15	7.7
2	2.04	2.13	17	9.0
4	2.09	2.22	19	9.5
6	2.19	2.36	27	10.5
8	2.22	2.41	29	11.0
16	2.37	2.62	36	12.0

A low-resolution surface structure of the tag free Sbp monomer was generated as an *ab initio* model from 2 mg/ml sample, as shown in figure 27.

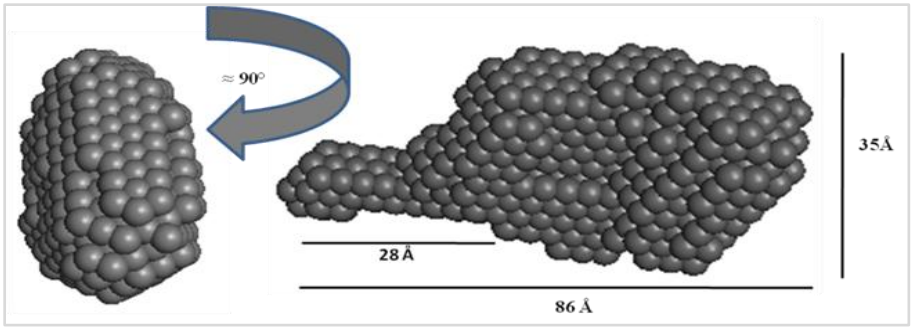

Figure 27: The average and volume corrected *ab initio* model of the Sbp monomer observed from *DAMMFILT* generated by *DAMMIF*.

The *ab initio* models of Sbp with 6 x His tag and tag free Sbp were superimposed over each other to compare their shapes. Both models were comparable to each other. However, the elongated tail-like region of Sbp with tag and linker region (violet bead model) was comparatively longer than the Sbp without tag (grey bead model), as shown in figure 28.

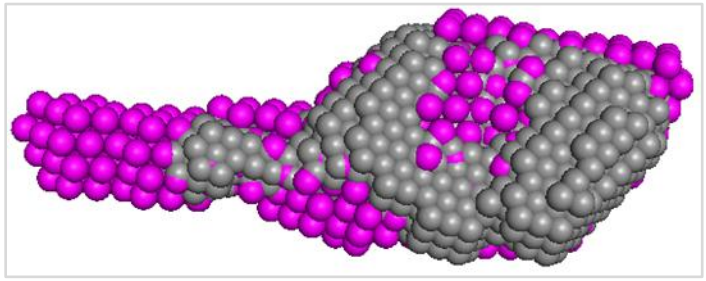

Figure 28: The superimposition of an I-Tasser predicted homology model of Sbp onto an *ab initio* model of Sbp, using *SUPCOMB* (Kozin and Svergun, 2001).

4.1.11. Native mass spectrometry

In order to analyze the aggregation/ oligomerization behavior of Sbp, the molecular mass of pure Sbp (10 µM) was determined by native MS in 250 mM ammonium acetate at pH 7.4. The obtained spectrum in figure 29, upper panel showed that there is a main species causing ~ 80 % of the signal and a mass of 16.9 kDa of Sbp representing monomer. However, there is also a dimeric form observed constituting ~20 % of 33.8 kDa of Sbp.

The lower panel in figure 29 shows a spectrum of a sample that has a three-fold (30 µM) increase of concentration compared to the one in the upper panel. It can be observed that the percentage of dimer species increases and some species of trimeric form of Sbp are also observed. The non-specific oligomerization of Sbp can be a result of a high concentration of the protein, which leads to clustering of molecules in the droplets, a common ESI artefact at concentrations higher than 20 µM. The other possibility is that Sbp oligomerizes with low affinity at increasing protein concentration. The origin of this dimer signal is inconclusive, due to jumping values in concentration determination via absorption spectroscopy. However, concentration dependent oligomerization in Sbp could also be observed in SAXS experiments, as described in section 4.1.5 and 4.1.10.

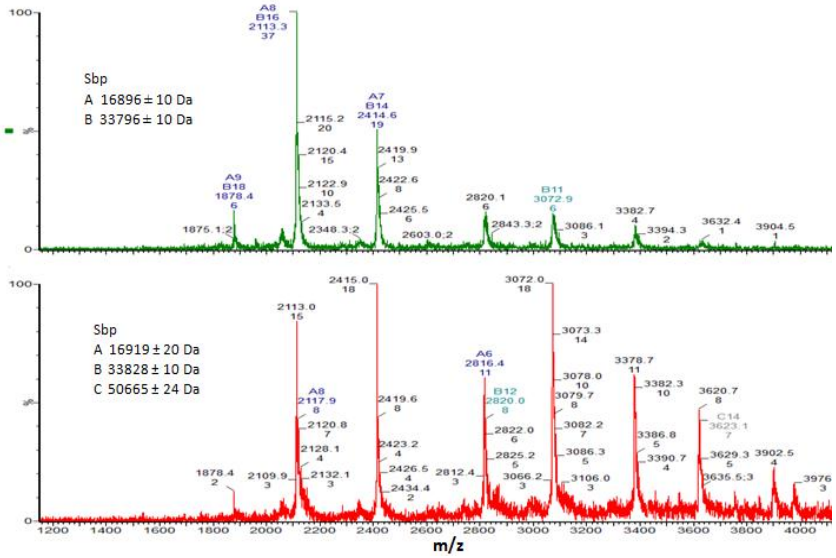

Figure 29: Native MS of Sbp. In the upper panel, the most abundant state of Sbp is a monomer. The lower panel represents the mass species of Sbp when the concentration was increased 3 folds, which was 30 µM. There are dimeric and trimeric species observed, but their abundance is very low and can be a result of crowding effect.

4.2. Accumulation associated protein

In order to achieve the second aim of the project, three constructs of sub domain G5 of Aap domain B ($G5_1EG5_2EG5_3$, $G5_1EG5_2$ and $G5_1E$) were cloned, expressed and characterized. The interaction analysis between Aap and Sbp was followed further.

4.2.1. Cloning of *aap* G5 constructs - $G5_1EG5_2EG5_3$, $G5_1EG5_2$ and $G5_1E$ in pET302-NT His vector

The PCR amplified products of three constructs of sub domain G5 of *aap*, domain B ($G5_1EG5_2EG5_3$, $G5_1EG5_2$ and $G5_1E$), using gene specific primers, were cloned into a linearized pET 302 NT-His vector applying the Gibson cloning method described in section 3.2.2. The N-terminus of each construct was supplemented with a 6 x His-tag and a TEV protease recognition cleavage site (ENLYFQG) during the synthesis of primers. Transformations of the clones were done in BL21 Star cells and plasmids were isolated by the Qiagen Plasmid Mini Kit (Qiagen, Hilden, Germany). The PCR amplification of the cloned constructs of *aap* G5 constructs using gene specific primers showed that all the bands corresponded well to their expected sizes. The amplified gene $G5_1EG5_2EG5_3$ appeared at 969 bp, $G5_1EG5_2$ at 591 bp and $G5_1E$ was at 387 bp when compared with DNA size marker, as shown in figure 30. Sequencing of the respective plasmids also confirmed successful cloning of all constructs.

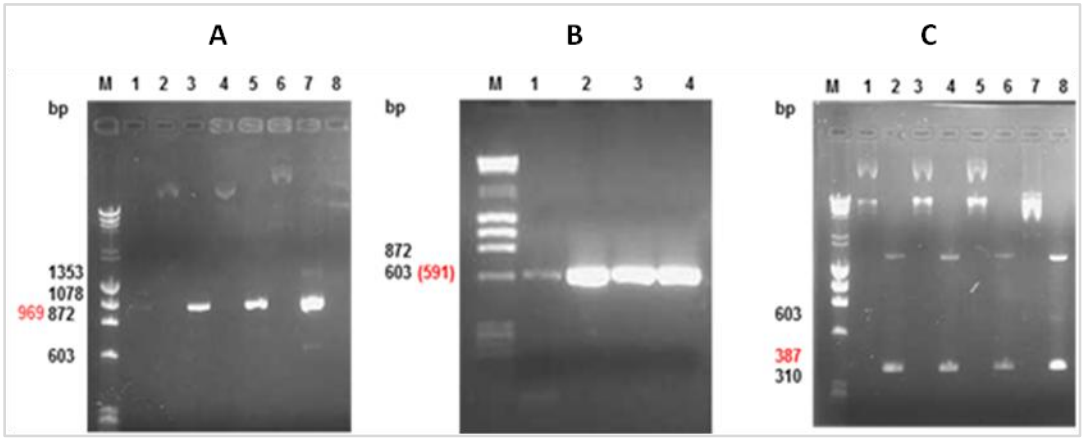

Figure 30: Gel electrophoresis of PCR amplified Aap constructs. (A) Lane M represents the marker, lane 1 is negative control and lanes 2, 4, 6 and 8 are plasmids of different clones of the construct $G5_1EG5_2EG5_3$, lanes 3, 5 and 7 represent the PCR amplification of $G5_1EG5_2EG5_3$ gene from clones. **(B)** Lanes 2-4 show the PCR amplified products of $G5_1EG5_2$. **(C)** Lanes 1, 3, 5 and 7 represent cloned plasmids before amplification. Lanes 2, 4 and 6 represent the PCR amplification of the $G5_1E$ gene.

4.2.2. Test expressions

In order to analyze the optimal cell density for induction of expression and the optimal duration of expression for all three recombinant G5 constructs (rG5), test expression experiments were performed, as described in section 3.3.1, but the temperature was set to 20 °C after induction of gene expression by IPTG.

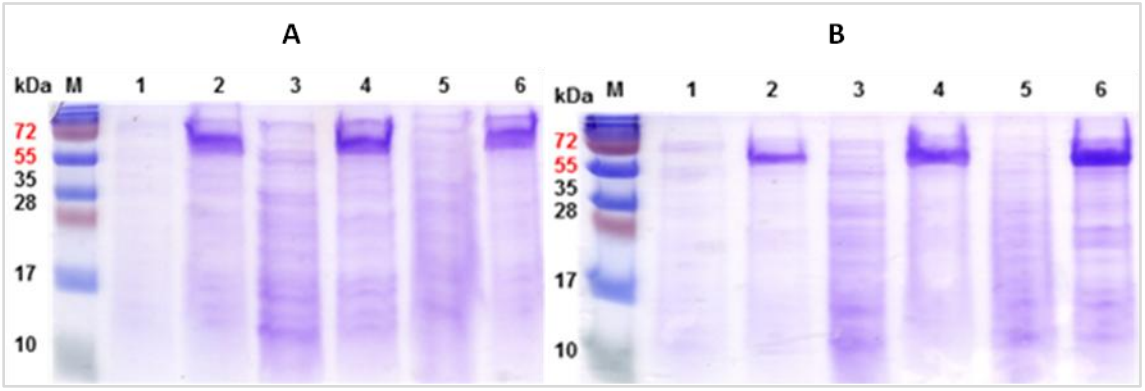

Figure 31: SDS PAGE analysis of test expression of $rG5_1EG5_2EG5_3$. (A) Lane M represents the marker, lanes 1, 3, 5: non-induced (NI) samples when the OD_{600} value reached 0.5, lanes 2, 4, 6: the expression of $rG5_1EG5_2EG5_3$ after induction of samples when the OD_{600} value reached 0.5 after 2 hrs, 3 hrs and 4 hrs respectively. **(B)** Lane M represents marker, Lanes 1, 3, 5: NI samples when the OD_{600} value reached 0.9 after 2 hrs, 3 hrs and 4 hrs respectively, Lanes 2, 4, 6: the expression of $rG5_1EG5_2EG5_3$ after induction of samples when the OD_{600} value reached 0.9 after 2 hrs, 3 hrs and 4 hrs respectively.

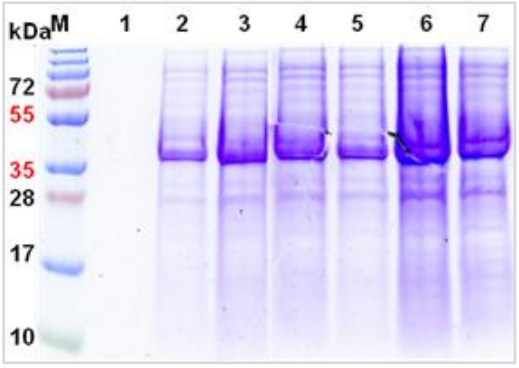

Figure 32: SDS PAGE analysis of test expression of $rG5_1EG5_2$. Lane M shows the marker, lane 1 is NI sample, Lanes 2-4: the expression of $rG5_1EG5_2$ after induction of samples when the OD_{600} value reached 0.5 lanes 5-7: the expression of $rG5_1EG5_2$ after induction of samples when the OD_{600} value reached 0.9 after 2 hrs, 3 hrs and 4 hrs respectively.

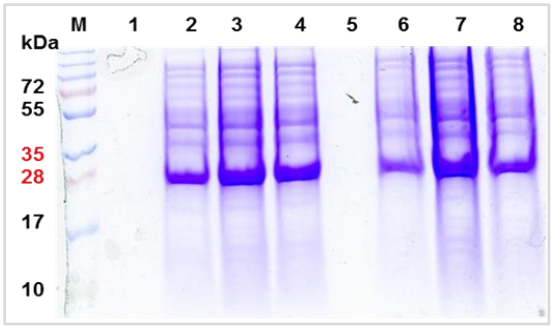

Figure 33: SDS PAGE analysis of test expression of rG5$_1$E. Lane M shows the marker. Lane1 and 5 represent NI samples when the OD_{600} value reached 0.5 and 0.9 respectively, lanes 2-4: the expression of rG5$_1$E after induction of samples when the OD_{600} value reached 0.5 after 2 hrs, 3 hrs and 4 hrs respectively, lanes 5-8: the expression of rG5$_1$E after induction of samples when the OD_{600} value reached 0.9 after 2 hrs, 3 hrs and 4 hrs respectively.

Surprisingly and interestingly, the expressed genes appeared at higher molecular size on SDS PAGE than expected, calculated theoretically from cloned gene sequences. For example, the molecular weight of rG5$_1$EG5$_2$EG5$_3$ along with 6 x His-tag and sequence of TEV protease cleavage site was predicted to be 37 kDa, according to the EXPASY program, ProtParam (Gasteiger, E *et al*., 2005). However, the protein band was observed at ~60 kDa on SDS PAGE, as shown in figure 31. rG5$_1$EG5$_2$ should appear at a molecular weight of 23 kDa and G5$_1$E should have a molecular weight of 16 kDa, while both appeared at higher molecular weight of ~37 kDa and ~21 kDa, respectively, as shown in figure 32 and 33. However, the maximum expression of rG5$_1$EG5$_2$EG5$_3$ was obtained 4 hrs after induction, when OD_{600} value reached 0.9 as shown in figure 31 (B) lane 6. rG5$_1$EG5$_2$ showed maximum expression level 3 hrs after induction when OD_{600} value reached 0.9 as shown in figure 32 lane 6. The expression of rG5$_1$E was also maximum 3 hrs after induction when OD_{600} value reached 0.9 as shown in Figure 33 lane 7. However, the expression of rG5$_1$E was good and same in almost every condition.

4.2.3. Purification and characterization

In order to confirm the identities of all rG5 constructs, they were purified by affinity chromatography using the His-tag, as described in the section 3.3.3. In order to remove the 6 x His-tag from each construct, cleavage reaction was performed, as described in section

3.3.6. TEV protease and tag were removed from the protein solution by affinity chromatography.

Each purified tag free protein sample was concentrated and size exclusion chromatography was performed in order to fractionize the samples according to size and to identify the molecular weight of the constructs, as described in section 3.3.5. The eluted fractions obtained for each construct were analyzed by 12 % SDS PAGE, as shown in figures below.

The construct $G5_1EG5_2EG5_3$ eluted at 12.3 ml which corresponds to 160 kDa molecular weight, as shown in figure 34 (A). Figure 34 (B) shows SDS PAGE result of the eluted fractions, which were same showing a band shift to a higher molecular weight, as described in section 4.2.2 during test expression.

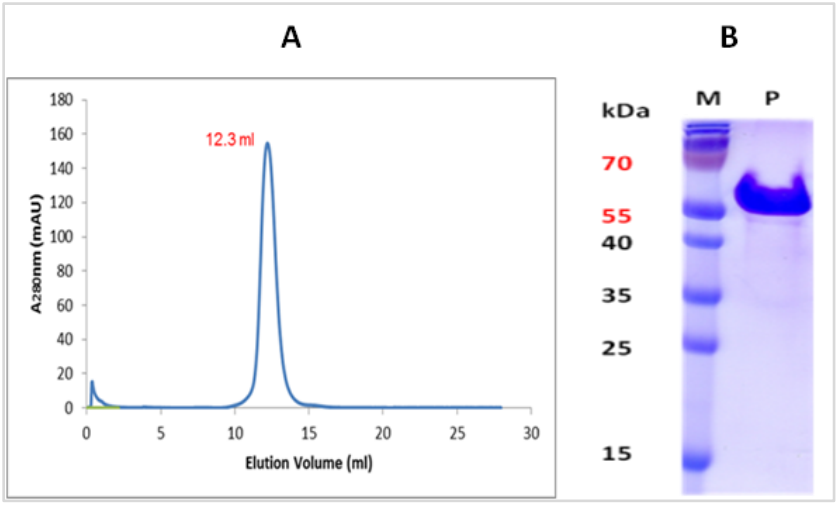

Figure 34: SEC and SDS PAGE analysis. (A) Chromatogram of $G5_1EG5_2EG5_3$. **(B)** SDS PAGE analysis of purified $G5_1EG5_2EG5_3$. M represents size marker, lane P is purified protein after SEC.

Figure 35 (A) shows the elution volume of $G5_1EG5_2$ through Superdex^{TM} 200 10/300GL column, which is 14.1 ml and corresponds to calculated molecular weight as 61 kDa. SDS PAGE results shown in figure 35 (B) were same, as described in section 4.2.2 during expression of respective protein.

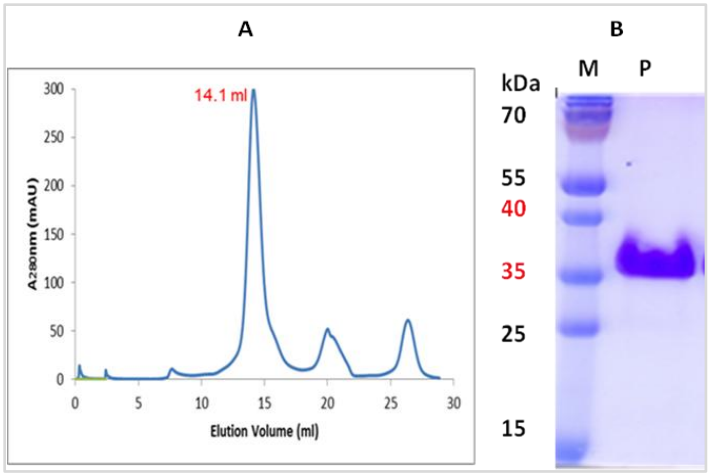

Figure 35: SEC and SDS PAGE analysis. (A) Superdex^{TM} 200 SEX profile of $G5_1EG5_2$. **(B)** SDS PAGE analysis of purified $G5_1EG5_2$. M represents size marker, Lane P represents purified $G5_1EG5_2$ after SEC.

$G5_1E$ eluted at 15.6 ml volume of the column [shown in figure 36 (A)], which corresponds to 21 kDa when compared to and calculated from the standard calibrated proteins, as described in section 4.1.1. The SDS PAGE analysis of eluted sample [figure 36 (B)] was same as described in section 4.2.2 during expression.

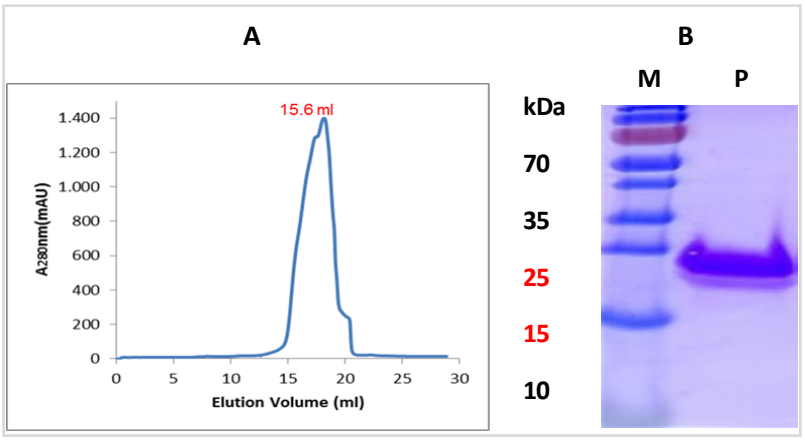

Figure 36: SEC and SDS PAGE analysis. (A) Superdex^{TM} 200 SEX profile of $G5_1E$**. (B)** SDS PAGE analysis of purified $G5_1E$. M represents size marker, Lane P is purified $G5_1E$ after SEC.

This anomalous behavior of G5 constructs was further checked, as described later.

The presence of purified G5 proteins was confirmed by western blot analysis. As all three constructs are the parts of domain B of Aap, a single antibody Rabbit anti-rDomB was used as primary antibody, which was detected by peroxidase labeled secondary antibody. The samples were blotted onto PVDF membrane from SDS PAGE gels. The membrane was treated with primary and secondary antibodies described in section 3.3.11. The signals of the respective bands of proteins $G5_1EG5_2EG5_3$, $G5_1EG5_2$ and $G5_1E$ were detected on X-ray films shown in figure 37. The detection of bands by western blot showed that the purified proteins shown in figure 34 (B), 35 (B) and 36 (B) of section 4.2.3, belong to Aap.

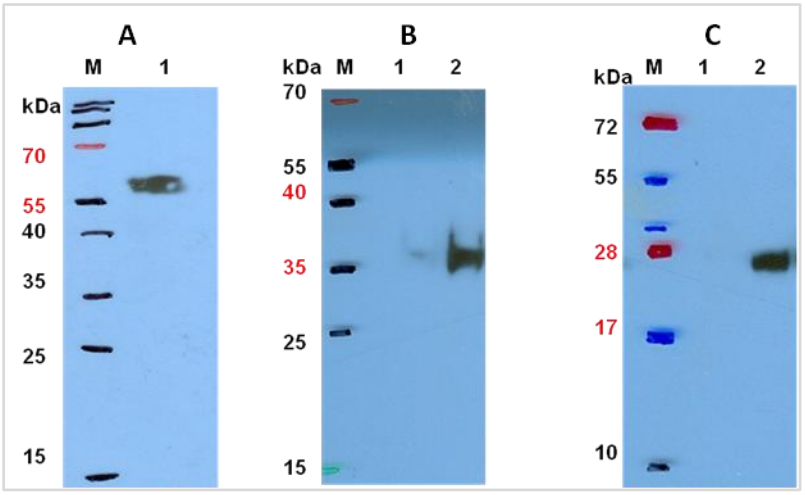

Figure 37: Western blot analysis of Aap constructs. X-ray film **(A)** Lane 1 depicts the chemiluminescence signal of bound antibody against $G5_1EG5_2EG5_3$, **(B)** lane 2 represents the detection of $G5_1EG5_2$ and **(C)** shows the presence of $G5_1E$.

In order to investigate the approximate sizes and oligomeric states of all the three rG5 proteins and to compare them with the results obtained from SDS PAGE, blue native PAGE was performed. The native state of proteins remains conserved during native-PAGE. The result showed that for $G5_1E$ and $G5_1EG5_2$ two distinct bands were observed, while $G5_1EG5_2EG5_3$ appears as a single compact band on native gel, as shown in figure 38.

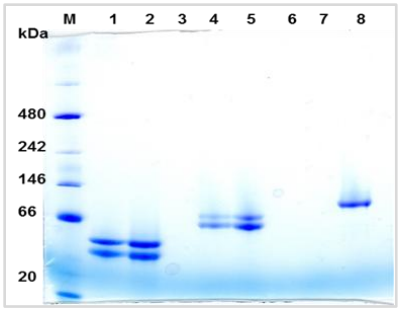

Figure 38: Blue native-PAGE of Aap constructs. Lane M stands as a reference marker. Lanes 1 and 2 show the purified $G5_1E$, lanes 4 and 5 represent the purified samples of $G5_1EG5_2$ and lane 8 represent purified $G5_1EG5_2EG5_3$.

4.2.4. Identification of Aap rG5 proteins

Each protein construct digested by trypsin resulted into a number of peptides, which were identified to be present in the original sequence of each construct. The cleavage of $G5_1EG5_2EG5_3$ construct by trypsin resulted into a total number of 34 peptides. These peptides were aligned to the original amino acid composition of the construct using CLC workbench software, as shown in the figure 39.

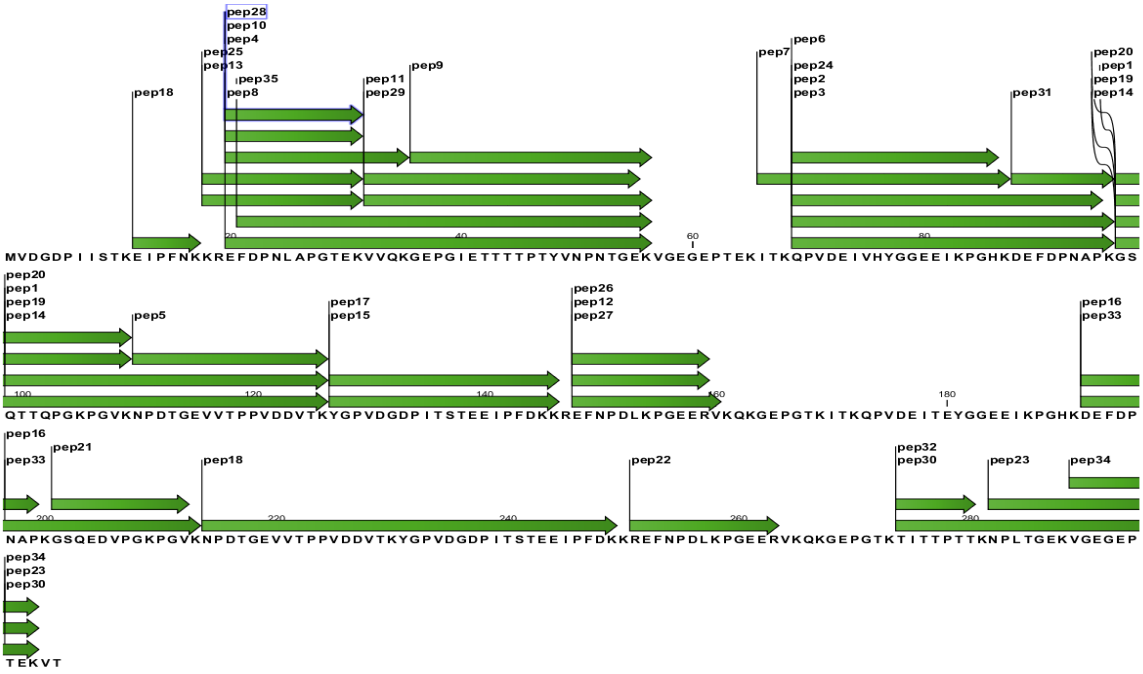

Figure 39: Peptide digestion of $G5_1EG5_2EG5_3$ for identification by mass spectrometry analysis. The identified peptides shown as green arrows are found within the amino acid sequence of $G5_1EG5_2EG5_3$.

Likewise, all the peptides obtained as a result of tryptic digestion of $rG5_1EG5_2$ and $rG5_1E$ were also identified to be present in the amino acid sequence of Aap.

The high mass accuracy of native MS was exploited to determine the molecular masses of G5 constructs in 250 mM ammonium acetate at pH 7.4. After purification, 6x His tag cleavage by TEV protease and removal of tag from each protein were carried out. The proteins were applied to ESI, as described in section 3.4.2. The obtained spectra, representing the relative abundance normalized to the base peak of charged proteins at particular m/z values, showed monomeric states for all constructs. The mass of each construct was determined from all peaks in the charge envelope.

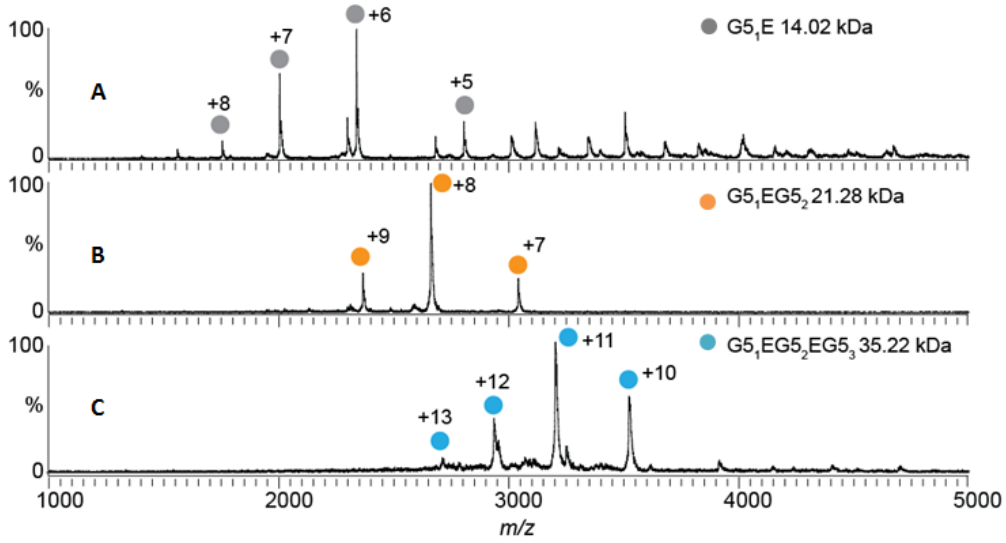

Figure 40: Native mass spectra of Aap constructs. Panel **(A)** represents the individual mass species of $G5_1E$ with 5-8 positive charges revealing monomeric state of protein with a mass of 14 kDa**. (B)** Panel B refers to the monomeric species of $G5_1EG5_2$. The molecular mass of $G5_1EG5_2$ was 21.3 kDa. **(C)** Referring to panel C, the calculated molecular mass of $G5_1EG5_2EG5_3$ was 35.2 kDa.

The calculated masses of each construct of Aap fits well to the theoretical masses of the respective proteins calculated by ProtPram (Gasteiger *et al.*, 2005), according to the amino acid sequence of the expressed genes. The above experimental results confirmed that molecular cloning of the respective genes was successful. The expression of the respective proteins was also up to the mark despite of the fact that these proteins appeared at higher molecular size on SDS PAGE and gel filtration.

4.2.5. Biochemical characterization of Aap G5 proteins

Purified Aap G5 constructs were subjected to DLS for checking their monodispersive nature under the given buffer conditions. The strong signals suggested that most of the part of each protein construct was monodisperse.

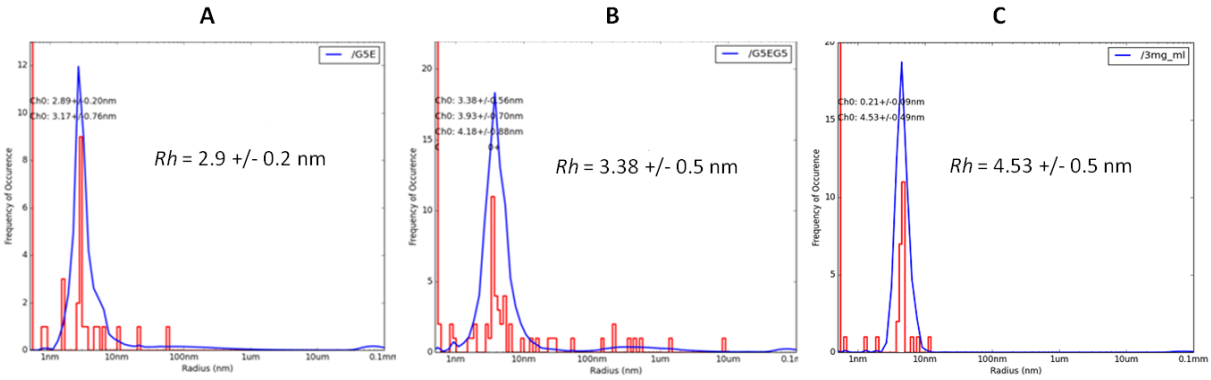

Figure 41: DLS measurements showing monodispersive nature of protein solutions with hydrodynamic radii as 2.9, 3.38 and 4.53 nm of $G5_1E$, $G5_1EG5_2$ and $G5_1EG5_2EG5_3$ in **(A)**, **(B)** and **(C)** respectively.

To estimate the secondary structure of G5 constructs, CD spectrometry measurement was performed for each protein sample in solution form CD spectrometry indicated the presence of β-sheets in all the three G5 constructs. The minima at 198 nm wavelength (figure 42) in each construct are representatives of strong β-sheets.

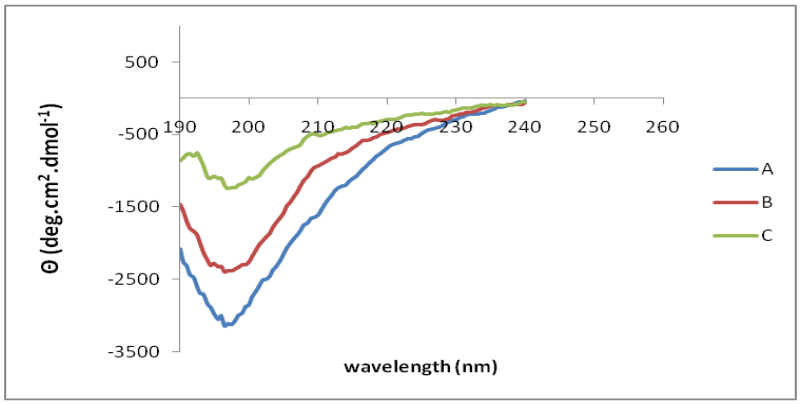

Figure 42: CD spectra of $rG5_1E$, $rG5_1EG5_2$ and $rG5_1EG5_2EG5_3$ represented as **(A)** blue, **(B)** red and **(C)** green, respectively

4.2.6. Sequence homology and structural alignment of G5 protein constructs

The crystal structure of a segment of a homologous protein SasG, $G5_1EG5_2$ (PDB ID: 4FUO), from *S. aureus* has already been solved (Conrady *et al*., 2013). In order to determine the secondary structure in G5 protein constructs, the multiple sequence alignment of G5 constructs and 4FUO was done using clustal omega. The secondary structure elements were predicted using EsPript tool.

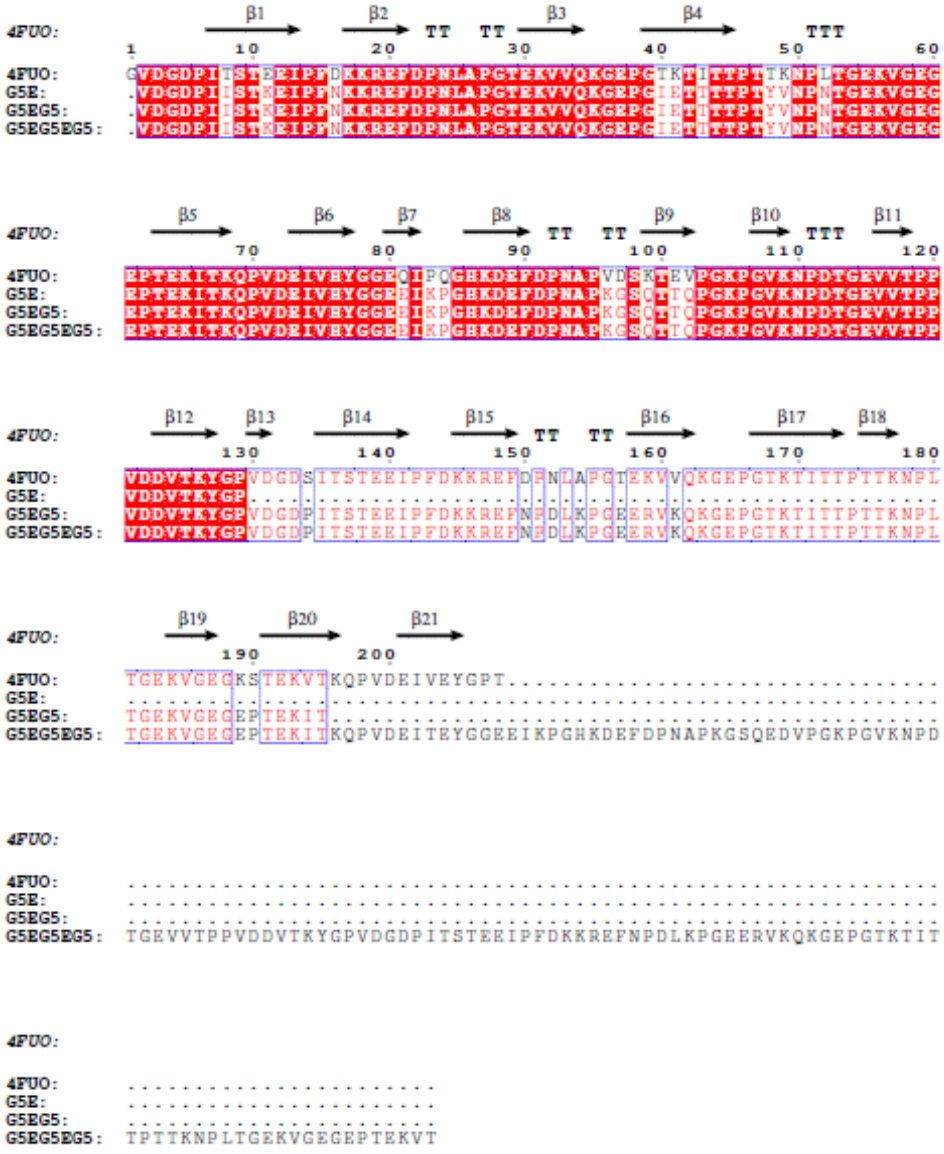

Figure 43: The amino acid sequence alignment of $G5_1E$, $G5_1EG5_2$ and $G5_1EG5_2EG5_3$ with $G5_1EG5_2$ (PDB ID: 4FUO). The red boxes represent the sequence identity among proteins. The arrow heads in

black colour show β-sheets and Ts represent turns. There are no alpha helices in the structure of proteins.

4.3. Crystallization of $G5_1EG5_2EG5_3$

Pre crystallization test (Hampton Research) was performed to optimize the concentration of G5 proteins for the most suitable crystallization condition. Initial screening of the concentrated protein was performed against a diversity of available screening conditions using the honeybee robot, applying the sitting drop method. A 400 nl of each protein sample (concentrated in the range of 20-25 mg/ml in the presence and absence of 1mM $ZnCl_2$, separately) was mixed with the same volume of precipitant solution and 45 µl of precipitant solution was used as reservoir. 480 different conditions were set in 5 Nextal Qial plates (Qiagen, Germany), as described in section 3.4.8. The plates were placed at RT and analyzed every week by a compound microscope. After 7 weeks, a brick like $G5_1EG5_2EG5_3$ crystal (shown in figure 44) appeared in condition A9 (see table 19) of the Morpheus suite in the absence of $ZnCl_2$. Some crystals were also observed in condition E5 (see table 19) of the PACT suite.

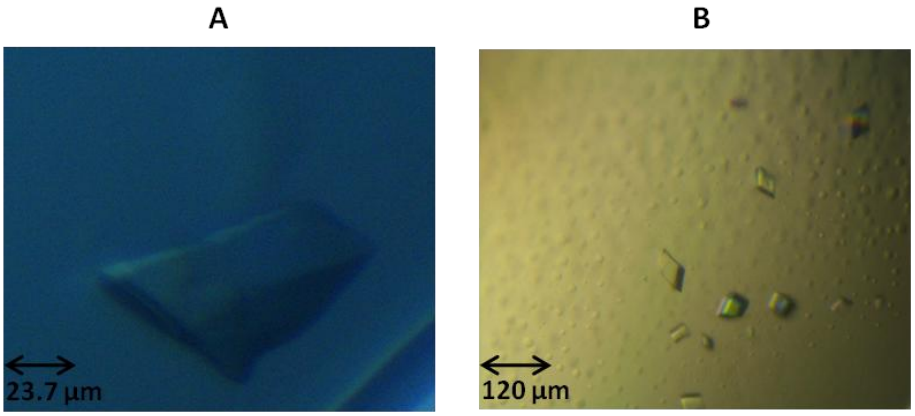

Figure 44: (A) $G5_1EG5_2EG5_3$ crystal appeared after 7 weeks in condition A9 of the Morpheus suite diffracted up to 3.5 Å. The crystal dimension was 60 x 89 x 97 µm. **(B)** Crystals in condition E5 of the PACT suite.

Table 19: The conditions A9 of the Morpheus suite and E5 of the PACT suite in which the crystals were grown, their compositions and the sizes of obtained crystals.

Fig. 44	Condition	Constituents	Dimension
A	Morpheus-A9	0.1 M bicine/trizma base pH=8.5, 0.03 M $CaCl_2$, 0.03 M $MgCl_2$, 10% w/v PEG 20,000, 20% v/v PEG MME550	60 x 89 x 97µm
B	PACT-E5	0.2 M Na citrate, 20 % PEG3350	50-130 µm

After two more weeks more crystals appeared in the conditions A8 and A12 of the Morpheus suite. The crystal dimensions were in the range of 50 to 170 microns, shown in figure 45 (A) and (B).

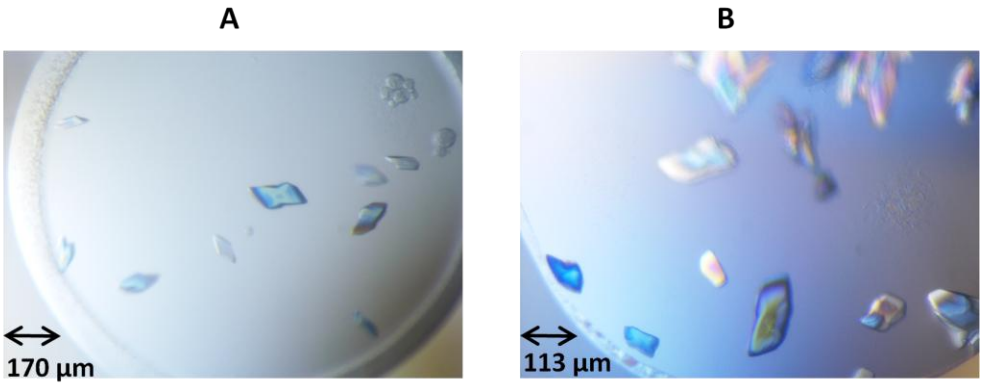

Figure 45: The shown crystals diffracted up to 2.7 Å.

Table 20: The conditions A8 and A12 of the Morpheus suite in which the crystals were grown, their composition and size of crystals.

Fig. 45	Condition	Constituents	Dimensions
A	Morpheus- A8	0.1M MOPS/HEPES pH=7.5, 12.5%w/v PEG1000, 12.5% w/v PEG3350, 12.5% v/v MPD, 0.03M $MgCl_2$& 0.03M $CaCl_2$	50-170 µm
B	Morpheus- A12	0.1M bicine/trizma base pH=8.5, 12.5% w/v PEG1000, 12.5% w/v PEG 3350, 12.5% v/v MPD 0.03M $CaCl_2$, 0.03M $MgCl_2$	50-170 µm

Note: Crystals were obtained only for the $G5_1EG5_2EG5_3$ protein. The other two G5 proteins $G5_1EG5_2$ and $G5_1E$ did not crystallize.

In order to obtain more crystals, further optimization of the conditions mentioned above was performed by varying the concentrations of the constituents in 24 well maxi plates, using the hanging drop method. In this experiment, crystallization drops were made using the purified tag free $G5_1EG5_2EG5$ and 6 x His tagged protein in separate crystallization set ups. However, the crystals were grown for His tagged $G5_1EG5_2EG5$ protein, as obtained previously. There were no tag free protein crystals observed in the same crystallization set ups. Figure 46 (A) and (B) represent the crystallization drops with the same constituent composition. The crystals of $G5_1EG5_2EG5_3$ with 6 x tag were grown [see figure 46 (A) and (B)].

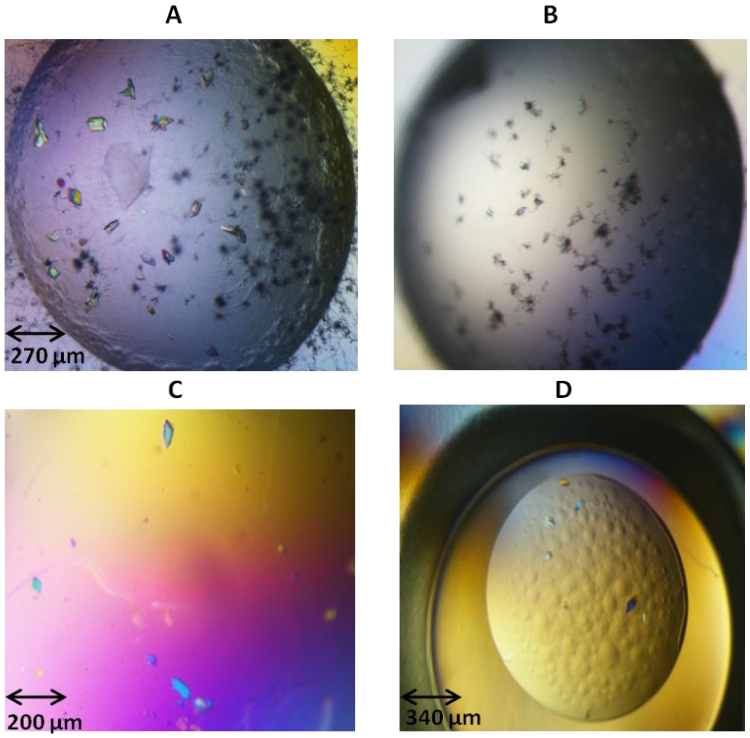

Figure 46 (A) $G5_1EG5_2EG5_3$ crystals grown in five weeks in the modified Morpheus-A9 condition. **(B)** The drop with the same condition but tag-free $G5_1EG5_2EG5_3$. No crystals observed. **(C)** $G5_1EG5_2EG5_3$ crystals in modified Morpheus-A8. **(D)** Crystals in modified Pact-E5 condition.

Table 21: Optimized conditions for crystallization

Fig. 46	Optimized Condition	Constituents	Dimensions
A & B	Morpheus- A9	0.1 M bicine/trizma base pH=8.5, 0.03 M $CaCl_2$, 0.03 M $MgCl_2$, 9.5 % w/v PEG 20,000, 20 % v/v PEG MME550	60-150 µm
C	Morpheus- A8	0.1 M MOPS/HEPES pH=7.5, 14 %w/v PEG1000, 12.5 % w/v PEG3350, 12.5 % v/v MPD, 0.03 M $MgCl_2$& 0.03 M $CaCl_2$	50-160 µm
D	PACT-E5	0.2 M Na citrate, 17.5 % PEG3350	50-130 µm

4.3.1. Data collection and processing

One of the two crystals with a size of 60 x 89 x 97 µm (from A9 condition, described in table 19, of the Morpheus suite) diffracted up to 3.5 Å resolution at the beamline P13, DESY, Hamburg. From the crystals in the A8 condition of Morpheus suite, the diffraction data up to 2.7 Å resolution were collected at 100 K at the beamline P13, DESY, Hamburg. PEG, which was already present in the crystallization condition (mother liquor) acted as cryo-protectant and prevented the formation of ice rings. The data frames were processed by XDS program (Kabsch, 2010) and an mtz file was created. The crystals belonged to orthorhombic group with unit cell parameters a = 52.28, b = 70.90, c = 235.94 Å corresponding to the space group $P2_12_12_1$. The Matthews Coefficient V_M (Matthews, 1968) indicated the presence of two molecules in the asymmetric unit.

Table 22: Statistics of data collection of $G5_1EG5_2EG5_3$crystal

Data collection	
Source	PETRA III P13
Wavelength (Å)	0.976
Temperature (K)	100
Space group	$P2_12_12_1$
Unit-cell parameters	

a, b, c (Å)	52.28, 70.90, 235.94
Resolution range (Å)	50– 2.7
Total reflections (outer shell)	327749 (48209)
Unique reflections	25061 (3594)
R_{merge} (%)	15.3 (67.3)
Average I/σ(I)	13.3 (3.3)
Multiplicity	13.1 (13.4)
Wilson's B factor	43.2
$CC_{1/2}$ (%)	99.9 (97)
Completeness (%)	100 (100)

To obtain phase information, Molecular Replacement (MR) was used applying Molrep (Vagin *et al*., 1997) and Phaser (McCoy *et al*., 2007), from CCP4i suite (Collaborative Computational Project, Number 4, 1994). Unfortunately, MR failed till now and phase information could not be obtained (see details in discussion). Next step was to solve the phase problem by Single Anomalous Dispersion (SAD), by introducing a heavy metal atom into the protein crystal.

Native PAGE and soaking experiments were done with different heavy metals including including platinum, mercury, europium, samarium, thallium, bromide and lead, but unfortunately no anomalous signal was observed till to date and the phase problem could not be solved.

4.4. SAXS measurements of G5 proteins

Aap G5 constructs were further characterized using SAXS. The distance distribution of SAXS data of $G5_1E$ indicated anR_g of 3 nm and D_{max} of 11.3 nm, as shown in figure 40 (A). The SAXS data of $G5_1EG5_2$showed anR_g of 4.2 nm and D_{max} of 16.5 nm, as represented in (B), while the R_g of $G5_1EG5_2EG5_3$is 6 nm and has a D_{max} of 26.9 nm, as shown in (C).

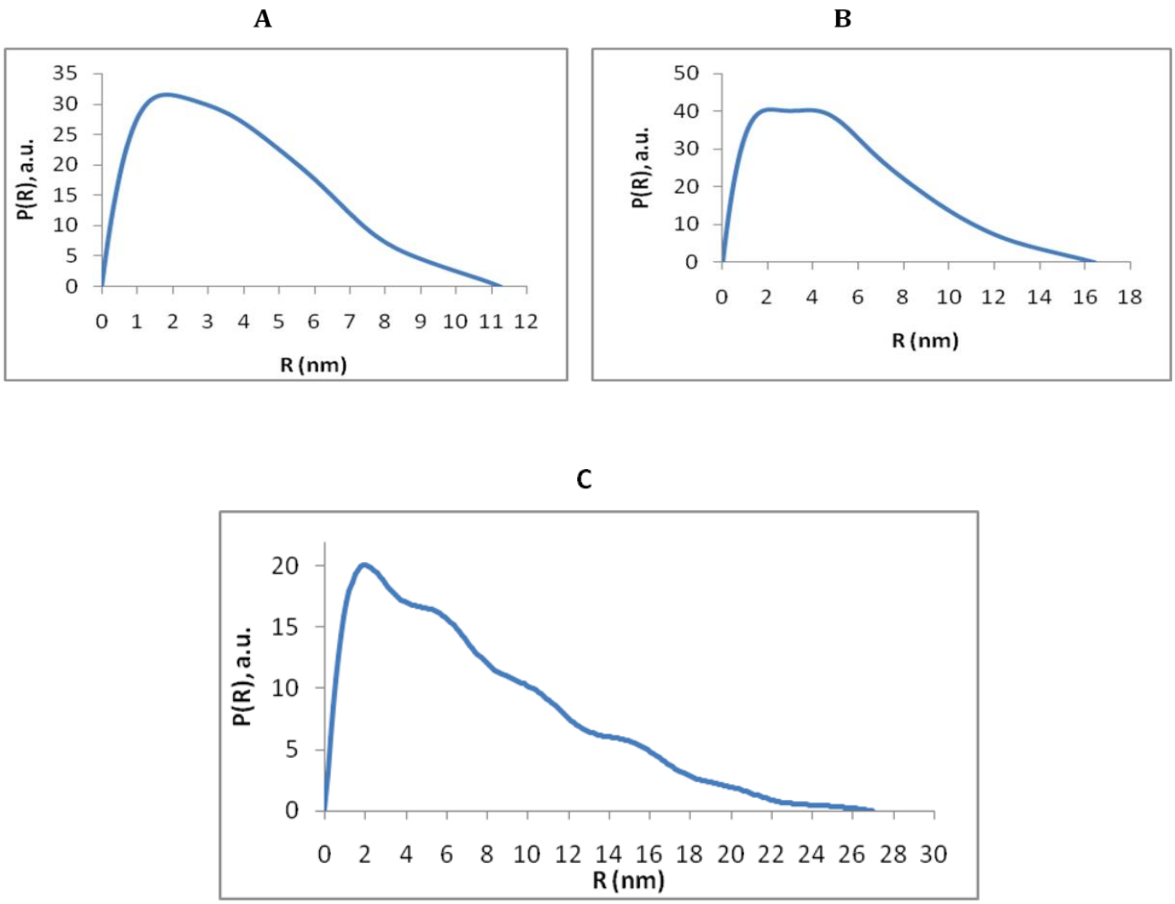

Figure 47: The *P*(*R*) profile of $G5_1E$, $G5_1EG5_2$ and $G5_1EG5_2EG5_3$ shown in (A), (B) and (C), respectively.

The *ab initio* models (shown in figure 48) of all the three Aap G5 proteins were generated using *DAMMIF*. At total of 10 bead models for each construct were modeled and the fitting of experimental data to their respective model indicated χ^2 values of 1. On spatial alignment and averaging, the NSD values were $G5_1E$ = 0.63; $G5_1EG5_2$ = 0.67 and $G5_1EG5_2EG5_3$ = 0.48. The results obtained from the extracted SAXS structural parameters, *P*(*R*) profile and modeling indicate a systematic increase in the R_g and length of the Aap G5 constructs as the number of domains increases. Overall Aap adopts a very elongated rod-like and aniosotropic structure in solution. The largest construct $G5_1EG5_2EG5_3$ extends to 21 nm, but is only ~3.1 nm in diameter (figure 48).

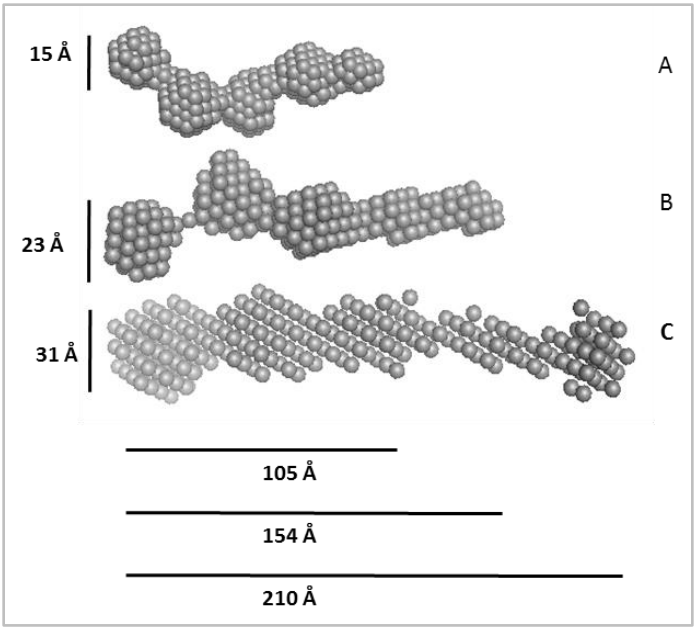

Figure 48: The *ab initio* models of **(A)** $G5_1E$ **(B)** $G5_1EG5_2$and **(C)** $G5_1EG5_2EG5_3$ represented as spheres.

4.5. Analysing interaction between Sbp and Aap G5 proteins

After purification of Sbp, $G5_1EG5_2EG5_3$, $G5_1EG5_2$and $G5_1E$, the interaction of each G5 construct with Sbp, in the presence of 1 mM $ZnCl_2$, was investigated by SEC. A molar ratio of 2:1 of Sbp and each G5 construct was used in the experiments. The obtained chromatograms showed that there is no interaction between Sbp and any of G5 constructs under these conditions. Sbp eluted at the volume corresponding to its respective molecular weight (see figures 49, 50 and 51) and each G5 construct eluted at the respective volume, as described in section 4.2.3.The chromatogram of Sbp with $G5_1EG5_2EG5_3$, shown in figure 49, revealed the presence of two distinct peaks. The left peak showed the elution of $G5_1EG5_2EG5_3$ at 12.3 ml, which is in good agreement with the result of SEC of pure $G5_1EG5_2EG5_3$ described in section 4.2.3. The peak at an elution volume of 16.8 ml is indicative of the presence of Sbp.

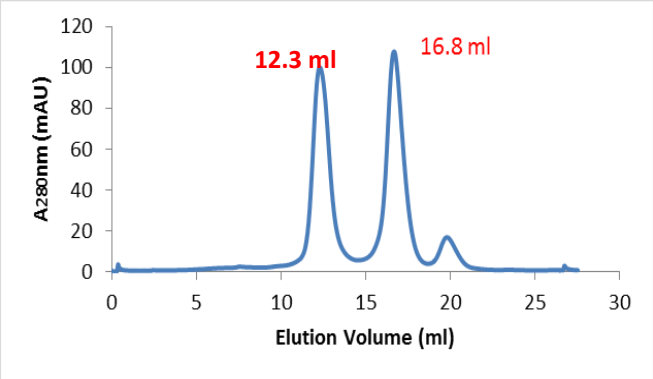

Figure 49: Superdex^{TM}200 SEC profile of mixture of Sbp and $G5_1EG5_2EG5_3$ in the presence of 1 mM $ZnCl_2$. The peak at 12.3 ml volume corresponds to $G5_1EG5_2EG5_3$, while the right peak (16.8 ml) corresponds to the molecular weight of Sbp (16.7 kDa).

The chromatogram in figure 50 shows that both proteins Sbp and $G5_1EG5_2$ elute at their respective sizes through Superdex^{TM} 200 10/300GL column. There is no additional peak observed to indicate the interaction between these two proteins.

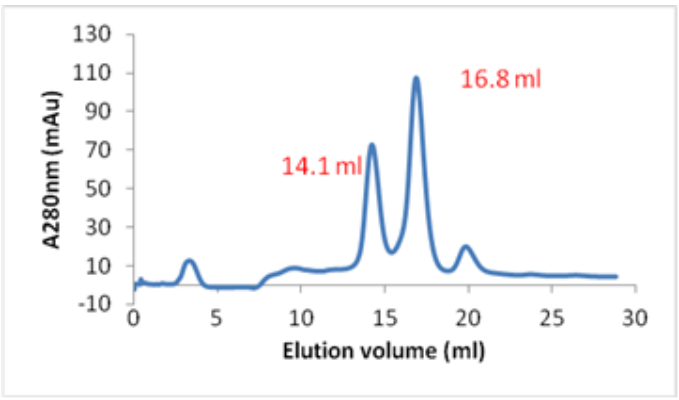

Figure 50: Superdex^{TM} 200 SEC profile of mixture of Sbp and $G5_1EG5_2$ in the presence of 1 mM $ZnCl_2$. The fractions of protein eluted at 14.1 ml volume correspond to $G5_1EG5_2$.The right peak corresponds to the molecular weight of Sbp.

The two peaks at the elution volumes 15.5 ml and 16.8 ml correspond to $G5_1E$ and Sbp respectively, as shown in figure 51.

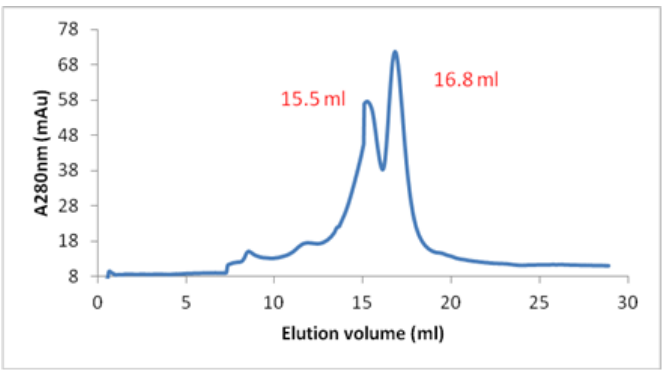

Figure 51: Superdex^{TM} 200 SEC profile of mixture of Sbp and $G5_1E$ in the presence of 1 mM $ZnCl_2$. The fractions of protein eluted at 15.5 ml volume correspond to $G5_1E$. The right peak corresponds to the molecular weight of Sbp.

4.5.1. Determination of interaction by native MS and MST

In order to determine the interaction, Sbp and $G5_1EG5_2EG5_3$ were examined separately from each other to determine their masses. The effect of Zn^{2+} on the oligomerization of each individual protein was also investigated. Figures 52 and 53 show the spectra of the individual proteins. Sbp has a mass of 16.9 kDa and is present as a monomer, as shown in figure 43. In the presence of 0.1 mM Zn (II) acetate (figure 52 upper panel), the main species is also the monomeric form of Sbp. The concentration of Sbp was 5 µM and10 V collisional energy was applied. In panel A, a peak broadening originating from Zinc can be observed; however, the peaks have baseline separation and can be clearly assigned to the respective mass species of Sbp. An influence of Zn^{2+}on the quaternary structure could not be determined at 5 µM.

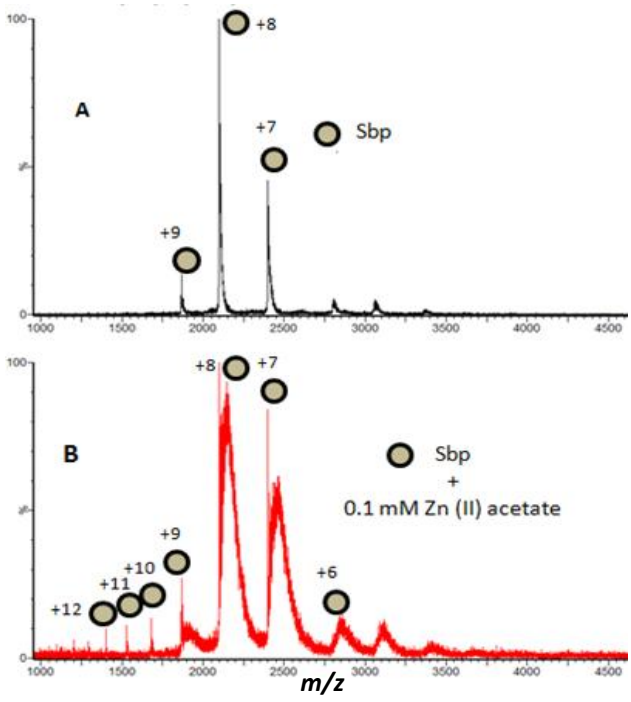

Figure 52: Native mass spectra of Sbp. Refering to upper panel A, the main species is a monomer. It shows that there is no influence of Zn^{2+}, when compared to the lower panel (Sbp in the presence of Zn (II) acetate). Both spectra represent peaks detected at the same *m/z* values.

The spectra of the individual measurements of $G5_1EG5_2EG5_3$ are presented in figure 53. This protein is also present as a monomer. It has a mass of ~35 kDa. Oligomerization has neither been observed in the absence (figure 53 panel A) nor in the presence (figure 53 panel B) of Zn^{2+}. The concentration of $G5_1EG5_2EG5_3$ was maintained as 5 μM.

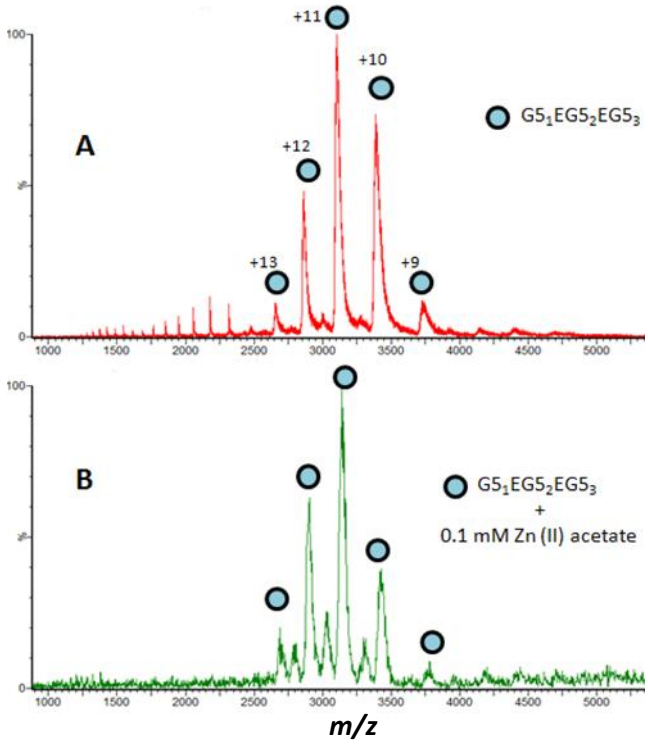

Figure 53: Native mass spectra of $G5_1EG5_2EG5_3$. The protein is mainly present as a monomer. It has a size of ~35 kDa. The distribution of peaks in both panels A and B (in the absence and presence of Zn (II) acetate respectively) is equivalent.

The binding assay of Sbp with $G5_1EG5_2EG5_3$ was performed under two pH conditions, pH 7.4 and pH 6.6 after mixing both proteins (1:1) and incubating them for 30 min at RT, as described in section 3.4.2. Figure 54 shows the spectra of a mixture of the two proteins. Panel A and B represent the obtained spectra at pH 7.4 in the presence and absence of Zn^{2+} respectively, while panel C and D show the obtained spectra at pH 6.6 in the presence and absence of Zn^{2+} respectively. The two individual proteins were readily detectable in each

case, but a complex was not observed. There are no additional peaks and masses, which could have been representatives of a protein complex.

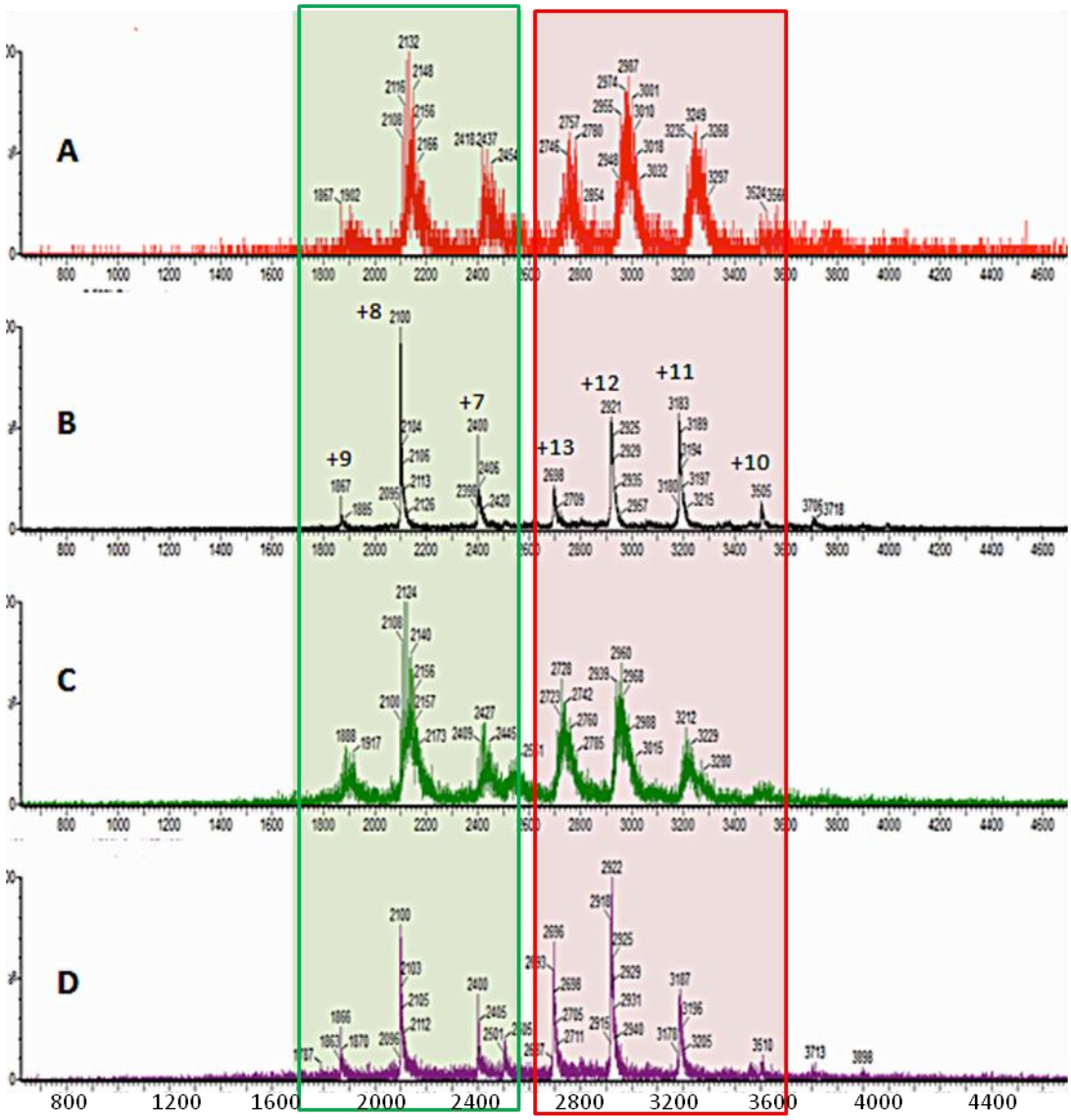

Figure 54: Native mass spectra of Sbp and $G5_1EG5_2EG5_3$. Panel A and B: G5EG5EG5 with Sbp in the presence and absence of 0.1 mM Zn (II) Ac respectively at pH 7.4. Panel C and D show $G5_1EG5_2EG5_3$ with Sbp in the presence and absence of 0.1 mM Zn (II) Ac respectively at pH 6.6. The peak series on left side of each panel in highlighted in green result from Sbp and the series in red on the right side of all panels result from $G5_1EG5_2EG5_3$.

Sbp and the construct $G5_1EG5_2$ were tested together in the presence of 0.1 m M Zn (II) Ac. Figure 55 shows no peaks that could be assigned to the calculated mass of the complex of proteins.

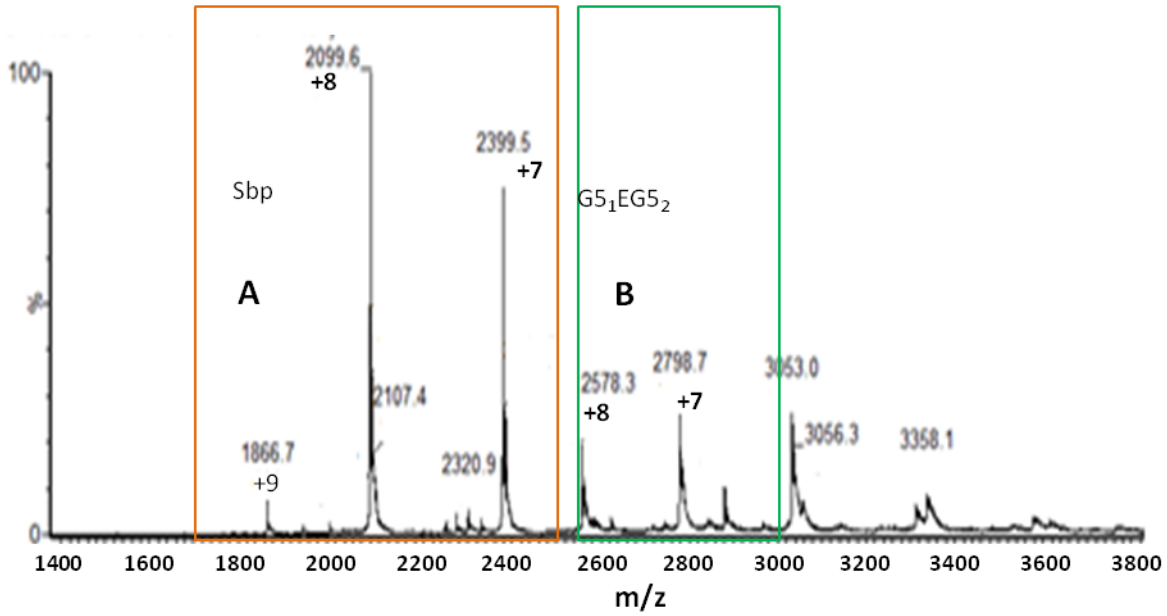

Figure 55: Native mass spectrum of mixture of $G5_1EG5_2$ and Sbp. The series of peaks on left side in a box A with orange colored boundary represent the charges on the molecules hitting at the m/z values, which correspond to the mass of Sbp of ~16.9 kDa. The right series of peak in the box B refers to $G5_1EG5_2$, which is ~21 kDa.

The G5 proteins in general were very unstable upon buffer exchange and during electro spray ionization, an aggregation behavior could be observed when spiking the protein with Zn^{2+} prior to buffer exchange, leading to loss of more than 90% of the sample. Also in electro spray, G5 proteins frequently aggregated and clogged the capillary, which resulted in low intensity signals and indicated a general instability of the protein. This instability could further be observed upon partial unfurling, which if not an intrinsic property of this protein occur during buffer exchange or conversion into the gas phase.

MST data also showed no binding of Sbp with any of G5 constructs. The labeled G5 proteins showed fluorescence in unbound state. The two models, Hill and Kd model, in-built- in Nanotemper monolith failed to find a fitting curve of unbound and bound states of protein.

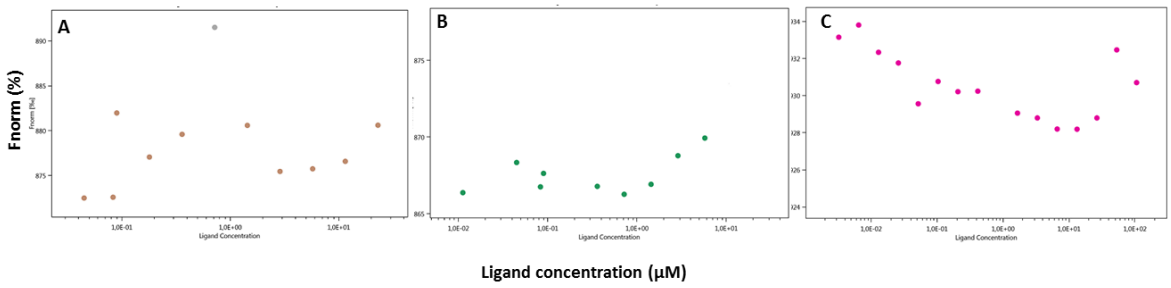

Figure 56: Panels **(A)**, **(B)** and **(C)** represent the labelled fluorescent $G5_1EG5_2EG5_3$, $G5_1EG5_2$ and $G5_1E$ proteins respectively. The absence of a fit model showing the bound and unbound states of two proteins confirmed the absence of interaction between G5 proteins and Sbp under specific described conditions.

5. DISCUSSION

Nosocomial pathogen *Staphylococcus epidermidis* proliferates by forming adherent biofilms on biotic and abiotic surfaces and results in chronic implant-associated infections. Biofilm communities rely on the production of a bacterial derived extracellular matrix, which serves as an adhesive to maintain the architecture of bacterial biofilm (Schommer *et al*., 2011). Small basic protein (Sbp) and Accumulation associated protein (Aap) are major and crucial components of biofilm matrix. Along with other constituents, including the extracellular matrix binding protein (Embp) and the polysaccharide intercellular adhesin (PIA), Aap plays its dynamic role in promoting cell to cell adhesion, aggregation and, hence, stabilizing the architecture of multilayered biofilm (Flemming and Wingender, 2010; Rohde *et al*., 2005).

Small basic protein

The spatial organization of Sbp suggests that it acts as a scaffold, which supports Aap and PIA-mediated biofilm formation. Sbp is the first extracellular structural and functional protein of *S. epidermidis* with considerable relevance to not only protein dependent biofilm formation but also provides a scaffold for polysaccharide dependent biofilm formation (Foster *et al*., 2011).

The *sbp* gene encoding for Sbp was already cloned in pDEST17 expression vector having 6 x His tag upstream of *sbp*. There was a linker region of 18 amino acids between 6 x His tag and *sbp* gene. In the present study, the expression of *sbp* in *E. coli* cells was optimized for large scale purification. The advantage of 6 x His at N- terminus of the gene was taken to purify the protein by immobilizing it on a high affinity metal matrix using affinity chromatography. In the first step of successful affinity purification, the purity of Sbp was ~80 %. The high purity of Sbp was obtained by purifying it by size exclusion chromatography (SEC). SDS PAGE analysis revealed that Sbp is ~ 95 % pure with a molecular mass of 19.8 kDa. Once the purification was done, it was highly challenging to solubilize and stabilize Sbp in solution. Sbp proved quite unstable because it used to aggregate and readily precipitate in a wide range of buffer conditions, varying in molar strengths of chemical components, pH and temperature. The aggregation behavior, homogeneity and monodispersity of the protein were studied using DLS. DLS measurements of Sbp showed severe aggregation in a wide range of

buffers. At low concentration, Sbp was relatively less aggregated as compared to the aggregation at higher concentrations. The aggregation of Sbp resulted into the inclusion bodies formation. Inclusion bodies formation may result from three probable mechanisms; aggregation of the unfolded part of protein, aggregation of native protein or aggregation of intermediate states that are partially folded (Wetzel, 1996). However, an ideal buffer condition for soluble and stable Sbp was optimized. Over a period of approx. 10 days, the stability and homogeneity of the protein solution was examined by dynamic light scattering (DLS) after removing aggregates by centrifugation in the ideal buffer 50 mM $NaPO_4$ and 150 mM NaF, pH=7.4. The hydrodynamic radius (*Rh*) of approx. 2.8 nm of a stable Sbp particle confirmed the monodispersity of protein.

In vitro folding of a protein may not necessarily reflect its folding state *in vivo*; however, it was necessary to see the secondary structure and folding state of Sbp before crystallization experiments. Different homology models of Sbp were created by bioinformatics tools. Although each model designed by a particular tool was different due to very low sequence homology with proteins of already known structures, but all of them suggested the presence of a high percentage of β-strands, some parts of α-helices and a large part consisting of coils. The stability of Sbp in solution, in terms of folding state, was examined by CD spectrometry, which revealed an overall partially folded protein with a composition of β-sheets (45 %), helices (2 %) and random coil regions (51 %), according to Yang's algorithm (Yang *et al*., 2015). Reed's algorithm (Greenfield, 2006) showed an approximation of 46 % β-sheets, 6 % α-helices and 33 % random coil. Both these references are not only in good agreement with each other but also support the credibility of the secondary structure prediction by bioinformatics tools to some extent. The aggregation behavior of Sbp can be correlated to the fact that β-sheet rich proteins mostly involve more intermolecular interactions resulting in aggregation of proteins (Fink, 1998)

High resolution X-ray crystallography provides structural details at atomic-resolution. In order to produce crystals of pure Sbp, after confirming the homogeneity of protein solution, a wide variety of crystallization conditions were used to screen a potential condition to obtain crystals of Sbp. Different possible methods of crystallization were used, including vapour diffusion sitting drop and hanging drop methods. Crystallization under oil method was also

applied. Unfortunately, even after trying different batches of purified protein, no potential condition for Sbp crystal formation was identified.

The *in solution* methods, such as SAXS, prove beneficial in estimating the oligomerization, flexibility and concentration dependent aggregation properties of proteins, so Sbp was characterized using SAXS. Different concentrations (2 mg/ml, 3 mg/ml, 4 mg/ml and 5 mg/ml) of Sbp samples were used. The SAXS data showed an increase in the radius of gyration (*Rg*), going from low to higher concentration of Sbp. *Rg* values observed for the above described Sbp samples were 2.9, 2.8, 3.0 and 3.2 nm respectively. The scattering intensities obtained from the 2 mg/ml sample were used to determine the model using *DAMMIF* (Franke *et al*., 2009), out to a maximum scattering vector of s = 3.5 nm^{-1}. *DAMMIF* calculated 10 *ab initio* models that were then spatially aligned and averaged using the *DAMAVER* set of programs (Volkov and Svergun, 2003). The normalized spatial discrepancy (NSD), which provides a measure of the spatial consistency between the individual models, was 0.526. This NSD value indicated that all the individual bead models of Sbp are similar to each other. Whereas, NSD < 0.7 represents a level of spatial consistency among the individual models (Volkov and Svergun, 2003). The shape of Sbp obtained from an *ab initio* model is elongated ellipsoid with a thick globular head and a small extended tail-like region. The spatial alignment of I-Tasser predicted model of Sbp with a SAXS calculated *ab initio* model fitted well to each other. The N-terminus of predicted model was extended from the globular region to the tail-like region of the *ab initio* model. There was a linker consisting of 18 amino acids, followed by a 6 x His tag at the N-terminus of Sbp upstream of the sequence of protein. The linker region and His tag, contributing to the formation of the tail region of Sbp may play a role in hindering the formation of Sbp crystals.

Unfortunately, no cleavage site was available to cut the tag off from Sbp after purification to obtain tag free Sbp, so the cloning of a new *sbp* construct having a TEV protease cleavage site was necessary. Successful cloning of *sbp* was done in a pDEST15 vector with glutathione *S*-transferase (GST) tag at the N-terminus, but the potential expression of the gene was not obtained. Sbp was highly insoluble and formed inclusion bodies. Several methods, e.g. temperature change in bacterial growth culture, change in time points of expression, different concentrations of inducer, media components and buffer conditions to

lyse the cells, were tried to increase the solubility of Sbp. The expression level of *sbp* was not improved, as most of the protein was detected in the cell pellet. Successful cloning of *sbp* gene in pDEST17 with a 6 x His tag at N-terminus and a TEV protease cleavage site also resulted in a poor gene expression. The solubility of Sbp was so low that only 1 mg/ml of protein was yielded from 6 l of culture sample. The cloning of *sbp* in the pET302 NT-His vector with a TEV protease cleavage site expressed a soluble Sbp. 6 x His tag was cleaved from Sbp using TEV protease. The stability of the construct was analyzed by DLS. The DLS measurements showed an increase in *Rh* with increasing concentration of Sbp in solution. To obtain the crystals to obtain an atomic resolution structure of Sbp, crystallographic methods were applied but the results were not different than those for the previous construct. Sbp did not crystallize, however, SAXS measurements of tag free Sbp were performed to see the aggregation behavior of the construct and compare the *ab initio* model with the *ab initio* model of the previous full length construct. Interestingly, there was also concentration dependent, oligomerization/aggregation observed. The aggregation property observed here also support the idea that Sbp acts a proteinaceous biofilm scaffold on artificial surfaces and contribute to cell aggregation mechanisms of *S. epidermidis* (Decker *et al*., 2015). The tail-like region of *ab initio* model was reduced as a result of removing tag and linker (18 amino acids) from the N-terminus of Sbp in the new construct. Native mass spectrometry also confirmed the non-specific oligomerization upon increase of protein concentration.

In short, the results obtained from CD spectrometry suggest the presence of a high proportion of β-sheets, which are normally linked to intermolecular interactions in a protein. The intrinsic flexibility property of a protein is linked to conformational changes in the protein, which ultimately rearrange domains or smaller fragments of the protein (Marsh *et al*., 2014). The partial folding of Sbp in CD spectrometry analysis also suggests that there is an intrinsic flexibility in the protein, which does not allow it to form a crystal. Sbp tends to show concentration dependent increase in *Rh* in DLS experiments, which is due to intermolecular interactions. The *ab initio* model, representing the shape of Sbp, suggests that Sbp is an elongated ellipsoidal shaped protein with a large globular head and a short tail-like extension at the N-terimus. This short tail can be deduced as another reason of hindrance in Sbp crystal formation. The models predicted from different bioinformatics tools show quite different

models due to the lack of availability of a high homology search model. However, the model predicted by I-Tasser fit best into the *ab initio* model of Sbp. Keeping the tail region and peptide bond length in consideration, the first 9 amino acids of Sbp from N-terminus should be removed to get rid of the tail region. This could be a next step to follow in future to get more insights into the structure of Sbp.

Interaction between sub domain G5 proteins of Aap domain B and Sbp

A physical contact between two or more proteins is termed as protein-protein interaction (PPI). The interaction is established as a result of a biochemical event and/or electrostatic and Van der Waals forces. The activity and function of a protein are affected and modulated by the protein with which it interacts. Proteins can interact with each other based on different parts of their composition. There could be interactions between different domains and peptides i.e. domain-domain contacts or domain-peptide interaction. The interactions can be stable or transient depending upon the nature of chemical interaction between proteins.

Based on the studies of co-localization of Aap domain B and Sbp in *S. epidermidis* biofilms, it was hypothesized that Sbp and Aap may interact with each other. Later on, it was reported that Sbp is a necessary co-factor for Aap mediated biofilm formation and the molecular interactions between matrix components of biofilm and *S. epidermidis* promote biofilm accumulation. Dose (concentration) dependent interactions between Sbp and Aap domain B have been observed (Decker *et al*., 2015). Moreover, it has also been reported that Domain B of Aap is a Zn^{2+} binding protein and Aap links bacterial cells together in the presence of Zn^{2+} (Gruszka DT *et al*., 2012 , Conrady *et al*., 2013).

In the present study, three constructs ($G5_1E$, $G5_1EG5_2$ and $G5_1EG5_2EG5_3$) of sub-domain G5 of Aap were cloned into pDEST17 expression vector and expressed in *E.coli* BL21 AI. To our surprise, all the G5 proteins showed a band shift on SDS PAGE at higher molecular size than expected (calculated from amino acid sequence). The protein constructs were purified, utilizing the 6x His tag by affinity chromatography. The anomalous behavior of protein constructs sustained during SEC and each construct eluted at a volume, specific for much higher molecular weights than expected. The elution of each construct through superdex^{TM} 200 SEC column remained consistent to a defined volume, so a defined peak corresponding

to a particular construct could be taken as a reference of that particular construct in further studies. Although the sequencing of the clones along with start and stop codon confirmed the in-frame cloning of the desired constructs and western blot analysis using antibodies against the Aap domain B and histidine, separately, also confirmed the presence of Aap constructs, yet it was thought that there could be some problem during protein translation as the constructs are repetitive sequences of G5 and spacer E regions. Again, successful cloning of all the three constructs was performed into pET302/NT-His expression vector and expressed in *E.coli* BL21 Star cells. Interestingly, the results were consistent to the ones, obtained from the previous clones.

Peptide digestion mass spectrometry based identification of Aap G5 constructs was done, which suggested that all the identified peptide lie in the amino acid sequence of the respective constructs. The anomalous behavior of proteins can be due to very large negative charges at neutral pH, which resulted in poor binding of proteins to SDS and, hence, low electrophoretic mobility through gel was experienced. This kind of anomalous behavior has been reported in β-lactamases from *Actinomadura* R39 (Andre *et al*., 1991). The other reason could be an unusual shape of SDS-protein complex, as some of the membrane proteins aggregate in SDS, which slows down their electrophoretic mobility through gels and a band shift is observed at higher molecular size than expected. However, the length of the complete amino acid sequence was not explainable by the data obtained from peptide digestion MS. Native MS determined the molecular masses of all three Aap G5 constructs accurately by converting the protein solutions into gas phase on the principle of electrospray ionization. The protein masses were in good agreement with the theoretical molecular masses calculated from amino acid sequences. In short, the sequence integrity of each construct was experimentally determined by native MS.

It has been reported that Aap accumulates bacterial cells together in the presence of Zn^{2+}. Biofilm growth assays demonstrate that there is Zn^{2+} dependent dimerization of G5 domains at physiological conditions (Conrady *et al*., 2012). The interaction between each G5 construct and Sbp was studied in the presence of $ZnCl_2$, using SEC, native MS and MST techniques. SEC studies of the possible complex formation (2:1 molar ratio of Sbp and each G5 construct separately) showed that there was no interaction between Sbp and any G5

construct, as both proteins eluted at their respective volumes though superdex^{TM} 200 SEC column in each case. The amount of $ZnCl_2$ was controlled from 0.1 to 1.0 mM. To see the difference between the elution profiles in each case, SEC profiles of each construct and Sbp were set as reference in the presence and absence of $ZnCl_2$ separately. Native MS analysis of possible complex formation (Sbp with each construct separately) also showed no interaction in the presence and absence of 0.1 mM Zn (II) acetate. Each protein appeared as a separate charged moiety at particular *m/z* value in its monomeric state. No peaks representing the mass of the complex were detected. The G5 proteins in general were very unstable upon buffer exchange and during electrospray ionization, an aggregation behavior could be observed when spiking the protein with Zn^{2+} prior to buffer exchange, leading to loss of more than 90% of the sample. Also in electrospray, G5 proteins frequently aggregated and clogged the capillary, which resulted in low intensity signals and indicated a general instability of the protein.

Structural characterization of Aap G5 constructs

G5 proteins were further characterized to determine their shapes *in solution* using SAXS. Before SAXS experiments, the pure solutions of G5 proteins were examined by DLS over a suitable period of time. The *Rh* of $G5_1E$, $G5_1EG5_2$ and $G5_1EG5_2EG5_3$ were observed as 2.89, 3.38 and 4.63 nm respectively, showing monodisperse protein solutions. The CD spectrum of each construct showed a single minimum at 198 nm, which confirms the presence of β-sheets. However, the lack of a positive peak showed that each construct was only partially folded and had a substantial unfolded part. This unfolding is the intrinsic property of G5 proteins, due to the spacer region E between two consecutive G5 regions. Studies on the SasG protein, a homologue of Aap G5 from *S. aureus*, report that E domain, in isolation or when preceded by G5 region, is disordered (Gruszka *et al.*, 2012). The SAXS data showed a regular increase in *Rg* and D_{max}, while going from small to large G5 construct, with step wise addition of G5.E units. *Rg* and D_{max} for $G5_1E$ were 3.0 nm and 11.3 nm, $G5_1EG5_2$; 4.2 nm and 16.5 nm and $G5_1EG5_2EG5_3$; 6.0 nm and 26.9 nm. On the other hand, there was a very slight change observed in the cross-section of each protein, which was 1.5 nm, 2.3 nm and 3.8 nm, while going from lower to higher size G5 construct protein. All these parameters suggest that G5 proteins are highly extended and slightly bent or coiled with respect to their cross-sections.

These results can be correlated to the reported facts about SasG proteins, that they are highly extended thin fibrils visible using electron microscopy and attached to the cell wall of *S. aureus* (Corrigan *et al*., 2007; Banner *et al*., 2007).

The monodisperse solutions of purified monomer G5 constructs were subjected to crystallization set ups. After 7 weeks the first crystals of $G5_1EG5_2EG5_3$ were observed in A9 condition of Morpheus suite, which consisted of 0.1 M bicine/trizma base pH=8.5, 0.03 M $CaCl_2$, 0.03 M $MgCl_2$, 10% w/v PEG 20,000 and 20% v/v PEG MME550. The crystal diffracted up to 3.5 Å resolution. The crystals appeared only in the crystallization set up, where no $ZnCl_2$ was added into the buffer after purification. However, Zn^{2+} mediated dimerization of G5 constructs depends upon the C-terminal G5 domain sequence (Shelton *et al*., 2017). After optimization of initial screening condition, a few more crystals were obtained but it took almost same time as the previous one. Another data set was collected from a different condition (E5 PACT suite) till 2.7 Å at PETRA III beamline P13 (EMBL, Hamburg) and initial processing was done. Initial processing of both data sets revealed the same space group, which was 19 and more or less same unit cell parameters with slight variations.

It is necessary to have phase information along with the coordinates to solve the structure of the protein, therefore, it was tried to obtain phase information by Molecular Replacement (MR) using programs Molrep (Vagin *et al*., 1997) and Phaser (McCoy *et al*., 2007) from CCP4i suite (Collaborative Computational Project, Number 4, 1994). Three search models, PDB ID: 4FUO (Conrady *et al*., 2013), 4WVE (Gruszka *et al*., 2015) and 3TIQ (Gruszka *et al*., 2012), were selected based on the homology to amino acid sequence of protein. Different strategies were applied to find phase information by molecular replacement. Each search model was used as a reference model for MR separately. The residual replacement by poly-alanines to avoid side chains was also tried but, unfortunately, the statistical values to obtain the phases were not satisfactory. Inspite of the availability of the homologous structure, MR failed to solve the structure of G5 till now. Hence, the solution of crystal structure by MR was unsuccessful. Next step was to solve the phase problem by Single Anomalous Dispersion (SAD) by introducing a heavy metal atom into the protein crystals. However, the diffraction

data obtained from heavy metal (platinum and europium) soaked crystals did not produce anomalous signal.

Initial processing of both data sets obtained from two different conditions revealed the same space group and same unit cell parameters with slight variations. SAXS data also confirm the twisting and bending of the linear shape of G5 proteins. Moreover, the diffraction data could not be improved better than 2.7 Å resolution. Keeping in consideration the time required to grow crystals, quality of crystals and flexibility in the cross-section (probably side chains) of protein, it is a quite challenging task to obtain the phases to solve the atomic structure of $G5_1EG5_2EG5_3$. More studies are needed to solve the phase problem.

6. SUMMARY

Staphylococcus epidermidis is a nosocomial pathogen, which infects biotic and abiotic surfaces by forming adherent biofilms. Biofilm is a multilayered microbial community formed by the production of an extracellular matrix. The constituents of extracellular matrix, consisting of proteins, polysaccharides and lipids, regulate different stages of biofilm formation. Once a biofilm is matured, the infection becomes chronic especially in case of implanted medical devices (IMDs). Among many other components of extracellular matrix, small basic protein (Sbp) and accumulation associated protein (Aap) play key roles in biofilm formation in *S. epidermidis*. Aap consists of two domains A and B. Domain B is reported to be involved in biofilm formation and consists of glycine rich repeats called G5 domains. The genes encoding for Sbp and three G5 domain constructs of Aap domain B ($G5_1E$, $G5_1EG5_2$ and $G5_1EG5_2EG5_3$) were successfully cloned and expressed in *E. coli*. The concentration dependent non-specific oligomerization/aggregation of purified Sbp was studied in solution, using SAXS after improving the solubility and stability of protein. A low-resolution structure *ab initio* model of Sbp monomer has been proposed to present its ellipsoidal shape using bead modeling approach by applying the interpretation of SAXS measurements. X-ray diffraction of $G5_1EG5_2EG5_3$ crystals, obtained by sitting drop vapour diffusion method, revealed some initial leads, including unit cell parameters, symmetry and number of molecules per asymmetric unit to solve the structure in future. The low resolution elongated rod-like *ab initio* models of three G5 constructs have been generated. The SAXS data showed a regular increase in *Rg* and *Dmax*, while going from small to large G5 construct with stepwise addition of G5.E units. Based on native MS, MST and SEC data, it was further suggested that there is no interaction observed between Sbp and G5 constructs under particular *in vitro* experimental conditions. This work contributes towards the better understanding of organization of proteins in the extracellular matrix produced during biofilm formation in *S. epidermidis*.

7. ZUSAMMENFASSUNG

Staphylococcus epidermidis ist ein nosokomiales Pathogen, das biotische und abiotische Oberflächen durch die Bildung adhärenter Biofilme infiziert. Biofilm ist eine mehrschichtige, mikrobielle Gemeinschaft, die durch eine zum Teil selbsterzeugte extrazelluläre Matrix zusammengehalten wird. Die Bestandteile dieser extrazellulären Matrix sind Proteine, Polysaccharide und Lipide, deren Synthese Einfluss in verschiedenen Stadien der Biofilmbildung reguliert ist. Sobald ein Biofilm gereift ist, verläuft eine solche Biofilm-assoziierte Infektion insbesondere bei implantierten medizinischen Geräten chronisch. Als eine unter vielen anderen Komponenten der extrazellulären Matrix spielen das kleine basische Protein(Sbp) und das Akkumulation assoziierte Protein (Aap) eine wesentliche Rolle bei der *S. epidermidis* Biofilmbildung insbesondere bei *icaADBC*-negativen Stämmen.. AAP besteht aus zwei Domänen A und B. Die Bedeutung der Domäne B an der Biofilmbildung konnte bereits hinreichend gezeigt werden. Die Domäne B besteht aus Glycin-reichen Wiederholungen, die G5-Domänen genannt werden. Die Gene, die für Sbp und drei G5-Domänenkonstrukte von Aap-Domäne B ($G5_1E$, $G5_1EG5_2$ und $G5_1EG5_2EG5_3$) codieren, wurden erfolgreich kloniert und in *E. coli* rekombinant exprimiert. Die konzentrationsabhängige unspezifische Oligomerisierung / Aggregation von gereinigtem Sbp wurde in Lösung untersucht, wobei SAXS nach Verbesserung der Löslichkeit und Stabilität der Proteine verwendet wurde. Ein ab-initio-Modell des Sbp-Monomers mit niedriger Auflösung wurde erarbeitet, um seine ellipsoide Form unter Verwendung des Wulstmodellierungsansatzes in dem die Interpretation von SAXS-Messungen angewendet wird, darzustellen. Aus der Röntgenbeugung von $G5_1EG5_2EG5_3$ Kristallen, die durch das Tropfen-Dampf-Diffusionsverfahren entstanden sind, konnten erste Strukturdaten, einschließlich der Elementarzellenparameter, der Symmetrie und der Anzahl der Moleküle pro asymmetrischer Einheit gewonnen werden, um die Struktur in Zukunft lösen zu können. Mithilfe nativer MS- und SEC-Daten konnte keine Interaktion von Sbp- und G5-Konstrukten unter den verwendeten in vitro-experimentellen Bedingungen nachgewiesen werden. Die langgestreckten, stabförmigen Ab-initio-Modelle mit niedriger Auflösung von drei G5-Konstrukten wurden erzeugt. Diese Arbeit trägt zum besseren Verständnis der Organisation von Proteinen in der extrazellulären Matrix bei, die während der Biofilmbildung in *S. Epidermidis* produziert wird.

8. REFERENCES

- Adal, K.A., Farr, B.M., 1996. Central venous catheter-related infections: a review. Nutrition 12, 208–13.

- Ahlstrand, E., Svensson, K., Persson, L., Tidefelt, U., Söderquist, B., 2011. Glycopeptide resistance in coagulase-negative staphylococci isolated in blood cultures from patients with hematological malignancies during three decades. Eur. J. Clin. Microbiol. Infect. Dis. 30, 1349–1354. doi:10.1007/s10096-011-1228-8

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

- Amini, S., Goodarzi, H., Tavazoie, S., 2009. Genetic Dissection of an Exogenously Induced Biofilm in Laboratory and Clinical Isolates of E. coli. PLoS Pathog. 5, e1000432. doi:10.1371/journal.ppat.1000432

- Arias-Palomo, E., O'Shea, V.L., Hood, I.V., Berger, J.M., 2013. The Bacterial DnaC Helicase Loader Is a DnaB Ring Breaker. Cell 153, 438–448. doi:10.1016/j.cell.2013.03.006

- Banat, I.M., De Rienzo, M.A.D., Quinn, G.A., 2014. Microbial biofilms: biosurfactants as antibiofilm agents. Appl. Microbiol. Biotechnol. 98, 9915–9929. doi:10.1007/s00253-014-6169-6

- Banner, M.A., Cunniffe, J.G., Macintosh, R.L., Foster, T.J., Rohde, H., Mack, D., Hoyes, E., Derrick, J., Up ton, M., Handley, P.S., 2007. Localized Tufts of Fibrils on Staphylococcus epidermidis NCTC 11047 Are Comprised of the Accumulation-Associated Protein. J. Bacteriol. 189, 2793–2804. doi:10.1128/JB.00952-06

- Becker, K., Heilmann, C., Peters, G., 2014. Coagulase-Negative Staphylococci. Clin. Microbiol. Rev. 27, 870–926. doi:10.1128/CMR.00109-13

- Bester, E., Kroukamp, O., Hausner, M., Edwards, E.A., Wolfaardt, G.M., 2011. Biofilm form and function: carbon availability affects biofilm architecture, metabolic activity and planktonic cell yield. J. Appl. Microbiol. 110, 387–398. doi:10.1111/j.1365-2672.2010.04894.x

- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information.

Nucleic Acids Res. 42, W252–W258. doi:10.1093/nar/gku340

- Boutet, S., Lomb, L., Williams, G.J., Barends, T.R.M., Aquila, A., Doak, R.B., Weierstall, U., DePonte, D.P., Steinbrener, J., Shoeman, R.L., Messerschmidt, M., Barty, A., White, T.A., Kassemeyer, S., Kirian, R.A., Seibert, M.M., Montanez, P.A., Kenney, C., Herbst, R., Hart, P., Pines, J., Haller, G., Gruner, S.M., Philipp, H.T., Tate, M.W., Hromalik, M., Koerner, L.J., van Bakel, N., Morse, J., Ghonsalves, W., Arnlund, D., Bogan, M.J., Caleman, C., Fromme, R., Hampton, C.Y., Hunter, M.S., Johansson, L.C., Katona, G., Kupitz, C., Liang, M., Martin, A. V., Nass, K., Redecke, L., Stellato, F., Timneanu, N., Wang, D., Zatsepin, N.A., Schafer, D., Defever, J., Neutze, R., Fromme, P., Spence, J.C.H., Chapman, H.N., Schlichting, I., 2012. High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. Science (80-.). 337.

- Büttner, H., Mack, D., Rohde, H., 2015. Structural basis of Staphylococcus epidermidis biofilm formation: mechanisms and molecular interactions. Front. Cell. Infect. Microbiol. 5, 14. doi:10.3389/fcimb.2015.00014

- Cerca, F., França, Â., Guimarães, R., Hinzmann, M., Cerca, N., Lobo da Cunha, A., Azeredo, J., Vilanova, M., 2011. Modulation of poly-N-acetylglucosamine accumulation within mature Staphylococcus epidermidis biofilms grown in excess glucose. Microbiol. Immunol. 55, 673–682. doi:10.1111/j.1348-0421.2011.00368.x

- Cheung, G.Y.C., Rigby, K., Wang, R., Queck, S.Y., Braughton, K.R., Whitney, A.R., Teintze, M., DeLeo, F.R., Otto, M., 2010. Staphylococcus epidermidis Strategies to Avoid Killing by Human Neutrophils. PLoS Pathog. 6, e1001133. doi:10.1371/journal.ppat.1001133

- Choong, S., Whitfield, H., 2000. Biofilms and their role in infections in urology. BJU Int. 86, 935–41.

- Christner, M., Franke, G.C., Schommer, N.N., Wendt, U., Wegert, K., Pehle, P., Kroll, G., Schulze, C., Buck, F., Mack, D., Aepfelbacher, M., Rohde, H., 2010. The giant extracellular matrix-binding protein of *Staphylococcus epidermidis* mediates biofilm accumulation and attachment to fibronectin. Mol. Microbiol. 75, 187–207. doi:10.1111/j.1365-2958.2009.06981.x

- Christner, M., Heinze, C., Busch, M., Franke, G., Hentschke, M., Bayard Dühring, S., Büttner, H., Kotasinska, M., Wischnewski, V., Kroll, G., Buck, F., Molin, S., Otto, M., Rohde, H., 2012. *sarA* negatively regulates *Staphylococcus epidermidis* biofilm formation by modulating expression of 1 MDa extracellular matrix binding protein and autolysis-dependent release of eDNA. Mol. Microbiol. 86, 394–410.

doi:10.1111/j.1365-2958.2012.08203.x

- Clarke, S.R., Harris, L.G., Richards, R.G., Foster, S.J., 2002. Analysis of Ebh, a 1.1-megadalton cell wall-associated fibronectin-binding protein of Staphylococcus aureus. Infect. Immun. 70, 6680–7. doi:10.1128/iai.70.12.6680-6687.2002

- Collaborative Computational Project, Number 4, 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763. doi:10.1107/S0907444994003112

- Conrady, D.G., Brescia, C.C., Horii, K., Weiss, A.A., Hassett, D.J., Herr, A.B., 2008. A zinc-dependent adhesion module is responsible for intercellular adhesion in staphylococcal biofilms. Proc. Natl. Acad. Sci. 105, 19456–19461. doi:10.1073/pnas.0807717105

- Conrady, D.G., Wilson, J.J., Herr, A.B., 2013. Structural basis for Zn2+-dependent intercellular adhesion in staphylococcal biofilms. Proc. Natl. Acad. Sci. 110, E202–E211. doi:10.1073/pnas.1208134110

- Corrigan, R.M., Rigby, D., Handley, P., Foster, T.J., Timothy Foster tfoster, C.J., n.d. The role of Staphylococcus aureus surface protein SasG in adherence and biofilm formation. doi:10.1099/mic.0.2007/006676-0

- Decker, R., Burdelski, C., Zobiak, M., Büttner, H., Franke, G., Christner, M., Saß, K., Zobiak, B., Henke, H.A., Horswill, A.R., Bischoff, M., Bur, S., Hartmann, T., Schaeffer, C.R., Fey, P.D., Rohde, H., 2015. An 18 kDa scaffold protein is critical for Staphylococcus epidermidis biofilm formation. PLoS Pathog. 11, e1004735. doi:10.1371/journal.ppat.1004735

- Dunne, W.M., 2002. Bacterial adhesion: seen any good biofilms lately? Clin. Microbiol. Rev. 15, 155–66.

- Evans, P., 2006. Scaling and assessment of data quality. Acta Crystallogr. Sect. D Biol. Crystallogr. 62, 72–82. doi:10.1107/S0907444905036693
-

- Fey, P.D., Olson, M.E., 2010. Current concepts in biofilm formation of *Staphylococcus epidermidis*. Future Microbiol. 5, 917–933. doi:10.2217/fmb.10.56

- Fink, A.L., 1998. Protein aggregation: folding aggregates, inclusion bodies and amyloid.

- Flemming, H.-C., Wingender, J., 2010. The biofilm matrix. Nat. Rev. Microbiol. 8,

623. doi:10.1038/nrmicro2415

- Foster, C.B., Sabella, C., TC, H., LA, G., P, P., MR, M., H, A., AM, I., JJ, S., NE, M., HH, B., 2011. Health Care–Associated Infections in Children. JAMA 305, 1480. doi:10.1001/jama.2011.449

- Foster, T.J., 2005. Immune evasion by staphylococci. Nat. Rev. Microbiol. 3, 948–958. doi:10.1038/nrmicro1289

- Franke, D., Svergun, D.I., 2009. *DAMMIF* , a program for rapid *ab-initio* shape determination in small-angle scattering. J. Appl. Crystallogr. 42, 342–346. doi:10.1107/S0021889809000338

- Franke, D., Kikhney, A.G., Svergun, D.I., 2012. Automated acquisition and analysis of small angle X-ray scattering data. Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip. 689, 52–59. doi:10.1016/j.nima.2012.06.008
-

- Franke, D., Jeffries, C.M., Svergun, D.I., 2015. Correlation Map, a goodness-of-fit test for one-dimensional X-ra scattering spectra. Nat Methods. 12(5):419-22.doi

- Frebourg, N.B., Lefebvre, S., Baert, S., Lemeland, J.F., 2000. PCR-Based assay for discrimination between invasive and contaminating Staphylococcus epidermidis strains. J. Clin. Microbiol. 38, 877–80.

- Ganeshnarayan, K., Shah, S.M., Libera, M.R., Santostefano, A., Kaplan, J.B., 2009. Poly-N-Acetylglucosamine Matrix Polysaccharide Impedes Fluid Convection and Transport of the Cationic Surfactant Cetylpyridinium Chloride through Bacterial Biofilms. Appl. Environ. Microbiol. 75, 1308–1314. doi:10.1128/AEM.01900-08

- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Analysis Tools on the ExPASy Server 571 571 Protein Identification and Analysis Tools on the ExPASy Server.

- Gerke, C., Kraft, A., Süssmuth, R., Schweitzer, O., Götz, F., 1998. Characterization of the N-acetylglucosaminyltransferase activity involved in the biosynthesis of the Staphylococcus epidermidis polysaccharide intercellular adhesin. J. Biol. Chem. 273, 18586–93.

- Gill, S.R., Fouts, D.E., Archer, G.L., Mongodin, E.F., DeBoy, R.T., Ravel, J., Paulsen, I.T., Kolonay, J.F., Brinkac, L., Beanan, M., Dodson, R.J., Daugherty, S.C., Madupu, R., Angiuoli, S. V., Durkin, A.S., Haft, D.H., Vamathevan, J., Khouri, H.,

Utterback, T., Lee, C., Dimitrov, G., Jiang, L., Qin, H., Weidman, J., Tran, K., Kang, K., Hance, I.R., Nelson, K.E., Fraser, C.M., 2005. Insights on Evolution of Virulence and Resistance from the Complete Genome Analysis of an Early Methicillin-Resistant Staphylococcus aureus Strain and a Biofilm-Producing Methicillin-Resistant Staphylococcus epidermidis Strain. J. Bacteriol. 187, 2426–2438. doi:10.1128/JB.187.7.2426-2438.2005

- Götz, F., 2002. Staphylococcus and biofilms. Mol. Microbiol. 43, 1367–78.

- Greenfield, N.J., 2006. Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. 1, 2876–90. doi:10.1038/nprot.2006.202

- Grice, E.A., Kong, H.H., Conlan, S., Deming, C.B., Davis, J., Young, A.C., NISC Comparative Sequencing Program, N.C.S., Bouffard, G.G., Blakesley, R.W., Murray, P.R., Green, E.D., Turner, M.L., Segre, J.A., 2009. Topographical and temporal diversity of the human skin microbiome. Science 324, 1190–2. doi:10.1126/science.1171700

- Gruszka, D.T., Whelan, F., Farrance, O.E., Fung, H.K.H., Paci, E., Jeffries, C.M., Svergun, D.I., Baldock, C., Baumann, C.G., Brockwell, D.J., Potts, J.R., Clarke, J., 2015. Cooperative folding of intrinsically disordered domains drives assembly of a strong elongated protein. Nat. Commun. 6, 7271. doi:10.1038/ncomms8271

- Gruszka, D.T., Wojdyla, J.A., Bingham, R.J., Turkenburg, J.P., Manfield, I.W., Steward, A., Leech, A.P., Geoghegan, J.A., Foster, T.J., Clarke, J., Potts, J.R., 2012. Staphylococcal biofilm-forming protein has a contiguous rod-like structure. Proc. Natl. Acad. Sci. U. S. A. 109, E1011-8. doi:10.1073/pnas.1119456109

- Heilmann, C., Schweitzer, O., Gerke, C., Vanittanakom, N., Mack, D., Götz, F., 1996. Molecular basis of intercellular adhesion in the biofilm-forming Staphylococcus epidermidis. Mol. Microbiol. 20, 1083–91.

- Henderson, R., Unwin, P.N.T., 1975. Three-dimensional model of purple membrane obtained by electron microscopy. Nature 257, 28–32. doi:10.1038/257028a0

- Hussain, M., Heilmann, C., Peters, G., Herrmann, M., 2001. Teichoic acid enhances adhesion of Staphylococcus epidermidis to immobilized fibronectin. Microb. Pathog. 31, 261–270. doi:10.1006/mpat.2001.0469

- Kabsch, W., 2010. XDS. Acta Crystallogr. D. Biol. Crystallogr. 66, 125–32. doi:10.1107/S0907444909047337

- Kaplan, J.B., Velliyagounder, K., Ragunath, C., Rohde, H., Mack, D., Knobloch, J.K.-M., Ramasubbu, N., 2004. Genes involved in the synthesis and degradation of matrix polysaccharide in Actinobacillus actinomycetemcomitans and Actinobacillus pleuropneumoniae biofilms. J. Bacteriol. 186, 8213–20. doi:10.1128/JB.186.24.8213-8220.2004

- Kikhney, A.G., Svergun, D.I., 2015. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS Lett. 589, 2570–2577. doi:10.1016/j.febslet.2015.08.027

- Kim, J., Choi, J.N., Kim, P., Sok, D.-E., Nam, S.-W., Lee, C.H., 2009. LC-MS/MS profiling-based secondary metabolite screening of Myxococcus xanthus. J. Microbiol. Biotechnol. 19, 51–4.

- Kloos, W.E., Musselwhite, M.S., 1975. Distribution and persistence of Staphylococcus and Micrococcus species and other aerobic bacteria on human skin. Appl. Microbiol. 30, 381–5.

- Knobloch, J.K.-M., Von Osten, H., Horstkotte, M.A., Rohde, H., Mack, D., 2002. Minimal attachment killing (MAK): a versatile method for susceptibility testing of attached biofilm-positive and -negative Staphylococcus epidermidis. Med. Microbiol. Immunol. 191, 107–14. doi:10.1007/s00430-002-0125-2

- Kong, K.-F., Vuong, C., Otto, M., 2006. Staphylococcus quorum sensing in biofilm formation and infection. Int. J. Med. Microbiol. 296, 133–139. doi:10.1016/j.ijmm.2006.01.042

- Kozin, M.B., Svergun, D.I., 2001. Automated matching of high- and low-resolution structural models. J. Appl. Crystallogr. 34, 33–41. doi:10.1107/S0021889800014126

- Kozitskaya, S., Olson, M.E., Fey, P.D., Witte, W., Ohlsen, K., Ziebuhr, W., 2005. Clonal analysis of Staphylococcus epidermidis isolates carrying or lacking biofilm-mediating genes by multilocus sequence typing. J. Clin. Microbiol. 43, 4751–7. doi:10.1128/JCM.43.9.4751-4757.2005

- Laskowski, R.A., 2001. PDBsum: summaries and analyses of PDB structures. Nucleic Acids Res. 29, 221–222. doi:10.1093/nar/29.1.221

- Leslie, 1992. Recent changes to the MOSFLM package for processing film and image plate data. oint CCP4 + ESF-EAMCB Newsl. Protein Crystallogr. 26.

- Līduma, I., Tračevska, T., Bērs, U., Žileviča, A., 2012. Phenotypic and genetic analysis of biofilm formation by Staphylococcus epidermidis. Medicina (Kaunas). 48,

305–9.

- Linnes, J.C., Ma, H., Bryers, J.D., 2013. Giant extracellular matrix binding protein expression in Staphylococcus epidermidis is regulated by biofilm formation and osmotic pressure. Curr. Microbiol. 66, 627–33. doi:10.1007/s00284-013-0316-7

- López, D., Vlamakis, H., Kolter, R., 2010. Biofilms. Cold Spring Harb. Perspect. Biol. 2, a000398. doi:10.1101/CSHPERSPECT.A000398

- Ma, J., Wang, S., Zhao, F., Xu, J., 2013. Protein threading using context-specific alignment potential. Bioinformaticss 29, i257–i265. doi:10.1093/bioinformaticss/btt210

- Macintosh, R.L., Brittan, J.L., Bhattacharya, R., Jenkinson, H.F., Derrick, J., Up ton, M., Handley, P.S., 2009. The Terminal A Domain of the Fibrillar Accumulation-Associated Protein (Aap) of Staphylococcus epidermidis Mediates Adhesion to Human Corneocytes. J. Bacteriol. 191, 7007–7016. doi:10.1128/JB.00764-09

- Mack, D., Fischer, W., Krokotsch, A., Leopold, K., Hartmann, R., Egge, H., Laufs, R., 1996. The intercellular adhesin involved in biofilm accumulation of Staphylococcus epidermidis is a linear beta-1,6-linked glucosaminoglycan: purification and structural analysis. J. Bacteriol. 178, 175–83.

- Mack, D., Rohde, H., Dobinsky, S., Riedewald, J., Nedelmann, M., Knobloch, J.K., Elsner, H.A., Feucht, H.H., 2000. Identification of three essential regulatory gene loci governing expression of Staphylococcus epidermidis polysaccharide intercellular adhesin and biofilm formation. Infect. Immun. 68, 3799–807.

- Mack, D., Siemssen, N., Laufs, R., 1992. Parallel induction by glucose of adherence and a polysaccharide antigen specific for plastic-adherent Staphylococcus epidermidis: evidence for functional relation to intercellular adhesion. Infect. Immun. 60, 2048–57.

- Manganelli, R., van de Rijn, I., 1999. Characterization of emb, a gene encoding the major adhesin of Streptococcus defectivus. Infect. Immun. 67, 50–6.

- Marsh, J.A., Teichmann, S.A., 2014. Protein Flexibility Facilitates Quaternary Structure Assembly and Evolution. PLoS Biol. 12, e1001870. doi:10.1371/journal.pbio.1001870

- Matthews, B.W., 1968. Solvent content of protein crystals. J. Mol. Biol. 33, 491–497. doi:10.1016/0022-2836(68)90205-2

- McCann, M.T., Gilmore, B.F., Gorman, S.P., 2008. Staphylococcus epidermidis device-related infections: pathogenesis and clinical management. J. Pharm. Pharmacol. 60, 1551–71. doi:10.1211/jpp/60.12.0001

- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. *Phaser* crystallographic software. J. Appl. Crystallogr. 40, 658–674. doi:10.1107/S0021889807021206

- Milne, J.L.S., Borgnia, M.J., Bartesaghi, A., Tran, E.E.H., Earl, L.A., Schauder, D.M., Lengyel, J., Pierson, J., Patwardhan, A., Subramaniam, S., 2013. Cryo-electron microscopy--a primer for the non-microscopist. FEBS J. 280, 28–45. doi:10.1111/febs.12078

- Mullis, S.N., Falkinham, J.O., 2013. Adherence and biofilm formation of Mycobacterium avium, Mycobacterium intracellulare and Mycobacterium abscessus to household plumbing materials. J. Appl. Microbiol. 115, 908–14. doi:10.1111/jam.12272

- Nakano, M.M., Zuber, P., 1998. ANAEROBIC GROWTH OF A "STRICT AEROBE" (*BACILLUS SUBTILIS*). Annu. Rev. Microbiol. 52, 165–190. doi:10.1146/annurev.micro.52.1.165

- O'Grady, N.P., Alexander, M., Dellinger, E.P., Gerberding, J.L., Heard, S.O., Maki, D.G., Masur, H., McCormick, R.D., Mermel, L.A., Pearson, M.L., Raad, I.I., Randolph, A., Weinstein, R.A., 2002. Guidelines for the prevention of intravascular catheter-related infections. Centers for Disease Control and Prevention. MMWR. Recomm. reports Morb. Mortal. Wkly. report. Recomm. reports 51, 1–29.

- O'Toole, G., Kaplan, H.B., Kolter, R., 2000. Biofilm formation as microbial development. Annu. Rev. Microbiol. 54, 49–79. doi:10.1146/annurev.micro.54.1.49

- Otto, M., 2014. Staphylococcus aureus toxins. Curr. Opin. Microbiol. 17, 32–37. doi:10.1016/j.mib.2013.11.004

- Otto, M., 2009. Staphylococcus epidermidis--the "accidental" pathogen. Nat. Rev. Microbiol. 7, 555–67. doi:10.1038/nrmicro2182

- Patel, J.D., Colton, E., Ebert, M., Anderson, J.M., 2012. Gene expression during S. epidermidis biofilm formation on biomaterials. J. Biomed. Mater. Res. A 100, 2863–9. doi:10.1002/jbm.a.34221

- Patel, J.D., Ebert, M., Ward, R., Anderson, J.M., 2007. S. epidermidis biofilm

formation: Effects of biomaterial surface chemistry and serum proteins. J. Biomed. Mater. Res. Part A 80A, 742–751. doi:10.1002/jbm.a.31103

- Reichhardt, C., Fong, J.C.N., Yildiz, F., Cegelski, L., 2015. Characterization of the Vibrio cholerae extracellular matrix: a top-down solid-state NMR approach. Biochim. Biophys. Acta 1848, 378–83. doi:10.1016/j.bbamem.2014.05.030

- Reiter, K.C., Sant'Anna, F.H., d'Azevedo, P.A., 2014. Upregulation of icaA, atlE and aap genes by linezolid but not vancomycin in Staphylococcus epidermidis RP62A biofilms. Int. J. Antimicrob. Agents 43, 248–253. doi:10.1016/j.ijantimicag.2013.12.003

- Robert, X., Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42, W320–W324. doi:10.1093/nar/gku316

- Rohde, H., Burandt, E.C., Siemssen, N., Frommelt, L., Burdelski, C., Wurster, S., Scherpe, S., Davies, A.P., Harris, L.G., Horstkotte, M.A., Knobloch, J.K.-M., Ragunath, C., Kaplan, J.B., Mack, D., 2007. Polysaccharide intercellular adhesin or protein factors in biofilm accumulation of Staphylococcus epidermidis and Staphylococcus aureus isolated from prosthetic hip and knee joint infections. Biomaterials 28, 1711–1720. doi:10.1016/j.biomaterials.2006.11.046

- Rohde, H., Burdelski, C., Bartscht, K., Hussain, M., Buck, F., Horstkotte, M.A., Knobloch, J.K.-M., Heilmann, C., Herrmann, M., Mack, D., 2005. Induction of *Staphylococcus epidermidis* biofilm formation via proteolytic processing of the accumulation-associated protein by staphylococcal and host proteases. Mol. Microbiol. 55, 1883–1895. doi:10.1111/j.1365-2958.2005.04515.x

- Rohde, H., Frankenberger, S., Zähringer, U., Mack, D., 2010. Structure, function and contribution of polysaccharide intercellular adhesin (PIA) to Staphylococcus epidermidis biofilm formation and pathogenesis of biomaterial-associated infections. Eur. J. Cell Biol. 89, 103–11. doi:10.1016/j.ejcb.2009.10.005

- Rohde, H., Kalitzky, M., Kroger, N., Scherpe, S., Horstkotte, M.A., Knobloch, J.K.-M., Zander, A.R., Mack, D., 2004. Detection of Virulence-Associated Genes Not Useful for Discriminating between Invasive and Commensal Staphylococcus epidermidis Strains from a Bone Marrow Transplant Unit. J. Clin. Microbiol. 42, 5614–5619. doi:10.1128/JCM.42.12.5614-5619.2004

- Rohde, H., Mack, D., Christner, M., Burdelski, C., Franke, G., Knobloch, J.K.-M., 2006. Pathogenesis of staphylococcal device-related infections: from basic science to new diagnostic, therapeutic and prophylactic approaches. Rev. Med. Microbiol. 17,

45–54. doi:10.1097/01.revmedmi.0000244134.43170.83

- Rożej, A., Cydzik-Kwiatkowska, A., Kowalska, B., Kowalski, D., 2015. Structure and microbial diversity of biofilms on different pipe materials of a model drinking water distribution systems. World J. Microbiol. Biotechnol. 31, 37–47. doi:10.1007/s11274-014-1761-6

- Rupp, M.E., Ulphani, J.S., Fey, P.D., Mack, D., 1999. Characterization of Staphylococcus epidermidis polysaccharide intercellular adhesin/hemagglutinin in the pathogenesis of intravascular catheter-associated infection in a rat model. Infect. Immun. 67, 2656–9.

- Sagane, Y., Hayashi, S., Matsumoto, T., Miyashita, S.-I., Inui, K., Miyata, K., Yajima, S., Suzuki, T., Hasegawa, K., Yamano, A., Nishikawa, A., Ohyama, T., Watanabe, T., Niwa, K., 2013. Sugar-induced conformational change found in the HA-33/HA-17 trimer of the botulinum toxin complex. Biochem. Biophys. Res. Commun. 438, 483–487. doi:10.1016/j.bbrc.2013.07.112

- Saw, J.H., Mountain, B.W., Feng, L., Omelchenko, M. V, Hou, S., Saito, J.A., Stott, M.B., Li, D., Zhao, G., Wu, J., Galperin, M.Y., Koonin, E. V, Makarova, K.S., Wolf, Y.I., Rigden, D.J., Dunfield, P.F., Wang, L., Alam, M., 2008. Encapsulated in silica: genome, proteome and physiology of the thermophilic bacterium Anoxybacillus flavithermus WK1. Genome Biol. 9, R161. doi:10.1186/gb-2008-9-11-r161

- Schaeffer, C.R., Woods, K.M., Longo, G.M., Kiedrowski, M.R., Paharik, A.E., Büttner, H., Christner, M., Boissy, R.J., Horswill, A.R., Rohde, H., Fey, P.D., 2015. Accumulation-associated protein enhances Staphylococcus epidermidis biofilm formation under dynamic conditions and is required for infection in a rat catheter model. Infect. Immun. 83, 214–26. doi:10.1128/IAI.02177-14

- Schell, R.F., Sidone, B.J., Caron, W.P., Walsh, M.D., White, T.F., Zamboni, B.A., Ramanathan, R.K., Zamboni, W.C., 2014. Meta-analysis of inter-patient pharmacokinetic variability of liposomal and non-liposomal anticancer agents. Nanomedicine 10, 109–17. doi:10.1016/j.nano.2013.07.005

- Scherr, T.D., Heim, C.E., Morrison, J.M., Kielian, T., 2014. Hiding in Plain Sight: Interplay between Staphylococcal Biofilms and Host Immunity. Front. Immunol. 5, 37. doi:10.3389/fimmu.2014.00037

- Schommer, N.N., Christner, M., Hentschke, M., Ruckdeschel, K., Aepfelbacher, M., Rohde, H., 2011. Staphylococcus epidermidis Uses Distinct Mechanisms of Biofilm Formation To Interfere with Phagocytosis and Activation of Mouse Macrophage-Like

Cells 774A.1. Infect. Immun. 79, 2267–2276. doi:10.1128/IAI.01142-10

- Seper, A., Pressler, K., Kariisa, A., Haid, A.G., Roier, S., Leitner, D.R., Reidl, J., Tamayo, R., Schild, S., 2014. Identification of genes induced in Vibrio cholerae in a dynamic biofilm system. Int. J. Med. Microbiol. 304, 749–763. doi:10.1016/j.ijmm.2014.05.011

- Shelton, C.L., Conrady, D.G., Herr, A.B., 2017. Functional consequences of B-repeat sequence variation in the Staphylococcal biofilm protein Aap: Deciphering the assembly code. Biochemical Journal (2017) DOI: 10.1024/BCJ20160675

- Simossis, V.A., Heringa, J., 2005. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. Nucleic Acids Res. 33, W289-94. doi:10.1093/nar/gki390

- Siuzdak, G., Bothner, B., Yeager, M., Brugidou, C., Fauquet, C.M., Hoey, K., Change, C.-M., 1996. Mass spectrometry and viral analysis. Chem. Biol. 3, 45–48. doi:10.1016/S1074-5521(96)90083-6

- Taylor, P.K., Yeung, A.T.Y., Hancock, R.E.W., 2014. Antibiotic resistance in Pseudomonas aeruginosa biofilms: towards the development of novel anti-biofilm therapies. J. Biotechnol. 191, 121–30. doi:10.1016/j.jbiotec.2014.09.003

- Vagin, A., Teplyakov, A., IUCr, 1997. *MOLREP* : an Automated Program for Molecular Replacement. J. Appl. Crystallogr. 30, 1022–1025. doi:10.1107/S0021889897006766

- Vergara-Irigaray, M., Maira-Litran, T., Merino, N., Pier, G.B., Penades, J.R., Lasa, I., 2008. Wall teichoic acids are dispensable for anchoring the PNAG exopolysaccharide to the Staphylococcus aureus cell surface. Microbiology 154, 865–877. doi:10.1099/mic.0.2007/013292-0

- Volkov, V. V., Svergun, D.I., 2003. Uniqueness of *ab initio* shape determination in small-angle scattering. J. Appl. Crystallogr. 36, 860–864. doi:10.1107/S0021889803000268

- Vuong, C., Kocianova, S., Voyich, J.M., Yao, Y., Fischer, E.R., DeLeo, F.R., Otto, M., 2004. A Crucial Role for Exopolysaccharide Modification in Bacterial Biofilm Formation, Immune Evasion, and Virulence. J. Biol. Chem. 279, 54881–54886. doi:10.1074/jbc.M411374200

- Walenta, E., 1985. Small angle x-ray scattering. Von O. GLATTER und O.

KRATKY. London: Academic Press Inc. Ltd. 1982. ISBN 0-12-286280-5. X, 515 Seiten, geb. £ 43,60; US $ 81.00. Acta Polym. 36, 296–296. doi:10.1002/actp.1985.010360520

- Wang, R., Khan, B.A., Cheung, G.Y.C., Bach, T.-H.L., Jameson-Lee, M., Kong, K.-F., Queck, S.Y., Otto, M., 2011. Staphylococcus epidermidis surfactant peptides promote biofilm maturation and dissemination of biofilm-associated infection in mice. J. Clin. Invest. 121, 238–248. doi:10.1172/JCI42520

- Wang, X., Preston, J.F., Romeo, T., 2004. The pgaABCD locus of Escherichia coli promotes the synthesis of a polysaccharide adhesin required for biofilm formation. J. Bacteriol. 186, 2724–34.

- Wetzel, R., 1996. For Protein Misassembly, It's the "I" Decade. Cell 86, 699–702. doi:10.1016/S0092-8674(00)80143-9

- Widerström, M., McCullough, C.A., Coombs, G.W., Monsen, T., Christiansen, K.J., 2012. A multidrug-resistant Staphylococcus epidermidis clone (ST2) is an ongoing cause of hospital-acquired infection in a Western Australian hospital. J. Clin. Microbiol. 50, 2147–51. doi:10.1128/JCM.06456-11

- Xu, F.-F., Morohoshi, T., Wang, W.-Z., Yamaguchi, Y., Liang, Y., Ikeda, T., 2014. Evaluation of Intraspecies Interactions in Biofilm Formation by Methylobacterium Species Isolated from Pink-Pigmented Household Biofilms. Microbes Environ. 29, 388–392. doi:10.1264/jsme2.ME14038

- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2014. The I-TASSER Suite: protein structure and function prediction. Nat. Methods 12, 7–8. doi:10.1038/nmeth.3213

- Yarwood, J.M., Schlievert, P.M., 2003. Quorum sensing in Staphylococcus infections. J. Clin. Invest. 112, 1620–5. doi:10.1172/JCI20442

- Zhang, X., Jin, L., Fang, Q., Hui, W.H., Zhou, Z.H., 2010. 3.3 Å Cryo-EM Structure of a Nonenveloped Virus Reveals a Priming Mechanism for Cell Entry. Cell 141, 472–482. doi:10.1016/j.cell.2010.03.041

- Zhang, Y.-Q., Li, Y.-Z., Wang, B., Wu, Z.-H., Zhang, C.-Y., Gong, X., Qiu, Z.-J., Zhang, Y., 2005. Characteristics and living patterns of marine myxobacterial isolates. Appl. Environ. Microbiol. 71, 3331–6. doi:10.1128/AEM.71.6.3331-3336.2005

- Zhang, Y.-Q., Ren, S.-X., Li, H.-L., Wang, Y.-X., Fu, G., Yang, J., Qin, Z.-Q., Miao,

Y.-G., Wang, W.-Y., Chen, R.-S., Shen, Y., Chen, Z., Yuan, Z.-H., Zhao, G.-P., Qu, D., Danchin, A., Wen, Y.-M., 2003. Genome-based analysis of virulence genes in a non-biofilm-forming Staphylococcus epidermidis strain (ATCC 12228). Mol. Microbiol. 49, 1577–93.

9. RISK AND SAFETY STATEMENTS

9.1. Commercial crystallization solutions and kits

Name	Supplier	Risk label	Risk phrases	Safety phrases
PCT	Hampton	**-**	-	-
Classic Suite	Qiagen	**T, N**	R10, R45, R46, R60, R61,R23/25,R36/37/38, R48/20/22, R51/53	S20, S26, S45, S53, S36/37/39
PACT premier	Molecular Dimensions	**T**	R23/25, R52/53	S20, S36, S45, S61
Morpheus	Molecular Dimensions	**T, N**	R10, R45, R46, R60, R61,R63, R23/25, R36/37/38,R48/20/22, R51/53	S20, S26, S45, S53, S61, S36/37/39
ComPAS Suite	Qiagen	**T**	R10, R45, R23/24/25, R36/38, R39/23/24/25, R51/53	S13, S26, S45, S53, S61, S36/37/39.
Stura/Footprint	Molecular Dimensions	**T, N**	R10, R45, R46, R60,R61, R25, R36/37/38, R48/20/22, R51/53	S20, S26, S45, S53, S61, S36/37/39
JCSG+ Suite	Qiagen	**T, N**	R10, R21, R41, R45, R23/25, R37/38, R51/53	S13, S20, S26, S45, S53, S36/37/39
Ammonium sulfate Suite	Qiagen	**T+, N**	R10, R25, R26, R45, R46, R60, R61, R48/23/25, R51/53	S45, S53, S61, S36/37.
Cryos Suite	Qiagen	**T, N**	R10, R45, R46, R60, R61, R23/25, R36/37/38, R48/20/22, R51/53	S20, S26, S45, S53, S61, S36/37/39.

9.2 Chemicals used (GHS classification)

Compound	CAS-No.	Supplier	GHS Hazard	Hazard statements	Precautionary
Acetic acid	64-19-7	Chem-solute	GHS02 GHS05	H226, H314	P280, P305+351+338, P310
Acrylamide 30%	79-06-1	Carl Roth	GHS06 GHS08	H301, H312, H316, H317, H319, H332, H340, H350, H361f, H372	P201, P280, P301+310, P305+351+338, P308+313
Agarose	9012-36-6	Serva	-	-	-
Ampicillin	69-52-3	Carl Roth	GHS08	H334, H317	P280, P261, P302+P352, P342+P311
APS	7727-54-0	Carl Roth	GHS03 GHS07 GHS08	H272, H302, H315, H317, H319, H334; H335	P280, P305+351+338, P302+352, P304+341, P342+311
Coomassie brilliant blue R250	6104-59-2	Serva	-	-	-
DTT	578517	Applichem	GHS07	H302, H315, H319, H335	P302+352, P305+351+338
EDTA	60-00-4	Sigma	GHS07	H319	P305+351+338
Ethanol	64-17-5	Carl Roth	GHS02	H225	P210
Hydrochloric acid >25%	7647-01-0	Merck	GHS05, GHS07	H314, H335	P261, P280, P310, P305+351+338

Imidazole	288-32-4	Carl Roth	GHS05, GHS06, GHS08	H301; H314; H361	P260, P281, P303+P361+P353, P301+P330+P331, P305+P351+P338, P308+P313
IPTG	367-93-1	Carl Roth	-	-	-
Isopropanol	67-63-0	Carl Roth	GHS02, GHS07	H225, H319, H336	P210, P233, P305+351+338
TEMED	110-18-9	Merck	GHS02 GHS05 GHS07	H225, H302, H314,	P261, P280, P305+351+338
Tris	1185-53-1	Fluka	GHS07	H315, H319, H225	P261, P305+351+338
Yeast extract	8013-01-2	Serva	-	-	-

9.3. GHS, risk symbols and information about hazards

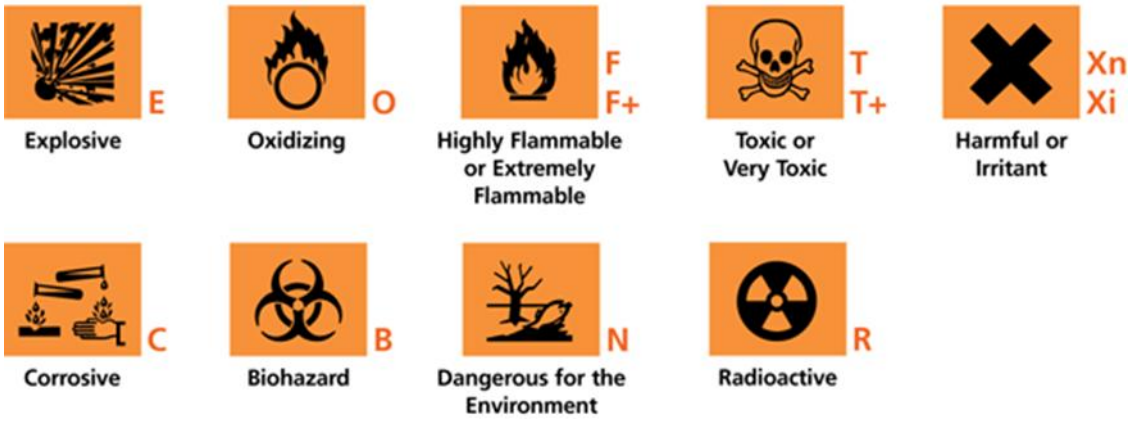

Figure 55: Hazard symbols according to (http://www.sigmaaldrich.com) for formulations and respectiverisk labels

Figure 56:GHS pictograms according to (http://www.evansvanodine.com)- Evans Vanodine International plc/ global Hygene Solutions/UK).

10. ACKNOWLEDGEMENTS

Submission before YOU, always gave me hope and light during desperate times of my journey to seek the truth. O' my LORD, it's you and only you. Special thanks to the Almighty, I am nothing without whom.

I am thankful to my parents and siblings for their support and unconditional love throughout my educational career. Special thanks to Maryam and Abu Bakar who always cheered me up during my hard times.

I pay my respect and gratitude to Higher Education Commission, Pakistan for providing me an opportunity of scholarship to start my Ph.D studies. Thanks to the German academic exchange service (DAAD) for organizing the scholarship and related issues very professionally.

I would like to thank my supervisor Prof. Betzel, who brought me into the field of crystallography. It would have been difficult for me to achieve my goal without his kind guidance and scientific suggestions. I pay my gratitudes to him for providing me this vast infra-structure to carry on my studies. I want to especially thank Prof. Dr. Rohde, my immediate supervisor, who provided me an opportunity to work in his lab. He was always there to discuss new ideas and took special interest in my work. I thank him for his trust and making every possible resource available to my project.I thank PD Dr. Perbandt, my co-supervisor, for his scientific guidance and technical advices throughout my research work.

I would like to thank Dr. Cy Jefferies from EMBL for helping me in comprehending SAXS data. I convey my thanks to Boris Krichel from Heinrich Pette Institute, UKE, with whom I performed native mass spectrometry experiments, which facilitated my research progress. Special thanks to Dr. Friedrich Buch for helping me in mass spectrometry.

I pay my thanks to Dr. Sven Falke, Dr. Henning Büttner, Rana, Gesche and all lab mates from whom I always learnt some technical points, which were worthy for my work.

Selbstständigkeitserklärung

Hiermit erkläre ich an Eides statt, dass ich die an der Universität Hamburg zur Promotion eingereichte Dissertation mit dem Titel " Structural Characterization of Small basic protein (Sbp) and Accumulation associated protein (Aap) – two Proteins involved in Biofilm Formation in *Staphylococcus epidermidis*" im Institut für Biochemie des Fachbereichs Chemie und im Institut für Mikrobiologie am Universitätsklinikum Hamburg Eppendorf der Universität Hamburg ohne sonstige nicht angeführte Hilfe durchgeführt und bei der Abfassung der Dissertation keine anderen als die dort aufgeführten Hilfsmittel benutzt habe

Ferner versichere ich, dass ich bisher an keiner in- oder ausländischen Universität ein Gesuch um Zulassung zur Promotion eingereicht und daher weder diese noch eine andere Arbeit als Dissertation vorgelegt habe.

.

Madiha Fayyaz