1 Introduction

Human body is constantly exposed to and challenged by microorganisms. Not all are harmful; some being part of normal gut flora, though a few will cause disease. Most of the infections are cleared from the body by innate immune response. When innate immunity is compromised, harmless and commensal microorganisms can turn infectious. On the other hand there are pathogenic microorganisms which during some phase of its life-cycle, advance disease and alter the health or behaviour of the host. Humans are beset by a huge number of possible invaders. *Shigellae*, which presents a global human health problem, is one such enteroinvasive pathogenic bacteria. Due to the global emergence of drug resistance [1], the choice of antimicrobial agents for treating shigellosis is limited. A better understanding of biology and mechanism of action will equip the community with varied tools.

1.1 *Shigella*

The genus *Shigella*, the prototypic invasive enteric pathogen, consists of four species, which are distinguished serologically by the O antigen of their lipopolysaccharide (LPS). These are named *S. dysenteriae* (serogroup A), *S. flexneri* (group B), *S. boydii* (group C), and *S. sonnei* (group D).

1.1.1 Properties

The bacterium was first described by Shiga Kiyoshi in the year 1896 [2, 3], hence the name *Shigella*. Organisms of the genus *Shigella* belong to the tribe *Escherichia* in the family Enterobacteriaceae. Shigellae are Gram-negative, nonmotile, facultatively anaerobic, non-spore-forming rods. *Shigella* is closely related to *Escherichia* but can be differentiated from on the basis of pathogenicity, physiology i.e., failure to ferment lactose or decarboxylate lysine and serology [4].

The disease causing properties of *Shigella* are encoded on a 220 kb virulence plasmid. Strains cured of this plasmid are non-pathogenic [5] but viable. If the *Shigella* virulence plasmid is transferred into a non-pathogenic *E. coli*, the plasmid confers invasiveness and cytotoxicity *in vitro* [6]. *Shigella* is highly infectious, 10-100 bacteria are sufficient to cause shigellosis [7] in humans.

Infection is initiated by ingestion of shigellae via faecal-oral contamination. An early symptom, diarrhoea is characterized by painful abdominal cramps, nausea

tenesmus, and frequent stools containing blood and mucus. The hallmarks of shigellosis are bacterial invasion of the colonic epithelium and inflammatory colitis.

1.1.2 Epidemiology

Humans are the only natural hosts for *Shigella*. Shigellosis is endemic in developing countries, with substandard sanitary facilities. A recent study [8] concluded that, of the estimated 165 million cases of Shigellosis that occur annually, 99% occur in developing countries, and of these 69% of episodes occur in children under five years of age. The prevalence of these infections decreases significantly after five years of life. Each year 1.1 million deaths are attributed to *Shigella* infections [8] worldwide.

The history of shigellosis has followed an interesting course. Following its original description in 1898, *S. dysenteriae* was the major isolate until replaced by *S. flexneri* after World War I, which itself was supplanted by *S. sonnei* in industrialized countries following World War II. Beginning in 1969, however, epidemic *S. dysenteriae* reappeared in Latin America, Asia, and Africa [9]. The transition from *S. flexneri* to *S. sonnei* is associated with economic development, though the association between the socioeconomic context and prevalence of a particular species of Shigella are currently unknown [10].

1.1.3 Pathogenicity

Infection with the *Shigella* follows a multistep process (Fig. 1). After passage through the stomach and small intestine, the bacteria reach the large intestine, where they establish an infection [11]. To gain access to the intestinal mucosa, *Shigella* crosses the intestinal epithelium barrier through microfold cells or M cells [12] located in the gut-associated lymphoid tissue (GALT) [13]. M cells are responsible for the initial processing of the luminal macromolecules, antigens and microorganisms [14-16]. M cells are the only port of entry as *Shigella* barely interacts with the apical surface of polarized epithelial cells and preferentially enters through the basolateral pole [17]. After transcytosis, *Shigella* is released into an intraepithelial pocket, where the bacteria encounter resident macrophage and epithelial cells. Although macrophage engulf the bacteria, the *Shigella* ensures its survival by rapidly inducing apoptosis [18-20]. This enables *Shigella* to escape the phagolysosome to the cytoplasm within minutes [21, 22]. Macrophage cell death is accompanied by the release of the proinflammatory cytokines interleukin-1β (IL-1β) and IL-18 [23, 24]. Both cytokines

incite acute and massive inflammatory response elicited by *Shigella*. While IL-1β signaling triggers the strong intestinal inflammation characteristic of shigellosis, IL-18 is involved in the generation of an effective antibacterial response [23, 25]. IL-18 activates natural killer (NK) cells and promotes the production of gamma interferon (IFN- γ), leading to amplified innate immune response [26, 27]. Thereafter, once released from the dying macrophage, it invades the epithelial cells through the basolateral side, escapes from the phagosome into the cytoplasm and replicates there. *Shigella* utilizes the cytoskeleton of the host cells by directed polymerization of actin for intra- and intercellular movement [28-30]. The invasion of epithelial cells triggers a strong inflammatory response. Peptidoglycan fragments released by the *Shigella* elicits Nod1-mediated activation of nuclear factor κB (NF-κB), which triggers the up-regulation and secretion of IL-8 [31-33]. The secretion of IL-8 leads to massive recruitment of polymorphonuclear neutrophil leukocytes (PMN) to the site of the infection [33, 34]. PMN damage the colonic mucosa by breaking the tight junctions to reach the bacteria in the intestinal lumen. This enables more luminal bacteria to reach the submucosa without the need of the M cells. Though macrophage killing, destruction of the epithelial layer, and the massive influx of PMN aggravate the bacterial infection, eventually PMN resolve the infection by a novel way of entangling and killing the bacteria [35-37].

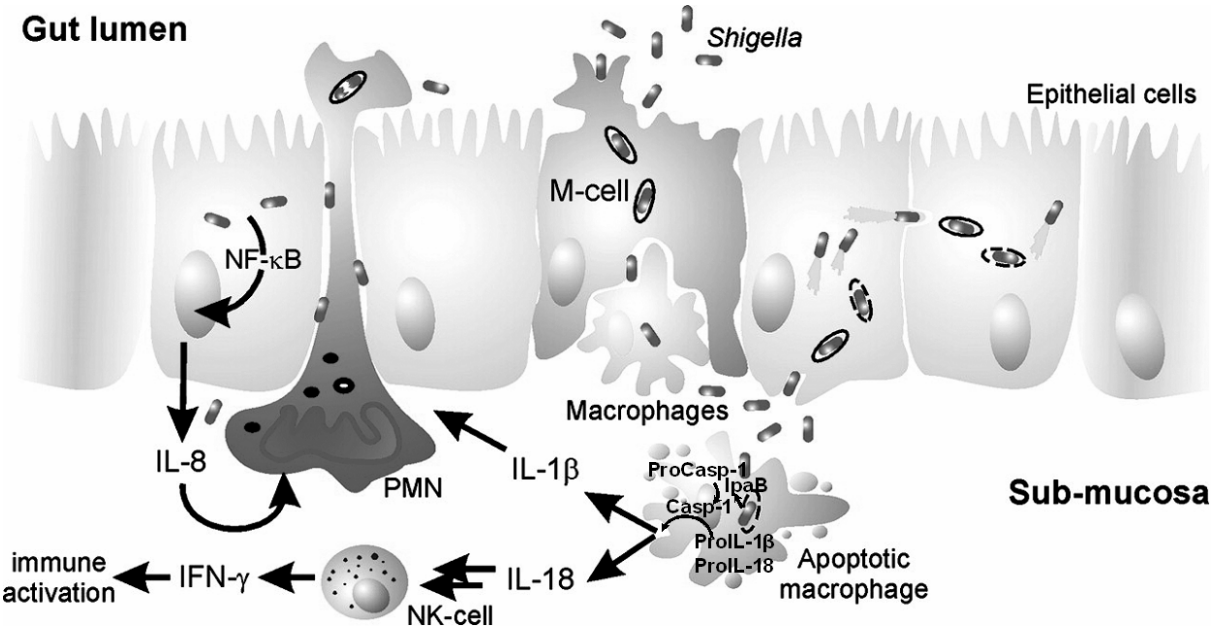

Fig. 1.1: *Shigella* pathogenesis. *Shigella* crosses the epithelial barrier by transcytosis through M cells. It is taken up by the resident macrophages where it induces apoptosis and invades epithelial cells. The dying macrophages release IL-1β and IL-18 while IL-8 is released by the epithelial cells. Eventually, the PMN are recruited and clear the infection. Fig. modified from [38].

1.2 Virulence Factors

The cellular pathogenesis and clinical presentation of shigellosis rely on a large repertoire of virulence factors. *Shigella* virulence factors are encoded on a 220 kb virulence plasmid which is essential for *Shigella* pathogenicity [5]. *S. flexneri* invasion genes are localized in a 31 kb 'entry region' (Fig. 1.2) of the virulence plasmid [39]. This region encodes the invasion plasmid antigen (*ipa*) operon, the membrane expression of *ipa* (*mxi*) and surface presentation of *ipa* antigen (*spa*) operons. The entry region is necessary and sufficient for epithelial cell invasion and macrophage killing [39-42].

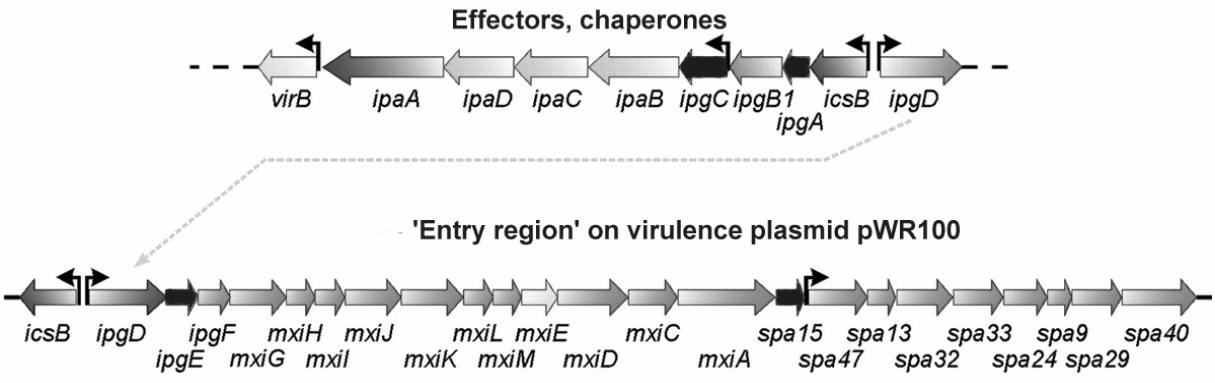

Fig. 1.2: Genetic organization of 31 kb 'entry region' on *S. flexneri* virulence plasmid. The genes encode Mxi-Spa TTSS, effector proteins, chaperones (darkened black arrows) and regulatory proteins. The region is essential for invasion. Fig. modified from [38].

Based on the function these genes can be broadly classified into four different groups. The first group consists of secreted proteins that act at the membrane or in the cytoplasmic compartment of host cells. The secreted effectors are capable of modulating various host functions. The effectors IpaA, IpaB, IpaC, IpgB1, IpgD and VirA secreted via type three secretion system (TTSS) are involved in stimulating the reorganization of F-actin and microtubule cytoskeletons, which trigger bacterial uptake by host cells [43]. IpaB, IpaC and IpaD apart from having effector functions also control the secretion and translocation of other effector proteins into eukaryotic host cells [44-46]. The second group comprises proteins encoded by the *mxi-spa* locus that are needed for the assembly and function of TTSS. The TTS apparatus is composed of at least 30 proteins and is conserved among enteropathogenic bacteria [47]. The transcriptional activators MxiE, Spa15 and VirB which regulate TTSS – associated genes located in the entry region or in the virulence plasmid [48-53], represent the third group. Finally efficient secretion of effectors and regulation of

TTSS requires the aid of chaperones (IpgA, IpgC, IpgE, and Spa15), which constitute the fourth group.

1.3 Type three secretion system (TTSS)

TTSSs are central to the virulence of many Gram-negative pathogenic bacteria. TTS apparatus is a complex multi-subunit membrane-bound structure (Fig. 1.3) composed of more than 20 proteins encoded by the *mxi* and *spa* genes [47, 54, 55]. The apparatus deliver proteins within the membrane or directly into host cells [55]. Additionally, a TTSS is also employed in bacterial flagellar assembly [56-59]. Proteins targeted to these apparatus are thought to travel through a narrow channel of ~25-30 Å in diameter that traverses the entire needle complex.

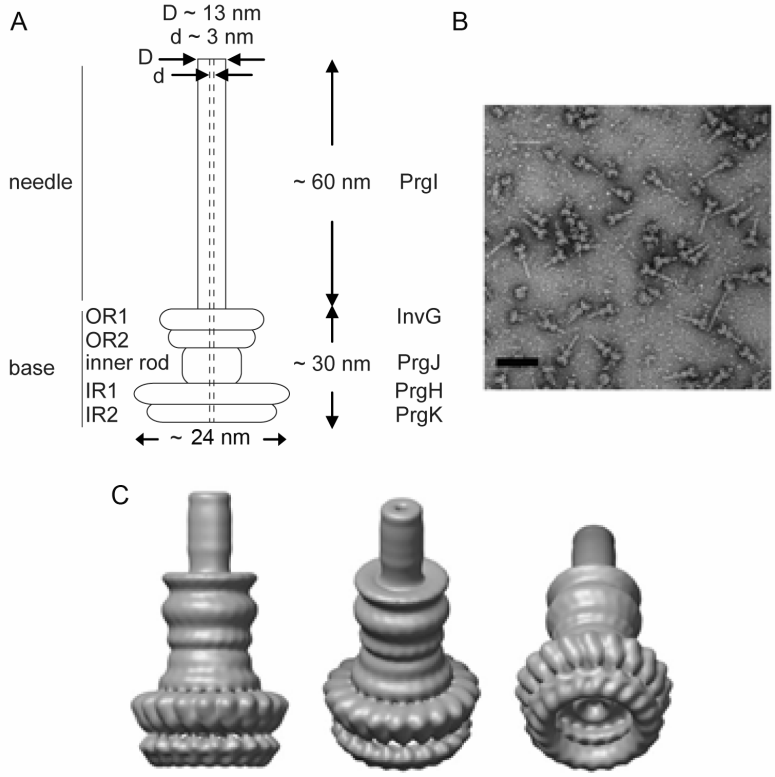

Fig. 1.3: The Needle complex (NC) of *S. typhimurium*.
A, The NC can be divided into two substructures: the membrane-bound base and the extracellular needle filament. The base is anchored to the outer membrane through the outer rings 1 and 2 (OR1, OR2) and to the inner membrane through the inner rings 1 and 2 (IR1, IR2). The inner channel is symbolized by dashed lines. Protein names are given on the right.
B, Electron micrograph of negatively stained isolated needle complexes.
C, Surface rendering of the needle complex structure obtained by 3D reconstruction at 20 Å. Shown are different views. Note that the needle is truncated and that the ATPase is not present Figures b and c were adapted from [60].

Characteristic features of this system are (i) the absence of a typical, cleavable, *sec*-dependent signal sequence in secreted substrates; (ii) the requirement of chaperones for secretion; (iii) the continuous transit of proteins through both the inner and outer bacterial membranes; (iv) the requirement of activating signals to initiate secretion; (v) the energy dependence.

The precise mechanism of secretion via the TTSS is still to be investigated thoroughly. An alternate mode for secretion via flagellar apparatus [58, 59, 61] as well as via bacterial outer-membrane vesicles [62] has been proposed.

1.4 Chaperones associated with TTSS

Many proteins require chaperones for their secretion by TTSS [63, 64]. Before being secreted the protein in the bacterial cytosol need to be stabilized, prevented from intra- and intermolecular association and maintained in secretion competent state. TTS chaperones are implicated in all these functions. Chaperone loss reduces secretion of their specific substrate protein(s), whereas export of other proteins remains unaffected [65]. The substrate without its specific chaperone accumulates intracellularly and is often rapidly degraded [65-67]. Chaperones role in presenting the proteins to be secreted to the TTS apparatus, imposing a temporal hierarchy of secretion and a role in regulation (Table 1.1) has been addressed by various studies.

PROPOSED FUNCTION	EXAMPLES	REFERENCE
Anti-aggregation, stabilization	SycE, IpgC, IpgE, SycN, YscB, SycD, Spa15, SicA, YscG,	[63, 65, 68-72]
Maintenance of a secretion-competent state, antifolding	SycE, SicP, Spa15	[70, 73, 74]
Secretion pilot	YopE, SicP	[75, 76]
Hierarchy of secretion	SycE, SycH, SicP	[75, 77, 78]
Regulation of TTS system	SycH, SicA, IpgC, LcrH	[52, 79-81]

Table 1.1: Different functions of chaperones associated with TTSS.

Although primary sequence identity between the chaperones is low, they are all less than 20 kDa, have acidic pI, and usually bind to one or two substrates; less frequently to more substrates (Table 1.2) [82].

CHAPERONE	SIZE (aa)	pI	SUBSTRATE(S) (aa)
VIRULENCE			
IpgC	155	4.6	IpaB (580), IpaC (382), MxiE (210)
SycD	168	4.9	YopB (401), YopD (306)
BicA	171	4.9	BipB (620), BipD (310)
SicA	165	4.8	SipB (593), SipC (409), InvF (216)
SycE	130	4.7	YopE (219)
SycN	123	5.2	YopN (293)
SycT	130	4.0	YopT (322)
SicP	116	4.0	SptP (543)
CesD	151	6.9	EspB (312), EspD (380)
CesT	156	6.9	Tir (550)
SpcU	137	4.6	ExoU (687)
FLAGELLAR			
FlgN	140	5.3	FlgK (552), FlgL (317)
FliT	122	5.0	FliD (466)
FliS	135	4.9	FliC (489)

Table 1.2: Properties of chaperones. Listed are few TTSS associated chaperones. They generally have a low pI, small size and associate with one or two substrates.

1.4.1 Classification of chaperones

TTSS chaperones have been classified according to effector binding into three categories.

1.4.2 Class I chaperones

This class of chaperones binds either one (class IA) or multiple effectors (class IB). Until now, the majority of structures solved fall into class IA with structures of SicP [74], SigE [83], SycE [84], SycH [85], SycN-YscB [86], CesT [83], AvrPphF ORF1 [87] and SycT [88, 89] reported. On the other hand structures of Spa15 [90], InvB-SipA [91] from class IB are reported. These chaperones display a similar dimeric structure, with each subunit being composed of three α helices and a five-stranded antiparallel β sheet (Fig. 1.4). The topology of each molecule is α-β-β-β-α-β-β-α, which positions the helices on one face of the β sheet. Dimer formation generally occurs via helix α2 and the loop and β strand that immediately follow.

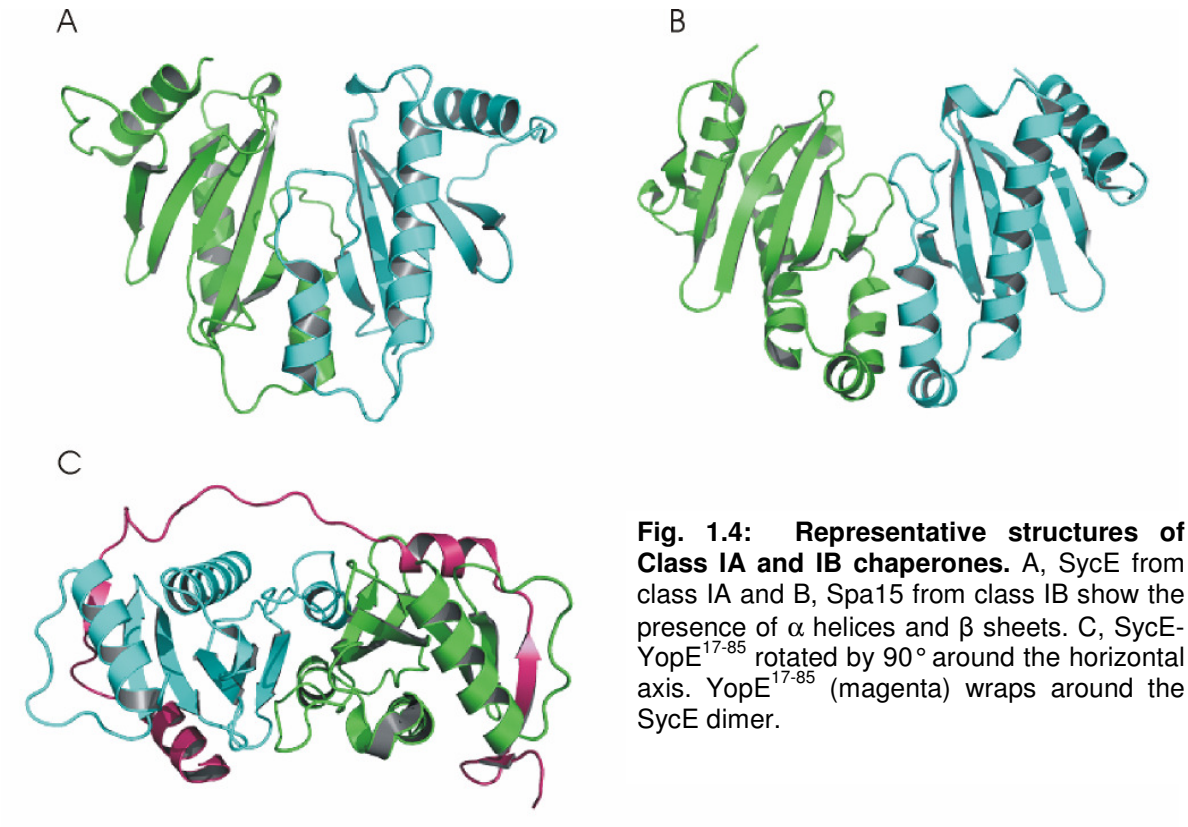

Fig. 1.4: Representative structures of Class IA and IB chaperones. A, SycE from class IA and B, Spa15 from class IB show the presence of α helices and β sheets. C, SycE-YopE$^{17-85}$ rotated by 90° around the horizontal axis. YopE$^{17-85}$ (magenta) wraps around the SycE dimer.

Structures of class I chaperone with their cognate effectors have revealed that the chaperone binding domains are mostly nonglobular and wrap around the surface of the chaperone in an extended conformation.

1.4.3 Class III chaperones

The export and assembly of components of the hook and the filament from the flagellar apparatus is regulated in part by their association with cognate export chaperones. These chaperones are grouped in Class III. Chaperones FlgN, FliT and FliS bind specifically to the structural components that polymerize outside of the bacterial membrane, i.e., FlkK or FlgL, FliD and FliC, respectively [92, 93]. These chaperones are structurally distinct to those of the TTS chaperones [94]. Moreover, chaperones of this class bind to the C-terminus of the substrate [95] unlike N-terminus in the case of TTS chaperones. However, binding of the substrate in an extended nonglobular conformation around the chaperone suggests that this mode of recognition is similar to class I chaperones. The structure of the chaperone FliS with its substrate FliC [94] (Fig. 1.5) suggests its function is to prevent the premature polymerization of FliC in the bacterial cytosol. The C terminus of FliC, which also corresponds to the chaperone binding site, is the key to the polymerization of this molecule [96].

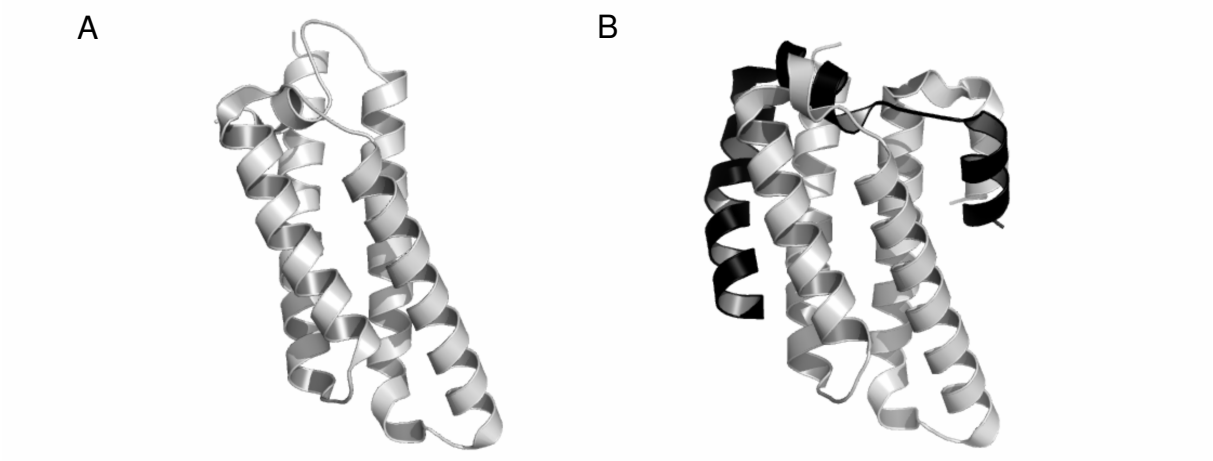

Fig. 1.5: The structure of Class III chaperone FliS and FliS-FliC464-518 complex. A, The structure of FliS shows antiparallel four-helix bundle with no semblance to other chaperones of TTSS. B, Structure of FliS (gray) with its cognate substrate FliC464-518 (black) shows binding of the latter in an extended, nonglobular conformation.

Though there are a number of superficial similarities between flagellar export chaperones and the TTS chaperones, they do not share a common evolutionary origin [94].

1.4.4 Class II chaperones

Chaperones of this class bind to substrate proteins that have been shown or proposed to associate in the eukaryotic membrane to form the pore allowing translocation of effectors [97]. The chaperones (grey arrows, Fig. 1.6) are encoded adjacently to their cognate effector genes in one large operon, suggesting strong selection for their co-existence in the genome.

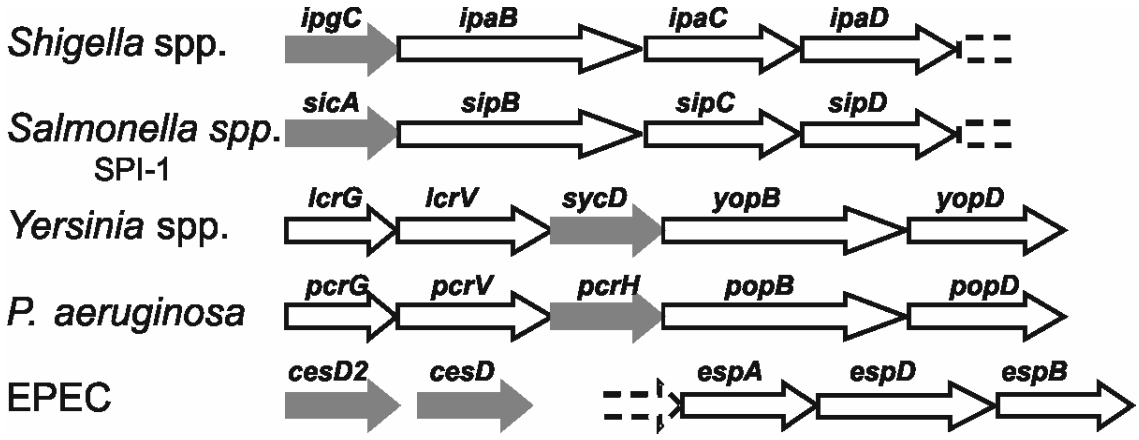

Fig. 1.6: Organization of the operon consisting of class II chaperones in different bacterial species. The chaperones shown in grey are upstream of their cognate effector genes.

The class II chaperones are characterized by the presence of tandem tetratricopeptide repeats (TPRs). TPRs are involved in a variety of activities such as cell cycle regulation, chaperone function and post-translation modifications where protein-protein interactions are involved. TPR motif form antiparallel helices [98] and are characterized by a 34-residue degenerate sequence with loosely conserved amino acids at canonical positions- 4, 7, 8, 20, 24, 27 and 32.

This group of chaperones proved recalcitrant to crystallization. Recently, in January 2008 first structure of a truncated SycD$^{21\text{-}163}$ from *Yersinia enterocolitica* (full length, 168 aa) from this class of chaperone was reported [99]. The structure of SycD$^{21\text{-}163}$ corroborates the presence of TPR made available by secondary structure predictions.

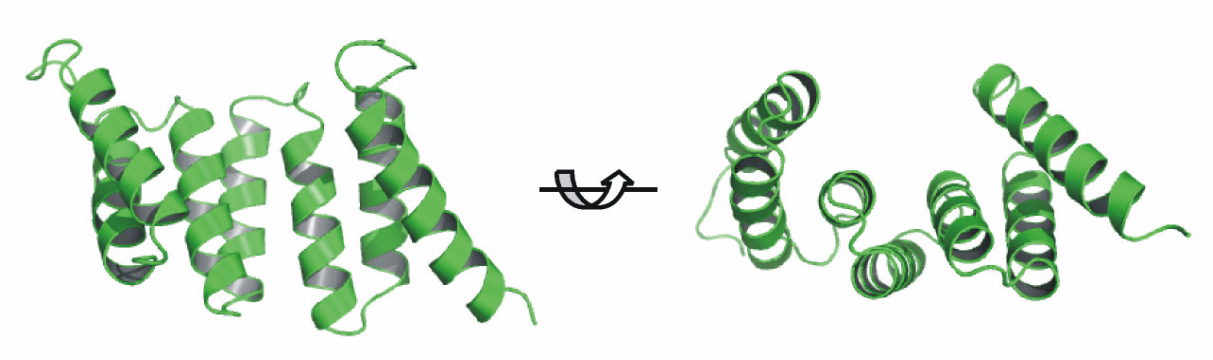

Fig. 1.8: Structure of Class II chaperone SycD$^{21-163}$. The structure shows all α-helical fold with 3.5 TPR. The right side image rotated by 90° reveals a concave groove.

The molecule contains 3.5 TPR motifs. Based on the crystal packing, the authors describe three possible dimers; (1) head-to-head arrangement interacting via TPR1, (2) back-to-back arrangement interacting via the convex face of the TPR repeats and (3) a double mutant S94E/Y95E revealing a head-to-head homodimer as well. Furthermore, to probe the physiological relevance of the dimerization of SycD, the authors carried out *in vivo* experiments. *sycD* containing head-to-head disrupting mutations showed a null mutant-like phenotype indicating that this arrangement is favoured *in vivo*.

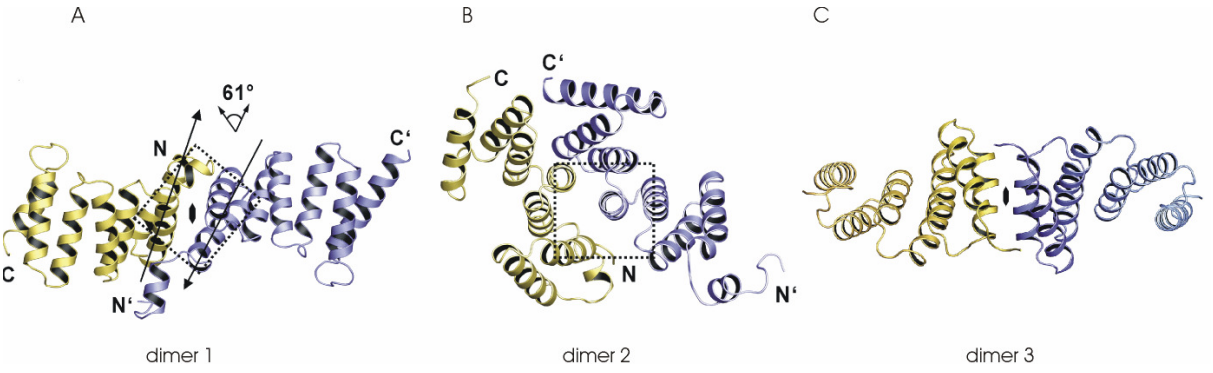

Fig. 1.9: Possible homodimer assemblies of SycD$^{21-163}$. A and C show head-to-head arrangement while B shows end-to-end arrangement of SycD molecules.

1.4.5 Substrates of the class II chaperone IpgC in *Shigella*

IpgC independently associates with two effectors IpaB and IpaC [65], which are encoded in one large operon (Fig. 1.2 and 1.6). Formation of IpaB-IpaC complexes in the bacterial cytoplasm is prevented by the association of each of these proteins with IpgC. The substrate proteins IpaB and IpaC along with IpaD were shown by transposon insertion and deletion mutagenesis to be essential for invasion, vacuolar escape, induction of macrophage apoptosis and virulence in animal models [41, 46, 100, 101]. IpaB and IpaC form a complex [65] that appears to be sufficient to invade epithelial cells. Though recently it has been shown that interaction between IpaB and IpaD at needle tips is key to host cell sensing, orchestrating secretion of IpaC [102].

IpaB is a 580 amino acid long, 62 kDa protein. In addition to its role in invasion, several lines of evidence indicate that the *Shigella* virulence factor IpaB is both necessary and sufficient to induce apoptosis in macrophages. First, mutant strains of *Shigella* that are invasive but do not express IpaB are not cytotoxic [101]. Second, microinjection of IpaB into the cytosol of macrophages efficiently triggers apoptosis [103]. Third, IpaB was shown to bind caspase-1 [also called IL-1β converting enzyme (ICE)] which plays an important role in *Shigella* induced apoptosis [104, 105]. Caspase-1 is a proapoptotic and proinflammatory cysteine protease. When activated, it cleaves the pro-inflammatory cytokines IL-1β and IL-18 to their biologically active forms [106, 107]. Together these data suggest a model where IpaB is secreted into the macrophage cytosol during infection.

Several structural features on IpaB have been mapped and linked to functions. By two-hybrid selection in yeast the chaperone binding domain has been mapped between residues 58 and 72 of IpaB [108]. The region from amino acids 110-170 has been proposed to facilitate trimerization [109]. The C-terminal part contains two hydrophobic domains, between residues 313 and 346 and between residues 398 and 423 [110]. Moreover, $IpaB^{418-580}$ differentially inhibits IpaB- directed liposome fusion and bacterial entry into cultured cells [109]. The N-terminal 75 amino acids are necessary for secretion while amino acids after 401 are dispensable for invasion and cytotoxic effects though Caspase-1 binding has been mapped to the region after amino acids 401 [111].

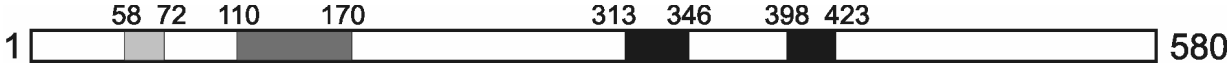

Fig. 1.10: Functional organization of IpaB. The N terminus contains chaperone binding region from 58-72 [108] and a region from 110-170 involved in oligomerization. C terminus possesses two hydrophobic regions from 313-346 and 398-423

IpaC is a 363 amino acid long, 42 kDa protein. In addition to its role in translocation, IpaC was also shown to mediate actin polymerization and filopodia formation in host cell [112, 113]. IpaC has been shown to trigger decrease in host phosphoprotein content [114]. Preincubation with purified IpaC has been demonstrated to result in the uptake of non-invasive *S. flexneri* [114] as well as enhanced invasion for wild type *S. flexneri* [115]. IpaC, at the molecular level, possesses a distinct functional organization. By two-hybrid screening, a region between residues 73 and 122 was shown to be the chaperone binding domain [108] while another study mapped the IpgC binding region between amino acids 50 and 80 [116]. The central region contains two hydrophobic domains, between residues 100 and 120 and between residues 150 and 170 and is necessary for interacting with phospholipids membranes [117]. The IpaC C-terminal region between amino acids 319 and 345 is predicted to form a coiled-coil structure. Together with IpaB, this IpaC region is also required for the formation of translocon pores in target cell membranes. Mutations within the C-terminal tail of IpaC defined by residues 345-363 have no effect on contact hemolysis but contribute significantly to IpaC's ability to trigger *S. flexneri* entry into cultured cells. IpaC invasion function requires its immediate C-terminus and is involved in its ability to trigger actin nucleation [118].

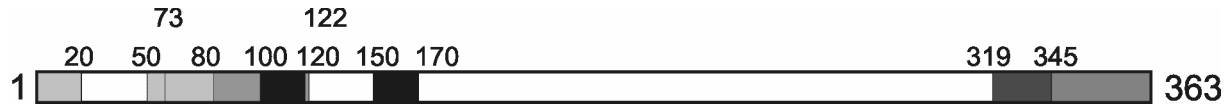

Fig. 1.11: Functional organization of IpaC. The N terminus contains export signal sequence with in 20 amino acids and harbours sequences for interaction with IpgC (residues 50-80 according to [116] and 73-122 according [108]). The central hydrophobic region (residues 100-120 and 150-170) is involved in penetration of phospholipid membranes. The C terminus comprises regions responsible for oligomerization (aa 319-345) and effector function (aa 345-363).

1.5 Protein X-ray crystallography

X-ray crystallography and NMR spectroscopy are the most prominent techniques in determining 3D structures of macromolecules at atomic level. In the present study the crystal structure of IpgC alone and in complex with the chaperone binding domain of IpaB was solved and is discussed.

In this chapter, a general introduction into structure solution by X-ray crystallography will be given which is extracted from different sources [119] [120] [121] [122].

X-ray diffraction analysis is by far the only technique that allows direct visualization of protein structure at the atomic or near-atomic level as applied to single crystals of pure proteins. This technique was successfully applied to solve the structures of IpgC and IpgC-IpaB.

X-ray crystallography has emerged as one of the most powerful and precise method for determining protein structures. As of 1st August 2008, over 85 % (~45,000 out of 52,103) of all protein structures deposited in the Protein Data Bank (PDB) were solved by this method. It requires a highly ordered single diffraction quality crystal. Certain bottlenecks along the way to successful structure determination are i) the high amounts of pure and soluble protein required, ii) the crystallization itself that might take time and demands laborious screening, and iii) the diffracting quality of the crystals. However, over the past decade, automation in X-ray crystallography regarding protein expression and purification, high-throughput micro-crystallization and documentation, high end synchrotron radiation beamlines and automated data collection by remote control from the home source have cut short the total time required for the whole process.

1.5.1 Principles of X-ray diffraction

When an electromagnetic wave interacts with a crystal, in which the spatial arrangement of the atoms is highly ordered, it is diffracted into secondary waves whose frequencies are either lower as that of the primary wave (inelastic scattering or Compton effect) or identical to it (elastic scattering or Thomson effect). Protein crystallography is based on elastic scattering of X-rays by protein crystals, which has been first discovered by Max von Laue in 1912, who predicted that diffraction would occur due to same dimensions of incident beam wavelength and atom distances in the crystal.

Elastic scattering is the outcome of the interaction between an oscillating electromagnetic field vector and electrons of atoms. By that, the electrons are excited and behave as oscillating electrical dipoles, and re-radiate secondary waves of the same wavelength. Due to phase differences between secondary waves, two phenomena are observed: constructive and destructive interference. Constructive interference occurs if waves have the same orientation and are in-phase. This results in an increased intensity. During destructive interference the waves extinct each other as they do not have the same orientation and are by that out-of-phase. Fig. 1.12 describes a diffraction event (2-dimensional section of a crystal lattice) with parallel planes within a crystal producing diffracted X-rays.

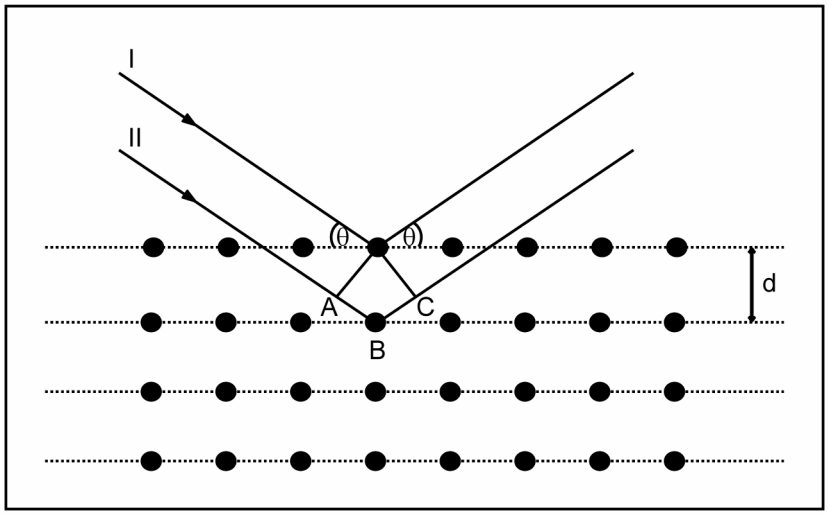

Fig. 1.12: Scheme of a 2-dimensional crystal lattice. Two incident X-rays (I and II, left side) are diffracted by the angle θ. Lattice points are displayed as dark circles, the distance between two lattice planes is given as d.

For mathematical reasons, X-ray diffraction is regarded as reflection on grid-planes in this geometric model: The crystal lattice is divided into sets of parallel planes, which are anchored on lattice points from each of the three crystallographic unit axes (a - along x, b - along y, c - along z) of the crystal's coordinate system. The crystalline lattice itself is composed of identical units called elementary cells. These are the smallest repeating units uniquely and completely defining the crystal, which itself can be generated by using periodic translations along the three crystallographic unit cell axes (a, b, and c).

For the two parallel incident waves (I and II in Fig. 1.13), the path difference of AB + BC has to be equal to 2 d sinθ (d: distance between planes, θ: angle of reflection).

Constructive interference of two waves can only occur if the path difference AB + BC is an integer number of their wavelength. This diffraction condition was first described in 1913 and is called Bragg's Law:

$$2d\sin\theta = n\lambda$$

The Ewald construction [122] offers a geometrical way of visualizing the conditions under which re-radiated X-rays will constructively interfere in reciprocal space Fig.1.13.

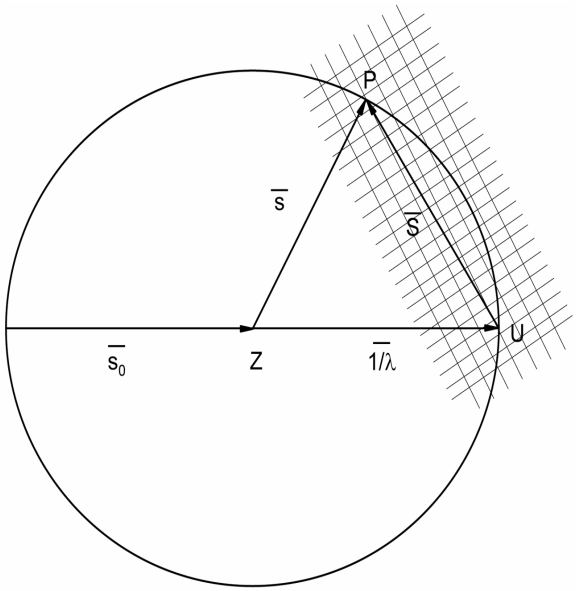

Fig. 1.13: Ewald construction (For description, see text).

Assume that the crystal lies in the centre of the Ewald sphere of radius 1/λ. The incident X-ray (s_0) is diffracted at the crystal lattice by 2 θ. The origin of the reciprocal lattice (U) lies on the surface of the Ewald sphere such that the incident X-ray passes through U. The difference vector S lies orthogonal to the lattice plane of the crystal lattice (Equations 2 to 4). If the endpoint of S intersects the Ewald sphere, P will be the correspondent point of the reciprocal lattice. In other terms, for each point of the reciprocal lattice, which is lying on the Ewald sphere, Bragg's law is satisfied. As the crystal is rotated in the X-ray beam, various reciprocal lattice points come into contact with this sphere and satisfy the diffraction or reflection conditions.

$$n\lambda = 2d\sin\theta \Leftrightarrow \frac{n}{d} = \frac{2\sin\theta}{\lambda} = \mid \vec{S} \mid \qquad (1)$$

$$|\vec{s}_0| = |s| = \frac{1}{\lambda},\ \vec{S} = \vec{s} - \vec{s}_0,\ \vec{S} \perp \qquad (2, 3, 4)$$

X-ray diffraction is caused by the electrons of an atom. Electrons are distributed over a finite volume element and it is assumed that this distribution displays spherical symmetry for a given atom. Thus, diffraction by an atom does not depend on the orientation of the difference vector S, but on its absolute value (Equation 1). This is the reason why diffraction amplitudes of single atoms, also known as structure factors ($f_{i,\theta}$), are generally plotted as a function of $\sin(\theta)/\lambda$, as shown in Fig. 1.14.

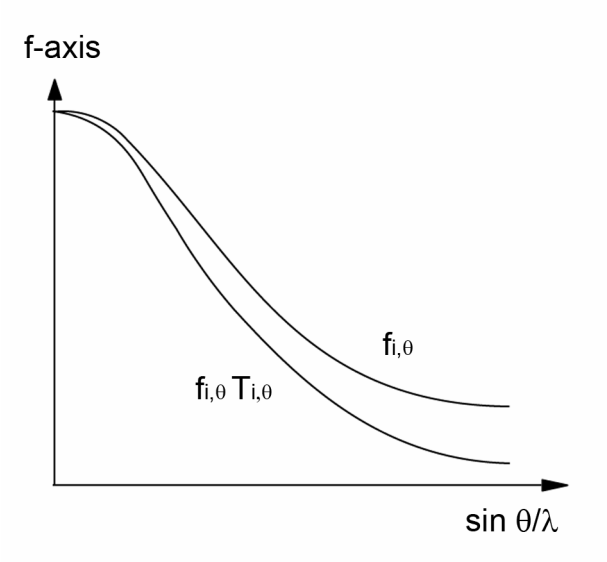

Fig. 1.14: Schematic diagram of the atomic structure factor $f_{i,\theta}$ and the thermal vibration corrected structure factor $f_{i,\theta}T_{i,\theta}$.

$f_{i,\theta}$ depends on the atom (number of electrons), the diffraction angle, the wavelength as well as the thermal vibrations (temperature factor) of the atom. Therefore, a second corrected curve is integrated in Fig. 1.14, which is obtained by multiplying the atom factor with the temperature factor ($T_{i,\theta}$). Although being anisotropic, $T_{i,\theta}$ is usually considered to be isotropic in protein crystallography due to the unfavourable ratio of observed to adjustable parameters (this reduces the parameters from 3 coordinates to 1). However, the correction factor includes further uncertainties such as inaccurate data scaling and statistical disorder of the crystal.

According to equation 5, the total diffraction by a crystal is the sum of the diffractive contributions from all atoms. For a unit cell containing n atoms, the structure factor is the sum of all atomic f values for individual atoms multiplied by their corresponding phase difference.

$$f(\vec{S}) = \int_v \rho(r) e^{(2\pi i \vec{r}\vec{S})} dv \tag{5}$$

$$F(S) = \sum_{j=1}^{n} f_j e^{(2\pi i \vec{r}_j \vec{S})} \tag{6}$$

Due to crystal mosaicity reflections are not sharp (point-shaped) and the lattice planes shown in Fig. 1.12 are in a “reflection position” over a finite angle. The resulting total intensity I(h) of an X-ray $I(_0)$ diffracted by a crystal, which is rotating uniformly about the angular velocity ω through the reflection positions, is given by the Darwinian formula [123].

$$I(\vec{h}) = \frac{I_0}{\varpi} \lambda^3 \frac{e^4}{m^2 c^4} \frac{1+\cos^2 2\theta}{2} \frac{LAV_x}{V^2} | F(\vec{h}) |^2 \tag{7}$$

In equation 7, h denotes a reflection, m the mass of an electron, c the light velocity and e the elementary charge. The term $(1+\cos^2 2\theta)$ is called polarisation factor, which gives the partial polarisation of the diffracted X-ray under the Bragg angle θ for an unpolarized incident X-ray. L, the Lorentz factor, considers the relative time that each reflection spends in the “reflection position”. A is the absorption factor, V_x the crystal volume and V the volume of the unit cell.

Thus, |F(h)| is the structure factor amplitude for a reflection h. Consequently, the intensity I(h) of the secondary wave is proportional to the square of the amplitude of the structure factor |F(h)|.

1.5.2 Calculation of the electron density

The structure factor |F(h)|, which contains the whole information of a diffraction experiment, is the Fourier transform of the electron density (ρ). Therefore, the electron density at any given position xyz within the unit cell can be calculated from the structure factors by performing a Fourier transformation according to equation 9:

$$\vec{F}(hkl) = | \vec{F}(hkl) | e^{i\alpha(hkl)} \tag{8}$$

$$\rho(xyz) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\vec{F}(hkl)e^{(-2\pi i[hx+ky+lz])} \quad (9)$$

The phase angle α(hkl) of any reflection hkl in a diffraction experiment is unassigned. Only the structure factor amplitudes |F(h)| can be calculated by integration of the reflection intensities (Equation 8). This is called the phase problem. It can only be solved using additional experimental methods.

1.5.3 Solution of the crystallographic phase problem

There are 4 possibilities to solve the phase problem:
a) Multiple Isomorphous Replacement (MIR)
b) Multiple Anomalous Diffraction (MAD)
c) Molecular Replacement (MR)
d) Direct methods
The phase problem is most often overcome by MIR using heavy atom derivatives. However, methods such as MAD, become more and more attractive [124] and are used in the past years without the need to label the protein (selenomethionine, selenotryptophan, etc), but directly using the sulphur of cyst(e)ins or methionines. For large proteins/protein complexes, MIR will still remain indispensable if no further structure informations are available [125] [126] [127]. With the increasing number of available protein structures, the chances of obtaining initial phase informations by Patterson search methods such as MR are also increasing for structure determination of homologous proteins [128]. Direct methods are successfully in use for small molecule crystallography.

The MAD method will be briefly described in the following.

1.5.4 Principles of MAD phasing

The method takes advantage of the tunability of synchrotron radiation X-ray sources (0.5-2.0 Å) and the presence of anomalous scatterers in the crystal that have absorption edges in the wavelength range around 1 Å. When the energy of incident X-rays approaches the absorption edge energies of an atom, resonance occurs which results in anomalous scattering. The absorption edges of C, O, N, S and H are far away from the accessible energy range and therefore are not suitable for MAD

phasing. MAD phasing can be carried out using proteins in which methionine residues are replaced by selenomethionine [129].

1.5.5 Choice of wavelengths

Choice of wavelength is quite critical when performing a MAD experiment as wavelengths are selected so as to maximize the f'' component of the anomalous scattering (equation 10). This in turn maximizes the f' component, which is the source of isomorphous differences between the data collected at different wavelengths. The total scattering factor can be described by two correction terms:

$$f = f_o + f' + f'' \tag{10}$$

Where f is total scattering, f_o is the normal or Thomson scattering and f' and f'' are real and imaginary components of the anomalous scattering.
The form and the position of the absorption edge depends on the chemical environment of the heavy atom in the crystal, therefore one cannot take the theoretically calculated values of f' and f'' of isolated atoms in vacuum. An X-ray absorption spectrum can be measured from the protein crystal containing the anomalous scattering element.

Typically diffraction data are collected at three wavelengths:

1. Absorption edge or Peak (λ_1) maximizes the f''component of the anomalous scattering and produces the largest differences in Bijvoet pairs.
2. Inflection point of the absorption edge (λ_2) minimizes the f' component.
3. Remote wavelength (λ_3) is usually collected above the absorption edge (smaller wavelength/higher energy).

1.5.6 Methods for locating anomalous scattering atoms

Most MAD structures have many anomalous scattering atoms and this makes the Patterson function too complicated for manual interpretation. Two methodologies have developed for solving the substructure. The first approach is based on automated interpretation of the Patterson function in combination with difference Fourier techniques and is implemented in the program SOLVE [130]. The other approach is to use crystallographic Direct Methods [131, 132].

1.6 Isothermal titration calorimetry (ITC)

All biochemical reactions involve recognition, binding, and the formation of noncovalent complexes. Many ligands act by selectively binding to protein or nucleic acid targets and altering their binding characteristics. Since all binding events are accompanied by the evolution or absorption of heat (a change in enthalpy, ΔH), a full thermodynamic characterization of the binding reaction provides fundamental information about the molecular interactions driving the association process.

ITC provides a direct route to the complete thermodynamic characterization of protein interactions. ITC is used mainly to measure affinity for protein-protein and protein-ligand interactions. ITC measures the enthalpy of a reaction; upon each titration the amount of heat released or absorbed is measured. In addition to the equilibrium dissociation constant (K_D), ITC titration allows to determine the number of binding sites (*n*), binding enthalpy (Δ*H*) and entropy (Δ*S*).

For a typical experimental setup, a syringe of ITC containing a ligand is titrated into a cell containing a protein solution. As the two elements interact, heat is released or absorbed in direct proportion to the amount of binding that occurs. When the protein in the cell becomes saturated with the added ligand, the heat signal diminishes until only the background heat of dilution is observed (Fig. 1.15).

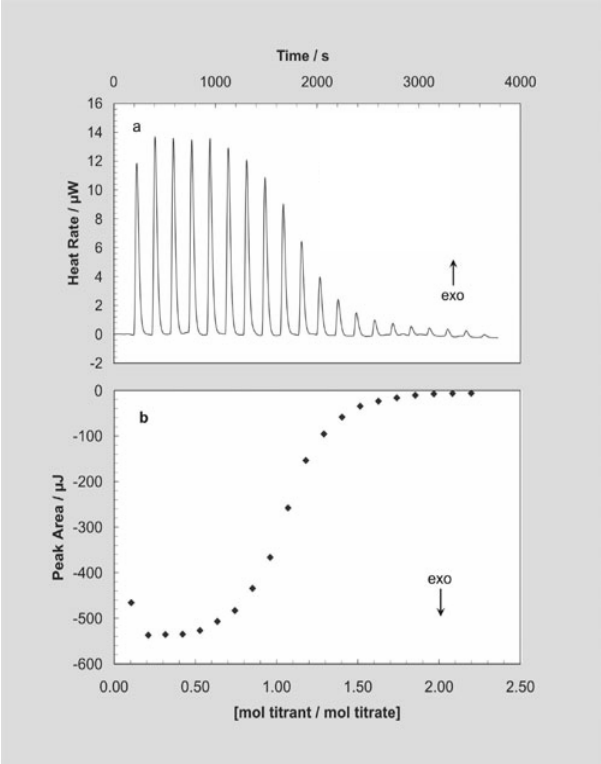

Fig. 1.15: ITC thermogram. A, By using reactant concentrations appropriate for the binding reactions, raw ITC data for the titration of a small molecule inhibitor into a protein yields a (B) sigmoidal area plot. Adopted from [133]

Thermodynamic calculations

From analysis of each generated thermogram, the heat of dilution ($\boldsymbol{h}_{dil}$) is used to determine the heat of reaction ($\delta\boldsymbol{h}_i$) for each injection by the equation

$$(\delta\boldsymbol{h}_i) = (\boldsymbol{h}_i) - (\boldsymbol{h}_{dil}) \quad (1)$$

where $\boldsymbol{h}_i$ is the heat released at a given injection calculated by the integration of the respective injection peak. The integration of an injection peak is used to calculate $\boldsymbol{h}_i$. The reaction enthalpy ($\Delta\boldsymbol{H}$) can be calculated from the equation

$$\Delta H = \sum_i \delta hi / n_{tp} \quad (2)$$

where $\boldsymbol{n}_{tp}$ is the total moles of protein in the cell. Plotting $\Delta\boldsymbol{H}$ against the molecular ligand:protein ratio produces a binding isotherm.

Nonlinear regression fitting to the binding isotherm (ORIGIN software; MicroCal Software, Inc.) gives the equilibrium dissociation constant of the ligand K_D. From the value of K_D, the free energy of binding ($\Delta G°$) and entropy of binding (ΔS) can be calculated from the following relationships

$$\Delta G° = -RT \ln (1/K_D) = \Delta H - T\Delta S \quad (3)$$

where T is 273 K and R is 1.9872 cal/Kmol.

Advantages:

Unlike other methods, it is a truly equilibrium solution method where KD is measured for native proteins;

is not limited by the ligand or protein size;

no labelling, modification and/or immobilization is required;

and is not affected by the optical properties of the samples.

Disadvantages:

ITC requires relatively high concentrations and volumes of sample. To address this new calorimeter (iTC200, MicroCal, LLC) requiring up to 7 times less sample volume i.e., 200 µl is recently introduced.

1.7 Goals of this study

Most of the current knowledge on mechanisms underlying Shigella pathogenesis is derived from studies of *S. flexneri*. Efficient type-III secretion utilized by pathogenic *Shigella* depends on cytosolic molecular chaperones. Based on the type of the effector binding, chaperones of TTSS have been classified into three groups. The mode of binding and crystal structures from class I and III are well characterized. However, little is known about the class II chaperones on the molecular level and no crystal structure was available until recently.

The goal of my study was to characterize the structure and function of a class II chaperone, IpgC from *Shigella flexneri*. In addition to structural characterization utilizing X-ray crystallography, functional analysis of the chaperone binding domains (CBDs) in its substrates- IpaB and IpaC was pursued. Further, functional relevance of CBDs on *Shigella* was attempted.