3 Results

This section is discussed in three parts.

Part-I will focus on characterization of the chaperone, IpgC. A progressive step by step approach in obtaining a good quality crystal is presented, as hitherto this class of chaperones have proved recalcitrant to crystallization. The structure of a functional chaperone is discussed. Further, the role of residues and especially, the N terminal helix in maintaining the quaternary structure is presented.

In Part-II, the characterization of IpaC, one of the two substrates that bind to IpgC is presented. IpaC is unstructured as revealed by limited proteolysis of IpaC-IpgC complex. Different novel approaches in precisely defining the chaperone binding domain (CBD) in IpaC are presented. Further, this information is applied in cocrystallization of IpaC-IpgC complex.

In Part-III, the characterization of IpaB, the other substrate that bind to IpgC is discussed. IpaB reveals two CBDs. The functional characterization of each these domains is discussed. Further, various strategies employed in obtaining cocrystal structure of CBDs of IpaB with IpgC is presented.

3.1 Part I: Structural characterization of the functional class II chaperone IpgC

3.1.1 Crystallization of IpgC

Wild type *ipgC* was cloned in pET28a having an N-terminal thrombin cleavable hexa-histidine tag. *ipgC* was recombinantly expressed in *E. coli* and purified utilizing 6x His tag. The tag was cleaved off and finally the protein was gel filtrated in 20 mM HEPES pH 7.4, 50 mM NaCl. Crystals were obtained in many conditions. Initial crystallization conditions were improved and the best diffracting crystals were obtained in condition containing 0.01 M hexadecyltrimethylammonium bromide, 0.5 M NaCl and 0.01 M MgCl$_2$. Diffraction data were collected up to a resolution of 3.5 Å at BESSY beamline 14.1, Berlin. The crystals belonged to the space group P3$_2$21 with unit cell dimensions $a = b = 105.26$ Å, $c = 165.77$ Å.
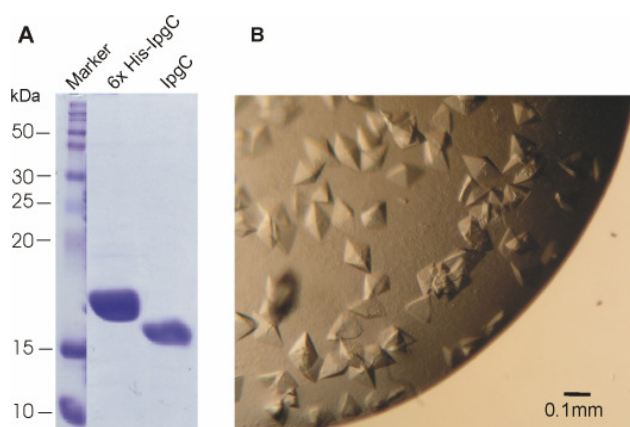

Fig. 3.1: **A**, shows the purified IpgC that was used for crystallization (**B**). The 6x His tag on the protein used for purifying the protein was cleaved off before crystallization trials.

For MAD phasing, selenomethionine (SeMet) labelled IpgC was crystallized in the condition mentioned above. Curiously, the crystals appeared overnight as in the case of unlabelled protein but disappeared by day two marked by increased precipitation (Fig. 3.2, B). Protein stored at 4 °C and loaded on a SDS-PAGE on day seven showed degradation which might explain instability of crystals.
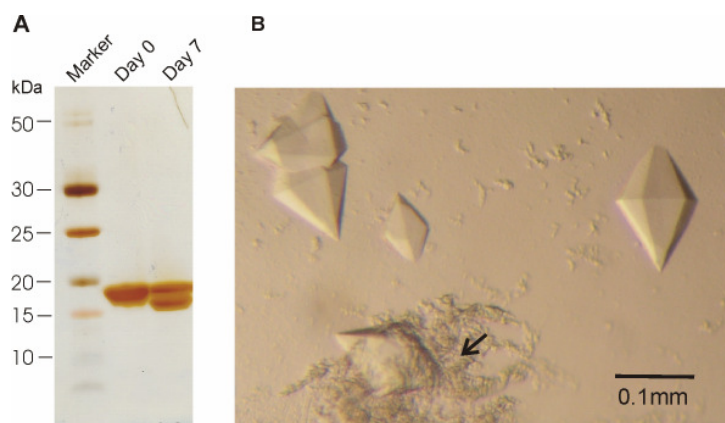

Fig. 3.2: **A**, shows silver-stained SDS-PAGE of SeMet IpgC loaded with sample from day 0 and day 7 stored at 4 °C. The latter sample shows degradation which explains instability of the crystals indicated by an arrow in **B**.

3.1.1.1 Overcoming the SeMet crystal instability

Close inspection of the sequence revealed presence of two methionines at the N terminus after cleavage of the 6x His tag. One methionine each contributed by the vector and the *ipgC* gene (Fig. 3.3, A) are present in close proximity. It was reasoned that presence of these two methionines in close vicinity at the N terminus destabilized SeMet crystals
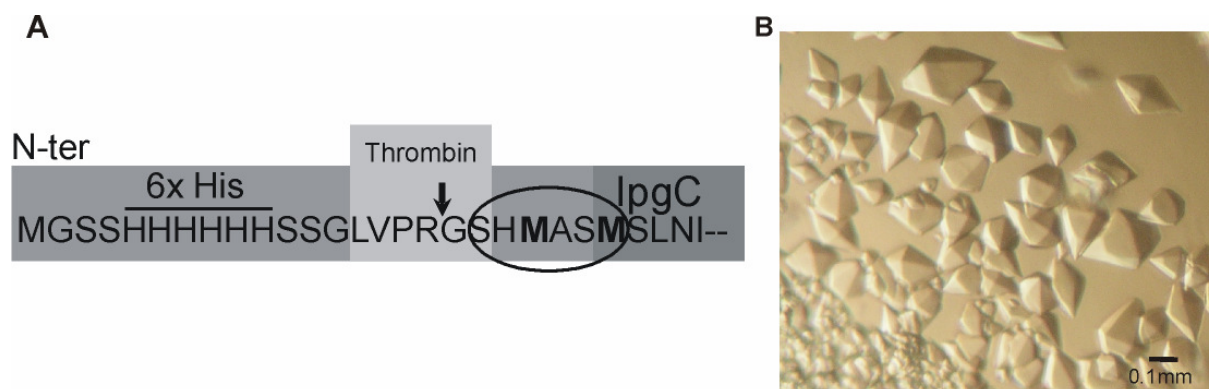

Fig. 3.3: N terminus of IpgC contained two methionines (**A**, in bold) after proteolytic cleavage of 6x His tag by thrombin (indicated by arrow). The residues encircled which include two methionines were deleted by site directed mutagenesis to obtain stable protein crystals shown in **B**. The crystal drop was devoid of any precipitation.

As a consequence, a construct much closer to the native protein having only a M1G substitution was generated (Fig. 3.3, A). Indeed, this protein yielded stable SeMet crystals devoid of any precipitation (Fig. 3.3, B). Diffraction data were collected up to a resolution of 3.4 Å at BESSY beamline 14.1, Berlin. The crystals belonged to the space group $P6_222$ or $P6_422$ with unit cell dimensions $a = b = 104.7$ Å, $c = 170.2$ Å. Poor diffraction of both the native and SeMet crystals compounded by failure in obtaining phase information solution called for a new approach.

3.1.1.2 Stabilization of the N-terminus

Having noticed the influence of manipulating the N-terminus in obtaining stable SeMet crystals, an eight amino acid long *Strep*-tag II [150] with amino acid sequence WSHPQFEK and having a secondary structure was introduced to stabilize the N-terminus. The secondary structure of *Strep*-tag II recognizes and selective binds to streptavidin. The *Strep*-tag fused IpgC was purified in two stages; initially by *Strep*-Tactin mentioned elsewhere (section 2.2.3.2) and finally gel filtrated in 20 mM HEPES pH 7.4, 50 mM NaCl (Fig. 3.4). The best diffracting crystals were obtained in

condition containing 2.0 M ammonium formate, 0.1 M NaCl and 20 mM HEPES pH 7.4. Diffraction data were collected up to a resolution of 5 Å at BESSY beamline 14.2.
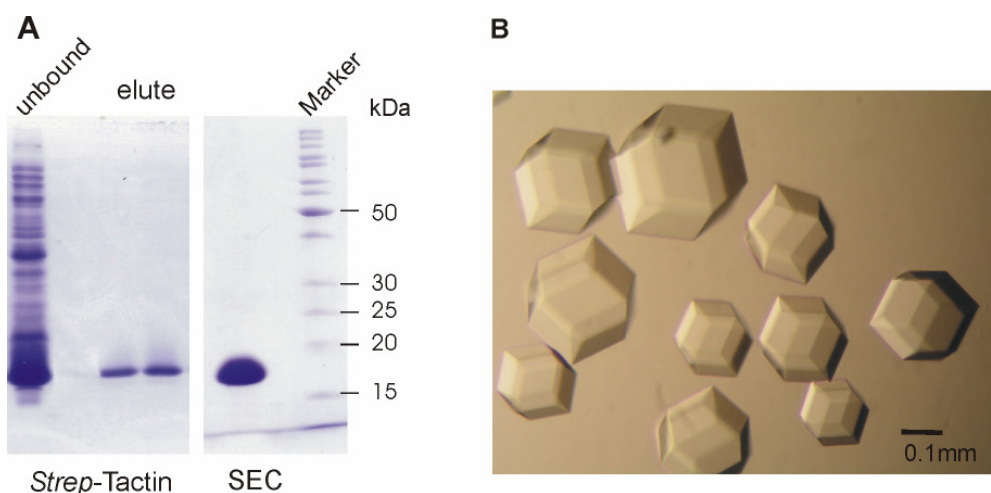

Fig. 3.4: Purification and crystallization of N terminus *Strep*-tag II fusion IpgC. N-terminus of IpgC was fused to 8 amino acid *Strep*-tag II. **A**, shows purification scheme using *Strep*-Tactin and size exclusion chromatography (SEC). **B**, shows crystal obtained which at best diffracted to 5 Å.

3.1.1.3 Utility of *in silico* tools

Two web based servers have been tried to find out alternate solutions to make IpgC amenable to crystallize in different conditions. First, Surface Entropy Reduction (SER) [163] was employed which can be accessed from NIH MBI Laboratory Servers maintained by UCLA [140]. This server suggested mutating surface exposed, high entropy amino acids residues such lysines and glutamates with residues that have small, low entropy side chains such as alanines. These are likely to enhance protein's crystallizability via generation of crystal contacts [164]. The proposed mutations were clustered in three categories: (1) E137A, K138A; (2) K101A; and (3) K122A, K124A, E125A. Next, IpgC was run through Predictor of Naturally Disordered Regions (PONDR) [165] accessible from Molecular Kinetics, Inc. [143]. Fig 3.5 shows the output of IpgC from PONDR. Disordered regions are entire proteins or regions of proteins which lack a fixed tertiary structure, essentially being partially or fully unfolded. Although disordered regions lack a defined 3-D structure in their native states, they frequently undergo disorder-to-order transitions upon binding to their partners. The PONDR use amino acid sequence data to predict disorder in a given region. The neural network trained on a specific set of ordered and disordered sequences, then outputs a value for the central amino acid in the window.

If a residue value exceeds a threshold of 0.5 the residue is considered disordered. PONDR predicted the C-terminus residues 150-155 i.e., IQDIKE to be disordered.
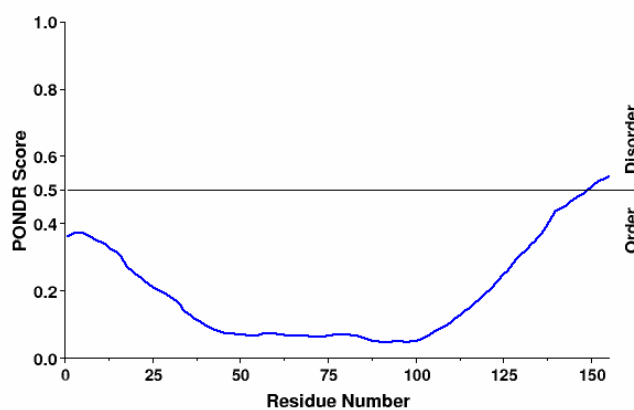

Fig. 3.5: PONDR prediction of IpgC shown as a graphic output of the residues those are ordered and disordered. The C-terminus residues 150-155 i.e., IQDIKE are predicted to be disordered.

The predictions from SER and PONDR suggested internal and terminal modifications respectively. The internal modifications from SER were ruled out as the suggested residues might be involved in intra- or intermolecular interactions. A careful decision based on PONDR prediction was taken. C terminus four amino acids instead of six suggested by PONDR were deleted to prompt IpgC crystallization in different condition. The predictions at best can be suggestive but not conclusive warranting human intervention to arrive at the best possible solution.

3.1.1.4 Crystallizing IpgC - a new approach

C terminus four amino acids- DIKE, all possessing flexible long side chains were deleted from IpgC that was used to generate stable and improved diffraction quality SeMet crystal. The recombinant protein was purified utilizing cleavable 6x His tag at the N-terminus. The tag was cleaved off and finally the protein was gel filtrated in 20 mM HEPES pH 7.4, 150 mM NaCl. The native crystals were grown against 0.1 M sodium N-(2-acetamidoimino) diacetic acid pH 6.5 and 1 M ammonium sulphate (Fig 3.6). While SeMet labeled IpgC was crystallized against 100 mM HEPES pH, 1.5 M LiSO$_4$. Diffraction data from SeMet labeled and native crystals were collected at BESSY (BL14.2), Berlin and ESRF (BL23.1), Grenoble respectively. Both native and SeMet labeled crystals belong to space group P3$_1$21, with two molecules in the asymmetric unit (a, b = 115.15 Å, c = 75.46 Å). The data statistics are listed in Table 3.1 and 3.2.
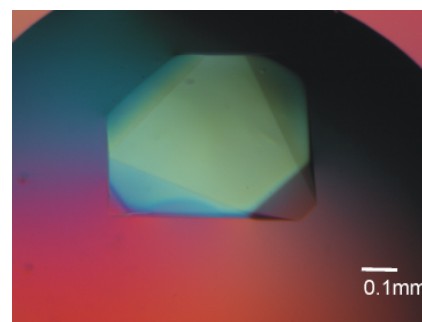

Figure 3.6: A single large crystal of native IpgC.

Table 3.1: X-ray-diffraction-data statistics of the IpgC crystal.

Data collection and processing					
	IpgC native	IpgCqd SeMet peak	IpgCqd SeMet inflection	IpgCqd SeMet high energy remote	IpgCqd SeMet low energy remote
wavelength (Å)	0.9762	0.97965	0.97987	0.97626	0.98793
resolution (Å)	38 - 2.10	40 – 3.10	40 – 3.10	40 - 3.10	40 – 3.10
resolution (Å) (last shell)	2.23 – 2.10	3.29 – 3.10	3.29 – 3.10	3.29 – 3.10	3.29 – 3.10
space group	P3$_1$21	P3$_1$21	P3$_1$21	P3$_1$21	
temperature (K)	100	100	100	100	100
X-ray source	ESRF BL 32.1	BESSY BL 14.2	BESSY BL 14.2	BESSY BL 14.2	BESSY BL 14.2
detector	ADSC Q315r	MARCCD 165mm	MARCCD 165mm	MARCCD 165mm	MARCCD 165mm
unit cell a (Å), α (°)	115.15, 90	113.5, 90	113.5, 90	113.5, 90	113.5, 90
parameters b (Å), β (°)	115.15, 90	113.5, 90	113.5, 90	113.5, 90	113.5, 90
c (Å), γ (°)	75.46, 120	76.7, 120	76.7, 120	76.7, 120	76.7, 120
total reflections (last shell)	296542 (19349)	75376 (12090)	71634 (11630)	68660 (11495)	71790 (11755)
mosaicity (°)	0.23	0.13	0.13	0.16	0.13
unique reflections (last shell)	33791 (5344)	19918 (3220)	19956 (3253)	19706 (3265)	19973 (3268)
$I/\sigma(I)$ (last shell)	9.33 (1.74)	10.18 (1.67)	10.0 (1.59)	8.97 (1.51)	10.22 (1.53)
% data completeness (last shell)	99.7 (96.6)	99.2 (98.3)	99.4 (99.3)	98.2 (99.6)	99.5 (99.7)
R_{meas}[a] (%) (last shell)	13.4 (78.6)	11.9 (87.6)	12.0 (89.3)	16.8 (92.4)	12.2 (95.0)
R_{sym}	12.7 (67.3)	10.3 (75.2)	10.2 (75.9)	14.3 (78.5)	10.4 (80.7)
r.m.s deviations					
bond length (Å)	0.008	n.a.	n.a.	n.a.	n.a.
bond angle (°)	1.2	n.a.	n.a.	n.a.	n.a.

	E (eV)	λ (Å)	f'	f''
Peak	12656.05	0.97965	-7.25	4.41
Inflection	12653.22	0.97987	-8.84	2.43
Remote (high)	12700	0.97626		
Remote (low)	12550	0.98793		

Table 3.2: Atomic scattering factors of SeMet-IpgC crystals at peak, inflection and remote wavelength. f' and f'' values were derived using the program CHOOCH [166].

For structure determination of IpgC, selenium peak-wavelength data to 3.1 Å resolution were used for multiple-wavelength anomalous diffraction (MAD) phasing. The Mathews coefficient suggested the presence of two IpgC molecules per asymmetric unit. All the 6 selenium atom sites were assigned using SHELXD. Refinement of the selenium atom was carried out by SHARP. Initial phases were calculated using SHARP, and density modification implemented in SOLOMON was used to improve phases. The initial map was obtained using Coot and CNS. The first model was built with ARP/wARP followed by multiple cycles of manual building using

Coot and refinement using CNS. The longest chain comprising 67 of the 151 residues was used to obtain a molecular replacement solution for the native IpgC crystals.

3.1.2 IpgC crystallized is fully functional

Complementation of a nonpolar *ipgC* deletion mutant *S. flexneri* ($\Delta ipgC$) [65] with either C-terminal deletion (*ipgC^{ct}*) or wild type (*ipgC^{wt}*) *ipgC* restored the secretion of effectors as seen in Fig. 3.7 (A). The amounts of cytoplasmic IpaB were greatly reduced in the *ipgC* mutant. This reduction in the amount of IpaB is not due to enhanced secretion of IpaB (Fig. 3.7 A) and indicated that significant part of IpaB was degraded in the absence of IpgC [65]. In contrast, IpaB accumulated to wild type levels when complemented with *ipgC^{ct}*. This indicated that *ipgC^{ct}* allowed the production and secretion of IpaB to a level comparable with *ipgC^{wt}*.

Next, the ability of *ipgC^{ct}* to restore invasion of epithelial cells in $\Delta ipgC$ mutant was evaluated by a gentamicin protection assay in HeLa cells (Fig. 3.7 B). Deletion of *ipgC* in *S. flexneri* ($\Delta ipgC$) led to abrogation of invasiveness. The invasiveness of $\Delta ipgC$/*ipgC^{ct}* was comparable to wild-type M90T showing that the *ipgC^{ct}* is fully functional.

Finally, the ability of *ipgC^{ct}* to restore cytotoxicity in $\Delta ipgC$ was tested (Fig. 3.7 C). The level of killing obtained with M90T was taken as a reference (100%) to calculate the cytotoxicity of $\Delta ipgC$ harbouring either *ipgC^{ct}* or *ipgC^{wt}*. Both *ipgC^{ct}* and *ipgC^{wt}* restored the ability to kill macrophages producing 100% and 97% cytotoxicity respectively.
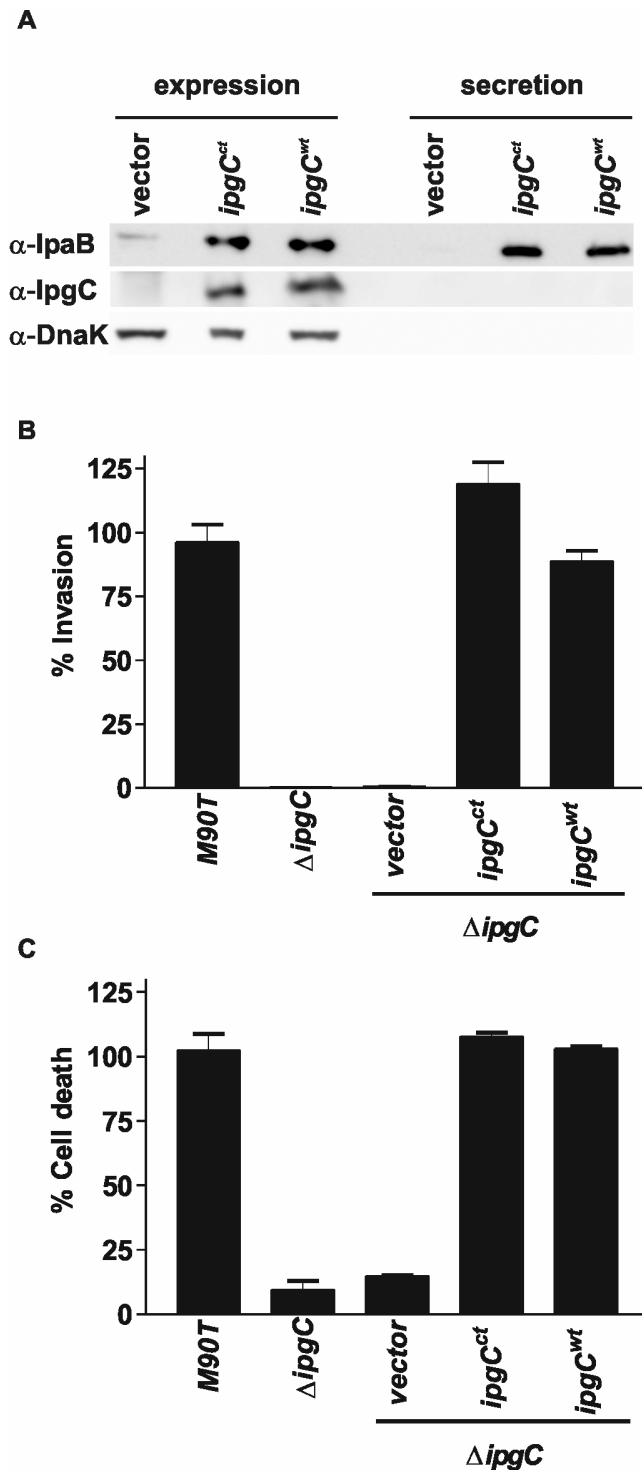
Fig. 3.7: **C-terminal truncated *ipgC* (*ipgC^{ct}*) is biologically active. A**, Expression of IpgC and expression as well as secretion of IpaB were analyzed by immunoblotting. For secreted protein bacteria-free culture supernatants were precipitated with TCA. DnaK was used as a control for bacterial membrane integrity and as a loading control. IpaB was detected in bacterial extracts and culture supernatants of Δ*ipgC* complemented with *ipgC^{ct}* or *ipgC^{wt}*. In the absence of *ipgC* (vector) IpaB is reduced in the cytoplasm and not secreted. *ipgC^{ct}* and *ipgC^{wt}* complement Δ*ipgC* for epithelial cell invasion measured with gentamicin protection assay (**B**) and macrophage cytotoxicity assayed by the release of LDH and measured as cytotoxicity percent (**C**). The results are the means ± s.e.m. of triplicate experiments.

Thus, complementation of a nonpolar *ipgC* deletion mutant (Δ*ipgC*) with either C-terminal deletion (*ipgC^{ct}*) or wild type (*ipgC^{wt}*) *ipgC* restored the secretion of effectors, epithelial cell invasion and macrophage cytotoxicity to similar levels (Fig. 1). This demonstrates that IpgC used in this study for structural characterization is fully functional.

3.1.3 The structure of IpgC reveals TPR domains

The crystallographic asymmetric unit contains two copies of IpgC (Fig. 3.8). Of the 151 residues, 9-151 could be traced in the electron density. The structure shows an all α-helical IpgC with 8 helices per subunit labeled H1 to H8. Helices H2 to H8 adopt a superhelical scaffold comprising tetratricopeptide repeat (TPR) domains with H2 to H7 forming three TPR motifs and H8 defining the first helix of a fourth motif. This arrangement creates a concave groove like surface (see Fig. below).
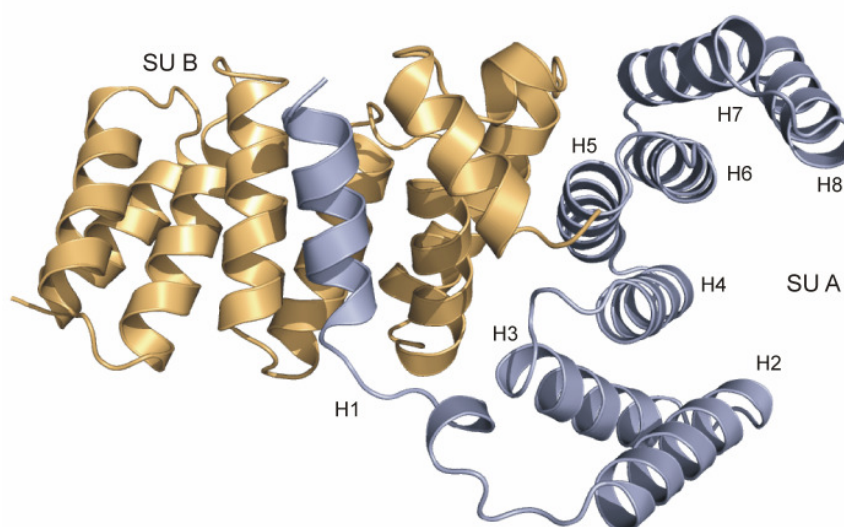

Fig. 3.8: Cartoon representation of the IpgC dimer found in the asymmetric unit with subunits (SUs) coloured in brown and blue. Helices of SU A are labeled from H1 to H8

TPR is a degenerate 34 amino acid anti-parallel α-helical repeat motif [167] described in diverse proteins with consensus sequence pattern. The motif occurs in tandem arrays and characterized by the presence of conserved small and hydrophobic residues at positions 8, 20 and 27 [168] (Fig. 3.9). Amino acids corresponding to positions 8 and 20 are situated where helices of TPR motif are in close contact while residues at position 27 are located at the interface of helices.

Fig. 3.9: Ribbon representation of an IpgC subunit highlighting the residues at canonical positions: 8 (A43, A77, G111), 20 (A55, A89, A123) and 27 (L62, A96, V130).
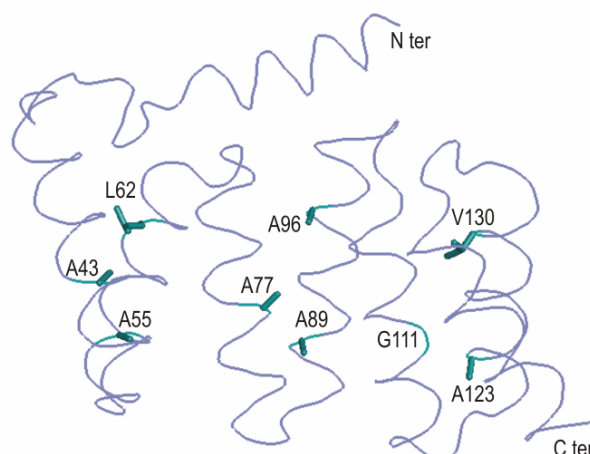
A multiple sequence alignment of IpgC orthologues from gram negative bacteria [169] (Fig. 6.1) validates the preference for a small amino acid at these positions. The structure of subunits designated A and B are similar for the TPR moiety with a RMSD of 0.7 Å but differ in the arrangement of the H1 and the loop connecting it which face away from each other when superimposed (Fig. 3.10, upper panel).
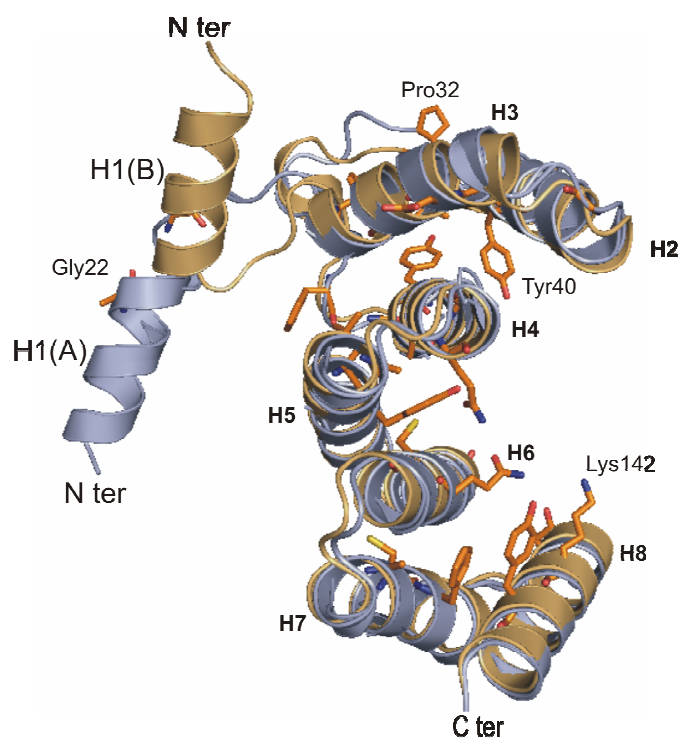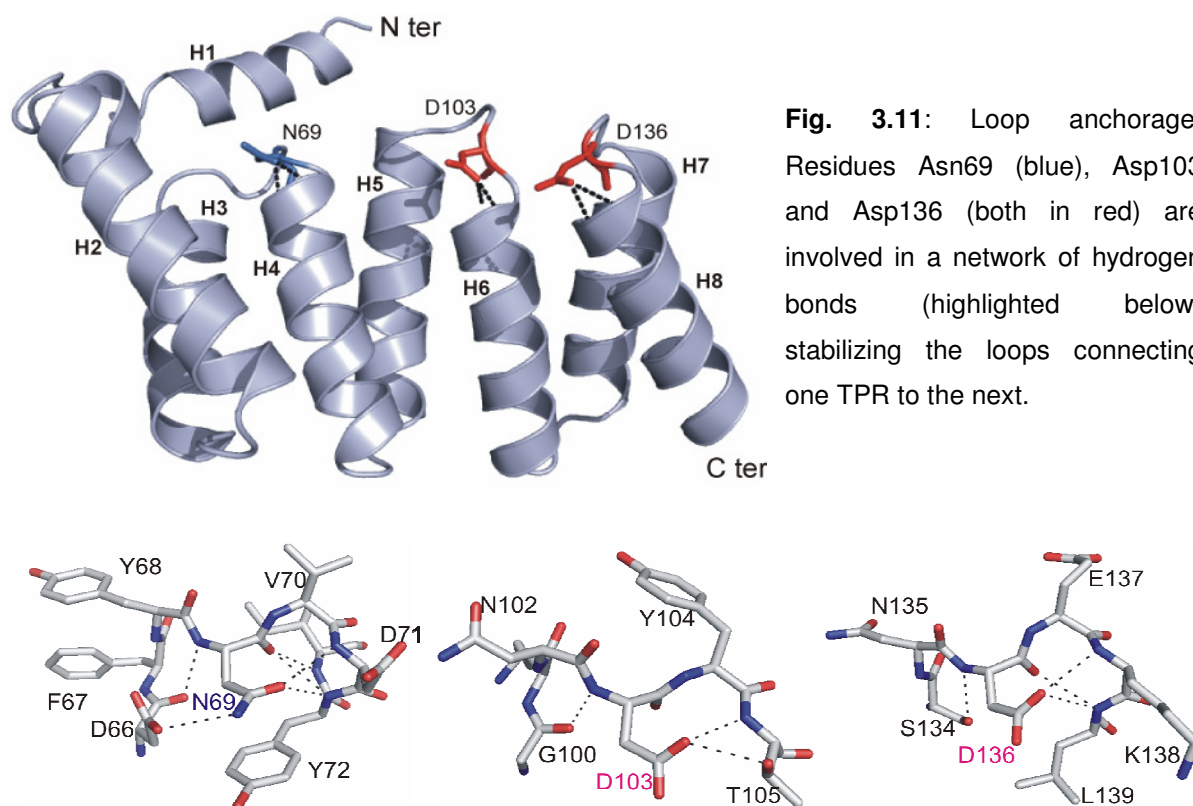

Fig. 3.10: Topological distribution of amino acids in IpgC and H1. The superimposition of the two IpgC subunits based on the Cα-atoms of the TPR domain is shown. SUs are marked in parenthesis for H1. Gly22 at the end of H1 and Pro32 before H2 are also labeled. Identical residues among IpgC homologue in SU A are coloured in orange and highlighted as stick model.

Most of the identical residues (highlighted in orange, Fig. 6.1) within the TPR domains of both subunits (SUs) are similarly positioned and with a few exceptions line the groove composed of every first helix of the TPR i.e., H2, H4, H6 and H8 (Fig. 3.10). In contrast, the opposite side of the cleft is lined by residues from both TPR α-helices [168]. Among the conserved residues, Tyr65 and Phe67 adopt different orientation in A and B. Tyr65 of A and B are involved in intra- and intersubunit polar contacts, while Phe67 of A and B contribute to inter- and intrasubunit hydrophobic interactions, respectively.

A network of hydrogen bonds stabilizes the loops linking one TPR motif to the next (Fig. 3.11). The inception of the helixes of TPR 2, 3 and 4 are capped by loop anchoring conserved residues Asn69, Asp103 and Asp136 respectively. Asn69

anchors the loop joining H3 to H4 to the backbone amide of Asp71 of H4. Asp103 and Asp136, in turn, anchor loops connecting H5 and H7 to H6 and H8 respectively (Fig. 3.11).

Fig. 3.11: Loop anchorage. Residues Asn69 (blue), Asp103 and Asp136 (both in red) are involved in a network of hydrogen bonds (highlighted below) stabilizing the loops connecting one TPR to the next.

An 'aspartic acid-array' at one end of the groove is formed by Asp27, Asp71, cluster of aspartates at positions 33, 34, 37 and 38 together with the loop anchoring aspartates i.e., Asp103 and Asp136 (Fig. 3.12). This highly negatively charged patch may provide distinctiveness to the substrate binding.
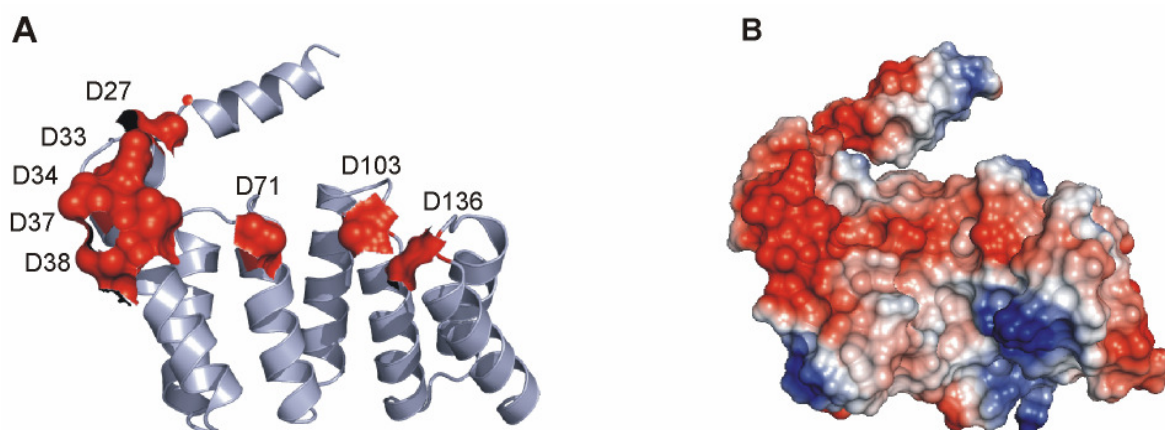

Fig. 3.12: **A**, Aspartic acid-array at one end of the groove comprising of aspartates at positions 27, 33, 34, 37, 38, 71, 103 and 136. **B**, shows electrostatic potential map on the protein surface in the same orientation. Positive and negative potential is colored blue and red, respectively.

3.1.3.1 Interactions at the dimer interface

Extensive intermolecular interactions are involved in the dimer contact. Over 30% of the theoretically accessible surface of both copies found in the asymmetric unit (4493.5/14917.2 [Å2]) is involved in dimer contacts. A combination of hydrophobic and polar interactions (Table 3.3) at the subunit interfaces between both chains formed by the first (H3) and the second TPR (H4, H5) of SU A and the first TPR (H2, H3) of SU B complement the extensive intermolecular contact.

subunit A	position of residue	subunit B	position of residue	type of interaction
Ile64	H3	Val57	H3	hydrophobic
Phe67	loop H3-H4	Val57, Phe58, Phe61	H3	hydrophobic
Tyr68 OH	**loop H3-H4**	**Arg60 NH2**	**H3**	**H bond**
Gln87 NE2	**H5**	**Asp38 OD1**	**H2**	**H bond**
Gln88 NE2	**H5**	**Tyr42 OH**	**H2**	**H bond**
Leu92	H5	Tyr42, Phe58	H2	hydrophobic
Ala94	H5	Met35, Ile39	H2	hydrophobic
Val95	H5	Ile39, Phe58, Phe61	H2, H3	hydrophobic
Phe97	H5	Ile31, Pro32, Met35	loop H1-H2, H2	hydrophobic
Ala98	H5	Ile31, Phe61, Tyr65	loop H1-H2, H3	hydrophobic
Ala98 O	**H5**	**Tyr65 OH**	**H3**	**H bond**
Leu99	H5	(Ile11), Phe61, (Tyr65)	(H1), H3	hydrophobic
Lys101 NZ	**loop H5-H6**	**Asn29 O**	**loop H1-H2**	**H bond**

Table 3.3: Residues involved in the intermolecular interactions. Residues from H3, H4 and H5 from SU (A) interact with loop H1-H2, H2 and H3 from SU (B). Both hydrophobic and polar interactions (bold) are involved. Residues in the parenthesis are likely to be involved in the interaction.

3.1.3.2 Helix 1 (H1) in dimerization

H1 and the loop connecting it to H2 show different arrangement in both SUs, as a consequence of the asymmetric interactions, (Fig. 3.10). In fact superimposition of the TPR domains of both SUs show that the N-terminal 32 amino acids face away from each other. The amphiphilic helix H1 is connected to H2 via an extended loop, the termini of which are defined by conserved residue Gly22 at one end facilitating flexibility together with Pro32 at the other end (Fig. 3.10). Over the length, H1(A) is stabilized by the hydrophobic pocket provided by H3, H4 and H5 of subunit B (Fig. 3.13). Additional hydrophobic residues line up the interface between the H1(A) and H1(B) with polar residues being oriented against the bulk. The backbone of conserved Ala18 and side chains of Ser21 of H1(A) which is H bonded to Gln88 of H5(B) fixes the H1(A) at the top of the loop (Fig. 6.4). On the other hand, intra-chain

backbone hydrogen bonds between conserved Ala18 and Ala23; Ala18, Ile19 and conserved Gly22 fixes H1(B) on top the loop supported by hydrophobic interaction in the interface between H1(B) and H1(A). Furthermore, a combination of hydrophobic and polar interactions at the subunit interfaces contributed by H2, H3 of subunit B and loop joining H3 and H4 of subunit A and H5 (A) mediate intermolecular interaction (Fig. 3.13).
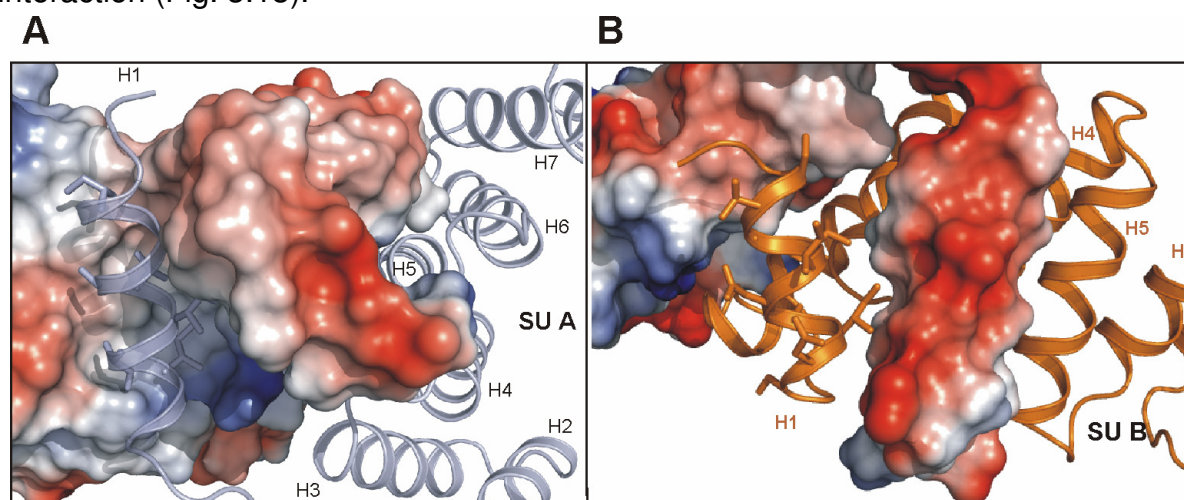

Fig. 3.13: The interactions of the amphiphilic H1. The asymmetric dimer organization allows of SU A and B to interact differently. **A**, shows the continuous and mostly hydrophobic interactions involving H1 (A) with SU B. **B**, shows the interaction of H1 (B) with SU A.

3.1.3.3 Functional IpgC is a dimer

The apparent molecular mass as determined by analytical size exclusion chromatography gave a value between 58.5 - 65 kDa which corresponds to 3.2 - 3.5 IpgC molecules (18.33 kDa). By more exact method i.e., multi-angle light scattering (MALS) the quaternary structure of the IpgC was established to be a dimer (Fig. 3.14). The larger value from former method could be attributed to the elongated conformation of the dimer and perhaps to high solvent content as well.
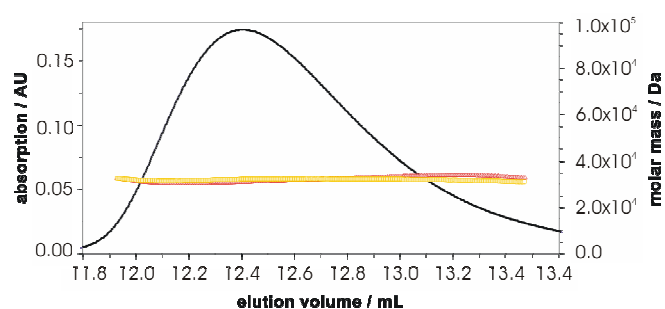

Fig. 3.14: On-line static laser light scattering reveals IpgC to be dimer. The black curve shows the protein absorption at 280 nm (left axis) versus volume on x axis. The yellow and orange lines refer to the measured molecular mass (right axis) for two independent experiments.

The asymmetry in the residues as well as different type of interactions involved at the subunit interfaces indicates that crystal packing may reflect the biological dimer found by analysis by multi-angle light scattering (MALS).

To assess the physiological role of IpgC dimerization, functional complementation of *ipgC* null mutant (SF619) with *ipgC* mutants lacking N terminus 21 (*ipgC*$^{\Delta 21}$) and 32 (*ipgC*$^{\Delta 32}$) amino acids were assayed in a gentamicin protection assay with HeLa human epithelial cells (Fig. 3.15). *ipgC*$^{\Delta 21}$ corresponds to IpgC lacking the H1 and *ipgC*$^{\Delta 32}$ corresponds to IpgC lacking H1 and the accompanying loop.

Deletion of either H1 alone or H1 and the accompanying loop from *ipgC* in *S. flexneri* led to an almost complete loss of invasiveness. Furthermore, heterologously expressed *ipgC*$^{\Delta 21}$ is soluble but prone aggregation. This demonstrates that H1 plays a crucial role in the functional dimerization of IpgC. Moreover dimerization of IpgC is absolutely required for virulence.
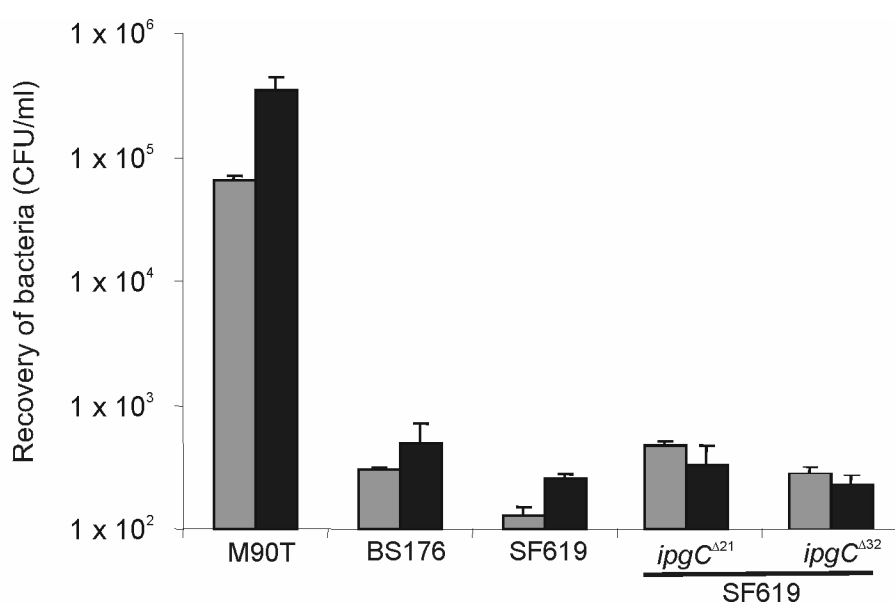

Fig. 3.15: Epithelial invasion assay. HeLa cell invasion was measured with the gentamicin protection assay, and recorded as number of colony-forming units per ml (CFU/ml). SF619, an *ipgC* mutant Shigella could not be complemented either by *ipgC*$^{\Delta 21}$ or *ipgC*$^{\Delta 32}$. Requirement of full length IpgC for restoring invasion is shown in Fig. 3.7. M90T, wild type *S. flexneri*; BS176, virulence plasmid cured *S. flexneri*; SF619, *ipgC* null mutant *S. flexneri*. Gray bars and black bars represent the time point at 1 h and 2 h respectively.

3.2 Part II: Characterization of the chaperone binding domain (CBD) of IpaC

3.2.1 IpaC is unstructured

IpaC though expressed in significant amounts (Fig. 5.1), remained mostly insoluble when standard expression conditions were employed [170]. To investigate the role of the chaperone IpgC in the expression and stability of IpaC, IpaC was coexpressed with IpgC. Better solubility of IpaC was achieved when coexpressed with IpgC indicating a role in maintaining stability and inhibiting unspecific aggregation. Hence, to further probe the role of IpgC in conferring conformational stability to IpaC as well as to map the IpgC binding region, limited proteolytic cleavage on the complex of IpaC with IpgC was performed. Treatment of IpaC-IpgC complex with either thermolysin (Fig. 4.1, A) or trypsin (Fig. 4.1, B) resulted in complete degradation of IpaC, suggestive of its unstructured nature.
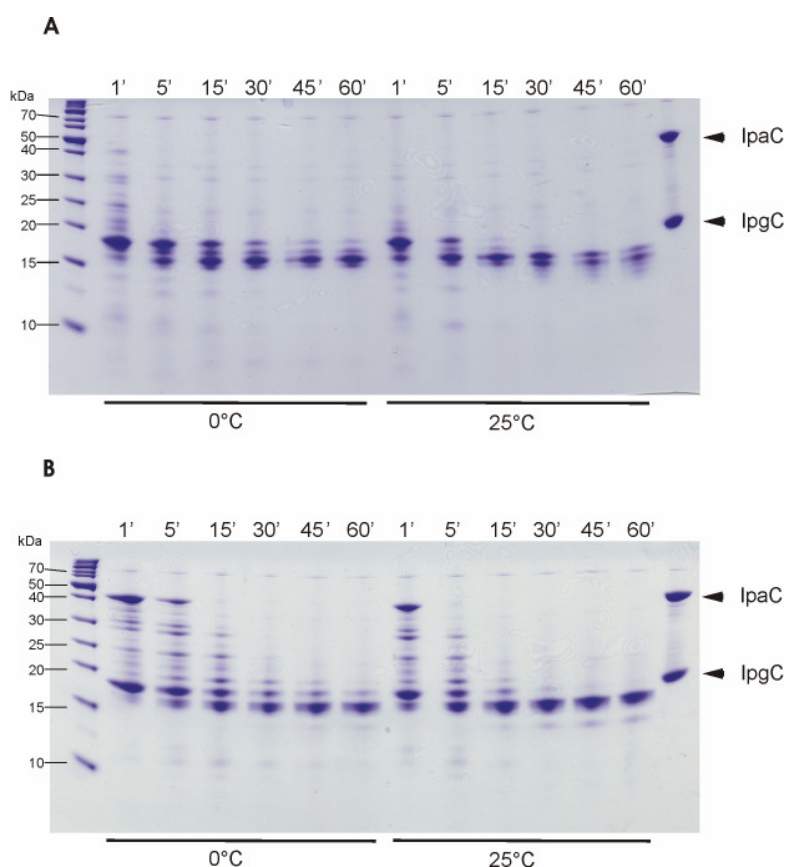

Fig. 4.1: Limited proteolysis assay of purified IpaC-IpgC complex with thermolysin (**A**) and with trypsin (**B**). Proteolysis was conducted on ice and at 25 °C. The E:S ratio was 1:100 (by weight) in each case. Aliquots were taken from the reaction mixture at time intervals indicated. The degradation was monitored by SDS-PAGE.

3.2.2 A novel approach to identify CBDs of IpaC

To circumvent the lack of any stable fragments of IpaC from proteolytic degradation and with an aim to characterize the potential chaperone binding region on IpaC, tryptic digestion of the IpaC-IpgC complex was carried out for 15 min on ice (Fig. 4.1, B), a time sufficient to obtain fragmented IpaC. Subsequently, the proteolysis was quenched by the addition of phenylmethylsulphonyl fluoride (PMSF) and immediately subjected to size exclusion chromatography to capture fragments of IpaC bound to IpgC and hence comigrate. The elution profile overlapped that of apo IpgC indicating degradation of full length IpaC and preservation of IpgC (Fig. 4.2). The eluate analyzed by MALDI-MS yielded fragments of IpaC which co-eluted with IpgC. Five fragments of IpaC were identified which were distributed throughout the length of the molecule i.e., $IpaC_{47-75}$, $IpaC_{127-148}$, $IpaC_{245-257}$, $IpaC_{299-316}$ and $IpaC_{339-350}$ (Fig. 4.2). However, there was not sufficient data to confirm the fragment $IpaC_{339-350}$ by MS/MS which was performed to confirm the data of fragments of IpaC obtained from MALDI-MS. The whole of IpgC starting from amino acid 2 and except for fragments 83-107 and 152-155 could be identified by mass spectrometry. Similar result was obtained upon repetition of the experiment with small variation in the fragment lengths of IpaC.
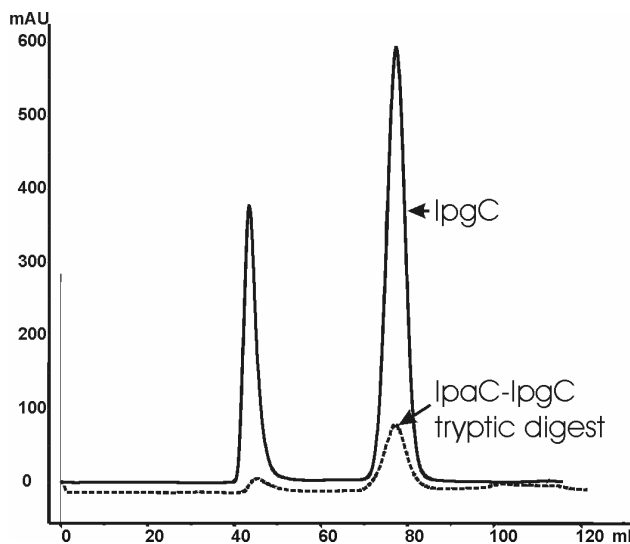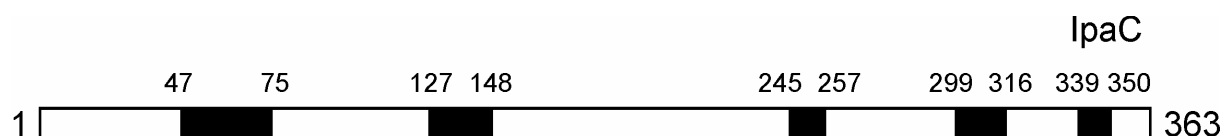

Fig. 4.2: Left panel, overlay of the purification profile of IpgC alone and IpaC-IpgC tryptic digest passed through Superdex 200 16/60. The peak around 40 ml is the void volume of the column which corresponds to >600 kDa. Below, schematic representation of the putative chaperone binding fragments of IpaC identified from mass spectroscopy.

3.2.3 IpgC binds to N-terminal moiety of IpaC

To probe and narrow down to the domain of IpaC that binds to IpgC, 7 constructs (see below, Fig. 4.3) addressing different putative chaperone binding regions individually or in combination were designed (Fig. 4.3).
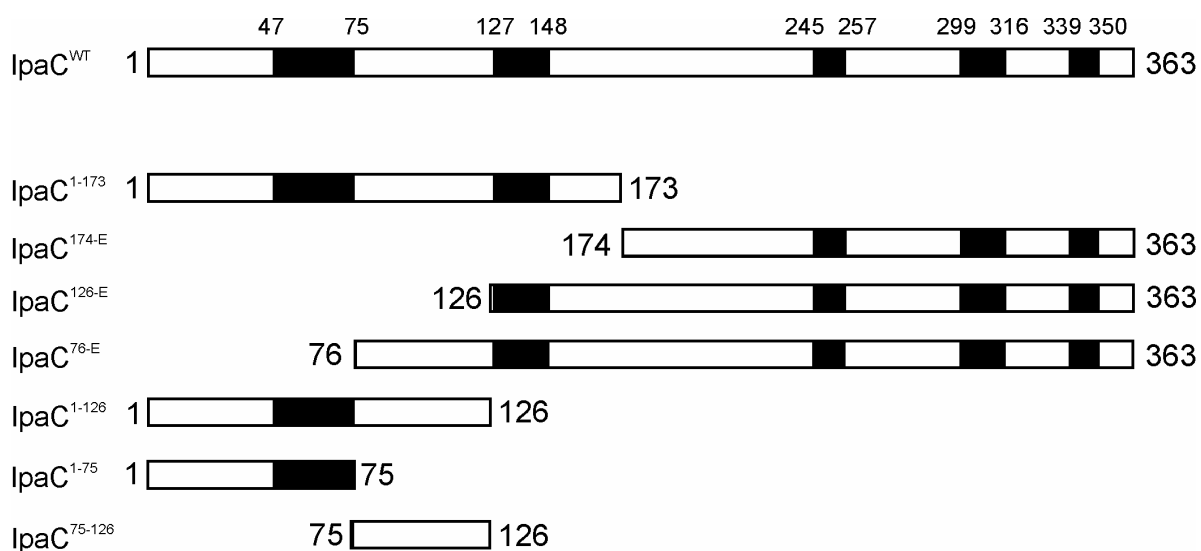

Fig. 4.3: Schematic representation of the different fragments of IpaC which were coexpressed and copurified with 6x IpgC.

As the putative IpgC binding domains on IpaC are distributed throughout the length of IpaC, the co-expression and co-purification of N-terminus and C-terminus moiety i.e., IpaC$^{1-173}$, IpaC$^{174-363}$ with IpgC was tested first. Both constructs were soluble, but only N-terminus moiety IpaC$^{1-173}$ could be co-purified with IpgC (Fig. 4.4 A). The binding of IpaC$^{1-173}$ was confirmed by immunoblot with IpaC antibody (Fig. 4.4 A1). Next, a C terminus construct, IpaC$^{126-363}$ containing an additional putative chaperone binding domain from amino acids 127 to 148 was tested (Fig. 4.4 B). As the construct IpaC$^{174-363}$ showed no binding; the binding, if any, of IpaC$^{126-363}$ would be solely contributed by the amino acids 127 to 148. This construct though soluble could not be copurified with IpgC. Thus, the fragment which included four of the five putative CBDs (Fig. 4.3), barring fragment 47-75, could not be verified by pull-down with IpgC. Interestingly, the construct IpaC$^{76-363}$ which included the region 73-122 identified in yeast-two hybrid screen [108] as the chaperone binding region could be co-purified with IpgC (Fig. 4.4 C).

Next, IpaC$^{1-126}$ consisting of the remaining putative CBD spanning amino acids 47 to 75 was generated. This construct exhibited the least solubility of all the constructs generated (Fig. 4.4 D) and hence no conclusion could be drawn. Therefore, two additional constructs addressing this region were generated i.e., IpaC$^{1-75}$ encompassing the putative CBD spanning amino acids 47 to 75 and IpaC$^{75-126}$. Both were soluble and could be copurified with IpgC (Fig. 4.4 E and 4.4 F). The binding of IpaC$^{1-75}$ was confirmed by immunoblot with IpaC antibody (Fig. 4.4 E1).

Table 4.1 summarizes the outcome of the pull-down of various constructs of IpaC with IpgC. These constructs of IpaC were generated based on the result from proteolytic degradation of IpaC-IpgC complex subjected to gel filtration and subsequently analyzed by mass spectrometry. Only the IpaC constructs containing the region comprising any combination of residues from 1-126 of IpaC showed binding to IpgC.

Sl. No.	Construct	Solubility	Binding
1	IpaC$^{1-173}$	++	+++
2	IpaC$^{174-363}$	++	-
3	IpaC$^{126-363}$	++	-
4	IpaC$^{76-363}$	++	++
5	IpaC$^{1-126}$	+	-
6	IpaC$^{1-75}$	++	+
7	IpaC$^{75-126}$	+++	++

Table 4.1: The solubility and binding of different IpaC constructs to IpgC. Different length fragments encompassing potential chaperone binding sequences were coexpressed and copurified with 6xHis IpgC. The degree of solubility and the ability of the fragments to bind to IpgC are indicated.
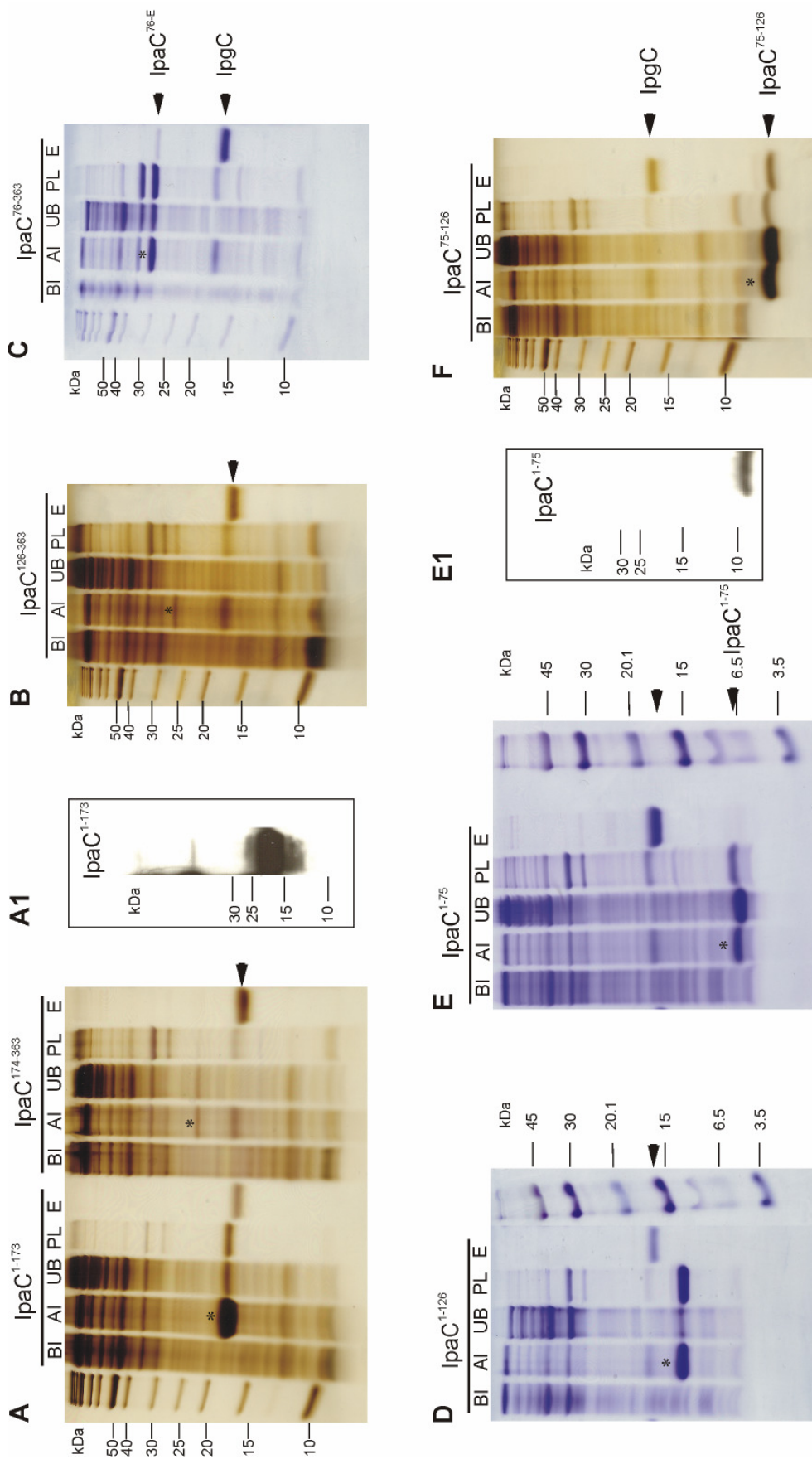
Fig. 4.4: Pull-down assays. The IpaC constructs were coexpressed with 6x His-IpgC. The expression (BI, before induction; AI, after induction) of the IpaC constructs is shown by asterisk (*). The cells were lysed and spun down. The insoluble fraction in the pellet is indicated as PL. The supernatant is subjected to Ni Sepharose column. The unbound (UB) and elute (E) are shown. The presence of IpaC$^{1-173}$ (A1) and IpaC$^{1-75}$ (E1) in elute is confirmed by immunoblot with IpaC Ab. IpgC in all the gels is indicated by an arrow. Molecular mass: IpaC$^{1-173}$, 19 kDa; IpaC$^{174-363}$, 21 kDa; IpaC$^{126-363}$, 26 kDa; IpaC$^{76-363}$, 31.6 kDa; IpaC$^{1-126}$, 14 kDa; IpaC$^{1-75}$, 8 kDa; IpaC$^{75-126}$, 5.5 kDa.

3.2.4 Defining the boundaries of the CBD in IpaC

Based on the aforementioned techniques, the region that binds to the chaperone IpgC in IpaC was narrowed down to the N terminus from amino acids 1 to 126. To further define the boundaries of the CBD in IpaC, a methodology developed by Schultz and co-workers [134, 171] was adopted. This involved introducing *p*-benzoyl-L-phenylalanine (*p*Bpa), a photoexcitable unnatural amino acid at specific positions in IpaC. The *in vivo* incorporation of the photocrosslinking amino acid was possible using a plasmid [134] that possess an orthogonal aminoacyl-tRNA synthetase/tRNA pair. In the presence of this plasmid, *p*Bpa incorporates with high translational efficiency into proteins in *E. coli* in response to the amber codon, TAG. The amber codon is specifically chosen as the frequency of usage per 100 codons is only 0.03% against 0.2% and 0.1% for TAA and TGA respectively [172]. The method has been described in detail elsewhere (section 2.2.3.3.).

Amber codon was introduced individually in IpaC for incorporation of *p*Bpa at amino acid positions- 29, 33, 36, 42, 50, 60, 64, 68, 73, 80, 84, 88, 99, 116, 120, 127, 146 and 154. These constructs were individually coexpressed with IpgC in the presence of *p*Bpa in the media. The protein complex was purified and covalent linkage between IpaC and IpgC upon UV excitation was observed. Crosslinked product was observed for IpaC having *p*Bpa at the positions- 36, 42, 50, 60, 64 and 68 (Fig. 4.5). A weak complex formation only on longer exposure to UV could be observed for IpaC having *p*Bpa at position 80. As a result the CBD in IpaC could be defined as a region from amino acids 36 and 68.
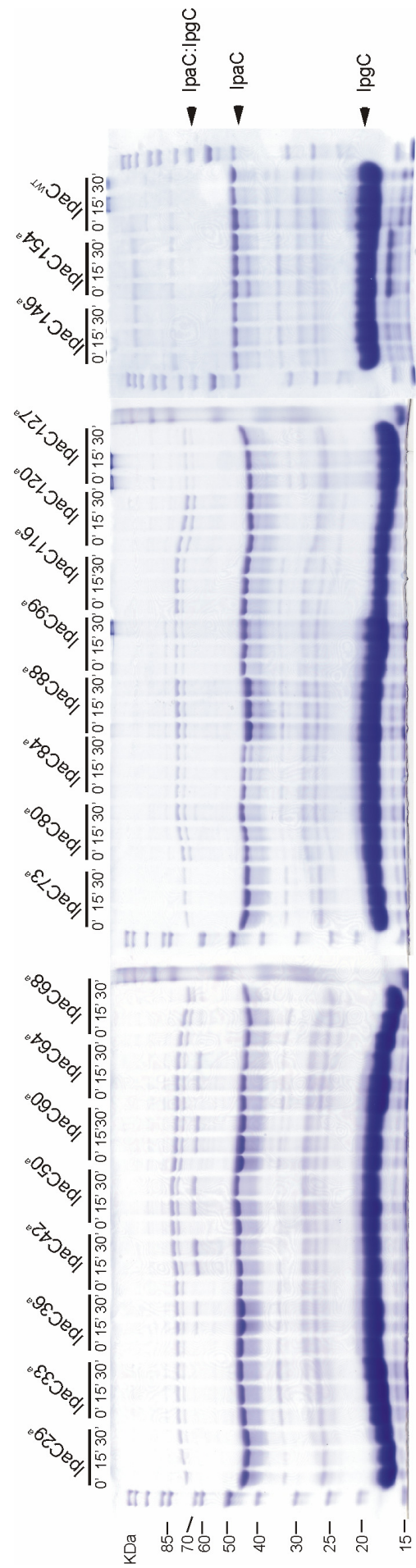
Fig. 4.5: Photocrosslinking of IpaC-IpgC complex. Photocrosslinkable unnatural amino acid, *p*Bpa is site-specifically incorporated in IpaC at positions indicated. Crosslinking at UV_{365} was performed on purified complex of IpaC-IpgC for the time indicated above the gel. Crosslinked product of IpaC-IpgC was observed for IpaC having *p*Bpa at positions 36, 42, 50, 60 and 68. A weak complex was observed for position 80 on long exposure.

3.2.5 Crystallization of the IpaC-IpgC complex

For crystallizing chaperone binding domain of IpaC together with IpgC, purified IpgC discussed in earlier section (section 3.1.1.4) was incubated with IpaC36-68 peptide encompassing the residues of CBD from amino acid 36-68. Crystallization was set up by mixing IpgC (12 mg/ml) and IpaC36-68 peptide at a molar ratio of 1:2. Crystals were obtained under a variety of conditions. Data were collected for three different crystals at 3.0 Å at BESSY (BL14.1), Berlin, at 3.4 Å at ESRF (BL23.1), Grenoble and at 3.85 Å at DESY (BW6), Hamburg. All the crystals belonged to space group $P3_221$ with two molecules in the asymmetric unit (a = 115 Å, b = 115 Å, c = 147 Å for the data collected at ESRF and DESY and a = 113 Å, b = 113 Å, c = 76 Å for the data collected at BESSY). All of them had high solvent content with 69% for the data collected at BESSY and 85% for data collected at DESY and ESRF. In none the electron density for the peptide could be traced.

As an alternate strategy, CBD of IpaC was fused to the C terminus of IpgC. A linker of 20 amino acids comprising amino acids 16-35 from IpaC separated IpgC and CBD of IpaC. The crystals were obtained in 0.1 M N-(2-Acetamido)iminodiacetic acid, pH 6.5, 1.0 M di-Ammonium hydrogen phosphate. Small crystals grew in three weeks time (Fig. 4.6). Currently, optimization to improve the size of the crystals is being pursued.
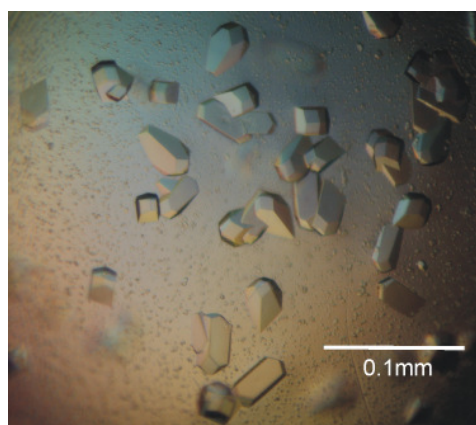

Fig. 4.6: Crystals of IpgC-IpaC CBD fusion protein.

3.3 Part III: Characterization of the CBD of IpaB

3.3.1 IpaB is toxic to *E. coli*

Individually, *ipaB* expression led to growth retardation in *E.coli* (Fig. 5.1) and was only detectable by immunoblot [170]. Enhanced and stable cytosolic yield of IpaB was achieved when coexpressed with the chaperone. Moreover, expression of *ipaB* was also achieved when coexpressed with *ipaC*, though the complex aggregated upon purification (data not shown). To further probe the role of IpgC in stable expression of IpaB and to map the regions that bind to IpgC in IpaB, limited proteolytic cleavage on the complex of IpaB-IpgC was performed.
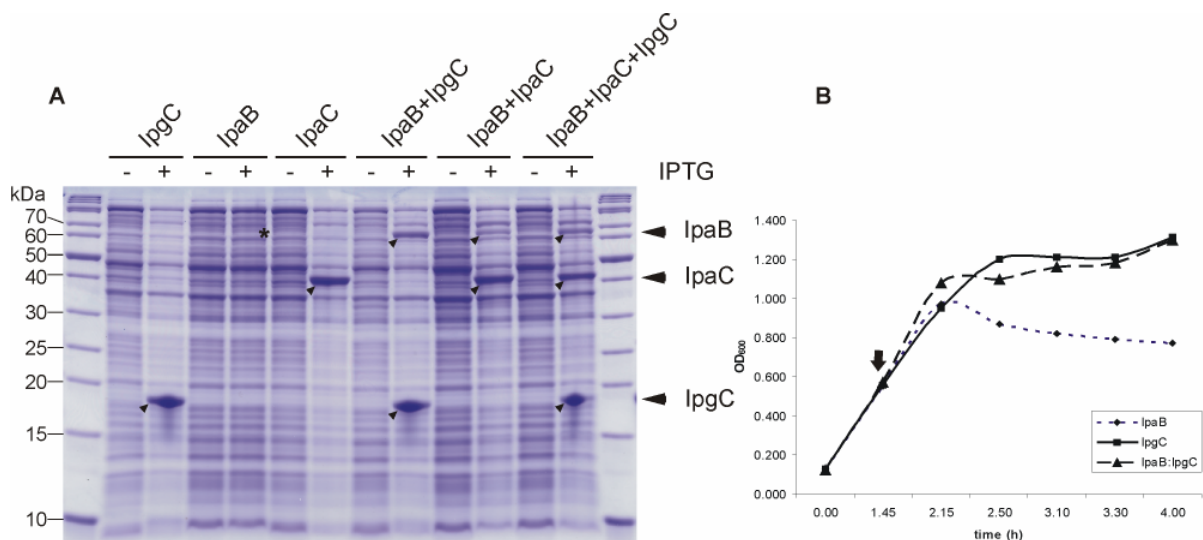

Fig. 5.1: Expression of different combinations of IpaB, IpaC and IpgC. Expression of IpaB alone (indicated by *, **A**) led to growth retardation in *E.coli* (dotted line, **B**). The constructs were expressed in *E.coli* and induced with 0.5 mM IPTG indicated by arrow in the right panel. The growth curve shows the expression of IpaB alone is toxic to the cells and can be rescued by coexpression with IpgC (dashed line, **B**).

3.3.2 IpaB yields a stable core upon proteolytic cleavage

To map the IpgC binding region limited proteolytic cleavage was performed on co-purified complex of IpaB-IpgC. Thermolysin treatment of purified IpaB-IpgC complex (Fig. 5.2) yielded a stable core of IpaB comprising residues 51-507 as detected by MALDI-MS and MS/MS. The amino terminus was additionally confirmed by Edman sequencing. In the presence of trypsin (Fig. 5.2), IpaB-IpgC complex showed reduced proteolysis with N-terminus of IpaB starting at amino acid 37. This indicated

that the chaperone bound IpaB restricts access of trypsin beyond amino acid 37 as the next cleavage site occurs at amino acid position 68 on IpaB. The stable core of IpgC started from amino acid 2 and except for the 6x His-tag and fragments 61-83 and 117-140, the whole of IpgC could be identified by MALDI-MS and MS/MS. The amino terminus was verified by Edman sequencing. Combining the results of thermolysin and trypsin digestion, it can be deemed that IpaB is shielded by bound chaperone and the IpgC binding site on IpaB starts from amino acid 51. This result is consistent with the finding from yeast two-hybrid system [108] that reports the chaperone binding site to be from amino acids 58 to 72 on IpaB.
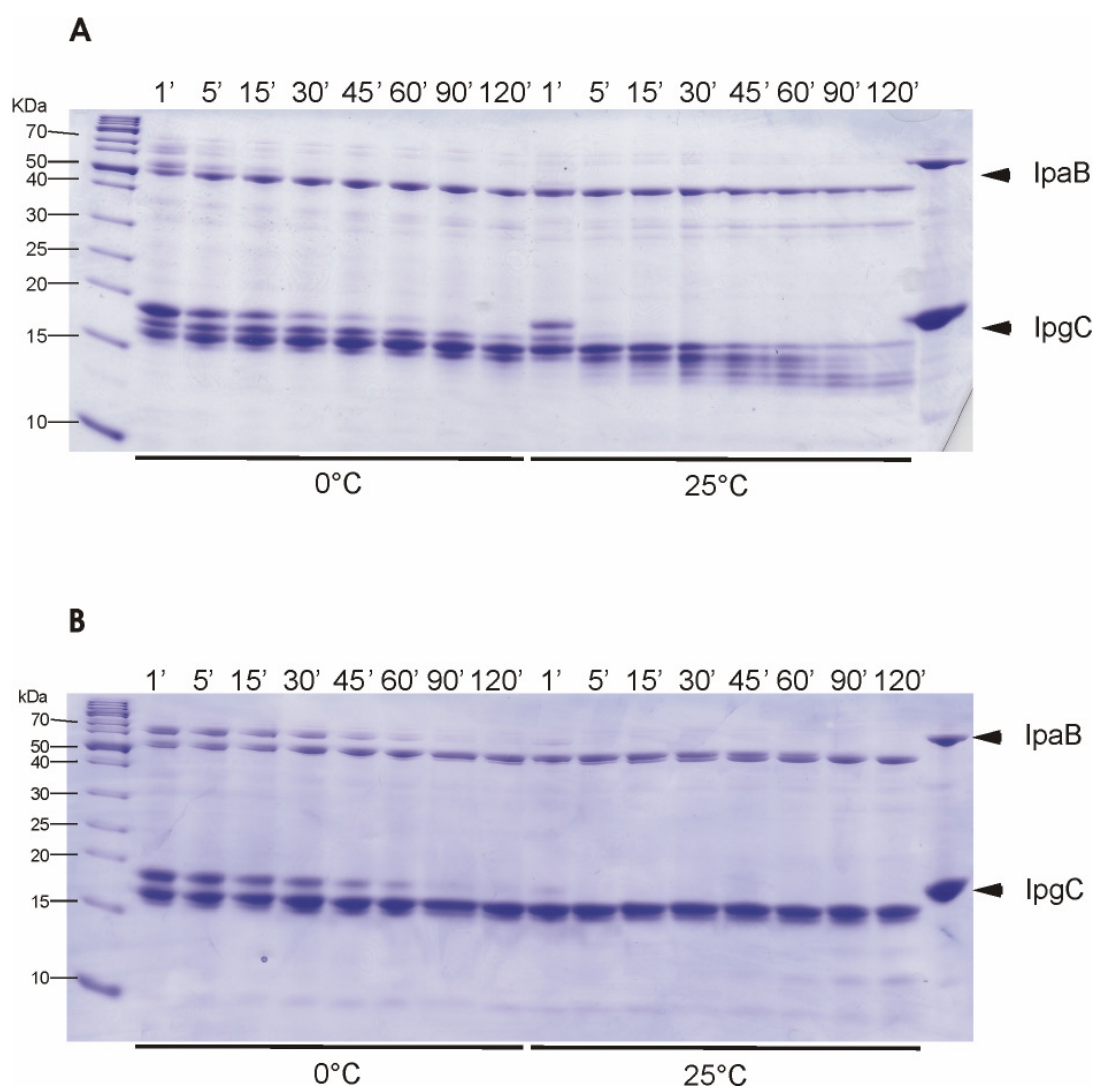

Fig. 5.2: Limited proteolysis assay of purified IpaB-IpgC complex with thermolysin (**A**) and with trypsin (**B**). Proteolysis was conducted on ice and at 25 °C. The E:S ratio was 1:100 (by weight) in each case. Aliquots were taken from the reaction mixture at time intervals indicated. The degradation was monitored by SDS-PAGE.

3.3.3 CBD on IpaB spans from amino acids 51 to 72

Next, the technique described in previous section (2.2.3.3) to site-specifically introduce artificial photoexcitable amino acid i.e., p-benzoyl phenylalanine (pBpa) to map the CBD of IpaB was employed. Combining the data from limited proteolysis assay which indicated the N-terminus of the stable core of IpaB to be from 51 and the data available from yeast-two hybrid system [108] several constructs of IpaB having pBpa at positions- 45, 48, 51, 55, 62, 65, 68, 74 and 79 were generated. IpaB having pBpa at positions 51, 55, 62, 65, 68 formed complex with IpgC implying the precise binding region of IpgC in IpaB to be from amino acids 51 to 72 (Fig. 5.3).
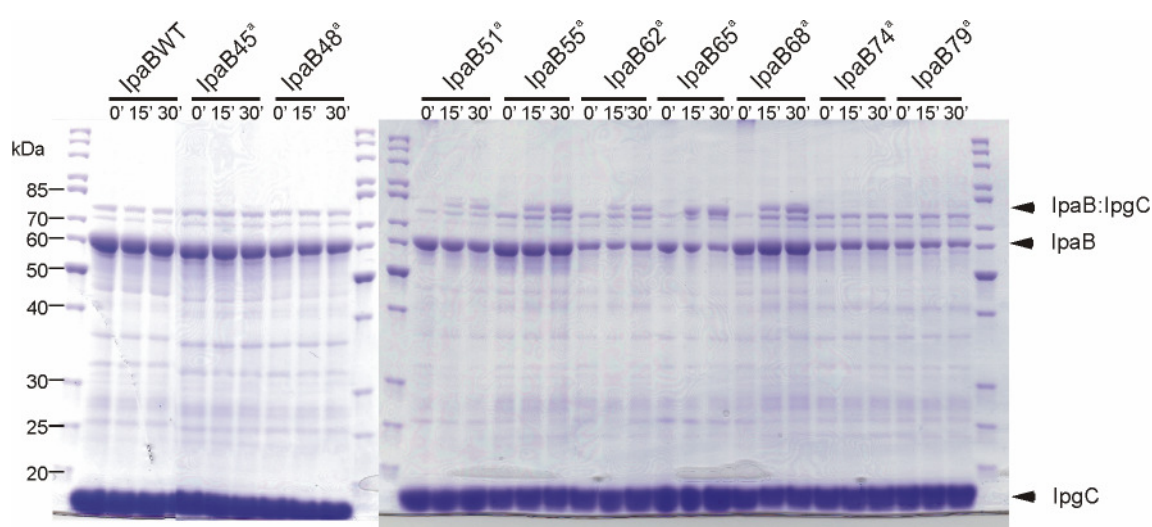

Fig. 5.3: Photocrosslinking of IpaB-IpgC complex. Photocrosslinkable unnatural amino acid pBpa is site-specifically introduced at the positions mentioned. Crosslinked product of IpaB-IpgC was observed for IpaB having pBpa at positions 51, 55, 62, 65 and 68.

3.3.4 CBD mutant IpaB$^{\Delta 51-72}$ shows similar activity as full-length IpaB

3.3.4.1 Co-purification of different mutants with IpgC

To further decipher the effect of the CBD on IpaB, three truncation mutants of IpaB i.e., IpaB$^{\Delta 1-50}$ lacking the N-terminus 50 amino acids, IpaB$^{\Delta 1-72}$ lacking the N-terminus 72 amino acids and IpaB$^{\Delta 51-72}$ lacking 51 to 72 amino acids were generated to be co-expressed with 6x His-IpgC. The latter two constructs lack the CBD on IpaB. *E. coli* coexpressing IpaB$^{\Delta 1-72}$ and IpgC exhibited growth retardation upon induction with IPTG and IpaB$^{\Delta 1-72}$ could not be visualized on copurification with IpgC. Interestingly, both IpaB$^{\Delta 1-50}$ and IpaB$^{\Delta 51-72}$ could be copurified with IpgC.
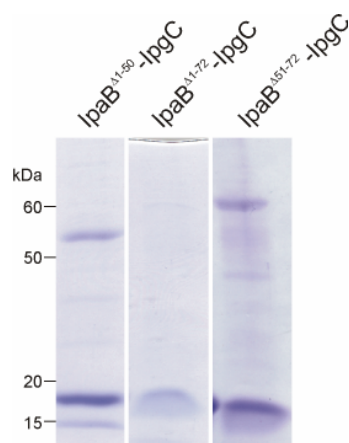
Fig. 5.4: Pull-down assay. IpaB$^{\Delta51-72}$ lacking CBD shows binding to IpgC. Cells expressing IpaB$^{\Delta1-72}$ lacking the CBD show growth retardation while IpaB$^{\Delta1-50}$ comprising CBD shows binding to the IpgC.

Stable coexpression and copurification of IpaB$^{\Delta51-72}$ but not of IpaB$^{\Delta1-72}$ with IpgC (Fig. 5.4) is indicative of a possible additional binding site N-terminus to amino acid 51 on IpaB and also putative role of this region in imparting stability to IpaB.

3.3.4.2 *In vivo* functionality of IpaB$^{\Delta1-50}$, IpaB$^{\Delta1-72}$ and IpaB$^{\Delta51-72}$

To determine the effect of chaperone binding domain on the secretion of IpaB, truncations bearing N-terminal (IpaB$^{\Delta1-50}$, IpaB$^{\Delta1-72}$) or internal (IpaB$^{\Delta51-72}$) mutations were cloned in a constitutively expressing strain, pUC19. The ability of these mutants to complement *S. flexneri* non-polar *ipaB* deletion mutant strain SF620 was tested. Additionally, the ability of these mutants to complement a non-polar *ipaB/ipgC* double mutant strain BJ007 was checked. Using gentamicin protection assay the capacity of the *ipaB* mutants to complement SF620 and BJ007 for invasion of HeLa cells was inspected (Fig. 5.5 A and B). N-terminal truncations, IpaB$^{\Delta1-50}$ and IpaB$^{\Delta1-72}$ were unable to restore invasion, while IpaB$^{\Delta51-72}$ showed invasiveness similar to wild type *Shigella*. In BJ007 none of the mutants including wild-type *ipaB* were invasive. However, invasion could be restored by complementing BJ007 with wild-type *ipaB* and *ipgC*.

Next, the bacterial extracts and culture supernatants were analyzed by immunoblotting for IpaB and IpaC to examine if these IpaB truncation mutants were stably expressed and secreted (Fig. 5.5 C). Immunoblotting for DnaK, an abundant cytosolic protein was performed as a control for bacterial membrane integrity. IpaB^{WT} was expressed and secreted in both SF620 and BJ007. Though in BJ007 IpaB^{WT} is not stably maintained in the cytosol. Similar observation was made for the IpaB$^{\Delta51-72}$. IpaB$^{\Delta1-50}$ though detected in SF620 extract was not secreted, while in BJ007 extract low amounts were detected. Expression of *ipaB$^{\Delta1-72}$* in both SF620 and BJ007 was severely abrogated.
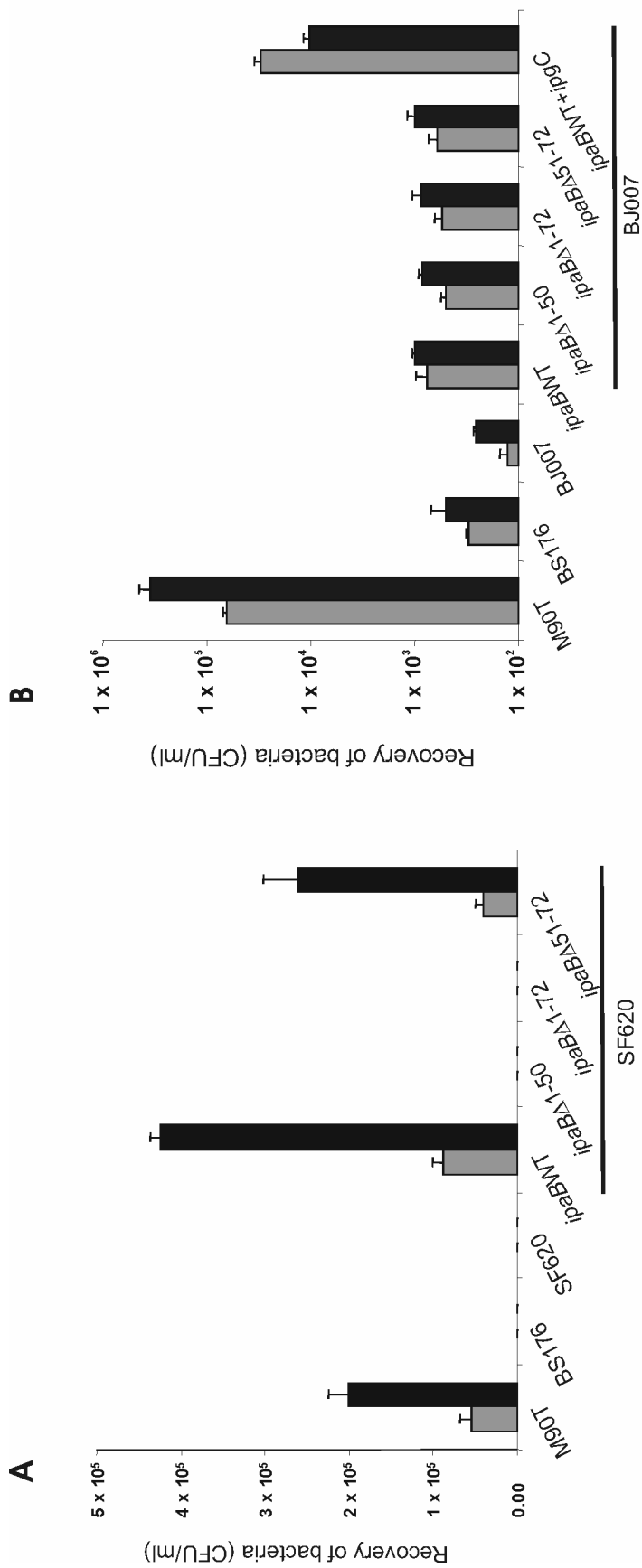
Fig. 5.5.Invasion and secretion assays. A, Epithelial cell invasion of SF620 and **B**, BJ007 with different truncation of IpaB. Gray bars and black bars represent the time point at 1 h and 2 h respectively. In both panels, error bars represent the means ± s.e.m. of triplicate experiments. Wild type S. *flexneri* (M90T) and a virulence plasmid cured strain, BS176 are also included. **C**, Expression and secretion of IpaB mutants was assayed in both SF620 and BJ007. Protein samples from the indicated strain were analyzed by immunoblot with mAb anti-IpaB. Experiments were done in triplicate.

3.3.5 An additional CBD in IpaB

3.3.5.1 IpaB chimera

To rule out the possibility of local structural rearrangement upon deletion of CBD contributing to the association of the CBD deletion mutant $IpaB^{\Delta 51-72}$ with IpgC, a chimeric IpaB, IpaB-PS was constructed. This chimera has PreScission cleavage site [173] comprising 8 amino acids with additional 4 amino acid linker on either side of the cleavage site in place of amino acids 51-72 in IpaB. The aim of generating this construct was to disrupt any local secondary structure contributed by the amino acids in the vicinity of CBD that could still mimic CBD upon its removal. This construct could answer i.) in case of stable co-expression and co-purification of IpaB-PS and IpgC, there could be an additional CBD within N-terminal 50 amino acids on IpaB, ii.) if or not IpgC associates with the N-terminus 1-50 fragment of IpaB upon subjection to PreScission protease and, iii.) finally, if protease treatment leads to destabilization of $IpaB^{72-580}$ C-terminus moiety.

IpaB-PS could be co-expressed and co-purified with IpgC (Fig. 5.6) which suggests the presence of an additional binding site N-terminus to amino acid 50 in IpaB. As the construct $IpaB^{\Delta 1-72}$ does not express stably even in the presence of IpgC any additional IpgC binding site downstream of amino acid 72 in IpaB is ruled out.
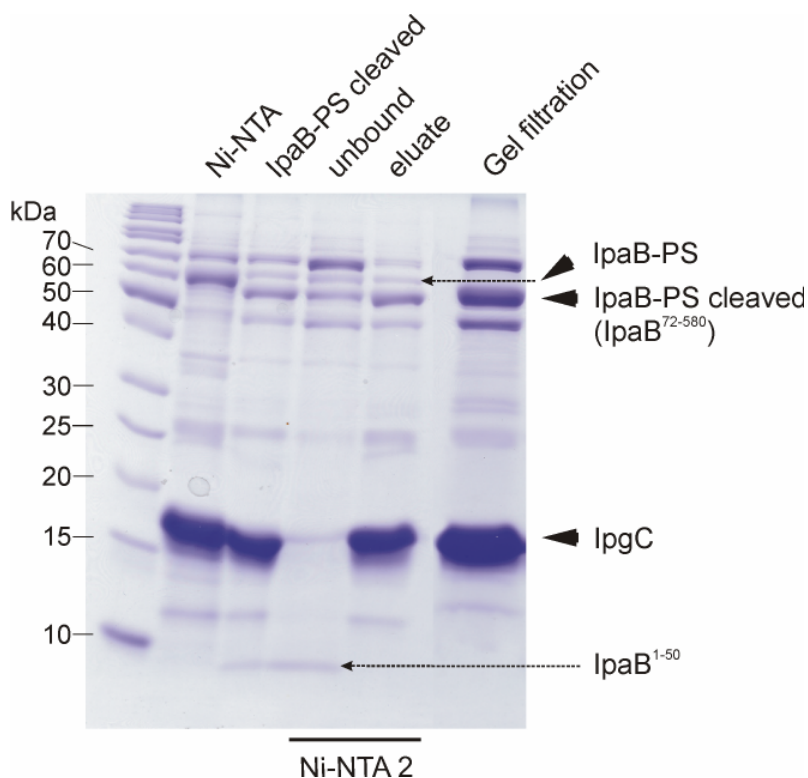

Fig. 5.6: Expression and purification of chimeric IpaB-PS with IpgC. The chimera lacking the CBD comigrated with IpgC (lane, Ni-NTA). Efficient cleavage of the proteolytic site was not observed plausibly due to masking of this site by bound chaperone (lane, IpaB-PS cleaved). As a consequence of proteolysis, the whole complex precipitated (lane, gel filtration).

Even after prolonged subjection of two days to PreScission protease the chimeric IpaB-PS was not processed completely plausibly due to masking of this site by bound chaperone (Fig. 5.6, lane, IpaB-PS cleaved). This protease subjected complex was passed through Ni-NTA column (Fig. 5.6, lanes, Ni-NTA2). IpaB-PS either processed or unprocessed along with the N-terminus fragment, IpaB$^{1-50}$ was observed in the unbound fraction. The identity of IpaB$^{1-50}$ was confirmed by MALDI analysis. It was surprising to observe that the eluted fraction also contained both processed and unprocessed IpaB-PS (lane, eluate). It became apparent on subsequent separation by the Superdex 200 that the whole complex precipitated as it eluted with in the void volume (lane, gel filtration). This indicates the destabilizing effect of cleavage of the N-terminal 50 amino acids from IpaB. Therefore, the N-terminal 50 amino acids not only impart stability to IpaB but also possess an additional IpgC binding domain.

3.3.5.2 Additional CBD (CBD1) spans amino acids 15 and 45

To precisely screen the additional chaperone binding site N-terminus to amino acid 50 in IpaB, constructs having amber mutations for incorporation of photo-excitable artificial amino acid pBpa at positions 10, 15, 18, 22 and 35 were generated. UV crosslinking was performed on the purified complex of corresponding IpaB mutant having pBpa at specific position and IpgC. IpaB having pBpa at positions 18, 22 and 35 formed complex with IpgC (Fig. 5.7). Thus, the presence of an additional binding site within amino acids 15 to 45 on IpaB is convincingly shown. Henceforth, this CBD will be referred to as CBD1.
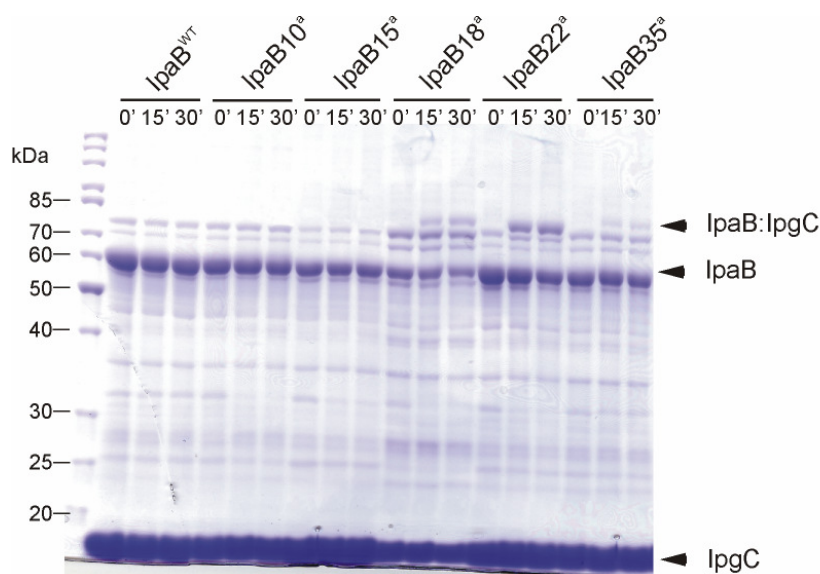

Fig. 5.7: Photo-crosslinking of IpaB-IpgC complex. Crosslinked product of IpaB-IpgC was observed for IpaB having pBpa at positions 18, 22 and 35.

3.3.6 Crystallization of IpaB-IpgC complex

For crystallizing chaperone binding domains of IpaB together with IpgC, purified IpgC discussed in earlier section (3.1.1.5) was incubated with peptides IpaB18-35 and IpaB51-72 encompassing the residues of both chaperone binding domains i.e., from amino acid 18-35 and 51-72 respectively. Crystallization was set up by mixing IpgC (12 mg/ml) and peptides at a molar ratio of 1:2. The best diffracting crystals were obtained in condition containing 0.1 M HEPES pH 7.4, 20% w/v PEG 8000 and 8% ethylene glycol. The crystal belong to space group $C222_1$ with two molecules in the asymmetric unit (a = 73.31 Å, b = 97.09 Å, c = 106.53 Å). Data were collected at 2.15 Å at BESSY (BL14.1 and BL14.2), Berlin and at 2.65 Å at ESRF (BL23.1), Grenoble, and merged to obtain good data quality and completeness (Table 5.1). Indexing, integration, merging and scaling were done using the program XDS [136].

Data collection and processing		
		IpgC+IpaB18-35+IpaB51-72
wavelength (Å)		0.9184
resolution (Å)		39.4 - 2.15
resolution (Å)	(last shell)	2.21 – 2.15
space group		$C222_1$
temperature (K)		100
X-ray source		BESSY BL 14.1, 14.2
detector		MARMOSAIC 225mm, MARCCD 165mm
unit cell parameters	a (Å), α (°)	73.3, 90
	b (Å), β (°)	97.1, 90
	c (Å), γ (°)	106.5, 90
total reflections	(last shell)	171660 (6125)
mosaicity (°)		0.23
unique reflections	(last shell)	19677 (1316)
$I/\sigma(I)$	(last shell)	13.17 (1.98)
% data completeness	(last shell)	93.4 (86.5)
R_{meas}[a] (%)	(last shell)	16.2 (75.3)
R_{sym}		15.4 (68.0)
r.m.s deviations		
	bond length (Å)	0.012
	bond angle (°)	2.0

Table 5.1: X-ray-diffraction-data statistics of the IpaB-IpgC crystal.

3.3.7 Structure of the IpaB-IpgC complex

The crystallographic asymmetric unit contains a dimer similar to the apo structure with one peptide, IpaB51-72 bound to each SU. 10 residues covering amino acids from Leu63 to Ser72 and 13 residues, covering amino acids Ser60 to Ser72, of the IpaB peptide are defined by the electron density in the cleft of IpgC. Both peptides have an antiparallel orientation to H2 with an overall contact surface area of 414 $Å^2$ (between peptide and SU B).
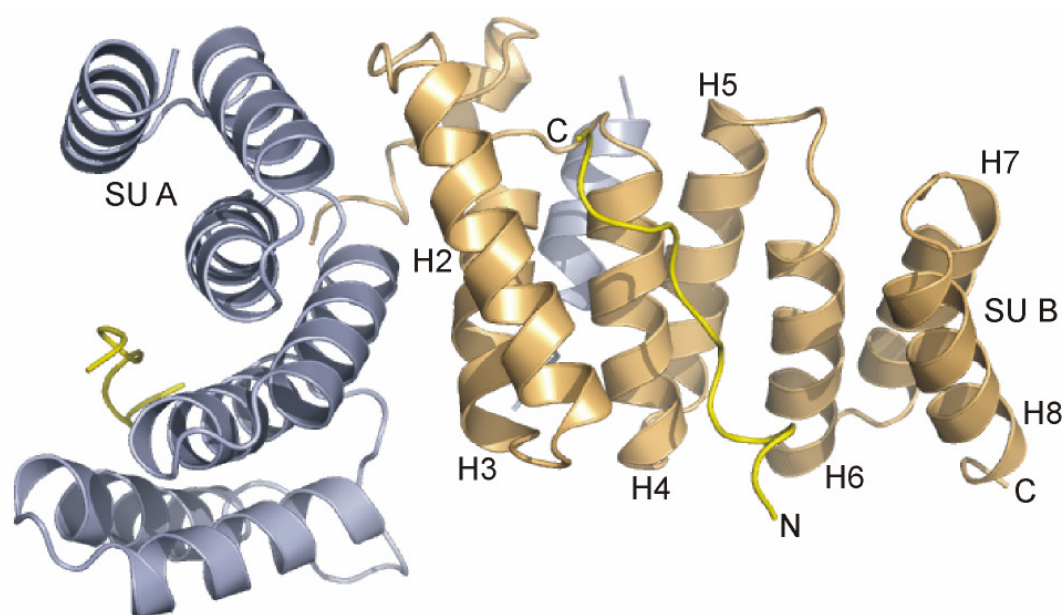

Fig. 5.8: IpaB-IpgC complex. Cartoon representation of the IpgC dimer found in the asymmetric unit with SU A colored blue and SU B colored in brown. Helices of SU B are labeled from H1 to H8. IpaB peptide is shown in yellow. The termini are indicated as N and C. The peptide is oriented antiparallel to the H2 of IpgC.

Though the corresponding SU of apo and IpaB complexed IpgC are superimposable in the TPR region with a RMSD of 0.5 Å, the SUs in the complex are tilted by about 15° and shifted by 5 Å relative to each other (Fig. 6.7). It is unclear whether this difference is the result exclusively of substrate binding or is influenced by crystal packing. The peptide is bound in the cleft surface which supports an extended conformation of the interacting peptide (Fig. 5.8). The surface of the cleft provides an amphiphilic surface, hence can support different mode of interactions with the binding ligand. The only structures of TPR domains in complex with their peptide ligand [174] present a positively charged surface in the groove limiting the kinds of interactions. On the other hand extensive interactions are made possible by the amphiphilic nature of the groove in IpgC.

The co-crystal structure analysis shows that there are three types of interactions that allow the binding between the IpaB peptide and IpgC. First, there is a salt bridge between side chain of amino acids Lys68 and Asp71 of $IpaB_{51\text{-}72}$ and IpgC. Additionally, Glu66 and Lys71 from IpaB51-72 forms salt bridge with Lys142 (SU A) and Asp37 (SU B) respectively (Fig. 5.11).
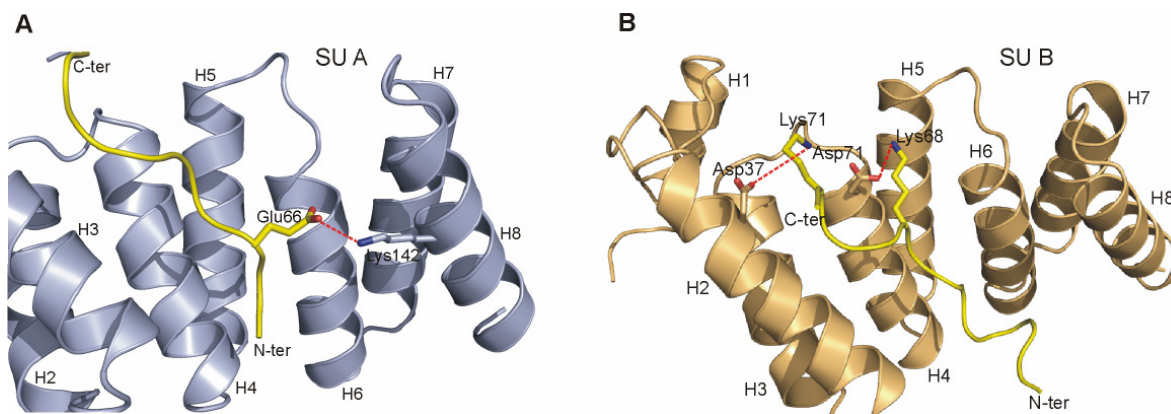

Fig. 5.11: Salt bridge interactions in IpaB-IpgC complex. **A**, shows interaction (red dotted line) of Glu66 from IpaB peptide with Lys142 of SU A of IpgC. **B**, shows Lys68 and 71 from IpaB peptide from IpaB peptide forming salt bridge with Asp71 and Asp37 respectively from IpgC. The termini of the peptide are indicated as N-ter and C-ter.

Second, there are strong, but not necessarily sequence specific, hydrogen bonds between the carbonyl of Ile62, Pro65, Lys68 in the IpaB peptide and the amide of Gln112, hydroxyl of Tyr47, Tyr40 in IpgC. These hydrogen bonds from IpgC to the IpaB peptide target the peptide backbone and not the side chains. Thus though the interactions are from the specific amino acids from IpgC, they do not exploit the sequence-specific features in the peptide.
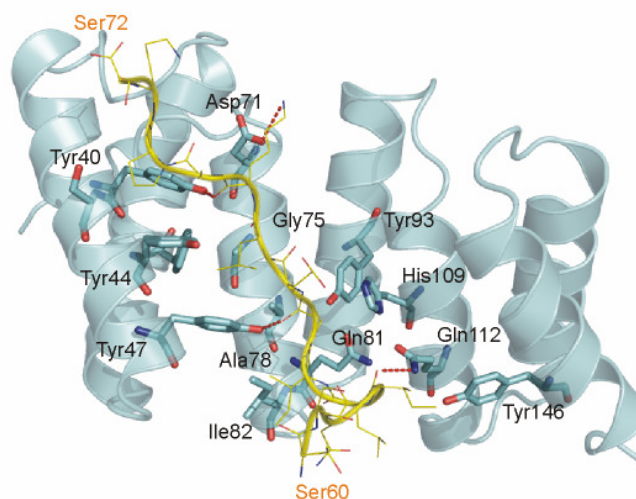

Fig. 5.12: Hydrogen bond interactions between IpaB-IpgC. Ile62-Gln112, Pro65-Tyr47, Lys68-Tyr40 from IpaB and IpgC respectively are involved in hydrogen bond interaction.

Third, there are hydrophobic and van der Waals interactions between the $IpaB_{51-72}$ and IpgC. These interactions are localized to three discrete pockets on the surface of IpgC (Fig. 5.13). The pockets, designated P1, P2 and P3, interact with side chains of conserved residues Pro65, Leu67 and Pro70 in the IpaB peptide. P1 is formed by conserved residues Tyr47, Ala78, Gln81, Tyr93, His109 and partially by Ile82. The predominantly hydrophobic central pocket, P2, is formed exclusively by conserved residues Tyr44, Tyr47, Phe59, Met74, Gly75 and Ala78 from H2 and H4. Interestingly, protein binding in the eukaryotic adaptor protein Hop [174] is also mediated through pockets similar to P1 and P2 to the molecular chaperones Hsp70 and Hsp90. The P3 pocket is formed by conserved residues Tyr40 and Tyr44 of the tyrosine ladder supported by Ser41 and partially by Asp45 from H2.
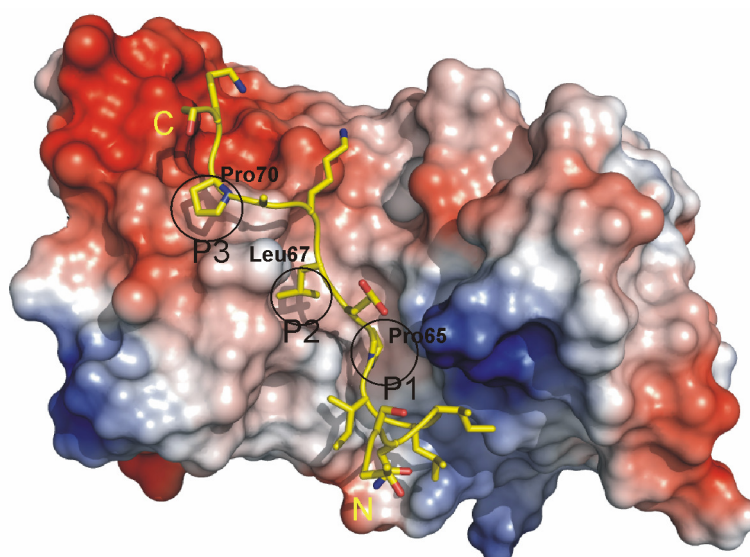

Fig. 5.13: Interaction of IpaB with pockets on IpgC. Pro65, Leu67 and Pro70 from IpaB interact with pocket P1, P2 and P3 on IpgC respectively.

3.3.8 Specificity of pocket binding residues in IpaB

The assembly of three pockets in the peptide binding groove of IpgC formed exclusively by conserved residues should in turn impose strict selection criteria on amino acid residues that interact with them. To demonstrate that it indeed is the case, residues in IpaB interacting with the pockets were selectively mutated and tested for binding to IpgC by isothermal titration calorimetry (ITC) (Fig. 5.14a). Short hexameric peptides covering the region interacting with the pockets in IpgC were generated. The peptide PELKAP representing the original sequence bound to IpgC with an affinity of 625 µM. Next, the residue Pro65 of IpaB interacting with pocket P1 of IpgC was substituted with arginine, a bulky and polar amino acid. Then the peptide

PENKAP having a polar residue aparagine substitution, to disrupt the hydrophobic interaction of Leu67 in IpaB with pocket P2 of IpgC was tested. Subsequently, a peptide PELKAD having a Pro70 to an acidic amino acid Asp substitution to disrupt the interaction with pocket P3 close to the acidic patch was checked for binding. All these peptides having unfavorable amino acid in the pocket-interacting positions showed markedly low to no affinity for IpgC binding (Fig. 5.14a).
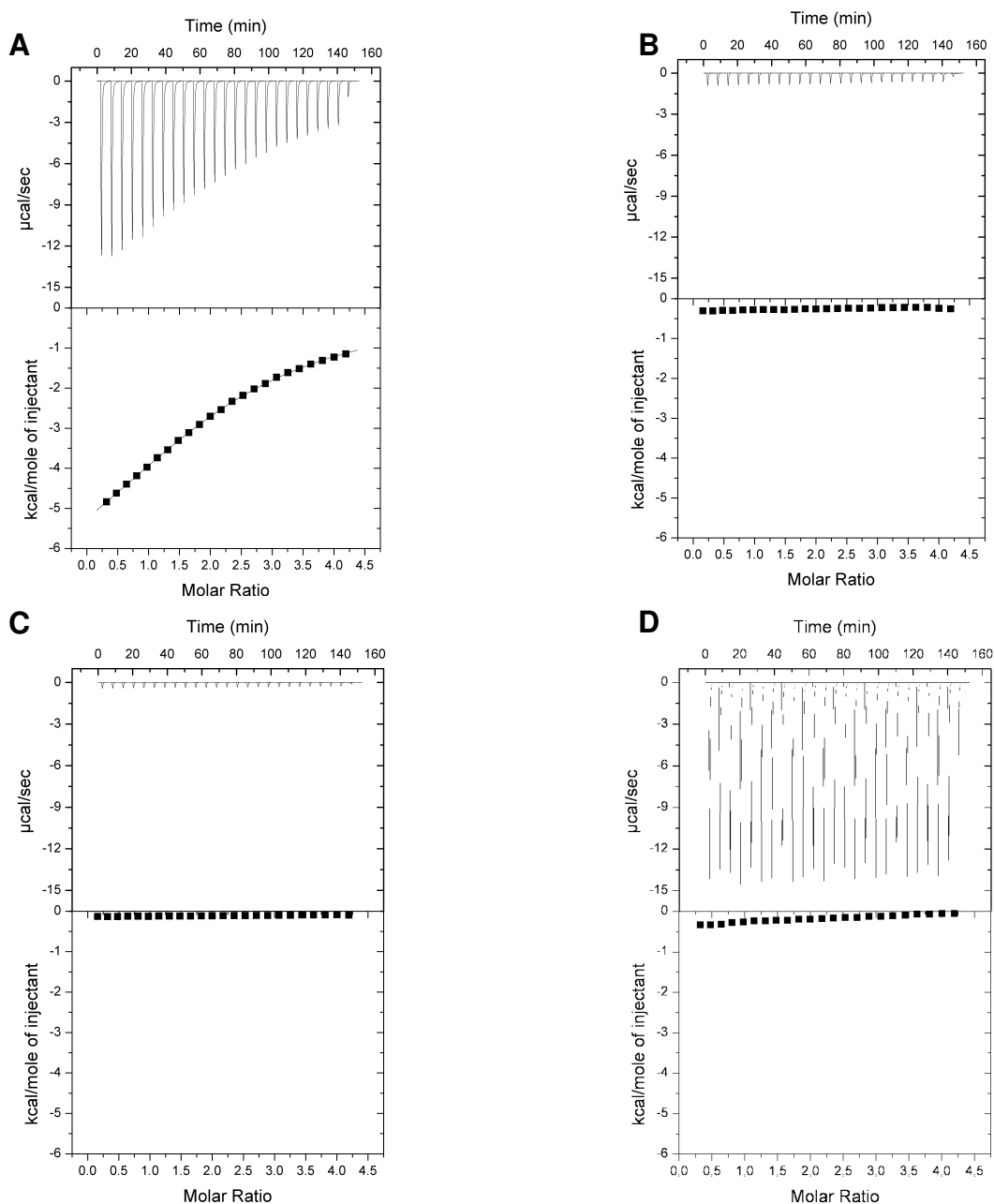

Fig. 5.14a: **A**, Binding of PELKAP, the original sequence from IpaB shows an affinity of 625 ± µM to IpgC. While, any peptide having unfavorable amino acid in the pocket-interacting positions, i.e., RELKAP, PENKAP or PELKAD (**B**, **C** or **D**) showed no binding to IpgC.

Furthermore, Lys68 in the hexameric IpaB peptide forms a salt bridge with Asp71 in IpgC. A mutation in the peptide K68A disrupting the salt bridge was checked for its affinity to IpgC. This peptide, PELAAP bound to IpgC with a reduced affinity of 935 µM (Fig. 5.14b). This is indicative of specificity imparted by the pocket on the sequence that binds to the groove.

Fig. 5.14b: Binding experiment of salt bridge disrupting peptide, PELAAP from IpaB with IpgC showed reduced affinity to IpgC.
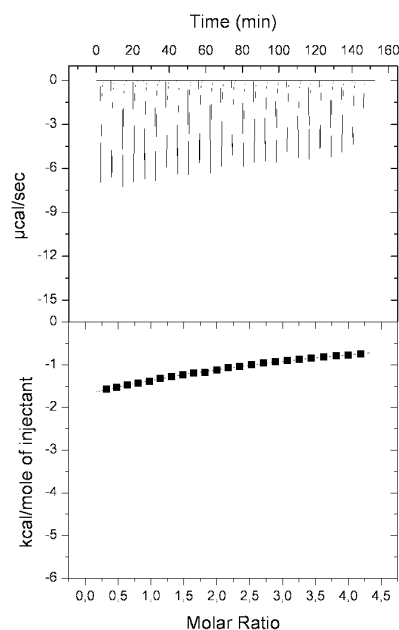
3.3.9 IpgC-IpaB forms a ternary complex

The multi-angle light scattering (MALS) profile for IpaB-IpgC is shown in the Fig. 5.15. The IpaB-IpgC complex peak is shown. The first peak which is not shown in the figure is protein aggregates eluting in the void volume of the column. By MALS the quaternary structure of the IpaB-IpgC was established to be a heterotrimer comprising two molecules of IpgC in complex with an IpaB molecule.
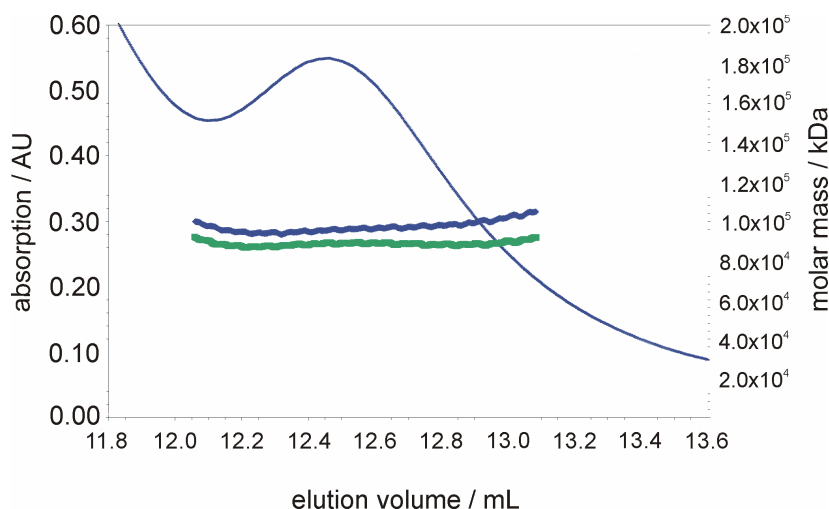
Fig. 5.15: On-line static laser light scattering reveals IpgC to be dimer. The blue curve shows the protein absorption at 280 nm (left axis) versus volume on x axis. The two lines refer to the measured molecular mass (right axis) for two independent experiments.
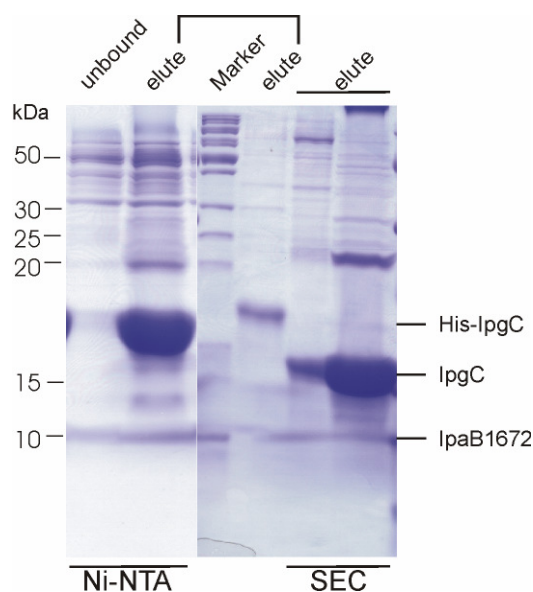
The heterotetramer arrangement in crystal is not in agreement with the ternary complex found in solution. The simultaneous binding of two CBDs might offer a clue to the correct quaternary structure. The strategy to further investigate the role of CBD1 (IpaB18-35) is explained in the next section.

3.3.10 Crystallization of IpaB encompassing both CBDs

The lack of density for IpaB1835 in the previously described cocrystal structure of IpaB-IpgC prompted an approach where the region in IpaB containing both CBDs i.e., amino acids 16-72 (IpaB1672) was expressed in different combinations described below.

3.3.10.1 Expression and purification

In order to achieve synchronous expression, *ipaB*1672 and *ipgC* were cloned in a bicistronic vector, pCDFDuet-1 (Novagen). *ipgC* was cloned in the multiple cloning site I while *ipaB1672* was cloned in multiple cloning site II. Additionally thrombin cleavage site was introduced between the N-terminal His-tag and IpgC. In the expression and subsequent pull down IpaB1672 could not be visualized. A common problem of achieving insufficient yield of small peptide or degradation by endogenous protease could be attributed to lack of IpaB1672 in the pull down with His-IpgC. Hence, *ipaB*1672 was cloned in pET21a and coexpressed with *ipgC* having N-terminal cleavable His tag cloned in pET28a. These plasmids having different antibiotics resistance i.e., ampicillin and kanamycin respectively were cotransformed in *E. coli*. Albeit stable expression and pull down of IpaB1672 was observed, the low level of purity (elute (SEC), Fig. 5.16) after removal of the His tag and gel filtration rendered it unsuitable for crystallization.

Fig. 5.16: Coexpression and purification of IpaB1672-IpgC complex IpgC.

Subsequently, IpaB1672 was cloned in pET28a having a thrombin cleavable N-terminal His tag and expressed and purified separately (Fig. 5.17c). As the IpaB1672 peptide does not contain tryptophan, tyrosine or cysteine it showed no absorbance at UV A_{280} (Fig. 5.17, dashed line). Therefore, protein purification was monitored at UV A_{230} (Fig. 5.17, solid line). Absorbance at UV A_{280} was also monitored to detect the presence of contaminating proteins. The peak of IB1672 (A_{230}) in the gel filtration was

devoid of any contaminating proteins as monitored by no absorbance at UV A_{280} (Fig. 5.17, B) and from SDS-PAGE (Fig. 5.17, C). Attempts to cleave the tag led to precipitation of the peptide.
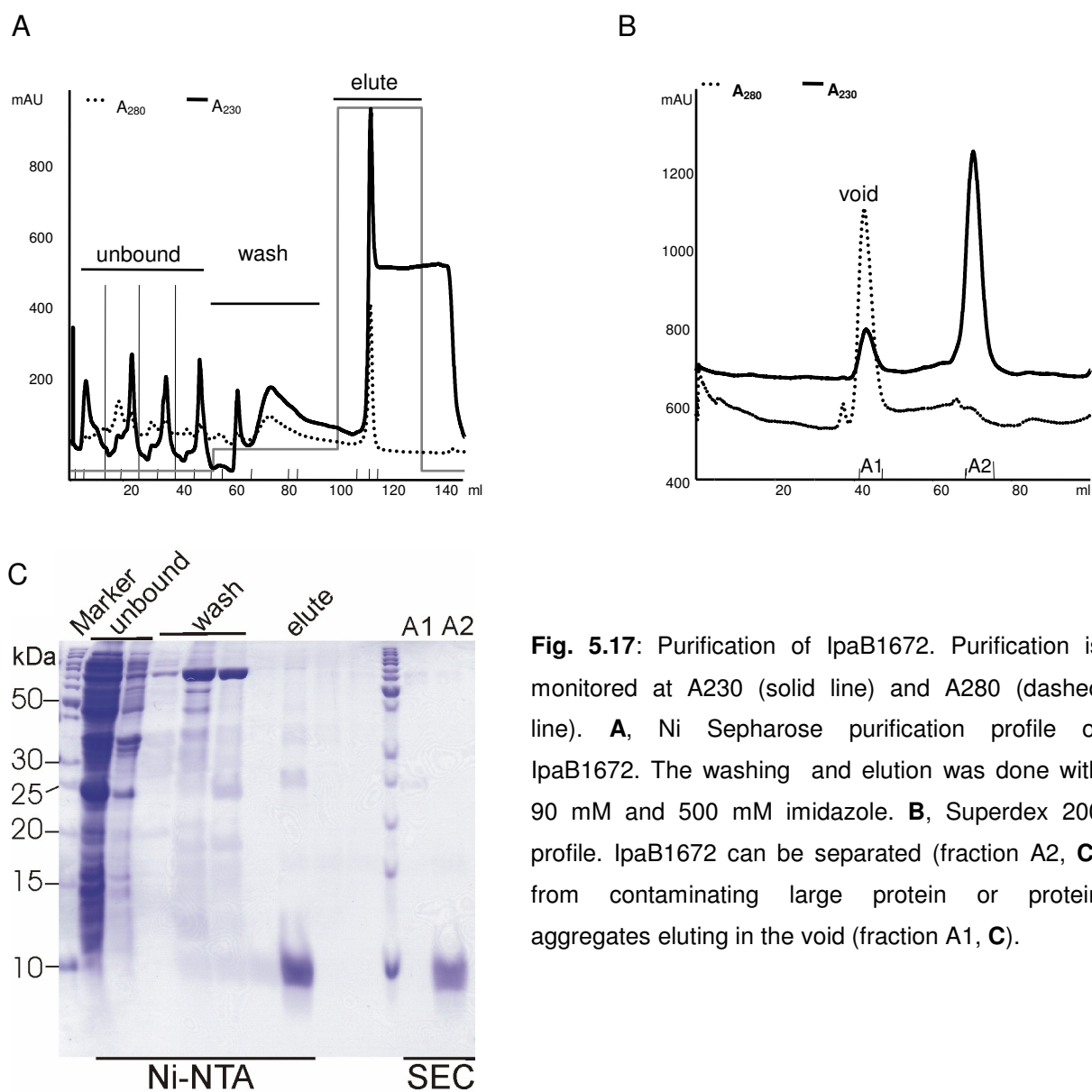
A

B

C

Fig. 5.17: Purification of IpaB1672. Purification is monitored at A230 (solid line) and A280 (dashed line). **A**, Ni Sepharose purification profile of IpaB1672. The washing and elution was done with 90 mM and 500 mM imidazole. **B**, Superdex 200 profile. IpaB1672 can be separated (fraction A2, **C**) from contaminating large protein or protein aggregates eluting in the void (fraction A1, **C**).

3.3.10.2 Co-crystallization of IpaB1672 with IpgC

Purified IpaB1672 peptide was incubated with IpgC (3.1.1.5) at a molar ratio of 2:1. A final concentration of 12 mg/ml was used for setting up cocrystallization trails. The best diffracting crystals were obtained in condition containing 0.1 M HEPES pH 7.5 and 0.8 M potassium sodium tartrate tetrahydrate. Diffraction data were collected up to a resolution of 2.65 Å at BESSY beamline 14.2, Berlin. The crystals belonged to

the space group P3$_1$21 with unit cell dimensions $a = b = 113.7$ Å, $c = 76.4$ Å. The diffraction data statistics are listed in Table 5.2.

Data collection and processing		
		IpgC+IpaB16-72
wavelength (Å)		0.9184
resolution (Å)		40 - 2.65
resolution (Å)	(last shell)	2.80 – 2.65
space group		P3$_1$21
temperature (K)		100
X-ray source		BESSY BL 14.2
detector		MARCCD 165mm
unit cell parameters	a (Å), α (°)	113.7, 90
	b (Å), β (°)	113.7, 90
	c (Å), γ (°)	76.4, 120
total reflections	(last shell)	270300 (31157)
mosaicity (°)		0.35
unique reflections	(last shell)	15633 (2288)
$I/\sigma(I)$	(last shell)	26.38 (3.40)
% data completeness	(last shell)	92.6 (90.4)
R_{meas} [a] (%)	(last shell)	7.1 (87.9)
R_{sym}		6.9 (84.5)
r.m.s deviations		
	bond length (Å)	not available
	bond angle (°)	not available

Table 5.2: X-ray-diffraction-data statistics of the IpaB1672-IpgC crystal.

3.3.10.3 Structure of the IpaB1672-IpgC complex

The crystallographic asymmetric unit contains a dimer similar to the apo structure with single peptide, IpaB16-72 bound to SU B. 10 residues covering amino acids from Asn61 to Pro70 of the IpaB peptide are traceable in the electron density in the cleft of IpgC. The peptide adopts a similar antiparallel orientation to the H2 of IpgC as seen in the cocrystal structure described in earlier section. The binding of a single peptide highlights the difference from the IpaB-IpgC structure described in previous section. This mimics the solution structure of ternary arrangement, though crystal packing aiding in such an arrangement cannot be ruled out. Lack of electron density for residues 16-50 limits from arriving at a definitive conclusion. In support of N-

terminal 50 amino acids maintaining the correct stoichiometry, attempts to copurify any combination of IpaB lacking the N-terminus 50 amino acids together with IpgC led to the aggregation of the complex (data not shown). Hence, plausibly the CBD IpaB18-35 plays a role in defining the right stoichiometry of IpaB-IpgC complex.
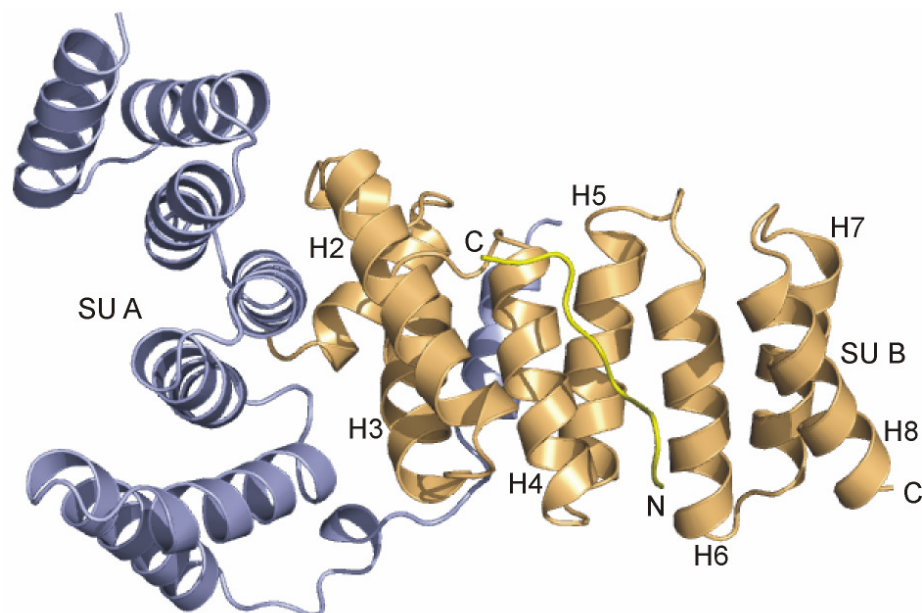

Fig. 5.18: IpaB1672-IpgC complex. Cartoon representation of the IpgC dimer found in the asymmetric unit with SU A colored blue and SU B colored in brown. Helices of SU B are labeled from H1 to H8. IpaB peptide is shown in yellow. The termini are indicated with N and C. The peptide is oriented antiparallel to the H2 of IpgC.