Chapter Four

*In silico* sequence analysis and homology modelling of Xanthomonas adhesin-like protein A (XadA) from *Xanthomonas oryzae pv. oryzae*
4.1 Abstract

Xanthomonas adhesin-like protein A (XadA) is an outer membrane-located protein involved in the initial attachment of Xoo to rice leaves and hydathodal entry. XadA has high amino acid sequence homology to adhesins of animal pathogenic bacteria. It possesses sequence features of a Trimeric Autotransporter Adhesin (TAA), containing a C-terminal transporter region and an N-terminal passenger domain. The putative transporter region of XadA was found to have high homology to the corresponding region of the collagen-binding Yersinia enterocolitica TAA YadA. In addition, XadA passenger domain can be divided into four distinct regions, each homologous to the YadA passenger domain. Homology models of the four regions of XadA passenger domain were generated and each domain has a ‘head’ made of left-handed parallel beta-roll repeats and a transition loop region called the ‘neck’, which connects two tandem heads. XadA structure was found to be considerably different from that of YadA since the ‘head-neck’ of YadA is followed by a coiled-coil region called the ‘stalk’, which is missing from three out of the four ‘head-neck’ domains of XadA. Overall, XadA can be speculated as being a long trimeric molecule with four distinct domains for binding cognate partner(s) in rice, intermediate loop regions aiding in this binding. Comparison of XadA orthologs from the genomes of related Xanthomonas species shows similarity in the overall architecture of the adhesins. However, variability in the number of ‘head-neck’ domains of XadA among different species might be indicative of subtle differences in binding partners in plant hosts and be instrumental in bringing about host and tissue specificity.
4.2 Introduction

The first stage of infection is colonization of the host by the pathogen. The most favoured portals of pathogen entry are the openings through which host tissues are exposed to the external environment. Surface molecules and organelles involved in microbial adherence to the host are pivotal to pathogenesis since proper adhesion determines the chances of infecting the correct target tissues. The number and complexity of the paraphernalia on a pathogen surface is astounding and considered essential for the successful conquest of host niches and establishment of disease (Pizarro-Cerda & Cossart, 2006).

Gram-negative bacterial pathogens assemble adhesive organelles called pili on their surfaces. Pili, also known as fimbriae, are hair-like appendages that can assume two different morphologies; Type I and P pili are rod-like fibres with a diameter of ~7 nm, whereas the other type consists of thin flexible 2-5 nm wide fibrillae (de Graaf and Mooi, 1986). The host receptor-binding adhesin is exclusively present on the pilus tip and is presented on the bacterial surface by a multi-repeat pili subunit. While the adhesins on the pili tips determine host, tissue and cell-type specificity, the pili rods have been implicated in binding host tissue matrix proteins non-specifically to increase the efficiency of the binding (Hultgren et al., 1993).

Apart from this type of ‘fimbrial’ adhesins present on polymeric pili, there are scores of afimbrial or non-pilus adhesins on the pathogen surface that are monomeric or oligomeric proteins but never heteropolymeric in nature. Well-studied examples of afimbrial adhesins are mostly from animal pathogenic bacteria infecting human or other mammalian cells. These include the adhesins AfaD and AfaE from Escherichia coli involved in its attachment to the urinary tract, the filamentous hemagglutinin (FHA) from Bordetella pertussis, responsible for attachment to the lung epithelial and phagocytic cells and the opacity proteins (Opas) from Neisseria that are considered responsible for cell-type specificity (Finlay & Cossart, 1997). Simultaneous expression of several types of fimbrial and afimbrial adhesins indicates that pathogenic bacteria optimize the host adhesion by binding to a large variety of host surface molecules. The host cells, however, are not mere inert surfaces and sense the adhesin-receptor binding by a series of signalling events. Most surprisingly, in some cases, bacterial pathogens bind to host surface proteins that are expressed downstream of activation of the host cells following detection of bacterial entry.
4.2.1 Adhesin-like functions of Xoo

*In silico* analyses show the presence of multiple fimbrial and afimbrial adhesin-like functions in the Xoo genome. *Xanthomonas* adhesin-like protein A (XadA) was identified as an afimbrial adhesin with high sequence homology to *Yersinia enterocolitica* afimbrial adhesin YadA (Ray et al., 2002). In a later study from our lab, a paralog of XadA called XadB, a protein homologous to the *Yersinia* virulence factor YapH and a type IV pilus were also identified as possessing adhesin functions (Das et al., 2009).

XadA-deficient mutant Xoo strain was found to be virulence-deficient when allowed to infect rice leaves through the hydathodes (epiphytic inoculation, where the rice leaves are dipped in the bacterial culture). However, upon infection through wounding (clip inoculation, where scissors dipped in bacterial cultures are used to clip the rice leaves), XadA-deficient mutant strain causes disease to an extent similar to the wild-type Xoo (Ray et al., 2002). An ingenious confocal microscopy assay was developed to assess the efficiency of rice leaf attachment and entry of EGFP-tagged Xoo strains. In this assay, Xoo cells were found to enter into rice leaves within the first 1h of infection and that the colonization and leaf entry occurs preferentially via the dorsal where the hydathodes are particularly concentrated (Das et al., 2009; Mew et al. 1984). Upon comparison of the ability of various Xoo mutant strains to attach to and enter rice leaves, the number of bacterial cells within the first 1mm of the leaf tip were counted. It is evident from the virulence assays that XadA-deficient mutants exhibited a deficiency in leaf attachment and entry in the confocal assay while XadB-deficient strain exhibited a less severe deficiency. A *xadA- xadB-* double mutant exhibited an even greater deficiency in leaf attachment and entry when compared with either the *xadA-* or the *xadB-* single mutants. This additive effect indicates that both genes are independently contributing to leaf attachment and entry. However, the XadB-deficient mutant strain of Xoo shows no virulence deficiency in any type of inoculation on rice leaves. A *xadA- xadB-* double mutant strain is more virulence deficient than *xadB-* mutant but less virulent than *xadA-* mutant strain. Together, these data suggest that XadA and XadB proteins promote Xoo virulence when Xoo enters through the natural mode of Xoo entry through the hydathodes and not when the bacterium can bypass this route by entering through wounds. In the same study, Xoo YapH-deficient strain and the type IV pilus protein PilQ-deficient mutants were found to be more important in Xoo attachment and entry through wounds and even the *in planta* migration of the bacterium inside the rice xylem vessels (Das et al., 2009).
Apart from establishing the role of several adhesins in a plant-pathogen interaction, this study also revealed that these adhesins have redundant functional roles since none of the Xoo strains deficient in one or more adhesins is completely virulence deficient. This opens up a possibility of the presence of other uncharacterized adhesins in the Xoo genome and underlines the requirement for a complete characterization of the structural and functional features of these molecules. The focus of this chapter is on the sequence features of XadA that indicate a great deal about its functional role.

4.2.2 Preliminary characterization of XadA

XadA is a 1265 amino acid long, outer membrane-located protein expressed only in a minimal medium and not the standard laboratory rich medium used for growing Xoo (Peptone-sucrose) (Ray et al., 2002). XadA orthologs have been reported from the genomes of the two sequenced strains of Xoo (KACC10331 and MAFF311018) as well as from the genomes of all the sequenced Xanthomonas strains such as Xac, Xcc and Xylella fastidiosa (Simpson et al., 2000; da Silva et al., 2002; Lee et al., 2005; Ochiai et al., 2005).

XadA sequence was observed to be rich in alanine (22.3%), glycine (15.3%), serine (10.1%), valine (9.6%) and threonine (8.1%). Surprisingly, XadA, being a protein of relatively large size, lacks cysteines. The presence of either a derivative of the sequence TAVG or a nine-amino-acid derivative of TDAVNVQL several times in XadA with little variation has also been noted. Homology of XadA to afimbrial adhesins of animal pathogenic bacteria, such as YadA of Yersinia enterocolitica and UspA1 of Moraxella catarrhalis was noted to be such that YadA/UspA1 homologous regions repeat several times in the XadA sequence. XadA N-terminal shows the presence of an unusually long putative signal peptide of 66 amino acids. The C-terminal has a stretch of hydrophobic amino acids that form the membrane anchor region in a class of afimbrial adhesins known as the autotransporters to which class belong YadA and UspA1 (Struyve et al., 1991; Henderson et al., 1998). In order to understand the relevance of these interesting observations, the structures of some of the autotransporter adhesins would be described first and then, compared with the known and more newly identified sequence features of XadA.
4.2.3 Afimbrial autotransporter adhesins

As the name suggests, ‘autotransporters’ are proteins that carry the whole information regarding their export and secretion through the bacterial cell envelope. Several proteins classified as afimbrial adhesins are known to be secreted through this elegant transport mechanism.

In Gram-positive bacteria, secreted proteins are commonly translocated across the single membrane by the Sec pathway that helps secrete unfolded/folding proteins or the two-arginine (Tat) pathway that secretes folded large complexes of enzyme-cofactors etc. However, Gram-negative bacteria have five distinct modes of protein secretion classified based on the mechanism of crossing the cell envelope that is made up of an inner membrane, a periplasmic space and an outer membrane. The type I, II, III and IV secretion systems have distinct mechanisms of secretion of proteins, polysaccharides, DNA etc. across the inner and/or outer membranes, through pores formed by multimeric complexes of several proteins (secretion systems reviewed in Tseng et al., 2009).

The type V secretion system is made up of ‘autotransporter’ proteins that contain the instructions for export and secretion in their amino acid sequences (reviewed in Dautin & Bernstein, 2007). Autotransporters are a large and diverse group, with almost 800 identified sequences to date, comprising of distinct N-terminal passenger domain and C-terminal membrane anchoring transporter domain. The transporter domain forms a β-barrel structure in the outer membrane (therefore called β-domain) and facilitates proper folding and secretion of the passenger domain, although the exact mechanism remains poorly understood (Pohlner et al., 1987; Bernstein, 2007). These proteins are generally dependent on the Sec pathway for inner membrane transport. The β-domain is the only common feature of all autotransporters and defines the family. The wide diversity in this family is due to the passenger domains that display a diverse range of sizes (<20 KDa upto >400 KDa) and functions such as adhesins, proteases, cytolysins, toxins and outer membrane proteins. These proteins are synthesized with an N-terminal cleavable signal peptide that directs their export into the periplasm via the Sec machinery. Passenger domains may be cleaved from the transporter domain after secretion or remain attached (Henderson et al., 2004).
Autotransporters have been classified into three categories, monomeric type Va proteins, two-component type Vb adhesins and trimeric type Vc adhesins (Desvaux et al., 2004a). The type Va proteins have a β-domain of ~30KDa size that can form the whole membrane-spanning translocator pore (Desvaux et al., 2004b). The type Vb passenger domain and the β-domain are translated as two separate proteins but retain the same mechanism of transport. The type Vc (referred to as Trimeric Autotransporter Adhesins; TAA) is an interesting group of proteins comprising of prominent bacterial virulence factors that are, to date, only adhesins in function. TAAs have a very small (~10KDa) β-domain and the translocator pore forms only when these proteins homotrimerize, trimerization then spreading to the passenger domains too (Linke et al., 2006).

### 4.2.4 Structural organization of trimeric autotransporter adhesins

TAAs from animal pathogens such as YadA from *Yersinia enterocolitica* (Fig.4.1), NadA and NalP of *Neisseria meningitides*, BadA from *Bartonella henselae*, Hia (Fig.4.2) and Hsf from *Haemophilus influenza* are very well characterized in terms of structural architectures and helped define TAAs from several other animal and plant pathogens. Initial scanning electron microscopic images revealed that these proteins form lollipop-shaped surface projections on the cell envelope (Hoiczyk et al., 2000). X-ray crystal structures of various regions of these proteins showed that the ‘lollipop’ comprises of head, neck, stalk that form the passenger domain and the membrane anchor β-domain.

**The Passenger Domain:**

X-ray crystal structures are available for only two TAA passenger domains, YadA and Hia. YadA forms a trimer of single-stranded left-handed β-helices in a novel β-roll fold (Fig.4.1a). The trimeric interface is held by strong hydrophobic interactions due to the periodically occurring conserved NSVAIGXXS sequence motifs (Fig.4.1b). These motifs or their variants are present in the heads of many TAAs, which suggests a structural (and probably functional) analogy in these regions. Following the head domain is a loop region called neck, which functions as an adaptor between the larger diameter of the head and the thinner stalk. The neck loops criss-cross below the trimers and create a ‘safety-pin’ like lock (Fig.4.1c) (Nummelin et al., 2004).
Figure 4.1. *Yersinia* YadA crystal structure. (A) The trimeric YadA head domain. The strands are drawn as arrows, helices as ribbons and the different monomers are in different colours. (B) One level of the β-roll in the trimer viewed along the z-axis showing the packing of the oligomeric core by large hydrophobic residues and the packing of the monomeric interior by small hydrophobic residues. The conserved ‘NSVAIG’ residues are marked. (C) Organisation of the neck region in the C-terminus of the head domain, viewed from the C-terminus along the z-axis. The safety-pin structures as well as the beginnings of the stalk domain helices are shown. (D) A model of the whole YadA trimer; note the long stalk and the transporter domain at the C-terminal. (Adapted from Nummelin et al., 2004; Linke et al., 2006)
Figure 4.2. *Haemophilus influenza* Hia crystal structure. (A) The trimeric Hia binding domain (HiaBD1 with which it binds to host cell receptors) structure. β strands and helices are represented with arrows and thick helices, respectively. (B) Organisation of the HiaBD1, as viewed from the N-terminus along the z-axis. (Adapted from Linke et al., 2006)

The Hia head-neck architecture is very different, assuming a novel β-prism fold (Fig.4.2). The proximal part of the head domain is formed by a four-stranded β-meander perpendicular to the fibre axis and the distal part is formed by a 5-stranded β-sheet. The XadA head-neck sequence has a higher homology to YadA, as would be described in detail in the results section and therefore, Hia-like head-neck architecture will not be discussed further.

The stalk domains of TAAs are fibrous, highly repetitive structures that are rich in coiled coils and extremely variable in length (Fig.4.1d). They function as spacers to project the head domains away from the bacterial cell surface and towards its host binding partners. TAAs from animal pathogens bind to the proteins abundant in extracellular matrix of the corresponding hosts, for example, YadA binds to collagen, fibronectin and vitronectin. However, the ligand-binding sites on the head domains are yet to be identified. Several YadA-like proteins have been identified in genome sequences of various pathogens where the head-
neck and stalk domains are generally found in multiple copies within the same protein. The architecture of such proteins, the functional relevance and the binding partners of such proteins is yet to be characterized.

The β-Domain:

The type Va autotransporter NalP β-domain from Neisseria meningitidis was the first translocator structure to be solved and revealed a 12-stranded β-barrel with a hydrophilic pore of 10 x 12.5Å that is filled by an α-helix belonging to its passenger domain (Fig.4.3b). This domain and β-domains from several type V-secreted proteins are shown to have translocation activity in vivo and in vitro (Oomen et al., 2004; Roggenkamp et al., 2003; Surana et al., 2004). The 12-stranded β-barrel structure of EspA (also type Va) transporter domain from E. coli revealed that the cleavage of the passenger domain occurs within the pore and that there is a post-cleavage conformational change that stabilizes the β-barrel (Barnard et al., 2007). However, the mechanism and location of the cleavage of the passenger domains may not be common since the pore inner-face electrostatic charges vary from negative in EspA to positive in NalP. The only TAA translocator domain structure to be solved is the Hia β-domain from Haemophilus influenza (Fig.4.3a; Meng et al., 2006). In this case the domain is only 70 residues long as compared to the 250-300 amino acid-long domain of type Va proteins. As expected, the trimerization of the domain forms the whole pore, with each monomer contributing four out of the 12 strands constituting the β-barrel.

The sequence features of XadA described in earlier studies from the lab indicated that XadA belongs to the TAA class of adhesins. However, the presence of multiple repeat sequences in XadA and their relevance was not clear. Therefore, to understand the sequence features of XadA and obtain domain information for crystallization and structure solution, an in depth study of the XadA protein sequence was undertaken. Apart from the domain organization, an interesting hypothesis towards the diversity generation using the varying domains present in XadA sequences across the genus Xanthomonas has been put forth from this bioinformatics study.
Figure 4.3. Structures of β-domains of autotransporter adhesins. (A) The trimeric β-domain of Type Vc *Haemophilus influenzae* adhesin Hia. The three monomers are coloured red, green and yellow. (Adapted from Meng et al., 2006) (B) The monomeric β-domain of Type Va *Neisseria meningitides* adhesin NalP. (Adapted from Oomen et al., 2004)

4.3 Methods

4.3.1 XadA protein sequence analysis

The XadA sequence from BX043 (lab strain of Xoo) is deposited in the National Center for Biotechnology Information (NCBI) database (Gene id. AF288222; Derived protein seq. id. AAG01335). This sequence was submitted to for whole protein BLAST searches (Altschul et al., 1997; http://blast.ncbi.nlm.nih.gov/Blast.cgi), PFAM server for identification of the XadA domains (Finn et al., 2008; http://pfam.sanger.ac.uk/) and CLUSTALW server for multiple sequence alignment of the domains of XadA with YadA and other autotransporter sequences (Larkin et al., 2007; http://www.ebi.ac.uk/Tools/clustalw2/). The sequence alignment of XadA domains with YadA was manually adjusted in the alignment editor software Jalview (Clamp et al., 2004). ALSCRIPT was used for formatting the alignments (Barton, 1993).
4.3.2 Manual alignments

A manual YadA structure-based alignment of the domains of XadA was performed to obtain more reliable alignments and secondary structure information about XadA. These features were not evident from the method described in section 4.3.1.

4.3.3 XadA homology modelling

The alignments obtained from the two methods described above were submitted to several homology-modelling servers that generate protein models based on the known structure and alignment of the query protein with the sequence of the protein with the known structure, YadA (PDB id: 1P9H), in this case. MODELLERv9 (Sali & Blundell, 1993), 3D-JigSaw (http://bmm.cancerresearchuk.org/~3djigsaw/), Swiss-Model (Schwede et al., 2003) and I-TASSER (Zhang, 2008) were used for this purpose. Structures of the homologous proteins were retrieved from Protein Data Bank (PDB; http://www.rcsb.org/pdb/). DALI server, freely available software in for structural alignments, was used for structural homology searches (Sali & Blundell, 1993). Structural models were generated as monomers for XadA in five different parts predicted to be domains. The transformation matrix for conversion of YadA monomer into trimer as provided in the PDB file on the RCSB server was used to generate XadA trimers for each monomer. The final model of the whole XadA trimerized molecule is a linear manual assembly of the different domain models.

Electrostatic surface potential of the individual trimers of each domain of XadA was analyzed using the software GRASP (Nicholls et al., 1991).

4.3.4 XadA sequence comparison across genus Xanthomonas

XadA sequences (protein and gene) from various members of genus Xanthomonas were retrieved from NCBI server (Xoo: MAFF 311018, KACC10331; Xcc str. ATCC 33913; Xcv str. 85-10; Xac str. 306). Sequences of the recently sequenced Xca and Xoo str. PXO99A were made available under a collaborative Xanthomonas genomes annotation project (courtesy Dr. Adam Bogdanove). XadA ORFs were annotated in the Xca (cabbage pathogen Xanthomonas campestris pv. armoraceae) and Xoo str. PXO99A genomes. XadA sequences from the seven genomes mentioned were submitted to the PFAM server to assess the number of YadA-like domains in each. Sequences from the same species of Xanthomonas having different number
of domains were analyzed further. Artemis Comparison Tool (ACT) was used to compare the whole coding regions and to find out whether there are any internal rearrangements within the corresponding genes (Carver et al., 2005).

4.4 Results and Discussion

4.4.1 XadA is a TAA with four distinct head-neck domains

A simple protein BLAST search with the whole XadA sequence identifies XadA-like proteins from other Xanthomonas members and closely related bacteria like Xylella fastidiosa, Stenotrophomonas maltophilia and Burkholderia sp. or putative hemagglutinin-like proteins from other Gram-negative pathogens like Shigella sp. and E. coli. Homology to autotransporters from other bacteria like H. influenza and Neisseria sp. is mainly to the C-terminal transporter domain. The only solved structure with which XadA shows significant sequence homology is that of YadA from Yersinia enterocolitica. The result of a BLAST alignment with YadA is shown in Fig. 4.4a. The peculiarity in this alignment is that the YadA homology is repeated five times in the XadA sequence, indicating the presence of repeats in XadA. Several ‘TAVG’ and ‘TDAVNVAQL’ sequence derivatives were identified in the XadA sequence as also reported by Ray et al., 2002. Interestingly, four out of five YadA homology regions were found to be centred around the ‘TDAVNVAQL’ stretches that constitutes a part of the ‘safety pin’-like structure of the YadA ‘neck’ region just below the ‘head’ (Fig. 4.1c). Moreover, XadA homology to Haemophilus influenza adhesin Hia is also at similar stretches. The ‘TAVG’ repeats were found to be recurring after about every ten residues and resembled the hydrophobic patches that form the core of the head trimer in YadA. The fifth region of YadA homology is to the C-terminal transporter end of XadA sequence. The presence of four ‘head-necks’ and a transporter region in XadA was confirmed by PFAM analysis (Fig. 4.4b & c). Despite the identification of four ‘head-neck’ domains, a careful observation of the alignments indicated that several regions were misaligned and many ‘TAVG’ derivatives were not aligned with the corresponding YadA regions of homology. The reasons for this misalignment could be that the lengths of the XadA head-neck domains were much different from the single YadA head-neck domain and that many repeats in the XadA head region were longer than the YadA repeats.
Figure 4.4. Preliminary sequence analysis of XadA. (A) The result of an NCBI-BLASTP alignment of XadA with YadA. Five discernible regions of homology and their respective E-values and percentage identities are mentioned. Note the presence of the highlighted highly homologous stretches; the ones marked in red correspond to the hydrophobic repeats found in the ‘head’ domains while the green correspond to the ‘neck’ domains. (B & C) PFAM analysis of the YadA (B) and XadA (C) sequences showing the presence of one or more ‘head’ domains (red), ‘neck’ domains (green) and autotransporter regions (yellow).
Therefore, a manual alignment of the whole XadA region was performed. The basis of this effort was that the YadA head is composed of a 14-residue repeat motif, of the form:

\[ \text{Turn (T1)} \rightarrow \text{Three-residue inner strand (IS)} \rightarrow \text{Turn (T2)} \rightarrow \text{Three-residue outer strand (OS)} \]

The ‘NSV AIG’ face is the inner strand and hydrophobic in nature. The outer strands of YadA head are rich in charged residues. The variation in the repeat length is from 13 to 16 residues, is caused by a variation in the T1 length. The other turn, T2, is always the same length. The start of every T2 is marked by an invariant Gly since the conformation of T2 does not allow other residues at this position. In the equivalent position in T1, Gly is not conserved because the conformation is that of a normal β-strand.

YadA also has Ser residues at the fourth position from the invariant Gly to stabilise turns by hydrogen bonding to Gly carbonyl oxygen. In a tight β-roll, there is no space within the core of the head monomers for large residues, and therefore there is a strong preference for small ones, especially Ser in the turns and Ala in the middle of the strands. Conversely, the inter-trimeric core is formed by the Val and Ile (Nummelin et al., 2004).

The residues constituting the T1, IS, T2 and OS were identified for the four head domains of XadA manually. In Fig.4.5a, the residues marked in green represent the XadA inner β-strand IS and the residues in purple are the outer β-strand OS (Fig. 4.5). The invariant Gly is also seen throughout the four heads, repeating after every 13 residues and is marked in red. This alignment was found to be a better indicator of the extent of similarity between YadA and XadA head-neck domains since it revealed the essential structural features of XadA that neither PFAM analysis nor other alignments could identify. With this alignment, it can be safely concluded that XadA head domains have a left-handed parallel β-roll structure.
Figure 4.5. Manual YadA-structure-based-alignment of XadA. XadA (A) and YadA (B) sequences were spaced based on the ‘Turn (T1)-Three-residue inner strand (IS)-Turn (T2)-Three-residue outer strand (OS)’ arrangement seen in YadA structure. Residues marked in green are the IS hydrophobic amino acids, the invariable Gly are marked in red and the polar residues of OS are highlighted in blue.
Although four head-neck domains are present, only one equivalent of the YadA coiled-coil stalk region was found in XadA. The lack of stalk domains after the first two head-neck domains is interesting and may indicate a plant ligand-specific adaptation. Several long stretches of XadA head repeats were found not conforming to the YadA structure. These were predicted as loop regions.

XadA head-neck domains were individually aligned to the YadA head-neck domain and submitted to various comparative homology-modelling servers to obtain structural models for XadA. However, since the sequence identities of these regions with YadA are very low (Fig.4.4), the models obtained were of very low scores (by the standards of individual programs) with large aberrations in the secondary structures and deviations from the predicted YadA-like left-handed parallel $\beta$-roll. In order to generate better and more reliable models, manual alignments were done for short stretches of XadA sequence with YadA using the software Jalview and modelled using MODELLER v9. Since the sequence identities of individual domains with YadA increased to $\sim$55% compared to only $\sim$30% before the manual alignment (Fig.4.4), the resultant models are expected to be more reliable, resembled the YadA structure closely and superimposed well with YadA (Fig.4.6). The r.m.s. deviations of each of the XadA domain models were in the range of 1.0-1.5 upon superposition with YadA.

As expected, the ‘TAVG’ hydrophobic patches formed the inner strand and the inner strands, in turn, formed the hydrophobic core upon trimer generation for each domain monomer. The predicted loop regions were found to constitute the T1 turns only and none was found to be present in the T2 turns. Some large loops have a propensity to form helices.

A working model of the whole XadA molecule was generated by organizing the trimeric form of each head-neck domain of XadA in a linear manner (Fig.4.7). The T1 regions of the head which were observed to loop out or form helices in the monomer models were found to be present away from the core $\beta$-roll, causing no obstructions in the trimerization and also, making XadA less compact than YadA. These loops and other secondary structures may have a role in mediating specific motif interactions with the plant ligand. It could be observed from the whole molecule model that the four head-neck domains can provide large ligand-binding areas and may help Xoo in binding to multiple ligands in the rice hydathodes at the same time, considering that XadA is involved in the very early entry of Xoo through rice hydathodes. The length of XadA molecule would allow it to cross the thick layer of exopolysaccharide surrounding the Xoo cells.
Figure 4.6. The homology models of the four head-neck domains of XadA. Models for (A) Head-neck 1, (B) Head-neck 2, (C) Head-neck 3, (D) a coiled-coil stretch between Neck 3 and Head 4 and (E) Head-neck 4 regions were generated as monomers by individually aligning the respective sequences to YadA. The individual monomers of each domain are shown in panels left of each arrow; Three monomers of each domain were generated based on the translation and rotation matrix of the three YadA monomers and used to create trimers (shown in the panels right of each arrow) with no Cα clashes. Note that the long loops between the β rolls in these domains might be artefacts of modelling in the regions of lower sequence identity. While each individual monomer is coloured in yellow, the three monomers of a trimer are coloured yellow, red and purple. β-strands are shown as arrows in the cartoon representation.
Figure 4.7. A homology model of the XadA passenger domain. This representation is the most probable structure of the XadA passenger as it might appear after being transported out of the Xoo cell outer membrane in a homotrimeric form. The trimers of individual head-neck domains shown in Fig.4.6 were joined to obtain this model. (A) Ribbon representation of the side-view; (B) Ribbon representation of the top-view (Note that XadA might also form a central pore similar to the YadA passenger domain; (C) Cartoon representation of the side-view. The three monomers of a trimer are coloured yellow, red and purple. β-strands are shown as arrows in the cartoon representation.
Figure 4.8. Electrostatic surface potential of the four trimeric domains of XadA. Surface potential of (A) Head-neck domain 1; (B) Head-neck domain 2; (C) Head-neck domain 3; (D) Head-neck domain 4; represented as Top-down view, front back faces and the bottom-up view. Red indicates negative charge and blue indicates positive charge.
Mapping the electrostatic surface potential of each trimeric domain of XadA revealed that while the second, third and fourth domains are highly negative in charge, the first domain appears to be positively charged (Fig. 4.8). This might be an indicator of the first and the most exposed domain functioning as a ligand ‘grabbing’ domain or that it may bind to a ligand that is chemically different from the ligand(s) for the rest of the domains. In addition, there are several clefts and pockets in the four domains, which are more negatively charged than the rest of the surface and might presumably be involved in conferring specificity to the ligand binding of XadA. It would be interesting to study the effects of deletion of these highly charged stretches on XadA function in Xoo.

4.4.2 Why should Xoo need a XadA protein?

The collagen binding ability of YadA is well-established (Tamm et al., 1993; Roggenkamp et al., 1996). Collagen is a hydroxyproline-rich protein with a triple helical structure forming long rigid fibrils in the extracellular matrix of higher animals. The central assembly of YadA head seems to bind collagen and a protein docking analysis was used to speculate that the most probable site of binding is a long groove formed by two adjacent T2 turns in the YadA trimer (Nummelin et al., 2004). Trimerization and the 14-mer head repeats of YadA have been shown experimentally to be essential for collagen binding (El Tahir et al., 2000). Proteins containing hydroxyproline are important structural components of plant cell walls. These proteins are termed hydroxyproline-rich glycoproteins (HRGPs) and fall into four groups: extensins, proline-rich proteins (PRPs), 4-hydroxyproline-rich lectins and arabinogalactan proteins (AGPs). HRGPs are often covalently cross-linked into large meshworks, providing tensile strength for the plant cell walls (Showalter, 1993; Cassab, 1988). HRGPs, AGPs specifically, are highly glycosylated molecules involved in cell-cell interactions and might be putative binding partners for XadA-like proteins. Some of the XadA loops might be involved in recognizing HRGP surface glycosylated moieties.

4.4.3 Diversity in XadA proteins from various Xanthomonads

As mentioned previously, XadA orthologs have been reported from the genomes of all the sequenced Xanthomonas species such as Xoo, *Xanthomonas axonopodis* pv. citrii (Xac), *Xanthomonas campestris* pv. vesicatoria (Xcv), *Xanthomonas campestris* pv. campestris (Xcc) and *Xylella fastidiosa* (Simpson et al., 2000; da Silva et al., 2002; Lee et al., 2005; Ochiai et
Chapter Four

05). XadA sequences from these genomes and the newly sequenced *Xanthomonas campestris* pv. armoraceae (Xca), *Xanthomonas oryzae* pv. oryzicola (Xoc) and Xoo PXO99A genomes were retrieved and analyzed. The three strains of Xoo viz., KACC10331, MAFF311018 and PXO99A were found to be identical and hence, were treated as one protein.

Multiple sequence alignment of the sequences revealed that the XadA proteins have lesser sequence homology amongst each other than expected from the highly homologous genomes of Xanthomonads. For example, although Xoo and Xoc genomes are considered about 90% identical, XadA proteins from the two are 70% identical. Interestingly, the variations in these sequences were found to be in the form of large stretches of sequences in the middle of the proteins and not as differences in residues, contributing to large differences in the lengths of XadA orthologs. PFAM analysis was also performed to assess XadA architecture among these bacteria.

The primary difference between the orthologs of XadA lies in the number of head-neck domains that they have and this leads to the different lengths of these proteins. For example, the Xoc XadA is 2117 aa long while Xoo XadA is 1265 aa long. The size difference is due to the deletion of two of the head-neck domains in Xoo XadA, possibly due to a recombination event involving sequences that are in/near the neck regions (Fig.4.9a). The putative site of recombination in Xoc beyond which the Xoc XadA is identical to Xoo XadA could also be traced (data not shown). Size differences between XadA proteins can also occur due to expansion of 14 aa repeats that comprise the head regions as observed between XadA of Xcc and Xca. Artemis comparison (Fig.4.9b) reveals that one region of the protein is very dissimilar between XadA orthologs of Xca and Xoc. Interestingly, this is the region (~2500bp or 850 aa long) that is present in Xoc XadA but absent from Xoo XadA.
Figure 4.9. Comparison of XadA across Xanthomonad species. A. PFAM analysis of XadA of different Xanthomonads. The red blocks represent the ‘Head’ structural motifs of Yersinia adhesin YadA consisting of several 14 amino acid repeats. The green blocks indicate the ‘Neck’ motifs of YadA and the yellow blocks mark the C-terminal domain of YadA. The region marked by arrows indicates the part of Xoc XadA deleted in Xoo XadA. B. xadA nucleotide sequence comparisons using Artemis Comparison Tool (ACT). The red blocks indicate stretches of identical/similar sequences and white blocks show stretches of deletions/dissimilarities. xadA from each genome has been marked as a grey block.

Generation of molecular diversity is a common feature found in TAAs of several animal pathogens, achieved by phase variation, repeat expansion-contraction and even loss-reacquisition of whole TAA molecules (Hoiczyk et al., 2000; Martin et al., 2003; Lafontaine et al., 2001). Especially interesting is the observed variations in the total size of XadA-like Bartonella adhesin BadA due to variations in the length of the repetitive neck/stalk sequences; the sizes of the head and membrane anchor sequences remaining equal in all Bartonella henselae strains investigated (Reiss et al., 2007). In animal pathogens, escape from detection by host immune surveillance is a major virulence determinant and hence the variability in surface molecules, which are antigenic and easily detected. A plausible reason for diversity in XadA proteins from genus Xanthomonas might not be to evade plant immune system but to arbitrate host or tissue specificity. Deletion of two head-neck domains of XadA from the
vascular pathogen Xoo XadA that enters rice through hydathodes as compared to the mesophyllic pathogen Xoc that enters rice leaves through stomata also indicates a role for XadA in tissue specificity. Similarly, the vascular pathogen Xcc and mesophyllic pathogen Xca infect cabbage plants and exhibit large variation in the size of one of the head repeats as the major difference in XadA proteins. Apart from the expansion and deletion in both the cases, the rest of XadA and the flanking regions are 96-100% identical. However, such large variations in sizes were not observed upon comparison of XadA proteins from Xanthomonas pathogens, which infect different hosts i.e. Xoc and Xcc (Fig. 4.9b). This finding suggests that these pathogens retain the more conserved ancestral versions and that the variations in the head-neck domains have led to evolution of varied XadA proteins to avoid competition with closely related species infecting the same host.

4.4.4 A bigger picture: Other afimbrial adhesins of Xanthomonads

Xoo, Xoc, Xac and Xcv have XadA and its paralog XadB, while Xcc and Xca encode only for XadA and lack XadB. XadB was also predicted to have XadA-like architecture, only shorter by one head-neck domain. It was found to be 97% identical between Xoo and Xoc while only 60% identical between Xoo and Xac/Xcv and again, 97% identical between Xac and Xcv. YapH, a member of type Va monomeric autotransporter family was also found to be similarly variable among Xanthomonads. A 38,766 bp genomic region containing genes fhaB, fhaX, fhaBl and fhaC that may encode predicted non-fimbrial adhesins of the two-component Type Vb system was found to be completely missing in the two Xoo strains BX043, KACC10331 and MAFF311018 while present in Xoo strain PX099^A and all the other Xanthomonas genomes (data not shown; Salzberg et al., 2008). The locus also includes several direct repeats of ISXo5 transposon elements, which in turn, are flanked by genes for a dual specificity phosphatase (DSP) and a DNA binding protein (DBP). In contrast, only one copy of the ISXo5 element is present between DSP and DBP in MAFF and KACC strains, indicating that the ISXo5 element was involved in the genomic rearrangement that led either to loss of the locus from BX043, MAFF and KACC or gain of the locus in PX099^A, the former being a likely case (data not shown; Salzberg et al., 2008). Together, this in silico analysis reveals that the orthologs of all the afimbrial autotransporter adhesins vary among the Xanthomonas genomes. These differences could have a major role in defining the host and/or tissue specificity among these pathogens.
4.5 Directions from the study

The analysis of Xoo XadA and other afimbrial autotransporter adhesins revealed several unique sequence features, distinct from other characterized adhesin structures and strongly advocated the need for detailed structural studies. Examination of XadA sequence convincingly identified four YadA-like head-neck domains with left-handed β-roll architecture and essential cues for trimer formation with a hydrophobic core. However, a possibility of XadA trimer taking up a more globular structure with the four domains interacting with each other to form a bulbous quaternary structure cannot be ruled out. It can be argued that a globular structure would form a very compact molecule that would be unable to cross the exopolysaccharide layer of Xoo cells and an elongated molecule with four distinct binding sites seems more appropriate as a structure for a surface adhesin. The homology models for XadA could be generated only in four parts and never as a complete molecule since overall identity and architectural similarity of XadA to the only characterized structural homolog YadA is very less. The manually generated working model of XadA (Fig.4.7) fits into the functional role of XadA well. Nevertheless, alternative quaternary structures for XadA could not be ruled out with this study and hence, X-ray crystal structure is necessary. As would be discussed in chapter 5, XadA was taken up for X-ray crystal structure analysis and the distinctive features identified in this chapter were used for cloning this membrane-anchored large protein.