Structure of the Tetragonal Surface Virulence Array Protein and Gene of Aeromonas salmonicida*

Shijian Chu, Sonia Cavaignac, Josiane Feutrier, Barry M. Phipps, Magdalena Kostrzynska, William W. Kay, and Trevor J. Trust†

From the Department of Biochemistry and Microbiology and Canadian Bacterial Diseases Network, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

(Received for publication, December 10, 1990)

The paracrystalline surface protein array of the pathogenic bacterium Aeromonas salmonicida is a primary virulence factor with novel binding capabilities. This species-specific structural gene (vapA) for this array protein (A-protein) was cloned into λgt11 but was unstable when expressed in Escherichia coli, undergoing an 816-base pair deletion due to a 21-base pair direct repeat within the gene. However, the gene was stable in cosmids pL2917 as long as expression was poor. A-protein was located in the cytoplasmic, inner membrane and periplasmic fractions in E. coli. The DNA sequence revealed a 1,506-base pair open reading frame encoding a protein consisting of a 21-amino acid signal peptide, and a 481-residue 50,778 molecular weight protein containing considerable secondary structure. When assembled into a paracrystalline protein array on Aeromonas the cell surface A-protein was totally refractile to cleavage by trypsin, but became trypsin sensitive when disassembled. Trypsin cleavage of the isolated protein provided evidence that both the NH₂ and COOH-terminal regions form distinct structural domains, consistent with three-dimensional ultrastructural evidence. The NH₂-terminal 274-residue domain remained refractile to trypsin activity. This segment connects a NH₂-terminal 78-residue linker region to a COOH-terminal 129-residue fragment which could apparently refold into a partially trypsin-resistant structure after cleavage at residue 323.

S-layers, or paracrystalline surface protein arrays, are regular two-dimensional assemblies of protein monomers that constitute the outermost layer of the cell envelope of a wide variety of bacteria (1-3). The majority of S-layers appear to be composed of a single species of protein or glycoprotein (3). S-layer proteins are generally the predominant cell protein and have precise ultrastructural morphologies. They self-assemble into a supramolecular structure covering the entire cell surface and are often anchored by interactions with the underlying wall or membrane. However, there is limited knowledge of the structural organization of S-layer proteins.

The best described S-layer in terms of biological activity and ultrastructure, as well as export and assembly, is the so-called A-layer of the bacterium Aeromonas salmonicida (4). This organism causes diseases of different pathogenesis in a wide variety of fish species worldwide (5, 6). Clinical isolates invariably produce A-layer which is primarily composed of a single protein (A-protein) which has been purified and subjected to biophysical, immunological, and biochemical characterization (7, 8). This protein is a tetragonal array whose three-dimensional structure has been reconstructed from a tilt series of electron micrographs of A-protein sheets (9). Transposon mutagenesis studies have also defined several steps in export of this protein, as well as structural requirements for anchoring of the layer to the cell surface (10). This protein array appears to be crucial for the ability of A. salmonicida to produce disease in fish, since isogenic mutants unable to produce A-layer display a >10⁶-fold reduction in virulence (11). A-layer also appears to be multifunctional in that it protects A. salmonicida from the bactericidal activity of both immune and nonimmune serum (12) and probably from the killing activity of phagocytic cells (13). Indeed A-layer appears to facilitate binding of A. salmonicida to macrophages (15) and in addition specifically binds immunoglobulins (14) and porphyrins (15), as well as certain tissue matrix proteins,' and thus represents a novel surface-exposed binding layer.

The A-layer of A. salmonicida is therefore an excellent model for studying the molecular basis of S-layer structure and function, as well as expression and export. Fundamental to understanding these phenomena in molecular terms is the elucidation of the primary structure of the protein. In this paper we report the cloning and expression of the A-protein gene of A. salmonicida (vapA = virulence array protein gene A) in Escherichia coli, determination of the nucleotide sequence, analysis of the structure of A-protein, and evidence for the presence of two structural domains in this tetragonal array forming protein.

EXPERIMENTAL PROCEDURES²

RESULTS

Cloning—The structural gene coding for the subunit protein of the A-layer of A. salmonicida was previously cloned from strain A449 using the phage λgt11 (23). The cloned gene was shown to code for a protein with an apparent subunit M₇.

1 P. Doig, L. Emody, and T. J. Trust, unpublished observation.
2 Portions of this paper (including "Experimental Procedures," Tables I and II, and Figs. 1, 4, and 5) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

The JOURNAL OF BIOLOGICAL CHEMISTRY
© 1991 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.
indistinguishable from that of the wild type protein. Plating of the recombinant phage on *E. coli* strain Y1090 always produced plaques unable to react with antibodies to A-protein. This instability also increased with storage, after conversion to lysogeny in *E. coli* strain Y1089, and after subculturing immunologically detectable A-protein was no longer produced. The A-protein gene in the lysogenic recombinant phage $\lambda$ 10g was localized using two oligonucleotide probes to the NH$_2$-terminal amino acid sequence (8). After subcloning into M13 and pTZ18R (Fig. 1), the nucleotide sequence was determined. This analysis showed that there were insufficient nucleotides to code for an intact A-protein, indicating that a rearrangement had occurred in the clone after conversion to lysogeny.

This finding necessitated recloning the gene coding for A-protein. On this occasion, strain A450 was used and the vector was the 21-kb$^2$ broad host range cosmid pLA2917 (21). From 20,000 clones, a single stable clone in *E. coli* HB101 which reacted weakly with anti-A-protein immunoglobulin was selected. Strongly reacting clones invariably were unstable when subcultured. Plasmid pBF501 contained a 27-kb Sau3A insert. A 3.4-kb SalI fragment was subcloned back into pLA2917 to yield pBF512. A-protein expression by subclone pBF512 was less than pBF501, so pBF512 DNA was digested by PstI, and the PstI fragments ligated with PstI-digested pTZ18R. After transformation into *E. coli* DH5$\alpha$, clones were screened for ampicillin resistance and *lac Z* gene complementation. One of 40 white colonies showed a strong reaction with anti-A-protein immunoglobulin, and Western blot analysis showed that the A-protein produced was indistinguishable from that of the wild type protein. This clone pTZ521 contained a 4.6-kb insert, and, in contrast to clone $\lambda$ 10g, A-protein production was stable to storage and subculture. The position of the A-protein gene in the pTZ521 insert was mapped with the aid of oligonucleotides synthesized during the sequencing of the A449 A-protein gene described above. Interestingly, when the 4.6-kb A-protein gene-containing insert was ligated into pTZ19R where the inverted polycloning site positioned the T; promoter downstream from the insert rather than upstream as was the case in pTZ521, there was no alteration in the production of A-protein, suggesting that A-protein synthesis was from a promoter on the *Aeromonas* DNA insert.

Restriction mapping of the insert in pTZ521 identified a 1.3-kb HindIII fragment at the 3' end of the 4.6-kb insert. The plasmid obtained after removal of this HindIII fragment, pSC150 (Fig. 1), contained the smallest insert still capable of producing A-protein in *E. coli*. Attempts to subclone smaller fragments containing the entire A-gene were uniformly unsuccessful, although subclones containing various regions of the gene were obtained and were stable. However, none appeared to produce immunologically detectable truncated peptides when assayed by Western blotting with anti-A-protein immunoglobulin; even plasmid pSC250 containing both the upstream region and approximately 80% of the A-protein gene (Fig. 1) was negative.

**Cellular Localization in *E. coli***—*E. coli* cells carrying plasmid pSC150 produced A-protein during growth at both 20 and 37°C. However, A-protein production in *E. coli* from the cloned *A. salmonicida* DNA was markedly lower than that of wild type *A. salmonicida* and detection required Western blot immunoblot assay. Comparison of growth curves of *E. coli* cells carrying plasmid pTZ18R with and without the A-protein gene insert showed that production of even this small amount of A-protein delayed the growth of *E. coli*. At 37 °C the delay in reaching the mid-exponential phase of growth was approximatively 2 h. Cell fractionation experiments showed that *E. coli* was capable of exporting A-protein to the periplasm (Fig. 2). This was shown using two procedures for the isolation of periplasmic proteins. A-protein was also detected in the cytoplasmic membrane, and cytosol fractions, but was not detected in the outer membrane fraction or the culture supernatant. Some proteolytic degradation of A-protein was apparent in the cytoplasmic fraction.

**DNA Sequence Analysis**—Sequence analysis of the subclones pSC250-pSC550 (Fig. 1) allowed the 3372-base pair insert in pSC150 to be sequenced in its entirety. The insert contained only a single complete open reading frame (ORF) greater than 300 bases in any of the six possible reading frames. This ORF was 1506 base pairs long, beginning at nucleotide 743 and terminated by a TAA stop codon at 2249 to 2251. Fig. 3 shows its sequence, as well as the 120 and 167 nucleotides immediately up and downstream. The ORF contained an overall G + C content of 48%, lower than the overall 55 ± 1.8% reported for the *A. salmonicida* genome (36). Codon usage shows a small bias for codons with T (58.9%) in the third position and showed significant differences from the codon usage of *A. hydrophila* genes encoding exported protein (Table I). This was clearly apparent for Ile, Val, Ser, Pro, Thr, Ala, Gin, Glu, Arg, and Gly. Codon usage for the *A. salmonicida* gene was more in accord with *E. coli* nonregulatory genes (41), with optimal codons being preferred with Phe, Leu, Ile, Val, Pro, Thr, Ala, Tyr, Asn, Glu, Arg, and Gly. For amino acids contained in the protein seven codons were not used. Three of these corresponded to rare codons in *E. coli* (42), although rare codons were only used at a frequency of 2.1% for the entire peptide.

The gene contains a direct 21-base pair repeat, GAAATCCAGGTGAnnGCCAAn separated by 795 bases, beginning at nucleotides 220 and 1036 and double underlined in Fig. 3. Upstream from the gene, there is an optimally positioned presumptive Shine-Dalgarno ribosomal-binding site (AGGA). Possible control elements in the 5'-untranslated region of the A-protein gene include a predicted -10 region promoter sequence with a typical consensus hexamer TATAAT and a poorly conserved -35 region (nnGAnA) ending 18 nucleotides further upstream. The 3' region contains a possible terminator structure beginning 17 bases after the termination codon with a GC-rich region of dyad symmetry followed by a cluster of thymine residues (43). Alignment of the sequence obtained from $\lambda$ 10g for the A-protein gene of strain A449 identified the position of the deletion suffered in this clone which resulted in loss of the ability to produce A-protein. The deletion was of 816 base

---

$^2$ The abbreviations used are: kb, kilobase(s); ORF, open reading frame; TPCK, 1,1-tosyl-amido-2-phenylthyl chloromethyl ketone; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

---

**Fig. 2.** Western immunoblot analysis of 7.5% SDS-PAGE showing the cellular localization of A-protein coded by pSC150 in *E. coli* DH5$\alpha$. Polyclonal affinity purified rabbit antibodies to *A. salmonicida* 450 A-protein were reacted at a dilution of 1:2,000. Lane 1, purified strain 450 A-protein; lane 2, culture supernatant; lane 3, outer membrane fraction; lane 4, periplasmic fraction; lane 5, cytoplasmic membrane fraction; lane 6, cytosol; lane 7, whole cell lysate.
pairs and included one of the repeats. Other than this deletion, the ORFs and the 483 bases of the A449 DNA sequenced downstream. The sequence of the DNA from the two strains was identical within immediate between two back-to-back SauIIIA restriction upstream A450 DNA is shown in italics and underlined in Fig. 3.

Partial amino acid sequences confirmed by Edman degradation are represented in Fig. 3. Nucleotide sequence of the A-protein gene and flanking DNA from A. salmonicida strains A449 and A450 (numbers on right) and translated amino acid sequence. The numbers above indicate nucleotide positions. Direct repeat sequences in the strain A450 is shown in italics and underlined. Vector nucleotide sequence in the region 3'-10-5' polymerase-binding site in front of gene are indicated, and the double underlined, vector DNA. Conservation of the Gene-Southern blot analysis using a numbers on right indicate nucleotide positions. The cleavage site for the signal sequence is indicated by Edman degradation are underlined. The cleavage site for the signal sequence is indicated (V).

The A-protein gene of strain A450 with chromosomal DNA isolated from ten A-layer producing strains of A. salmonicida A-protein gene and flanking DNA from strains A450 and A449 indicated that the A449 sequence representing the three subspecies of the species, showed that the A449 sequence represented vector DNA. Conservation of the Gene—Southern blot analysis using a 595-base pair sequence from the central region of the A-protein gene of strain A450 with chromosomal DNA isolated from ten A-layer producing strains of A. salmonicida, representing isolates from diverse species of fish, diverse geographical regions, from diseases with different pathogenesis, and representing the three subspecies of the species, showed that the gene was carried on a 2.5 kb BamHI restriction fragment.

![Fig. 3. Nucleotide sequence of the A-protein gene and flanking DNA from A. salmonicida strains A449 and A450 (numbers on right) and translated amino acid sequence. The numbers above indicate nucleotide positions. Direct repeat sequences in the strain A450 is shown in italics and underlined. Vector nucleotide sequence in the region 3'-10-5' polymerase-binding site in front of gene are indicated, and the double underlined, vector DNA. Conservation of the Gene-Southern blot analysis using a numbers on right indicate nucleotide positions. The cleavage site for the signal sequence is indicated by Edman degradation are underlined. The cleavage site for the signal sequence is indicated (V).](image_url)
in each strain (Fig. 4). No hybridization was seen with chromosomal DNA's of the four other tetragonal S-layer producing strains of *A. hydrophila* and *A. sobria*. Identical results were obtained with oligonucleotide probe 5'-GCTGACGTT GGTGCTTC 3' to nucleotides 1348-1364 at the 3' end of the A-protein-gene.

Protein Structure—The A-protein gene encodes a protein of 502 residues shown in Fig. 3. Comparison of the predicted sequence with the NH2-terminal sequence of the mature protein shows the presence of a classical 21-residue amino-terminal signal peptide. The predicted amino acid sequence was confirmed in various regions of the mature protein by automated-Edman degradation of peptides produced by CNBr hydrolysis of purified A450 A-protein, TPCK-trypsin, and endoproteinase Glu C cleavage. The confirmed residues underlined in Fig. 3 represent 31% of the sequence of the mature A-protein.

The predicted *M* of the mature peptide at 50,778 is consistent with that calculated from SDS-PAGE and amino acid composition (Table II). One noteworthy difference between the predicted and measured composition is the presence of only 2 methionine residues by sequence analysis, rather than the 4 by compositional analysis. In overall composition, the predicted mature A-protein contained 43.7% nonpolar/hydrophobic amino acids. The Kyte and Doolittle (44) hydropathic index using an interval of nine amino acids gives an average hydrophobicity score of -0.45. A-protein also contained 37.2% polar, 10.2% acidic, and 8.9% basic amino acids and displayed an overall negative charge. The predicted overall pI was 4.79 compared to the measured pI of 5.7-6.0 (7).

The secondary structure of processed A-protein predicted by combining the methods of Chou and Fasman (45) and Garnier et al. (46) was compared with the secondary structure previously measured by circular dichroism studies in the far UV (7). The predicted β-structure content of 26.2% is consistent with the measured 19-28% content, and this β-structure is predicted to be distributed along the length of the first 424 residues of the 481-residue sequence (Fig. 5). The predicted α-structure content of 24.5% was significantly higher than the 14% content measured for A-protein under non-denaturing conditions, but was similar to the 29% content obtained in the presence of 0.12% SDS. This α-helix content of A-protein was confined to residues after 125. The protein also contains a small content of turns (6.7%) which are distributed evenly among the NH2-terminal 164 and COOH-terminal 146 amino acids, but are lacking in the central region of the sequence (Fig. 5).

The peptide cleavage studies used to confirm the predicted amino acid sequence of A-protein also provided information on the conformation of A-protein. When assembled into native A-layer on the surface of *A. salmonicida* cells, A-protein was resistant to TPCK-trypsin even when cells were treated with trypsin/cell wet weight ratio as high as 1:100 (Fig. 5). However, when purified A-protein was treated with TPCK-trypsin at trypsin/protein ratios between 1:200 and 1:10 SDS-PAGE showed that A-protein was rapidly degraded to a major peptide of approximate *M*, 37,600, via a series of intermediates in the approximate *M*, range 48,000-40,000, and a second major peptide of approximate *M*, 16,700, via a *M*, 21,300 peptide which was unstable at a ratio of 1:100 (Fig. 5). The *M*, 16,700 peptide was ultimately digested at trypsin/protein ratios of >1:20, but the *M*, 37,600 peptide was resistant to further degradation even on prolonged incubation with the highest concentration of trypsin tested (1:5). Automated Edman degradation of intermediate *M*, 48,000 and 40,000 peptides and the *M*, 37,600 trypsin-resistant peptide showed the same NH2-terminal sequence as the mature protein. The first 25 residues of the *M*, 21,300 peptide corresponded to the deduced A-protein sequence beginning at residue 275, while the first 22 residues of the *M*, 16,700 peptide corresponded to the deduced sequence beginning at amino acid 324. Hydrolysis of A-protein by CNBr also resulted in two major peptides (Fig. 6), the *M*, 31,000 and 13,400 cleavage products. The remaining *M*, 6,500 CNBr-peptide was also produced, but was poorly stained by Coomassie Blue, and peptides of approximate *M*, 37,500 and 20,000 were present as partial cleavage products. Peptides of approximate *M*, 23,000 and 17,500 resulted from the formic acid hydrolysis.

**DISCUSSION**

Homogenous S-layers are composed of subunit proteins in the *M*, range 50,000-250,000 assembled into arrays with hexagonal, tetragonal, oblique, and linear morphologies (1-3). With the recent exception of the surface array protein of the pathogenic bacterium *Campylobacter fetus* (47), most molecular attention has been focused on hexagonal array proteins and S-layers with poorly defined biological roles (48-51). Tetragonal arrays, which are as common as hexagonal arrays, have received scant attention. This study has addressed this deficiency and has provided the sequence and structural predictions of a tetragonal array forming protein which uniquely performs a variety of defined biological roles, including a major role in virulence. The A-protein gene is the smallest S-layer gene sequenced to date, and this should facilitate a detailed analysis of the structure-function relationships of this novel protein.

This is also the first report of the sequence of a structural gene from *A. salmonicida*, a gene which appears to be conserved as shown by its presence in a diverse selection of strains of *A. salmonicida* representing its three subspecies, as well as strains isolated from different geographic sources, from different species of both marine and freshwater fish, and from diseases with markedly different pathogenesis. This is in agreement with earlier NH2-terminal sequence comparisons of A-protein (8). No significant homologies were demonstrable with the genes coding for morphologically similar tetragonal surface arrays produced by strains of *A. hydrophila* (43) and *A. sobria*, and searches of the GenBank (release 67), EMBL (release 25), SWISS-PROT (release 17), PIR (release 8), and searches of the GenBank (release 67), EMBL (release 25), SWISS-PROT (release 17), PIR (release 8).

**Fig. 6.** SDS-PAGE analysis stained by Coomassie Blue of *A. salmonicida* A450 A-protein structure by TPCR-trypsin cleavage and CNBr hydrolysis. A-protein (lane 1), A-protein treated at trypsin/protein ratios of 1:200 (lane 2), 1:100 (lane 3), 1:50 (lane 4), 1:20 (lane 5), 1:10 (lane 6), A-protein treated twice with CNBr (lane 7). Also shown is A-protein in glycine extracts of cells before (lane 8) and after treatment (lane 9) with a trypsin/cell ratio of 1:500. In contrast to the cleavage seen in the case of isolated A-protein, no cleavage is seen when A-protein is assembled as A-layer on the cell surface. Identical results were obtained when whole cell lysates were examined directly by SDS-PAGE. *M*, markers (∗1,000) are indicated at right.

![Fig. 6.](image-url)
A. salmonicida A-protein Sequence

23), and GenPept (release 64.3) databases by the FASTA protocol of Pearson (52) indicated that the gene is unique to A. salmonicida. The A-protein gene was particularly difficult to clone, and problems with gene stability were experienced. The instability encountered when the gene was cloned in λgt11, for example, appeared to be related to the presence of two direct repeats in the sequence, and the deletion of almost half of the gene can be explained as a cross-over event involving these repeats. A final cloning attempt provided a gene which was stable, but poorly expressed in E. coli. Poor expression is unlikely to result from the codon usage (41). The codon adaptation index, which is a measure of codon usage bias (54), was 0.516 suggesting that the gene should be relatively well expressed in E. coli. Moreover codon usage was significantly more closely related to E. coli than to A. hydrophila exported protein genes which are well expressed in E. coli (Table I, 37–39). Poor expression likely resulted from the

−35 region which has substitutions shown to have serious effects on promoter activity in E. coli (55) and perhaps the loss of additional control sequences upstream of the −35 region by vector replacement with vector sequence during the cloning events.

Interestingly, while E. coli was clearly a reluctant host for the A-protein gene, this species was capable of A-protein synthesis at 37 °C, a temperature which is lethal for A. salmonicida. Indeed in the case of the natural host of the gene, growth at temperatures above 25 °C results in a selection of variants which are permanently impaired in their ability to synthesize A-protein (11). Cell fractionation studies provided one likely explanation of the problem E. coli has when called upon to produce A-protein. Although a proportion of the A-protein produced in E. coli clearly reaches the periplasm by virtue of the functionality of the A-protein signal sequence in this foreign host, transfer across the cytoplasmic membrane appears to be inefficient with some A-protein being retained in the cytoplasm, as well as in the inner membrane. Such an accumulation of A-protein may be toxic for E. coli. Certainly the small amount of A-protein that was produced did delay growth. The inability of E. coli to export A-protein across the outer membrane is not surprising since this species is not regarded as a protein exporter. In contrast, Aeromonas is regarded as a protein exporting species (37–39). Indeed in its native background, A-protein likely has a unique export pathway, since transposon insertion mutants which are unable to translocate A-protein across the outer membrane and accumulate A-protein in the periplasm (10) still export other proteins with proteolytic and hemolytic activity. This suggests that even in A. salmonicida at least one additional gene product is required for translocation of A-protein across the outer membrane.

Analysis of the deduced composition of A-protein using the algorithm of Barrantes (56) gave a ratio value of 1.06, consistent with a peripheral protein, and analysis of the deduced amino acid sequence by the method of Klein et al. (57) also classified A-protein as a peripheral rather than a integral membrane protein. As was the case with the DNA sequence, searches of data banks containing protein sequences showed that A-protein was unique to A. salmonicida. When the overall sequence similarity of A-protein was compared with other S-layer proteins using the algorithm of Myers and Miller (58), best identity (82 (17%)) was shown with the S-layer protein of Deinococcus radiodurans. The significance of this putative sequence similarity is unclear.

The composition of the mature A-protein was similar to other S-layer proteins in terms of polar, acidic, and basic amino acid content. However, the hydrophobic content of A-protein is the highest yet reported for an S-layer protein, consistent with the surface hydrophobicity conferred upon A. salmonicida cells by A-layer (12). It seems probable that much of the hydrophobicity is located in surface exposed regions of the subunit and may be important in the interaction of A. salmonicida with cells such as macrophages (12) and in the binding activities of A-layer.

In addition to the cleavage at the two central sites to produce a COOH-terminal M, 21,335 peptide beginning at residue 275 and its subsequent cleavage to a M, 16,680 peptide beginning at residue 324, trypsin also sequentially cleaves the purified protein at a series of sites beginning from the COOH-terminal end of A-protein to produce a stable NH₂-terminal peptide of approximate M, 37,600, probably by cleavage after arginine 349 or lysine 353. The resulting overlapping NH₂- and COOH-terminal polypeptides both appear to adopt compact tertiary structures refractile to trypsin. The lesser mass COOH-terminal compact peptide unit displayed intermediate resistance to trypsin even though it contains 14 potential trypsin cleavage sites, while the NH₂-terminal M, 37,600 compact peptide unit with 27 potential trypsin sites appeared to be totally refractile to trypsin activity. Indeed, the 23 potential trypsin sites carried in the NH₂-terminal 274 residue sequence of A-protein appeared to be inaccessible until the protein was denatured. The difference between the predicted PI value of 4.3, and measured PI values of 5.7–6.0 (7), is also a likely consequence of folding and inaccessibility. The total refractivity of the protein to trypsin digestion when assembled as A-layer on the cell surface suggests that none of the arginine or lysine residues in the protein are located or accessible on the surface of the layer. Similarly, the inability of the iodinating reagent iodoagen (59) to label A-layer on the cell surface (60) suggests that tyrosine residues 178, 306, 334, 336, and 390, and histidine residues 88, 117, 368, and 393 are not located on the surface of the native layer.

Another bacterial supramolecular structure which is resistant to trypsin cleavage is the flagella filament. Like A-protein, Salmonella flagellin has a major internal sequence of approximate M, 40,000 which is relatively stable to trypsin digestion despite the presence of numerous potential cleavage sites for the enzyme (61). This 40-kDa sequence appears to comprise two compact structural units, one of which is an internal 27-kDa peptide which under non-denaturing conditions is totally refractile to trypsin, while the terminal regions of flagellin are disordered and mobile in solution and are susceptible to trypsin (62). These terminal regions of flagellin are predicted to be α-helical and are thought to interact with the corresponding regions of adjacent monomers upon polymerization into the flagella filament, and in so doing lose their susceptibility to trypsin (63).

Secondary structure predictions indicate that the sequence which forms the larger mass trypsin resistant compact unit after trypsin cleavage of A-protein can also be subdivided into two structurally defined regions: a β-sheet (28%)-β-turn (18%) NH₂-terminal 125-residue region followed by a region of 227 residues with alternating α-helix (37%) and β-sheet (26%), largely devoid of turns (3%). The remaining COOH-terminal portion of the molecule constitutes a M, 13,400 CNBr fragment which is incorporated into the lower mass compact unit of intermediate resistance to trypsin upon cleavage after lysine 323. It can also be separated into two structurally defined regions: the first 72 residues containing predominantly β-sheet (43%) (9% α-helix, 3% β-turns), followed by the COOH-terminal 57 residues which are devoid of β-

---

1 T. J. Trust, unpublished observation.
sheet but like the COOH-terminal region of flagellin is rich in α-helix (49%).

The presence of two major morphological domains appears to be a motif shared by all S-layers (64). Three-dimensional image analysis of A-protein arrays has also shown that the subunits contain a heavy mass domain with a linker arm to a domain of lesser mass. These subunits constitute an array with a lattice constant of 12.5 nm containing a major tetragon at one 4-fold axis of symmetry and a minor tetragon at the second 4-fold axis of symmetry (9, 65). The major tetragonal core of the array is composed of the heavy mass domains of four subunits, contains a large depression, and is located toward the inside of the layer (9). The minor tetragon which is composed of the lesser mass subunits of four subunits provides connectivity within the layer, and is raised and located at the outer surface of the layer. This structure provides the surface of the layer with considerable three-dimensional architecture which may facilitate its manifold binding activities, especially since as a monomer A-protein appears to lack binding ability (14, 15).

The present findings support these morphological findings by providing the first structural evidence for a two domain structure for A-protein. The COOH-terminal end of the isolated molecule is accessible to trypsin but capable of forming a 157-residue domain-like structure of M, 16,680 with intermediate resistance to trypsin. Analogy with flagellin suggests that this α-helix-containing region of the protein may be involved in interactions with adjacent monomers upon polymerization into the tetragonal A-layer array. The NH2-terminal region of the molecule is also capable of forming a domain-like structure of M, 37,500. Residues at the COOH-terminal end of this M, 37,500 peptide are accessible to trypsin in the native A-protein, but the M, 29,508–274 residue sequence at the NH2-terminal end of the native protein appears to be totally inaccessible to trypsin. It is tempting therefore to speculate that this region comprises the inner heavy mass core forming domain of A-layer.

Acknowledgments—Sharon Cubbage, Sandy Kieland, and Patricia Guerry are thanked for their assistance and advice.

REFERENCES
A. salmonicida A-protein Sequence

**Supplemental Material**

Structure of the tetragonal surface virulence array protein and gene of A. salmonicida.

Shijian Chen, Sonja Cavanagh, Jamiee Freiheit, Barry M. Phipps, Magdalena Kosteyszka, William W. Kay, and Trevor J. Trust

**Experimental Procedures**

**Bacterial strains, vectors, and growth conditions.** A layer producing A. salmonicida strains from the culture collection of J. Trust were grown at 30°C, while E. coli strains (RBSn for pFG2 and pUC recombinants), HBSn (for pLA927 recombinants), and Y1169 and 1196 (for p1, pG1, and M13 recombinants) were grown at 30°C and 27°C in Luria or on L-agar plates. When appropriate, media were supplemented with blood (20%), while E. coli XLI-M15 (for M13 recombinants) was maintained on minimal agar supplemented with 250 μg/ml of tetracycline. Phage of the M13 type, termed pM13mp8, and phage of the bl22 and bl23 series have been previously described (20, 21, 22). Plasmids were maintained in medium supplemented with ampicillin (50 μg/ml).

**DNA techniques.** Chromosomal DNA was isolated as previously described (23, 24). To construct a somatic library, A. salmonicida A450 DNA was partially digested with SalI, fractionated by 15% sucrose density gradient centrifugation, the 22-29 kb DNA fraction was ligated into the Smal site of pUB404, packaged into λ phage (Restrunggco), transformed into E. coli (HB101) (25), and transformants selected at 30°C and 37°C on tetracycline. Lysozyme columns on acid-precipitated filters were reacted with affinity purified anti-A. salmonicida A450 A-protein. Single and double strand sequencing was performed using the dideoxy method (25) with Sequenase (United States Biochemical Corp., Cleveland, OH) and dNTP and ddNTP mixtures. The sequencing reaction was denatured by boiling, and the DNA strands were separated by electrophoresis through 6% or 7.8% polyacrylamide gels. Peptides produced by CNBr hydrolysis and TPCK-trypsin digestion, the N-terminal 274 residue region of A. salmonicida A-protein, and the A-protein was subjected to radioiodination. Amersham's N56 plus 125I-Iodogen (Amersham Corp., Arlington Heights, IL) was used to prepare the probe. The probe was then hybridized with the labelled probe. Hybridization with labeled DNA was at 65°C for 9 h and hybridization in 3× SSC was at 65°C for 12 h. Bound probe was then hybridized using the protocols and procedures of the kit manufacturer. All other procedures were performed as described by Sambroak et al. (25).

**Protein and peptide techniques.** A-protein from A. salmonicida A450 was purified by ion exchange purification methods. Small scale ion exchange was purified using the previously described chromatography procedure (24). CNBr cleavage was performed by the procedure of Gross (27), with a 100 mM EDTA with respect to methionine of CNBr. The cleavage was then reprotonated with TCEP and subjected to trypsin digestion. The reaction was stopped by addition of 125I-Iodogen and 0.1% SDS. Samples were incubated at 29°C for 9 h. To determine the tryptic digestion, tryptic peptides were separated by gel electrophoresis and radioactivity was measured using a 2-D gel image analyzer to determine the reaction. When using the peptide sequence was determined by Edman degradation using the Applied Biosystems, F-9000 (for M13 recombinants). 10 μl TCEP (0.1% TCEP) (pH 7.8) was used in the reaction. The reaction was stopped by adding 0.125M EDTA and 25mM phenylmethylsulfonyl fluoride and boiling in electrophoresis sample buffer. N-terminal sequence analysis was performed by using TCEC-trypsin treatment at 37°C with weight with Lufen (100, 1, 500, 1, 1,000, 1, 2,000) as described above. SDS PAGE in either 12.5% or 16% polyacrylamide was performed by the method of Laemmli (26). When reported, separated peptides or proteins were electroeluted to microcrystals paper for immunological detection with a 1:20,000 dilution of affinity purified anti-A. salmonicida A450 A-protein rabbit immunoglobulin. (27) by the method of Towbin et al. (28) (previously described (29, 30).) or the Immobilon (Millipore Corp., Bedford, MA) for N-terminal sequence analysis, following the method of LeGendre and Maehama (31). Peptides were stained with Coomassie blue, cut from the Immobilon membrane (Millipore), and amino acid sequencing was performed as previously described (31).

**Cell fractionation.** Cells were collected from 12 h shaker cultures, treated with centrifugation at 4,000 g for 30 min. The supernatant was concentrated by evaporation and used in the culture supernatant fraction. The cells were sonicated by passage through a French pressure cell, and the cell envelope, outer membrane, cytoplasmic membrane and cytoplasmic fractions were prepared as previously described (30, 32). Each step was performed to isolate the periplasmic fraction. The first was the cationic shock method of Willis et al. (33). To ensure that the material collected in this procedure was not contaminated by small membrane particles, the cationic shock was centrifuged at 200,000 g for 60 min, and the supernatant was retained as the periplasmic protein fraction. The second method was the chloroform shock method of Anes et al. (33).

**Results**

![FIG. 1](https://example.com/f1.png)

**FIG. 1.** Physical map of the cloned A. salmonicida chromosomal DNA containing the A-protein gene (A450) (I. Brown, W. Kay, and T. Trust). The arrows indicate the DNA fragments. The A-protein gene is indicated, and vectors (labeled box) carrying A. salmonicida DNA are identified, as are the restriction sites used for subcloning. The recombinant plasmids designated are shown to the right. The identical sites in the deleted genes of strain A449 and the complete gene of strain A450 is essential to facilitate comparison. Asc = AscI; Bam = BamHI; Bgl = BglII; Eco = EcoRI; Hin = HindIII; Kpn = KpnI; Pvu = PvuII; Sal = SalI; Sau = Sau3A, and Spa = SpaI.

![FIG. 5](https://example.com/f5.png)

**FIG. 5.** (A) Predicted secondary structure of A. salmonicida A450 antigen based on combined Chou-Fasman (44) and Robes-Garnier (47) methods. (B) Map of A-protein showing location of major peptides produced by CNBr hydrolysis and TPCK-trypsin digestion. The tryptic resistant N-terminal 374 residue region of A-protein is shaded. The C-terminal CNBr peptide is cross-hatched. The location of tyrosine residues (●) and histidine residues (●) which are reactive to indole-radiolysis when cells are surface labeled with [35S]sulfate (11) are indicated.

**25 kb →**

1 2 3 4 5 6 7 8 9 10 11 12 13
A. salmonicida A-protein Sequence

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Codon usage A. salmon A. salmon codon usage</th>
<th>% codon usage A. salmon A. salmon (X)</th>
<th>Amino acid</th>
<th>Codon usage A. salmon A. salmon codon usage</th>
<th>% codon usage A. salmon A. salmon (X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe</td>
<td>UUU b</td>
<td>10 56.21</td>
<td>Ala</td>
<td>GCA b</td>
<td>10 16.50</td>
</tr>
<tr>
<td></td>
<td>UUG b</td>
<td>18 64.79</td>
<td></td>
<td>GCU b</td>
<td>4 6.24</td>
</tr>
<tr>
<td>Leu</td>
<td>UUA b</td>
<td>1 2.43</td>
<td>Thr</td>
<td>GCU b</td>
<td>3 6.20</td>
</tr>
<tr>
<td></td>
<td>UUG b</td>
<td>8 37.72</td>
<td></td>
<td>GUU b</td>
<td>3 6.20</td>
</tr>
<tr>
<td></td>
<td>CUG b</td>
<td>5 11.42</td>
<td>Val</td>
<td>GCA b</td>
<td>2 5.20</td>
</tr>
<tr>
<td></td>
<td>CUU b</td>
<td>4 9.26</td>
<td></td>
<td>GAC b</td>
<td>2 5.20</td>
</tr>
<tr>
<td></td>
<td>UCA b</td>
<td>0 0.26</td>
<td></td>
<td>GUC b</td>
<td>0 0.26</td>
</tr>
<tr>
<td></td>
<td>CUG b</td>
<td>26 61.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>AAU w</td>
<td>11 25.43</td>
<td></td>
<td>GAA b</td>
<td>8 1.82</td>
</tr>
<tr>
<td></td>
<td>AUC w</td>
<td>15 38.95</td>
<td></td>
<td>GAG b</td>
<td>8 1.82</td>
</tr>
<tr>
<td></td>
<td>AUA w</td>
<td>0 0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>AUG</td>
<td>3 7.60</td>
<td>Lys</td>
<td>AAA b</td>
<td>8 1.82</td>
</tr>
<tr>
<td></td>
<td>GUA b</td>
<td>3 7.60</td>
<td></td>
<td>GAG w</td>
<td>8 1.82</td>
</tr>
<tr>
<td></td>
<td>GUG b</td>
<td>5 12.47</td>
<td></td>
<td>GUC b</td>
<td>5 1.82</td>
</tr>
<tr>
<td>Ser</td>
<td>UCU w</td>
<td>11 25.43</td>
<td></td>
<td>GAU w</td>
<td>8 1.82</td>
</tr>
<tr>
<td></td>
<td>UCC w</td>
<td>3 6.38</td>
<td></td>
<td>GAC w</td>
<td>8 1.82</td>
</tr>
<tr>
<td></td>
<td>UCA w</td>
<td>0 0.25</td>
<td></td>
<td>GGG w</td>
<td>8 1.82</td>
</tr>
<tr>
<td></td>
<td>AGU w</td>
<td>1 2.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGC w</td>
<td>1 2.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>CUC w</td>
<td>4 10.25</td>
<td></td>
<td>GCC w</td>
<td>3 6.20</td>
</tr>
<tr>
<td></td>
<td>CUC w</td>
<td>1 10.25</td>
<td></td>
<td>GCG w</td>
<td>3 6.20</td>
</tr>
<tr>
<td></td>
<td>CCA w</td>
<td>1 10.25</td>
<td></td>
<td>GCG w</td>
<td>3 6.20</td>
</tr>
<tr>
<td></td>
<td>GCG w</td>
<td>8 26.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>ACC w</td>
<td>25 62.05</td>
<td></td>
<td>GCC w</td>
<td>3 6.20</td>
</tr>
<tr>
<td></td>
<td>AAC w</td>
<td>1 10.25</td>
<td></td>
<td>GCG w</td>
<td>3 6.20</td>
</tr>
<tr>
<td></td>
<td>ACG w</td>
<td>1 10.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>GCU w</td>
<td>42 86.11</td>
<td></td>
<td>GCG w</td>
<td>7 1.82</td>
</tr>
<tr>
<td></td>
<td>GGU w</td>
<td>8 16.80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Data from (37-39), b codons corresponding to the most abundant sRNA species in E. coli (41). c rare codons whose corresponding sRNA species in E. coli occur with an abundance of 0.5 or less on a scale of 0 to 1.0 (41); d whose percentage usage is approximately 10% or less (41).

TABLE II: Measured and predicted amino acid composition of mature A. salmonicida A-protein.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>No. of residues</th>
<th>Measured</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asx</td>
<td>66</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>66</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>23</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>37</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>12</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>57</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>63</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>23</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>59</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>25</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>24</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>16</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Top</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Residues/mol 479 457
Mr 45.1 45.2
Hydrophobic 43.2 43.1
Polar residues (%) (G, E, T, C, Y, N, Q) 37.2 35.5
Acidic residues (%) (D, E) basic 8.9 8.9
pl 5.76 4.79

* Data from (7)