A comparison of the amino acid sequence of the serine protease of the fish pathogen *Aeromonas salmonicida* subsp. *salmonicida* with those of other subtilisin-type enzymes relative to their substrate-binding sites

GEOFFREY COLEMAN* and PAUL W. WHITBY

Department of Biochemistry, Nottingham University Medical School, Queen's Medical Centre, Nottingham NG7 2UH, UK

(Received 23 July 1992; accepted 19 October 1992)

The amino acid sequence of the so-called 70 kDa (actually 64 kDa) serine protease secreted by the Gram-negative fish pathogen *Aeromonas salmonicida* has been determined. It shows a high degree of homology with the complete sequence of other bacterial serine proteases which, with molecular masses of approximately 30 kDa, are less than half its size. This homology is particularly marked in regions adjacent to the catalytic triad Asp32, His64 and Ser221 of subtilisin BPN'. Significant features of the *A. salmonicida* enzyme, a new member of the group of cysteine-containing subtilisin-type serine proteases, are the presence of six cysteine residues in the mature enzyme, a 37 amino acid extension at the N-terminus and 215 amino acids at the C-terminus when compared with subtilisin BPN'. In addition to a number of smaller peptide insertions there is a non-aligned 32 amino acid sequence in a position corresponding to its introduction between Lys213 and Tyr214 of subtilisin BPN'. This sequence is highly hydrophilic, with Asp/Asn accounting for 10 of the 32 amino acids. Further, the possession of two Cys residues separated by 24 amino acids provides the capacity for stabilizing the peptide as an externalized loop.

Introduction

Furunculosis is a disease of salmonid fish caused by the Gram-negative bacterium *Aeromonas salmonicida*. A characteristic of the disease is the appearance of furuncles in the form of elongated swellings along the side of the infected fish which contain liquefied muscle tissue. This liquefaction is caused by a serine protease secreted by all typical strains of *A. salmonicida* subsp. *salmonicida* (Fyfe *et al.*, 1986, 1987), in which it is considered to be one of the two most important extracellular virulence factors (Ellis, 1991).

A number of other bacterial species have been shown to produce extracellular serine proteases, of which the best known is subtilisin BPN', secreted by *Bacillus amyloliquefaciens*, for which a three-dimensional model is available showing the arrangement of all the amino acids (Wells & Estell, 1988). The substrate-binding site consisting of a triad of amino acids has been characterized and similarities have been identified between all the

bacterial serine proteases studied (Siezen *et al.*, 1991), the majority of which have molecular masses in the region of 30 kDa, considerably less than the estimated 70 kDa, based on SDS-PAGE data, of the *A. salmonicida* enzyme. In view of this dissimilarity it was of interest to carry out sequence studies in order to compare the arrangement of amino acids, particularly in the region of the catalytic triad, with those of the other bacterial serine proteases to determine how the extra amino acids are accommodated and how the substrate-binding site might be affected by the doubling of the amino acid chain length of the mature protein.

Methods

The predicted amino acid sequence of the serine protease of *A. salmonicida* was derived from a nucleotide sequence (EMBL accession no. X67043) (Whitby *et al.*, 1992). Sequence alignments were made using the CLUSTAL program.

Results and Discussion

The complete sequence of the mature *A. salmonicida* serine protease is shown in Fig. 1. With a molecular mass

* Author for correspondence. Tel. (0602) 709362; fax (0602) 422225.

Abbreviation: VR, variable region.

```

1
NESCTPLTGKEAGLDTGRSSAVRCLPGINPLQDL
      50
LNSGQNAFSPRGGMAGNDLNLWWAHRTEVLGQG
      74      100
INVAVVDGGLAIAHPDLADNVRPGSKNVVTGGSD
      113
PTPTDPDRCPRHVSVGIIAAVDNSIGTLGVAPRV
      150
QLQGFNLLDDNIQQLQKDWLYALGQRRHRRQPG
      200
QPELRMSLVDPGANGLDQVQLDRLFEQRTQHAS
AAYIKAAGTAFSRIAAGNYVAQPHRNLKLPFEN
      250
SNIDPSNSNFWNLVVRRAINADGVRSSYSVSGSNV
      300
FLSAPGGEGYGTADAPAMVTTDLPGCDMGYNRVDDP
      330
STNRLHNNPQLDASCDYNGVMNGTSSATPNTTGA
      350
MVLMAPYPDLSDVRDLRLDARNATRLDANQGPVQ
      400
INYTAANGERRQVTGLEGWERNAAGLWYSPSYGFG
LVDVNKTQPCSRQPRTAATTGAVALAKGKGNGRS
      450
PSAPSRVYVGSSPTRSSTQVDQPLTVEAVQVMVSL
      500
DHQRLPDLLIELVSPSGTRSVLLNPNNSLVGQSLDR
QQLGYVRTKGLRDMRMLSHKFYGEPAHG EWRLEVT
      500
DVANAAAQVSLLDRTNTRSTLTEGNSQPGQLLD
      597
WSRGYSVLGHDAARS*

```

Fig. 1. Amino acid sequence of the *A. salmonicida* mature serine protease. The numbers refer to the amino acids immediately below and the bold letters denote the amino acids corresponding to those of the catalytic triad of subtilisin BPN'. The 597 amino acids were identified from the nucleotide sequence of the *aspA* gene (Whitby *et al.*, 1992).

of 64173 Da it consists of 597 amino acids, compared with the 275 in subtilisin BPN' (Fig. 2). Thus, the 37 and 215 amino acid sequences Asn1–Ser37 and Glu382–Ser597, respectively, project beyond the regions of homology with the other serine proteases shown in Fig. 2. There are no obvious outstanding features of these projections and they bear no homology with any other sequences in the EMBL Database.

The alignment of amino acids Gly38 to Gly381 in the *A. salmonicida* serine protease (Fig. 1) with the complete sequence of subtilisin BPN' is shown in Fig. 2, compared with a number of other bacterial serine proteases of the subtilisin type (Kwon *et al.*, 1988). A high degree of homology can be seen with the smaller molecular species over the region of the catalytic triad of subtilisin BPN', with sequences being particularly well-conserved around the key amino acids Asp32, His64 and Ser221.

Sequence homology was observed not only for the catalytic triad but Ser125-Leu126-Gly127, which forms one wall of the active-site crevice of subtilisin BPN', is closely similar to the corresponding sequence Ser-Leu-Val in the *A. salmonicida* enzyme, whilst Ala152-Ala153-Gly154, which forms the opposite wall of the surface crevice of subtilisin BPN', is faithfully conserved (Robertus *et al.*, 1972). All the sequences shown in Fig. 2, excluding the *A. salmonicida* enzyme, contain Asn155 at the end of this latter conserved segment, which helps to stabilize the oxyanion generated in the tetrahedral transition state (Carter & Wells, 1990). However, in the 64 kDa protease the same conserved sequence is duplicated and Asn does, in fact, appear at the end of the duplicate sequence closer to the primary binding site, Ala160-Ala161-Gly162-Asn163. These observations suggest a similarity of tertiary structure with other serine proteases sufficient to allow the use of presently available three-dimensional models in the study of substrate specificity, i.e. the nature of the preferred amino acid side chains adjacent to the site of hydrolytic cleavage (Siezen *et al.*, 1991; Fersht, 1977).

It is interesting to note that replacement of Met222 with one of the non-oxidative residues Ala, Ser or Leu led to mutant subtilisin BPN' with activity reduced by 47–88% but with complete resistance to oxidation (Estell *et al.*, 1985). The *A. salmonicida* enzyme is the only one shown in Fig. 2 in which Met222 is replaced, the replacement being Ser.

The most significant features of the *A. salmonicida* enzyme revealed from optimal alignment of conserved sequences are non-aligned 4, 6, 14 and 32 amino acid peptide sequences between Gln185–Arg186, Val143–Ala144, Ala179–Val180 and Lys213–Tyr214 of subtilisin BPN', respectively. Such features are not uncommon among the serine proteases, in which they exist as externalized peptide loops connecting α -helices and β -sheets, referred to as variable regions (VRs) which may vary in length from a single residue to 151 residues as in *Lactococcus lactis* SK11 cell wall protease (Estell *et al.*, 1985). The structural non-equivalence of these loops may result from amino acid additions or deletions or they may result from loop flexibility or thermal motion (Gros *et al.*, 1990).

However, the novelty of the *A. salmonicida* protease, as shown in Fig. 3, compared with other VR-containing bacterial proteases of both Gram-positive and Gram-negative organisms (Siezen *et al.*, 1991) is not that its 32 residue VR is a somewhat larger continuous sequence than those which appear in other bacterial serine proteases in the same position in relation to the α -helix of the primary substrate binding site and the adjacent extended β -sheet structure but, more significantly, it contains two cysteine residues separated by a 24 amino

```

AQ:      ATQSPAPWGLDRI DQRDLPLSNSYTYTATQRVNVYVI DTQI RTTHREF-Q----QRA
PR:      AAQTNPWGLARI SSTSPSTSTYYYDESAQQQSCVYVI DTQI EASHPEF-E----QRA
TH:      YTPNDPYFSS-RQYQPQKI Q-----APQAW-DIAEQSQAKI AI VDTQVQSNHPDLAQKVY-QQW
BA:      A-QS-VPYQYSQIK-----APALHSQQYTQSNYKVAVI DSQI DSSHPDL--KVA-QGA
AM:      A-QS-VPYQISQIK-----APALHSQQYTQSNYKVAVI DSQI DSSHPDL--NVR-QGA
CA:      A-QT-VPYQIPLIK-----ADKYQAQQYKQANVKVAVLDTQI QASHPDL--NVV-QGA
DY:      A-QT-VPYQIPLIK-----ADKYQAQQYKQANVKVGI DTQI AASHTDL--KVY-QGA
AS:      -----GQNAFSPQGMAGNDLNLW-WA-HRTEVLQQQI NVAVYDDQLAI AHPDLADNVRPQSK
              10          20          30          40

AQ:      R-VGYDALQGNQ-D-CNQHQTHVAQTQ-----QVTYQVAKAVNLYAVRYLDCNQSQSTQ
PR:      QMVKTYYYSS--R-D-QNGQHQTHCAQTVQ-----S-QRTYQVAKKTQLFQVKYLDQNQSQQYST
TH:      DFVNDSTP---Q-N-QNGQHQTHCAQIAAVTNNSTQI -QATAPKASI LAVRYLDNQSQSTWTA
BA:      SMVPSETP--NFQ-D-DNSQHQTHVAQTVAAL-NSQI QV-LQVAPSSALYAVKYLDQAQSQQYSW
AM:      SFVPSETP--NYQ-D-QSSQHQTHVAQTIAAL-NSQI QV-LQVAPSSALYAVKYLDSTQSQQYSW
CA:      SFVAQEAY--N-T-D-QNGQHQTHVAQTVAAL-DNTTQV-LQVAPSVSLYAVKVLNQSQSQYSQ
DY:      SFVSQESY--N-T-D-QNGQHQTHVAQTVAAL-DNTTQV-LQVAPNVSLYAI KYLNQSQSQSTYSA
AS:      NVVTQSQSDPTPTDQRCPRHSYS--QI IAAV-DNSIQ-TLQVAPRVQLQFNLLQDNQI QQLQKD
              50          60          70          80          90          100

AQ:      VIAGVDWY-T-RN-HRRPA---VAN-MSLQQQV-STALDNQV--KNSI-----AAGVYVYAA
PR:      IIAQMDFYASDKN-NRNCQKQVYAS-LSLQQQYSSSYNSAAA--RLQ-----SSQVMYVYAA
TH:      VANGITY-AAD-Q-QAK-----VIS-LSLQQQTVQNSQLQQAV--NYAW-----NKQSVVYAAA
BA:      IINGIEW-AIA-N-NMD-----VIN-MSLQQQPSQSAALKAQV--DKAV-----ASQVVYVYAAA
AM:      IINGIEW-AIS-N-NMD-----VIN-MSLQQQPSQSTALKTYV--DKAV-----SSQI VVYVYAAA
CA:      IYSQIEW-ATT-N-QMD-----VIN-MSLQQQASQSTAMKQAV--DNAY-----ARQVVYVYAAA
DY:      IYSQIEW-ATQ-N-QLD-----VIN-MSLQQQPSQSTALKQAV--DKAY-----ASQI VVYVYAAA
AS:      WLYALGQRRHRRQPQLQ-----PELRMSLYDQPEQANQLQ-VQLDRLFEQRTQHASAAYIK-AA
              110          120          130          140          150

AQ:      QNDN-ANAC-NYS--PARVAEALTVQA-----TTSSDA---RASFSNYQSCVDL
PR:      QNNN-ADAR-NYS--PASEPSYCTVQA-----SDRYDR---RASFSNYQSVLDI
TH:      QNAQ-NTAP-NY--PAYYSNAIQAVS-----TDQNDN---KSSFSTYQSVYDY
BA:      QNEQ-STQSSSTVQYPQKYPSVIQAVQA-----VDSSNQ---RASFSQSVQPELDY
AM:      QNEQ-SSQSSSTVQYPQKYPSQTIQAVQA-----VNSSNQ---RASFSQSAQSELDY
CA:      QNSQ-NSQSTNTIQYPQKYDSVIQAVQA-----VDSNSN---RASFSQSVQAELVY
DY:      QNSQ-SSQSQNTIQYPQKYDSVIQAVQA-----VDSNKN---RASFSQSVQAELVY
AS:      QTAQFSRI AAGNYVAQPHRNLQPK-LPFENSNI DPSNSNFWNLVYRAI NADQVRSQSSVQSNVFL
              160          170          180          190

AQ:      FQPQASI PSAWYTSQ-T-----AT-QLNQTSQMATPHY
PR:      FQPQTDQSL--WIQQ--S-----T-RSQTSQMATPHY
TH:      AAPQSWIYSTYPTQ--T-----Y-ASLSQTSQMATPHY
BA:      MAPQVSI QSTLPQN--K-----Y-QAYNQTSQMASPHY
AM:      MAPQVSI QSTLPQQ--T-----Y-QAYNQTSQMATPHY
CA:      MAPQAGVYSTYPTN--T-----Y-ATLNQTSQMASPHY
DY:      MAPQSVSYSTYPSN--T-----Y-TSLNQTSQMASPHY
AS:      SAPQQEY-QTDAPAMVTTDLQCDQMYNRYDDPSTNRLHNNQLQASCDYNGYMNQTSATPNT
              200          210          220

AQ:      AQVAAALYLEQNPSATPAS-VASAILQ-ATT-QRLSQI--QSQSPNRLLYSLSSQSQ
PR:      AQLAAYLMTLQKTTA-ASACR-YIADT-ANK-QDLSNI--PFQTVN-LAYNNYQA
TH:      AQVAGLLASQQRS---ASNI RAAIQENT-ADKI Q-TQTYWAKQRVN--AYKAVQY
BA:      AQAAALILSKHPNWT-NTQVRSQSLQNT-TTKLQD-SFYQ-QKGLINQVQA--AAQ
AM:      AQAAALILSKHPTWT-NAQVQDRLEST-ATYLQD-SFYQ-QKGLINQVQA--AAQ
CA:      AQAAALILSKHPNLS-ASQVQRNRLSST-ATYLQS-SFYQ-QKGLINQVEA--AAQ
DY:      AQAAALILSKYPTLS-ASQVQRNRLSST-ATNLQD-SFYQ-QKGLINQVEA--AAQ
AS:      TQAMVLMAPY-PQLS----VQRD-LQDLLARNATRL-DANQQPVQI NYTA--ANQ----
              230          240          250          260          270

```

Fig. 2. Comparison of the amino acid sequence of *A. salmonicida* serine protease with those of other subtilisin-type serine proteases. The sequence of the *A. salmonicida* enzyme (AS) is aligned with those of aqualysin 1 (AQ), proteinase K (PR), thermitase (TH), subtilisin BPN' (BA), subtilisin Amylosacchariticus (AM), subtilisin Carlsberg (CA) and subtilisin DY (DY). The numbering refers to the sequence of subtilisin BPN', and the active-site residues D32, H64 and S221 are indicated by asterisks. Identical amino acids between the *A. salmonicida* enzyme and others are in bold letters (Kwon *et al.*, 1988).

Fig. 3. Comparison of variable regions adjacent to the primary substrate binding site, Ser221, of a number of serine proteases relative to subtilisin BPN'. The abbreviations are related to the following: *Bacillus amyloliquefaciens* subtilisin BPN' (BA), *Aeromonas salmonicida* serine protease (AS), *Enterococcus faecalis* cytolysin component A (EF), *Staphylococcus epidermidis* epidermin leader protease (SE) and *Serratia marcescens* extracellular serine protease (SM). The numerical positions of amino acids in subtilisin BPN' are indicated (Siezen *et al.*, 1991).

acid sequence which provide the capacity for loop stabilization by the formation of a disulphide bridge. Computer-generated Kyte–Doolittle hydrophathy plots (Kyte & Doolittle, 1982) over a 52 amino acid sequence, upstream from positions corresponding to amino acid 225 of the subtilisin BPN' sequence (see Fig. 2), of the enzymes compared in Fig. 3 showed that the *A. salmonicida* peptide sequence possessed by far the greatest hydrophobicity. This may be attributed to the fact that Asp/Asn account for 10 of the 32 residues of this VR. Two of the smaller VRs have even greater proportions of dibasic amino acids, thus, 3 of 6 are Glu/Gln and 6 of 14 are Asp/Asn. It is also of interest, and perhaps significant, that the peptide from which the DNA probe was constructed (Whitby *et al.*, 1992) appears in this 32 residue sequence – a reflection of accessibility?

The *A. salmonicida* enzyme is clearly a new member of the group of cysteine-containing subtilisin-type serine proteases, having six cysteine residues along its length with only three in the region of the catalytic domain, of

which two are in the hydrophilic peptide. The sequence shown in Fig. 2 accounts for 36 kDa of the 64 kDa molecular mass of the mature enzyme, of which 6 kDa is accountable to variable regions, the remainder being distributed 4 kDa at the N- and 24 kDa at the C-terminus.

It is interesting to note that whilst the enzyme is stabilized by disulphide bridges, with optimal activity at 50 °C and a pH of 9.5, its natural environment is in the cold, and the producing organism does not grow at temperatures above 25 °C. The preferred substrate of the enzyme is casein but in an infected animal it destroys musculature and produces furuncles (Finley, 1983).

In view of the role of this *A. salmonicida* enzyme as a virulence factor in furunculosis and its application as a vaccine constituent it is important that a structure–antigenic function picture be built up with the identification of key immunoaccessible epitopes; in this regard the hydrophilic 32 amino acid loop presents itself as one obvious candidate.

P. W. Whitby wishes to thank the Science & Engineering Research Council and the Scottish Office Agriculture & Fisheries Department for a studentship.

References

- CARTER, P. & WELLS, J. A. (1990). Functional interaction among catalytic residues in subtilisin BPN'. *Proteins Structure, Function and Genetics* **7**, 335–342.
- ELLIS, A. E. (1991). An appraisal of the extracellular toxins of *Aeromonas salmonicida* ssp. *salmonicida*. *Journal of Fish Diseases* **14**, 265–277.
- ESTELL, D. A., GRAYCAR, T. P. & WELLS, J. A. (1985). Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *Journal of Biological Chemistry* **260**, 6518–6521.
- FERSHT, A. (1977). *Enzyme Structure and Mechanism*, pp. 16–18. Reading & San Francisco: W. H. Freeman.
- FINLEY, A. (1983). *Studies on the extracellular protease of Aeromonas salmonicida*. PhD thesis, Nottingham University.
- FYFE, L., FINLEY, A., COLEMAN, G. & MUNRO, A. L. S. (1986). A study of the pathological effect of isolated *Aeromonas salmonicida* extracellular protease on Atlantic salmon, *Salmo salar* L. *Journal of Fish Diseases* **9**, 403–409.
- FYFE, L., COLEMAN, G. & MUNRO, A. L. S. (1987). Identification of major common extracellular proteins secreted by *Aeromonas salmonicida* strains isolated from diseased fish. *Applied and Environmental Microbiology* **53**, 722–726.
- GROS, P., VAN GUNSTEREN, W. F. & HOL, W. G. J. (1990). Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics. *Science* **249**, 1149–1152.
- KWON, S.-T., TERADA, I., MATSUZAWA, H. & OHTA, T. (1988). Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine protease) of *Thermus aquaticus* YT-1 and characteristics of the deduced primary structure of the enzyme. *European Journal of Biochemistry* **173**, 491–497.
- KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydrophobic character of a protein. *Journal of Molecular Biology* **157**, 105–132.
- ROBERTUS, J. D., ALDEN, R. A., BIRKTOFT, J. J., KRAUT, J., POWERS, J. C. & WILCOX, P. E. (1972). An X-ray crystallographic study of the binding peptide chloromethyl ketone inhibitors to subtilisin BPN'. *Biochemistry* **11**, 2439–2449.

- SIEZEN, R. J., DE VOS, W. M., LEUNISSEN, J. A. M. & DIJKSTRA, B. W. (1991). Homology modelling and protein engineering strategy of subtilisases, the family of subtilisin-like serine proteinases. *Protein Engineering* **4**, 719–737.
- WELLS, J. A. & ESTELL, D. A. (1988). Subtilisin – an enzyme designed to be engineered. *Trends in Biochemical Sciences* **13**, 291–297.
- WHITBY, P. W., LANDON, M. & COLEMAN, G. (1992). The cloning and nucleotide sequence of the serine protease (*aspA*) of *Aeromonas salmonicida* ssp. *salmonicida*. *FEMS Microbiology Letters* **99**, 65–72.