Characterisation of the Adhesive Proteins of a Major Fouling Organism, the Barnacle.

PART IV

Michael J. Naldrett

US Army RD & E Center, Kansas St,

Natick, MA 01760, USA

REPORT No 4

(25th March 1994)

US Army Contract Number "R&D 6791-BC-09"

CONCLUSIONS

Despite only finding the N-terminal sequences of two proteins so far, a 55 and a 36 kD protein, there also appear to be another protein of about 22 kD in *B. eburneus* cement. All other bands found on gels might come from any of these three sources or indeed from other yet undiscovered proteins. The key problem in analysing the cement is that there appears to be at least 2 different groups of protein which both have different solvent requirements. The 36 kD protein is quite easily dissolved in SDS but the 55 and the 22 kD proteins are not. The fact that *B. eburneus* cement is soluble in 70 % formic acid and 6 M guanidine hydrochloride shows that other denaturants can be used but it has also been noticed that the solubility window is very small and very specific depending on the cement being analysed. *B. crenatus* cement, for instance, dissolves only in SDS but can be pyridylethylated in guanidine thiocyanate whereupon it will dissolve. In this state it now becomes soluble in acetonitrile whilst even digested, pyridylethylated cement of *B. eburneus* remains insoluble in this solvent.

As the 55 kD protein has been difficult to blot onto a PVDF membrane further work with this protein has been slow compared to the work done on the 36 kD protein. Initial findings suggest that the differences in solubilty of these two proteins may be due to post-translational modification. So far at least one unknown has been encountered whilst sequencing the 55 kD protein and another possibly exists at residue 2 of fragment 2 from the 36 kD protein. This is hardly surprising since the vast majority of proteins undergo some covalent modification during their lifetime. However if these modifications are essential to the adhesive ability of the cement they require close attention.

Modified residues are not easily recognised even during analysis of complete amino acid sequences. This is because the majority of the naturally occurring derivatives are acid-labile and thus undetected by amino acid analysis after conventional acid hydrolysis. Unless alerted by some other factor, maybe weird spectral characteristics for example, the presence or nature of a post-translational modification may go unnoticed. Many derivatives are however detected by continual attention to anomalous phenylthiohydantoin (PTH) derivatives produced during the Edman chemistry used in protein sequencing or by attention to the continued anomalous behaviour of peptides produced during peptide mapping with enzymes. Certainly sequencing has started to show some anomalies and report 3 (Naldrett, 1994a) revealed that the 55 kD protein though containing MET residues was not suceptible to digestion with CNBr which is specific for these residues. As of yet too little data

exists on these unrecognised residues but the list of possible modified amino acids found in proteins to date is extensive (see Appendix). Probably the best way to characterise further those in the cement is by digesting the protein and subjecting a known peptide with the modification to analysis by mass spectrometry.

Another area liable to cause confusion in sequencing these peptides is the differences between methodologies used by different machines. As already stated, Edman degradation provides one of the best opportunities for locating unusual amino acids. However, harsh conditions are employed *eg* the protein is washed and cleaved each cycle using 100 % trifluoroacetic acid. You would therefore expect any amino acid which survives repetitive exposure to such extremes of degradation and which does not interfere with the Edman chemistry to be identifiable as its PTH derivative.

However, despite surviving these other extremes, poor or zero extraction of the ATZ complex from the reaction mixture may of course also prevent analysis of the amino acid. This may not be such a problem with newer systems using ethyl acetate as the solvent *eg* Beckman systems, but certainly in systems where 1-chlorobutane is used *eg* Applied Biosystems models, this is an area to be wary of. Machines using 1-chlorobutane are ideal for analysing small peptides *eg* pentapeptides because the solvent has a lower tendency to wash the peptide off the glass fibre supports than ethyl acetate does. However, if an oligosaccharide-linked residue exists this will not be soluble in 1-chlorobutane and so will not be injected onto the column. Further problems can be encountered with phospho-derivatives even when using radio-labels. for instance, the ^{32}P-labelled *O*-phosphoserine derivative ß-eliminates to yield poorly extracted inorganic [^{32}P]phosphate, the remainder of the serine is then extracted as the PTH of dehydroalanine. Unfortunately this can also be derived from unmodified serine (or unlabelled phospho-serine) or cysteine confusing the identification. One of the key pieces of work to be carried out here is the resequencing of the 55 kD protein on a Beckman sequencer using ethyl acetate as the solvent.

REFERENCES

Naldrett, M. J., 1993a. Characterisation of the adhesive proteins of a major fouling organism, the barnacle - Part I. US Army Contract number R&D 6791-BC-09.

Naldrett, M. J., 1993b. The importance of sulphur cross-links and hydrophobic interactions in the polymerization of barnacle cement. *Journal of the Marine Biological Association of the United Kingdom*, **73**, 689-702.

Naldrett, M. J., 1993c. Characterisation of the adhesive proteins of a major fouling organism, the barnacle - Part II. US Army Contract number R&D 6791-BC-09.

Naldrett, M. J., 1994a. Characterisation of the adhesive proteins of a major fouling organism, the barnacle - Part III. US Army Contract number R&D 6791-BC-09.

Schägger, H. & Von Jagow, G., 1987. Tricine-SDS PAGE for the separation of proteins in the range from 1 to 100KDa, *Analytical Biochemistry*, **166**, 368-379.

APPENDIX

Modifications of the common 20 amino acids which have been found in proteins

modified single α-amino acid

modification of the type:

naturally-occurring α-a.a. + naturally-occurring α-a.a.

naturally-occurring α-a.a. + α-a.a. linked through 1) sulphur 2) nitrogen 3) carbon. Both amino acids start off naturally-occurring but enzyme modification creates a bond which modifies one of the naturally-occurring amino acids.

L-Alanine
- *N*-Acetyl-
- *N*-Methyl-
- *N,N,N*-Trimethyl-
- Didehydro-

L-Arginine
- N^{ω}-(ADP-ribosyl)-
- N^{ω},N^{ω}-Dimethyl-
- $N^{\omega},N^{\omega'}$-Dimethyl-
- N^{ω}-Methyl-
- N^{ω}-Phospho-
- C^{G}-Deimino-C^{G}-oxo- (=Citrulline)
- N^{5}-Deamidino- (=Ornithine)

L-Aspartic acid
- *N*-Acetyl-
- 1-Amido-
- 3-Hydroxy-
- O^{4}-Methyl-

L-Asparagine
- N^{4}-Glycosyl-
- Iso- (=1-Amido-aspartic acid)

L-Cysteine
- 2-Dehydro-3-demercapto- (=didehydroalanine)
- *S*-Galactosyl-
- *S*-Glucosyl-
- *S*-(γ-Glutamyl)-
- *S*-(1-Glycero)-
- *S*-Mercapto-
- Seleno-

L-Glutamic acid
O^5-(ADP-ribosyl)$_n$-
1-Amido-
4-Carboxy-
O^5-Methyl-
Pyro-

L-Glutamine
1-Amido-

Glycine
N-Acetyl-
Amido-
N-Formyl-
N-Glucuronyl-

L-Histidine
1-Amido-
2-[(3-Carbamoyl-3-trimethylammonio)-propyl]-
2-[(3-Carbamoyl-3-trimethylammonio)-propyl]-τ-(ADP-ribosyl)-
4-Iodo-
π-Methyl-
π(τ)-Phospho-

L-Lysine
N^6-Acetyl-
N^6-(4-Amino-2-hydroxybutyl)-
N^6N^6-Dimethyl-
N^6-Fructosyl-
5-Glycosido (and oligoglycosido compounds)
5-Hydroxy-
5-Hydroxy-N^6,N^6,N^6-trimethyl-
N^6-Methyl-
N^6-Phospho-
N^6,N^6,N^6-Trimethyl-

L-Methionine
N-Acetyl-
1-Amido-
N-Formyl-
N-Methyl-

L-Phenylalanine
1-Amido-
3,4-Dihydroxy-
β-Glucosido-
β-Hydroxy-

L-Proline
- Amido-
- 4-Arabinosido-
- 3,4-Dihydroxy-
- 4-Galactosido-
- *N*,*N*-Dimethyl-
- 3-Hydroxy-
- 4-Hydroxy-
- 5-Oxo-

L-Serine
- O^3-Glycosyl-
- O^3-Phospho-
- O^3-(Glycosylphospho)-

L-Threonine
- O^3-Glycosyl-
- O^3-Phospho-

L-Tyrosine
- O^4-Adenylyl-
- 3,5-Dihalo-
- β-Glycosido-
- 3-Halo-
- O^4-Phospho-
- O^4-Sulfo-
- O^4-Uridylyl-
- 3-Hydroxy- (=DOPA)

Cross-links involving 2 of "the 20 amino acids"

- π-(3-Alaninyl)histidine
- τ-(3-Alaninyl)histidine
- *S*-(3-Alaninyl)cysteine
- N^6-(β-Aspartyl)lysine
- *S*-(*S*-Cysteinyl)cysteine (=cystine)
- 5-(*S*-Cysteinyl)glutamate
- 2-(*S*-Cysteinyl)histidine
- N^6-(*S*-Cysteinyl)lysine
- N^6-(γ-Glutamyl)lysine
- O^3-(γ-Glutamyl)serine
- 3-(O^4-Tyrosinyl)tyrosine (=isodityrosine)
- 3-(3-Tyrosinyl)tyrosine
- 3-[5-(3-Tyrosinyl)-3-tyrosinyl]tyrosine (=3,3'/5',3"-tertyrosine)

Di-(α-Amino Acids) where one is not one of "the 20"

1. Connected through a nitrogen atom

Lysinonorleucine
- Bis(6-norleucinyl)amine
- 5-Hydroxy-
- 5,5'-Dihydroxy-
- N^6:6'-Didehydro-
- N^6:6'-Didehydro-5-hydroxy-
- N^6:6'-Didehydro-5'-hydroxy-
- N^6:6'-Didehydro-5,5'-dihydroxy-

2. Connected through carbon-carbon bond

Di(allysine)
- Didehydro-
- Hydroxy-
- Dihydroxy-
- Merodesmosine
- Desmosine
- Isodesmosine

None of these di(α-amino acids) are connected through the α-amino or α-carboxyl groups and so either component can be a member of a conventional polypeptide.

ABSTRACT

The cement of *B. eburneus* can be dissolved and digested by using cyanogen bromide in 70 % formic acid. By careful evaporation of the reagents the resulting proteins and peptides can be redissolved in SDS and analysed by SDS-PAGE. Many of the darkest staining proteins come from the digestion of a 36 kD adhesive protein (AP). These bands are most easily semi-dry blotted from the gels onto polyvinylidene difluoride (PVDF) membranes from where they can be sequenced. Several other bands are not easily blotted and little data is available for these at the moment. Three fragments of the 36 kD protein have been sequenced: the N-terminus of the 36 kD protein NH_2-TYYPY LKTRH FGGID LTRY (F_5) which by amino acid analysis is shown to be tyrosine-rich; and two other fragments, F_3 & F_2 respectively, with sequence NH_2-LFPRL PLIVS KLRTY RFAP and NH_2-(?)(?)NLV APRR(?) VDFRP FYDRY. Comparing the amino acid analyses of these three fragments and one other for which we have no sequence shows that F_1 is ASX-, SER- and LEU-rich, F_5 is GLX-, PRO-, TYR-, CYS- and LEU-rich, F_2 is SER-, GLY- and LEU-rich, and F_3 is SER-, PRO- and LEU-rich. The high levels of LEU in all these peptides is reflected in the overall composition of the cement of *B. eburneus* showing that the majority of the cement is made of this 36 kD element. Two other fragments, F_4 and F_6, are not derived from the 36 kD protein and do not seem to be related to the previously analysed 55kD cement protein.

INTRODUCTION

The analysis of barnacle adhesives is complex. Earlier reports have shown that the adhesives of two so-called related species of barnacle behave very differently during their analysis. All the species of balanid barnacle looked at so far, with the exception of *B. crenatus*, have cements which behave very similarly when analysed by SDS-PAGE (Naldrett, 1993b). That the cement of *B. crenatus* behaves so differently to the cement of other balanid barnacles may provide helpful insight. It is possible that *B. crenatus* is a cement which contains weaker interactions making it easier to break down some or all of the matrix into its individual components. If this is the case then studying two species, one of which should be *B. crenatus*, should provide us with two perspectives on the same adhesive system.

The current work continues the search for sequence data of peptides produced by chemical digestion of the cement.

METHODS

SDS-PAGE

Electrophoresis was performed using a Novex mini gel system (Novex Corp, Encinitas, California) with precast gradient slab gels (10-20 % acrylamide; dimensions: 180 x 160 x 1 mm) and premade buffers. This tricine-SDS discontinuous buffer system is based on that described by Schägger & Von Jagow (1987). Sample buffer (2X) was made as follows: 3.0 M Tris-HCl, pH 8.45 (3.0 ml), glycerol (2.4 ml), SDS (0.8g), 0.1 % Serva Blue G (1.5 ml), 0.1 % Phenol Red (0.5 ml) and distilled water (to 10.0 ml). For reducing buffer 0.5 ml of 2-mercaptoethanol (2-ME) was added per 10 ml of sample buffer. In this series of experiments samples were always boiled in sample buffer for 2 to 5 min prior to electrophorsis. Following electrophoresis, the gels were electroblotted and then stained with 0.1 % Coomassie brilliant blue in a methanol, glacial acetic acid, water mix (5 : 1 : 5).

Protein sequencing

Proteins were wet-blotted onto polyvinylidene difluoride (PVDF) membrane (Immobilon™, Millipore) using a simple carbonate buffer made up as follows: 3mM sodium carbonate, 10 mM sodium bicarbonate, pH 9.9 in 10 % methanol. Blotting was carried out for 90-150 min at 30 volts constant voltage before the membranes were removed and rinsed extensively in deionised water. Protein bands were then stained with 0.1 % w/v Coomassie brilliant blue in 50 % (v/v) methanol and destained in 100 % methanol. Relevant bands were excised, dried and frozen in preparation for N-terminal sequencing. Sequence analyses were carried out by Mr T. Thannhauser of the Analytical and Synthetic Facility, Cornell University, NY, USA, on a Model 470A gas-phase sequencer and Model 120 PTH analyser fitted with the PTH C18 column: dimensions 220 mm x 2.1 mm (Applied Biosystems).

Amino acid analysis.

Amino acid analyses were carried out by Mr R Sherwood of the Analytical and Synthetic Facility, Cornell University, NY, USA using 6 N HCl hydrolysis and performic acid oxidation prior to hydrolysis. Samples of ground primary cement were subjected to gas-phase hydrolysis with 6 N HCl for 1 h at 150 °C according to the Waters Pico-Tag® Work Station manual (Waters, Milford, MA). Proteins blotted onto PVDF membrane could not be gas-phase hydrolysed. Instead these were hydrolysed by direct addition of propionic/HCl (Pierce). The addition of the organic acid helps thoroughly wet the hydrophobic membrane.

Performic acid oxidation of solid cements was carried out according to the following method. The samples were redried three times from 20 μl TFA and oxidised for ~60 min at 50°C in 50 μl of 4:1 88 % formic acid: hydrogen peroxide (v/v). The samples were dried and then redried from 20 μl water before carrying out the HCl hydrolysis.

Hydrolysed amino acids were derivatised with phenylisothiocyanate (PITC) to produce phenylthiocarbamyl (PTC) amino acids which were analysed and quantitated on a Pico-Tag® Work Station with a 30 cm Pico-Tag™ column (Waters, Milford, MA).

Protein digests using cyanogen bromide

The method for digesting cement proteins has been improved subtly since the recognition that the freeze-drying step was in fact making much of the digested material insoluble. If digests were completely dried the material would not even redissolve in the original solvent *eg* formic acid. This problem was overcome simply by drying the digested samples down to 10-20 μl before diluting and redrying. The samples were never allowed to reach absolute driness. Also the amount of CNBr added per milligram of protein was standardised to 5mg CNBr in 100 μl 70 % formic acid/0.125 mg protein (wet weight). The tubes were flushed with nitrogen and incubated in the dark for 24 h at room temperature. Samples were then diluted 10x with water and dried in a Savant speed vac with the above precautions.

RESULTS

For reference, part of the gel shown in Figure 8 of report 3 (Naldrett, 1994a) has been reproduced as a schematic (Figure 1) to illustrate which fragments of the CNBr digest have been analysed further. Fragments 1, 2, 3 and 5 are all derived from the 36 kD AP. The origin of fragments 4 and 6 is unknown but their amino acid compositions do not show such high levels of LEU, PRO or TYR as we might expect from peptides derived from the 36 kD AP. However, nothing about them is characteristic of the 55 kD AP which has high levels (> 10 %) of ALA, SER, GLY and THR. The amino acid analyses are given in Figures 2a & b along with analyses obtained in earlier works for comparison.

The following are N-terminal sequences of the fragments shown there.

Fragment 2

NH_2-(L)(S/F)NLV APRR(C*) VDFRP FYDRY

* this is most likely a CYS residue.

Fragment 3

NH_2-LFPRL PLIVS KLRTY RFAP

Fragment 5

NH_2-TYYPY LKTRH FGGID LTRY

This fragment is clearly the N-terminal sequence of the 36 kD AP which has been presented previously (Naldrett, 1994a). The sequence of the first 19 amino acids has now confirmed.

Fragment 4

NH_2-(Y/K)(N/N)(V/L)(N/Q)(T/Y)
(F/I)(N/L)(A/V)(E/H/R)(T/?)
(F/I)(F/L)(G/I)(L/T)(Q/?)

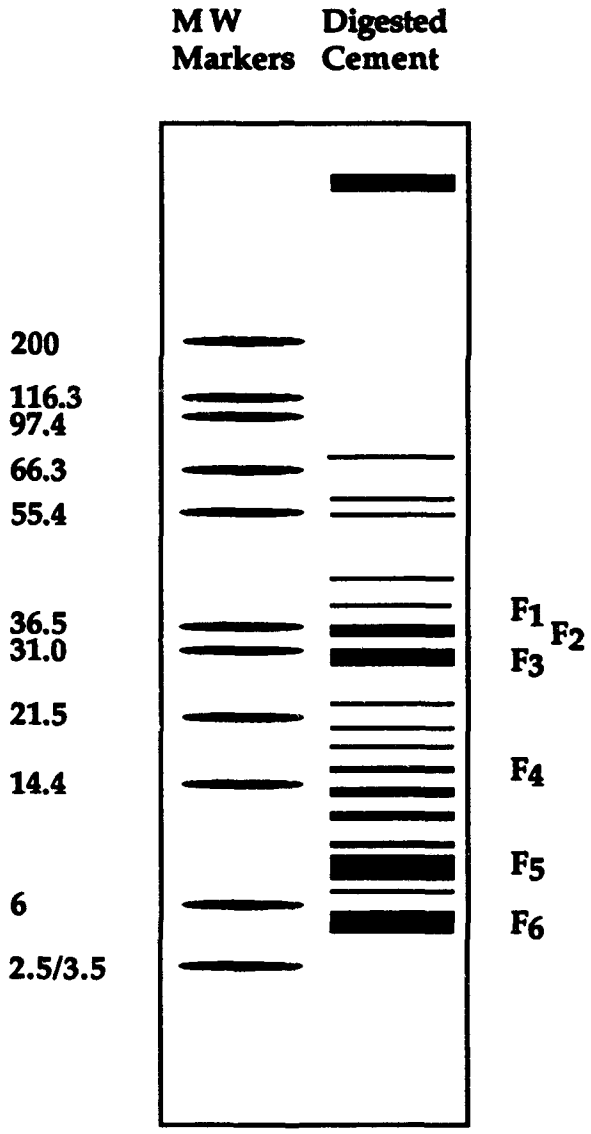

Figure 1: Schematic of the fragments produced on digestion of the cement of *Balanus eburneus* with cyanogen bromide. Taken from Figure 8 Report No 3 (Naldrett, 1994a)

Amino Acid	whole cement	36kD protein	Fragment 5	Fragment 3	Fragment 2	Fragment 1	Calc'd from seq.
ASX	5.6	5.37	7.06	7.25	7.68	12.40	6.56
GLX	6.24	4.37	9.42	5.37	6.59	7.60	1.64
SER	8.15	9.12	7.14	9.81	11.23	11.42	1.64
GLY	6.47	4.66	5.68	6.25	9.84	7.47	3.28
HIS	2.3	2.78	3.53	2.83	2.68	3.40	1.64
ARG	7.35	6.35	7.37	6.75	7.16	8.53	14.75
THR	5.58	6.75	5.00	5.54	5.48	6.24	6.56
ALA	4.84	3.67	3.48	5.59	5.62	4.52	3.28
PRO	8.44	9.44	11.90	9.70	8.31	7.88	11.48
TYR	7.39	10.04	11.21	5.91	5.98	6.59	13.11
VAL	6.09	4.6	5.46	3.66	4.34	2.74	4.92
MET	1.05	1.75	0.00	0.00	0.00	0.00	3.28
CYS	2.62	2.15	2.05	0.47	0.95	0.00	1.64
ILE	6.77	3.77	3.80	3.32	3.23	1.82	3.28
LEU	10.76	13.26	11.89	13.99	10.60	10.03	11.48
PHE	5.18	6.35	2.95	6.90	5.10	4.15	8.20
LYS	5.17	5.58	2.07	6.66	5.19	5.22	3.28

Figure 2a Amino acid compositions of the 4 digest fragments (F1, 2, 3 & 5) from the 36kD adhesive protein.

Amino Acid	Fragment 4	Fragment 6
ASX	9.62	12.03
GLX	8.96	9.13
SER	10.35	10.08
GLY	7.02	9.49
HIS	2.43	1.92
ARG	7.80	6.78
THR	4.98	3.65
ALA	5.77	3.23
PRO	6.29	6.78
TYR	5.56	7.41
VAL	4.94	5.55
MET	0.00	0.00
CYS	0.85	0.00
ILE	9.06	6.44
LEU	8.70	9.23
PHE	4.27	3.83
LYS	3.40	4.44

Figure 2b Amino acid compositions of two fragments (F4 & F6) from the digest of whole cement.

Fragment 4 appears as a hairline band on the gel from which it was blotted and yet sequencing reveals that there are two peptides present. Furthermore the two sequences cannot be separated from each other as the ratio of each protein present appears to be roughly the same. Since the double sequence starts in cycle one a single peptide has not been cleaved into two. The usual explanation is simply that two proteins with the same MW were not resolved by SDS-PAGE although a hetero-dimer could give this type of result, too.