Structural and Functional Studies of Protein Complexes

David J. Clarke

PhD Thesis
The University of Edinburgh
2005
Declaration

I, David J. Clarke, hereby certify that this thesis has been composed by myself, and that it is a record of my work, and that it has not been accepted in partial or complete fulfillment of any other degree or professional qualification.

David Clarke

University of Edinburgh

2005
Acknowledgements

I would like to acknowledge Dr Dominic Campopiano for the opportunity to carry out this PhD and I would like to thank him for his continued support and supervision.

I would also like to thank Prof. Robert Baxter and Dr Perdita Barran for their additional support, advice and guidance during the course of this work.

Without the following people this work would not have been completed-

Sally Shirran and Robert Smith who are responsible for mass spectrometry services with the Chemistry dept.

Dr. Julia Dorin (MRC Human Genetic Unit, Edinburgh)

Antimicrobial assays were performed in the CF lab under the guidance of Prof. John Govan (Medical Microbiology, University of Edinburgh) and with the help of Dr. Ross 'Desperados' Langley.

Finally, all members of Lab 229, past and present who have contributed to a friendly working environment especially Nick, Rachael, Paul, Ross and even Erik.
Abstract

We have characterised two biologically important protein complexes. Firstly, we have captured and analysed a complex of the biotin protein ligase (BPL) and the biotinylation domain (BCCPA67) of acetyl-CoA carboxylase (ACC) from the hyperthermophile *Aquifex aeolicus*. The genes encoding both BPL and BCCPA67 were overexpressed in *E. coli* and purified to homogeneity. Isolation of milligram amounts of both recombinant proteins allowed us to perform kinetic analysis of the BPL enzyme with its substrates using steady-state techniques. Furthermore, a chemically crosslinked complex of BPL and BCCPA67 was isolated and a comprehensive mutational study has identified a salt bridge between the two proteins which is important for heterodimerisation. The role of a conserved ‘thumb-like’ motif in BCCPA67 was also investigated by mutagenesis. Our results suggest an interaction between the biotin moiety and the ‘thumb’ reduces the propensity of BPL and holo-BCCPA67 to heterodimerise.

In a separate project we have investigated the relationship between the tertiary structure and biological activity of a novel β-defensin related peptide (Defr1). Defensins are cationic antimicrobial peptides which have a characteristic six cysteine motif and are important components of the innate immune system. Defr1 is a polymorphism of mouse β-defensin 8 and contains only 5 cysteines. Against a panel of pathogens, we found that oxidised synthetic Defr1 had significantly higher activity than its reduced form, and the oxidised and reduced forms of its six-cysteine containing analogue (Defr1 Y5C). Using non-denaturing gel electrophoresis and high
resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) we observed Defr1 and Defr1 Y5C dimers. Two complementary fragmentation techniques- collision induced dissociation (CID) and electron capture dissociation (ECD) - revealed that Defr1 Y5C dimers form by non-covalent, weak association of monomers which contain three intramolecular disulfide bonds. In contrast, Defr1 dimerisation is mediated by an intermolecular disulfide bond. Proteolysis and peptide mass-mapping revealed that Defr1 Y5C monomers fold with the canonical β-defensin disulfide bond connectivity whereas oxidised Defr1 is a complex mixture of dimeric isoforms with as yet unknown inter and intramolecular connectivities. We conclude that, compared with its six-cysteine analogue, the enhanced activity and stability of this mixture of Defr1 dimeric isoforms is due to the presence of an intermolecular disulfide bond.

In a related project, we have developed a method for the recombinant expression of defensins in *E. coli*. This method has been successfully used to express and purify wild-type and mutants of Defr1.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aa</td>
<td>Amino acid(s)</td>
</tr>
<tr>
<td>ACC</td>
<td>Acetyl-CoA carboxylase</td>
</tr>
<tr>
<td>AMP</td>
<td>Adenosine monophosphate</td>
</tr>
<tr>
<td>AMP</td>
<td>Antimicrobial peptide</td>
</tr>
<tr>
<td>ASL</td>
<td>Airway surface liquid</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>BC</td>
<td>Biotin carboxylase</td>
</tr>
<tr>
<td>BCCP</td>
<td>Biotin carboxyl carrier protein</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair(s)</td>
</tr>
<tr>
<td>BPL</td>
<td>Biotin protein ligase</td>
</tr>
<tr>
<td>CF</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CFTR</td>
<td>Cystic fibrosis transmembrane regulator</td>
</tr>
<tr>
<td>CFU</td>
<td>Colony forming units</td>
</tr>
<tr>
<td>CID</td>
<td>Collision-induced dissociation</td>
</tr>
<tr>
<td>CNBr</td>
<td>Cyanogen bromide</td>
</tr>
<tr>
<td>CO₂</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>CoA</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>CT</td>
<td>Carboxyltransferase</td>
</tr>
<tr>
<td>CTP</td>
<td>Cytidine 5'-triphosphate</td>
</tr>
<tr>
<td>Da</td>
<td>Daltons</td>
</tr>
<tr>
<td>DC</td>
<td>Dendritic cell</td>
</tr>
<tr>
<td>Defr1</td>
<td>Defensin-related peptide 1</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
</tbody>
</table>
Abbreviations

DNA - Deoxyribonucleic acid
DTT - Dithiothreitol
ECD - electron-capture dissociation
EDC - 1-ethyl-3-(dimethylamino-propyl)-carbodiimide
EDTA - Ethylene diaminetetracetic acid
ESI-MS - Electrospray ionization mass spectrometry
FAS - Fatty acid synthase
FT-ICR - Fourier transform ion cyclotron resonance
GTP - Guanosine 5'-triphosphate
HBD - Human β-defensin
HCS - Holocarboxylase synthase
HEPES - N-(2-Hydroxyethyl)piperazine-N’-(2-ethanesulfonic acid)
HPLC - High performance liquid chromatography
HNP - Human neutrophil peptide (α-defensin)
IL - Interleukin
IPTG - Isopropyl-1-thio-β-D-galactopyranoside
KSI - Ketosteroid isomerase
LB - Luria Bertani
PLG - Lysylphosphatidylglycerol
LPS - Lipopolysaccharide
MALDI - Matrix-associated laser desorption ionization
MBC - Minimum bactericidal concentration
NF-κB - Nuclear factor-κB
Abbreviations

ORF - Open reading frame
[O] - oxidised
PAGE - polyacrylamide gel electrophoresis
PCR - polymerase chain reaction
PMN - polymorphonuclear leukocytes
Q-TOF - quadrupole time of flight
[R] - reduced
RNA - ribonucleic acid
RP - reverse phase
Rpm - revolutions per minute
SORI - sustained off resonance irradiation
TCEP - tris carboxyethyl phosphine
TFA - Trifluoroacetic acid
TLR - Toll-like receptor
Tris - Tris [hydroxymethyl] aminomethane
TNF-α – Tumour necrosis factor-α
UTP - Uridine 5'-triphosphate
# Table of Contents

Declaration .................................................................................................................................. ii  
Acknowledgements .................................................................................................................... iii  
Abstract ....................................................................................................................................... iv  
Abbreviations ............................................................................................................................... vi  
Table of Contents ......................................................................................................................... ix  
List of Figures ............................................................................................................................... xvii  
List of Tables .................................................................................................................................. xx

## Chapter 1: Biotinylation: An Enzyme-Dependant Posttranslational Modification

1.1 The structure and properties of biotin .................................................................................. 2  
1.2 Biotin-dependant enzymes - The biotin enzyme family ..................................................... 3  
1.3 Acetyl-CoA Carboxylase (ACC) ......................................................................................... 5  
1.3.1 The Biological Role of Acetyl-CoA Carboxylase ............................................................ 5  
1.3.2 Heterodimeric Acetyl-CoA Carboxylases ...................................................................... 6  
1.3.3 The *E. coli* Biotin Carboxyl Carrier Protein ................................................................. 8  
1.3.4 Biotin Carboxylase (BC) ................................................................................................. 10  
1.3.4.1 BC Structure .............................................................................................................. 10  
1.3.4.2 BC Mechanism .......................................................................................................... 13  
1.3.5 Carboxyltransferase (CT) ............................................................................................... 15  
1.3.5.1 CT Structure .............................................................................................................. 15  
1.3.5.2 CT Mechanism .......................................................................................................... 17  
1.3.6 A Working Model of an ACC complex ........................................................................... 18
Table of Contents

1.4 Biotin Protein Ligase (BPL) .................................................................19
  1.4.1 The Biotinylation Reaction ...............................................................19
  1.4.2 Enzyme Classification .....................................................................21
  1.4.3 E. coli Biotin Protein Ligase (BirA) ..................................................22
    1.4.3.1 The Bifunctional Nature of BirA - Controlling Biotin Biosynthesis..22
    1.4.3.2 Structure of BirA .......................................................................23
    1.4.3.3 The BirA-BCCP Complex ............................................................25
  1.4.3 Aims ..............................................................................................27

Chapter 2: Enzymatic and Chemical Analysis of the BPL from Aquifex aeolicus.
......................................................................................................................28
  2.1 Analysis of the A. aeolicus genome .....................................................29
  2.2 Cloning, expression and purification of BPL ........................................31
  2.3 Cloning, expression and purification of BCCPΔ67 ...............................33
  2.4 Biological properties of BPL ...............................................................35
  2.5 Kinetic Analysis of BPL ......................................................................36
  2.6 Limited Proteolysis of BPL .................................................................39
  2.7 Discussion ..........................................................................................40

Chapter 3: Characterisation of the BPL:BCCP Complex from Aquifex aeolicus
......................................................................................................................44
  3.1 Chemical crosslinking of BPL and BCCP ............................................45
  3.2 Identification of Potential Residues Involved in BPL:BCCPΔ67
Crosslinking ...............................................................................................47
3.2.2 Crosslinking BCCPΔ67 Mutants with BPL ........................................ 51
3.2.3 Cloning, Expression and Purification of a bpl K19A mutant .......... 52
3.2.4 Chemical crosslinking BPL K19A with BCCPΔ67. ...................... 53

3.3 The Role of the "Thumb" Region of BCCPΔ67 in BPL:BCCPΔ67

3.3.1 Cloning, expression and purification of A. aeolicus AccBΔ67
"thumbless" mutants................................................................. 56
3.3.2 Chemical biotinylation and purification of BCCPΔ67ΔTh 4KR ...... 57
3.3.3 Chemical crosslinking BPL and apo- and holo- BCCPΔ67ΔTh 4KR 
58

3.4 Discussion.................................................................................. 59

Chapter 4: Defensins - Important Antimicrobial Peptides of Innate Immunity 63

4.1 The Diversity and Classification of Antimicrobial Peptides......... 64
4.2 Vertebrate Defensins .................................................................. 65
4.2.1 Classification .......................................................................... 65
4.2.2 Biological Role of Mammalian Antimicrobial Defensins - The Innate
Immune system............................................................................ 67
4.2.3 Defensin Genes, Biosynthetic Pathways, and Posttranslational
Processing...................................................................................... 69
4.2.4 Genetic Regulation of Defensins............................................ 70
4.2.5 Defensin Structure .................................................................. 72
4.2.6 Antimicrobial Activity of Defensins....................................... 75
4.2.7 Mechanism of Antimicrobial Activity................................. 76
4.2.8 AMP Resistance - Bacterial Countermeasures....................... 79
### Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.9 Other Biological Roles of Defensins</td>
<td>82</td>
</tr>
<tr>
<td>4.3 Defensins in human health</td>
<td>84</td>
</tr>
<tr>
<td>4.3.1 Lung Disease - Cystic Fibrosis</td>
<td>84</td>
</tr>
<tr>
<td>4.3.2 Dermatological Diseases - Psoriasis and Atopic Dermatitis</td>
<td>85</td>
</tr>
<tr>
<td>4.3.3 Inflammatory Bowel Diseases</td>
<td>86</td>
</tr>
<tr>
<td>4.4 Defensin-Related Peptide - Defr1</td>
<td>88</td>
</tr>
</tbody>
</table>

**Chapter 5: Structural and functional characterisation of defensin related peptide-1 (Defr1)**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Antimicrobial Properties of Defr1 and Defr1 YSC</td>
<td>90</td>
</tr>
<tr>
<td>5.1.1 Antimicrobial Activity of Defr1 and Defr1 YSC</td>
<td>90</td>
</tr>
<tr>
<td>5.1.2 Salt Sensitivity of Defr1 and Defr1 YSC</td>
<td>94</td>
</tr>
<tr>
<td>5.2 Structural properties of Defr1 and Defr1 Y5C</td>
<td>95</td>
</tr>
<tr>
<td>5.2.1 Native Gel Electrophoresis</td>
<td>95</td>
</tr>
<tr>
<td>5.2.2 Characterization of the nature of dimerisation of Defr1 and Defr1 Y5C using FT-ICR mass spectrometry</td>
<td>96</td>
</tr>
<tr>
<td>5.2.3 Dissociation of defensin dimers</td>
<td>99</td>
</tr>
<tr>
<td>5.2.4 Reverse-phase high-pressure liquid chromatography</td>
<td>103</td>
</tr>
<tr>
<td>5.2.5 Determination of Disulfide Connectivity of Defr1 and Defr1 Y5C</td>
<td>105</td>
</tr>
<tr>
<td>5.3 Discussion</td>
<td>112</td>
</tr>
</tbody>
</table>

**Chapter 6: Recombinant production of Defr1 and a Defr1 derivative**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Cloning the <em>Defr1</em> gene and its derivatives</td>
<td>115</td>
</tr>
<tr>
<td>6.1.1 Cloning the <em>Defr1</em> gene</td>
<td>115</td>
</tr>
<tr>
<td>6.1.2 Designing the <em>Defr1</em> lcys gene, a one-cysteine analogue of <em>Defr1</em></td>
<td>116</td>
</tr>
<tr>
<td>6.2 Overexpression and purification of KSI-defensin fusion protein</td>
<td>117</td>
</tr>
</tbody>
</table>
6.3 Cyanogen bromide cleavage of KSI-defensin fusion protein...........118
6.4 Deprotection and refolding recombinant defensins ..................119
  6.4.1 Deprotection and refolding Defr1.......................................119
  6.4.2 Deprotection and refolding Defr1 1cys ..............................120
6.5 Antimicrobial activity of recombinant defensins.....................121
  6.5.1 Antimicrobial activity of recombinant Defr1 ......................121
  6.5.2 Antimicrobial activity of recombinant Defr1 1cys ................121

Chapter 7: Materials & Methods ..................................................123

7.1 General materials......................................................................124
  7.1.1 General reagents .................................................................124
  7.1.2 Solutions and buffers ...........................................................124
  7.1.3 Media .................................................................................126
7.2 Molecular Biology – DNA manipulation ....................................127
  7.2.1 Bacterial cell lines ...............................................................127
  7.2.2 Oligonucleotide primers .......................................................128
  7.2.3 Storage of bacterial stocks .....................................................128
  7.2.4 Transformation of E. coli competent cells with recombinant plasmid DNA 129
  7.2.5 Preparation of plasmid DNA .................................................129
  7.2.6 Digestion of DNA with restriction endonucleases ................129
  7.2.7 Electrophoresis of DNA .......................................................129
  7.2.8 Purification of DNA .............................................................130
  7.2.9 Amplification and megaprimer mutagenesis of DNA ..........130
  7.2.10 Recursive PCR .................................................................130
Table of Contents

7.2.11 Direct cloning of PCR products .................................................131
7.2.12 DNA sequencing ........................................................................131
7.2.13 Cloning into plasmid vectors .......................................................132

7.3 Molecular Biology - Protein Expression & Purification ......................132

7.3.1 Polyacrylamide gel electrophoresis (PAGE) ..................................132
7.3.2 Large Scale Expression of *A. aeolicus* BPL and mutant ..............132
7.3.3 Purification of *A. aeolicus* BPL and mutant ..................................133
7.3.4 Large scale expression of His6-tagged *A. aeolicus* BCCPΔ67 and mutants 134
7.3.5 Purification of His6-tagged *A. aeolicus* BCCPΔ67 and mutants ....134
7.3.6 Overexpression of Defrl-KSI and Defrl 1Cys-KSI fusion protein as inclusion bodies .................................................................135
7.3.7 Purification of Defrl-KSI and Defrl 1Cys-KSI fusion protein ......136
7.3.8 Protection of KSI fusion proteins by sulfite addition ................137
7.3.9 Cleavage and separation of KSI fusion protein............................137
7.3.10 Deprotection and refolding Defrl ...............................................138
7.3.11 Deprotection and oxidation of Defrl 1Cys ................................138
7.3.12 Final purification of Defrl and Defrl 1Cys ..................................138

7.4 Protein Chemistry ............................................................................139

7.4.1 Assay of *A. aeolicus* BPL ..........................................................139
7.4.2 Limited proteolysis of *A. aeolicus* BPL .......................................140
7.4.3 Chemical crosslinking of *A. aeolicus* BPL and BCCPΔ67 mutants 140
7.4.4 Chemical Biotinylation ...............................................................141

7.5 Protein Characterisation ..................................................................141
Table of Contents

7.5.1 High performance liquid chromatography (HPLC) ........................................141
7.5.2 Liquid chromatography-mass spectrometry (LC-MS) ..................................141
7.5.3 FT-ICR Analysis.........................................................................................142
7.5.4 Protease digestion and peptide mass mapping of defensins ......................143
7.5.5 Collision induced dissociation (CID) of peptides ......................................144
7.6 Microbiology ...............................................................................................145
    7.6.1 Bacterial isolates....................................................................................145
    7.6.2 Preparation of defensin for antimicrobial assays.................................146
    7.6.3 Broth dilution antimicrobial assay of defensins ....................................146
    7.6.4 Time-kill assay of defensins.................................................................147
7.6.5 Salt sensitivity of defensins......................................................................147

Chapter 8: References ..........................................................................................148

Appendix 1: Recombinant proteins and synthetic peptides used in this study.173

Appendix 2: Publications ....................................................................................176


List of Figures

Chapter 1: Biotinylation - An Enzyme-Dependant Posttranslational Modification

1.1. D-(+)-biotin.
1.2. The carboxylation of acetyl-CoA.
1.3. The two-step reaction mechanism of AAC.
1.4. The NMR structures of both apo- and holo \textit{E. coli} BCCP.
1.5. The crystal structure of \textit{E. coli} BC subunit.
1.6. The chemical mechanism of BC.
1.7. The crystal structure of the CT domain of yeast ACC including bound CoA.
1.8. The biotinylation of BCCP.
1.9. Biotin Protein Ligase Classification.
1.10. The structure of the \textit{E. coli} BirA dimer with biotin bound at each active site.
1.11. A model of the \textit{E. coli} BirA: BCCP complex.

Chapter 2: Enzymatic and Chemical Analysis of the BPL from \textit{Aquifex aeolicus}.

2.1. Sequence alignments of BCCP and BPL from \textit{E. coli} and \textit{A. aeolicus}.
2.2. Expression vector pET28a/bpl.
2.3. Purification of \textit{A. aeolicus} BPL.
2.4. Expression vector pET6H/acchΔ67.
2.5. Purification of \textit{A. aeolicus} BCCPΔ67.
2.6. Biological properties of \textit{A. aeolicus} BPL.
2.7. Steady-state kinetic analysis of BPL substrate binding.

2.8. Proteolysis of BPL.

Chapter 3: Characterisation of the BPL:BCCP Complex from Aquifex aeolicus

3.1. SDS-PAGE analysis of chemical crosslinking assays.

3.2. Purification of the chemically crosslinked BPL:apo-BCCPΔ67 complex by size exclusion chromatography.

3.3. A model of the E. coli BirA:BCCP-87 complex.

3.4. SDS-PAGE analysis of purified BCCPΔ67 mutants.

3.5. Chemical crosslinking of BPL with BCCPΔ67 mutants.


3.7. Chemical crosslinking of BPL with apo- and holo-BCCPΔ67ΔTh4KR mutants.

Chapter 4: Defensins - Important Antimicrobial Peptides of Innate Immunity

4.1. The disulfide bridging connectivity in α, β, and θ-defensins.

4.2. Structure of prototypical defensin gene and peptide.

4.3. Sequence alignment of human β-defensins HBD1-3.

4.4. The crystal structures of representatives of α- and β-defensins.

4.5. The proposed mechanisms of the antibacterial activity of defensins.

4.6. Examples of lipid A from the Gram-negative bacterium Salmonella enteritica and Lipoteichoic acid from the Gram-positive bacterium Staphylococcus aureus.

4.7. Sequence of Defb8, Defr1 and Defr1 Y5C.
Chapter 5: Structural and functional characterisation of defensin related peptide-1 (Defr1).

5.1. Time-Kill Assay.

5.2. The effect of NaCl concentration on the antimicrobial activity of Defr1 and Defr1 Y5C.

5.3. Colloidal Coomassie-stained non-denaturing 16% Tricine gel of oxidized and reduced Defr1 and Defr1 Y5C.

5.4. Defr1 Y5C FT-ICR isotopic envelope.

5.5. Defr1 FT-ICR isotopic envelope.

5.6. Collision-induced dissociation (CID) spectra of Defr1 Y5C.

5.7. Collision-induced dissociation (CID) spectra of Defr1.

5.8. Electron capture dissociation (ECD) of Defr1.

5.9. HPLC analysis of the oxidised and reduced forms of Defr1 and Defr1 Y5C.

5.10. Proteolysis of Defr1 Y5C.

5.11. CID Spectrum of Defr1 Y5C tryptic fragment 1.

Chapter 6: Recombinant production of Defr1 and a Defr1 derivative.


6.2. Expression, purification and cleavage of KSI-defensin fusion peptide.

6.3. Cleavage and separation of Defr1-KSI.
List of Tables

Chapter 1: Biotinylation - An Enzyme-Dependant Posttranslational Modification
1.1. The biotin enzyme family.
1.2. Subunits of *E. coli* ACC.

Chapter 3: Characterisation of the BPL:BCCP Complex from *Aquifex aeolicus*
3.1. Analysis of BCCPΔ67 mutants by LC-MS.

Chapter 4: Defensins - Important Antimicrobial Peptides of Innate Immunity
4.1. Overview of antimicrobial peptides - Representatives from each class.

Chapter 5: Structural and functional characterisation of defensin related peptide-1 (Defr1).
5.1 Antimicrobial activity of Defr1 and Defr1 Y5C.
5.2. Determination of the disulfide bond connectivity of oxidized Defr1 Y5C.
5.3 Determination of the disulfide bond connectivity of oxidized Defr1.

Chapter 6: Recombinant production of Defr1 and a Defr1 derivative.
6.1. Refolding of recombinant Defr1.
6.2. Activity of recombinant Defr1 1cys.
Chapter 1: Biotinylation: An Enzyme-Dependant Posttranslational Modification
1.1 The structure and properties of biotin

Biotin (Vitamin H, Co-enzyme R) is a water-soluble B vitamin produced in small quantities by microorganisms and plants. Unlike bacteria, mammals are unable to synthesise the cofactor and thus rely on dietary intake to satisfy their biotin requirements.

Biotin was first isolated by Kögl and Tönnis in 1936 (1), who isolated 1.1 mg of the vitamin as its methyl ester from more than 500 pounds of egg yolk and its structure was later deduced by du Vigneaud et al. (2). X-ray analysis revealed that the molecule consists of an imidazolidone ring fused to a tetrahydrothiophene ring with a valeryl side chain. Of the eight optical active forms only one, D-(+)-biotin, is biologically active (Fig 1.1).

Today, biotin is widely used along with avidin and streptavidin, the specific biotin binding proteins, in the analysis biochemical phenomena. The chemical or enzymatic biotinylation of biomolecules allows easy detection/purification from complex mixtures using avidin affinity, and has great utility in many areas of scientific research. A structural and mechanistic understanding of the enzymes which carry out the post-translational addition of biotin to proteins may aid in the design of new regents to be used in biotechnology.

![Figure 1.1. D-(+)-biotin. The biologically active isoform of biotin.](image-url)
1.2 Biotin-dependant enzymes - The biotin enzyme family

In vivo biotin is utilised as a coenzyme in numerous essential enzymatic carboxylation reactions (3). The cofactor is covalently bound to these enzymes via an amide linkage between its valerate side chain and the ε-amino group on a specific lysine residue of the enzyme (4). The biotin prosthetic group then acts as a "swinging arm" and is used to transfer activated CO\(_2\) between active sites within the enzyme (5). The small group of biotin dependant enzymes share a remarkably similar catalytic mechanism. The first step involves fixation of CO\(_2\) by biotin, which occurs specifically at FN to form a carboxybiotin complex. This is followed by decarboxylation of carboxybiotin and proton transfer to different acceptor substrates (6).

Biotin dependant enzymes are involved in a diverse range of biochemical processes and can be divided into three classes (Table 1.1). Class I enzymes are ATP-dependant carboxylases which transfer CO\(_2\) from hydrogen carbonate to specific acceptors, and are exemplified by acetyl-CoA carboxylase (ACC, see section 1.3). Class II and III enzymes are transcarboxylases and decarboxylases respectively, and do not require ATP. In these enzymes the CO\(_2\) unit is transferred from either an acyl derivative or keto-ester to the enzyme before being transferred to either an acceptor or released as free CO\(_2\). To date, decarboxylases have only been found in anaerobic bacteria and employ the free energy of decarboxylation to pump Na\(^+\) across a membrane. The resulting Na\(^+\) gradient is then utilised for the synthesis of ATP (7,8).
<table>
<thead>
<tr>
<th>Class</th>
<th>Enzyme</th>
<th>Reaction</th>
<th>Biochemical Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Carboxylases</td>
<td>Pyruvate Carboxylase</td>
<td>Pyruvate to Oxaloacetate</td>
<td>Gluconeogenesis, lipogenesis</td>
</tr>
<tr>
<td></td>
<td>Acetyl-CoA Carboxylase</td>
<td>Acetyl-CoA to Malonyl-CoA</td>
<td>Fatty acid biosynthesis</td>
</tr>
<tr>
<td></td>
<td>Propionyl-CoA Carboxylase</td>
<td>Propionyl-CoA to Methylmalonyl-CoA</td>
<td>Propionate metabolism</td>
</tr>
<tr>
<td></td>
<td>3-Methylcrotonyl-CoA Carboxylase</td>
<td>3-Methylcrotonyl-CoA to 3-</td>
<td>Catabolism of leucine</td>
</tr>
<tr>
<td></td>
<td>Geranyl-CoA Carboxylase</td>
<td>Geranyl-CoA to Carboxygeranyl-CoA</td>
<td>Catabolism of isoprenoids (microbial)</td>
</tr>
<tr>
<td></td>
<td>Urea Carboxylase</td>
<td>Urea to N-carboxyurea</td>
<td>Catabolism of urea (microbial)</td>
</tr>
<tr>
<td>II. Transcarboxylase</td>
<td>Methylmalonyl-CoA Carboxyltransferase</td>
<td>Methylmalonyl-CoA and Pyruvate to oxaloacetate and Propionyl-CoA</td>
<td>Fermentation of carbohydrates to propionate (propionibacteria)</td>
</tr>
<tr>
<td>III. Decarboxylases</td>
<td>Methylmalonyl-CoA Decarboxylase</td>
<td>Methylmalonyl-CoA to Propionyl-CoA and CO2</td>
<td>Lactate fermentation (Micrococcus lactis)</td>
</tr>
<tr>
<td></td>
<td>Oxaloacetate Decarboxylase</td>
<td>Oxaloacetate to Pyruvate and CO2</td>
<td>Catabolism of citrate (Aerobacter aerogenes)</td>
</tr>
</tbody>
</table>

Table 1.1: The biotin enzyme family. The above table outlines biotin-dependent enzymes and the reactions they catalyse.
1.3 Acetyl-CoA Carboxylase (ACC)

1.3.1 The Biological Role of Acetyl-CoA Carboxylase

Acetyl-CoA carboxylase (ACC) catalyses the conversion of acetyl-CoA to malonyl-CoA, the first committed step of fatty acid synthesis (see Fig. 1.2) (9). The malonyl-CoA is then used as a unit for repeated rounds of condensation by fatty acid synthases (FAS).

Two physically distinct types of enzyme are found in nature. Heterodimeric ACCs, usually composed of four subunits, are found in prokaryotes; and homomeric ACCs consisting of a single large polypeptide are found in eukaryotes. The majority of plants express both types of ACC (10). Heterodimeric ACC is expressed in the plastid and is used in de novo fatty acid biosynthesis, whereas the ACC expressed in the plant cytosol is a homomeric and is required for fatty acid elongation and the biosynthesis of several secondary metabolites- examples include flavonoids and anthocyanins (11). In mammals, ACC exists in two isoforms, ACC1 and ACC2,
which are encoded by different genes. ACC1 is a cytosolic enzyme and regulates fat synthesis in lipogenic tissues, whereas ACC2 is associated to the mitochondrial membrane and controls the rate of lipid oxidation (12,13). In addition to fatty acid production, the malonyl-CoA produced by bacterial ACCs is also the building block for a variety of natural products - such as the vitamins, biotin and lipoic acid (14,15); polyketide antibiotics (16); and the acylated homoserine lactone pheromones used as quorum sensors (17). Furthermore, malonyl-CoA has been shown to play major roles in regulating the expression of other FAS enzymes (18).

Since ACCs play critical roles in the synthesis and regulation of fatty acid biosynthesis they have become attractive target areas for the development of novel therapeutics treatments for obesity, metabolic syndrome, type 2 diabetes, and antibiotic treatment of bacterial infection (19-23); and the first report of a potent bacterial ACC inhibitor was published this year (24).

**1.3.2 Heterodimeric Acetyl-CoA Carboxylases**

Heterodimeric ACCs found in bacteria and plant plastids are multi-subunit enzymes consisting of a biotin carboxylase (BC), a carboxyltransferase (CT), and a biotin carrier protein (BCCP). They are exemplified by the ACC from the bacterium *E. coli* (Table 1.2) (25). The enzyme complex readily dissociates in solution, however each subunit retains the ability to catalyse distinct partial reactions (9), and it is clear that the overall ACC reaction proceeds via two steps (Fig. 1.3). The first step involves the carboxylation of biotin prosthetic group which is attached to the BCCP subunit and is catalysed at the active site of the BC subunit. The activated CO$_2$ is then transferred to the active site of the CT subunit which catalyses the transfer of CO$_2$ to acetyl-CoA.
Hence, ACC catalyses the formation of a new carbon-carbon bond and the energy required for this new bond is derived from the hydrolysis of ATP (6).

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Encoding gene</th>
<th>Monomeric Protein MW (Da)</th>
<th>Solution Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biotin Carboxylase (BC)</td>
<td>accC</td>
<td>49.4</td>
<td>Dimer</td>
</tr>
<tr>
<td>Carboxyltransferase, a subunit (CT-α)</td>
<td>accA</td>
<td>35.1</td>
<td>Dimer plus dimer of CT-β</td>
</tr>
<tr>
<td>Carboxyltransferase, b subunit (CT-β)</td>
<td>accD</td>
<td>33.2</td>
<td>Dimer plus dimer of CT-α</td>
</tr>
<tr>
<td>Biotin Carboxyl Carrier Protein (BCCP)</td>
<td>accB</td>
<td>16.7</td>
<td>Tetramer</td>
</tr>
</tbody>
</table>

Table 1.2. Subunits of *E. coli* ACC. The above table outlines the subunit constituents of *E. coli* ACC. Together the homodimeric BC, the heterodimeric CT and the tetrameric BCCP would constitute an enzyme with an overall mass of >300kDa. However, isolation of such a complex has not been achieved.

Figure 1.3. The two-step reaction mechanism of AAC. The covalently bound biotin of BCCP was as a carrier of CO₂ and shuttles the carboxylate moiety between the BC and CT subunits.
1.3.3 The *E. coli* Biotin Carboxyl Carrier Protein

The biotin carboxyl carrier protein (BCCP) subunit of ACC has been extensively studied. The full-length protein is 156aa long and was found to be difficult to work with, so most studies have been carried out using an 82aa C-terminal fragment of the protein (which contains the biotin domain). Consequently very little is known of the structure and function of N-terminus. It is thought to mediate dimerisation and interact with the other subunits of ACC.

![Figure 1.4](image.png)

*Figure 1.4. The NMR structures of both apo- (A) and holo (B) *E. coli* BCCP. The biotinyl-domain of *E.coli* BCCP consists of a β-barrel with the biotin moiety exposed on a tight β-turn. Notice the characteristic thumb region of BCCP which is close in space to the active site. (PDB codes 2BDO and 3BDO).*
The structure of the biotin domain has been studied by both NMR and X-ray crystallography and is a remarkably symmetrical structure consisting of two sets of four antiparallel β sheets that form a flattened β-barrel (Fig. 1.4) (26-28). The biotin cofactor is covalently attached to Lys122 which is exposed on a tight β-turn at one end of the molecule, with the N- and C-terminals close together at the other. The active lysine residue is situated in a highly conserved biotinylation motif (Ala-Met-Lys-Met) (29,30). The biotin domain of BCCP shows striking similarity to the lipoic acid-modified domains of pyruvate dehydrogenases. Elegant studies by Reche and Perham have demonstrated that specific BCCP mutants can be lipoylated in vitro, suggesting these two carrier proteins share a common evolutionary origin (31). However, *E. coli* BCCP contains a "thumb-like" protrusion, comprising residues 94-101, which is not found in the lipoyl domain. It has been demonstrated that residues in the "thumb" interact with the biotin ring, and the "thumb" region becomes less mobile upon biotinylation of the protein, suggesting the thumb plays a role in orientating the biotin moiety in space (32). Indeed, the "thumb" is required for the acetyl-CoA carboxylation reaction in vivo (33).

Heteronuclear NMR studies of the apo- and holo-forms of the *E. coli* BCCP domain indicate that there are no major conformation changes accompanying post translational modification. However, dynamic changes have been noted, with some tightening of the structure around the "thumb" and more favourable side-chain packing (27,28,34). This change in dynamics explains why Cys116 can react with thiol reagents in the apo- but not the holo-form of biotinyl domain and why apo-BCCP is more susceptible to limited proteolysis (35).
1.3.4 Biotin Carboxylase (BC)

1.3.4.1 BC Structure

The biotin carboxylase (BC) subunit of ACC catalyses the carboxylation of the biotin uriedo ring, a reaction obligatory in all biotin-dependant carboxylases (6). The crystal structure of the BC subunit of ACC from *E. coli* has been solved to 1.9 Å and consists of dimers containing three domains (Fig. 1.5) (36,37). The N-terminal domain forms a dinucleotide binding motif consisting of five strands of parallel β-sheets flanked on either side by α-helices. The "B-domain" extends from the main body of the subunit where it folds into two alpha-helical regions and three strands of β-sheet. The C-terminal domain forms part of the globular portion of the protein and contains an eight-stranded antiparallel β-sheet, a small three stranded antiparallel β-sheet and seven α-helices. The proposed active site pocket is located between the B- and C-domains and is enclosed by the residues His 209-Glu 211, His 236-Glu 241, Glu 276, Ile 287-Glu 296, and Arg 338. The active sites are located away from the dimer interface and are separated by 62 Å. However, it has been found that the BC catalytic activity is dependant on the presence both subunits and the two subunits must be in communication during catalysis (38). More recently the structure of a BC mutant with ATP bound within the active site has been published, revealing the subunit belongs to the ATP-grasp superfamily (39,40). The crystal structure reveals that upon ATP binding the enzyme undergoes a large conformational change and the "B-domain" rotates by approximately 45° to cover the active site. Residues Lys-116, His-236 and Glu-201 are highlighted as important for nucleotide binding, consistent with site directed mutagenesis data (41). Modelling studies superimposing structures
of the BC domain with biotin and ATP bound separately have revealed the relative orientation of biotin and ATP within the active site. The γ-phosphate of ATP was adjacent to the 1’N of biotin, exactly as expected from the chemistry of the reaction (39).
Figure 1.5. The crystal structure of *E. coli* BC subunit. 

A. Apo-BC. The BC subunit is dimeric and each monomer contains 3 domains. One monomer is shaded grey, while the second monomer is coloured to highlight the 3 distinct domains. The N-terminal domain is shaded green, the B-domain is shaded blue, and the C-terminal domain is shaded gold (PBD code 1DV1).

B. A BC mutant (E288K) with ATP bound. Notice the large rotation of the B-domain Lys 116 upon ATP binding (PBD code 1DV2).

C. Residues involved in ATP binding.
1.3.4.2 BC Mechanism

Inhibition studies were performed in 1999 to probe the mechanism of BC, using an inhibitor of BC which incorporated structural aspects of all three substrates: ATP, biotin and bicarbonate (42). Results suggest that the reaction proceeds via an ordered addition of substrates, with ATP binding first followed by bicarbonate and biotin (43). Product release was shown to proceed with carboxybiotin released first, followed by ADP and then phosphate.

\[
\text{HCO}_3^- + \text{MgATP} \rightarrow \text{BCCP} + \text{BCCF}
\]

Figure 1.6. The chemical mechanism of BC. The carboxylation of biotin proceeds via the formation of carboxyphosphate. This is followed by direct attack on the carbon of carboxyphosphate by the IN of biotin.

The chemical mechanism of BC is thought to begin with the production of carboxyphosphate intermediate (Fig. 1.6)- labelling studies have demonstrated that one oxygen atom from bicarbonate is channelled to the phosphate product, indicating that bicarbonate and ATP are involved in a covalent interaction (44). This is also consistent with the observation that the catalytic cleavage of ATP by BC only requires bicarbonate (i.e. BC has bicarbonate dependent ATPase activity) (45). To
date however, the carboxyphosphate intermediate has not been trapped and characterised. Very little is known about the reaction of carboxyphosphate with biotin, but by comparison with other members of the ATP-grasp superfamily it is believed that there is direct attack on the carbon of carboxyphosphate by the 1’N of biotin (Fig 1.6). For this to occur, the carboxylation of biotin requires two key structural elements: a base for the abstraction of the proton from 1’N of the ureido ring; and an amino acid to stabilise the negative charge of the enolate anion. However, exhaustive mutational analysis of BC and other ATP-grasp enzymes have been unable to identify amino acid residues which fulfil these roles (41,46-48). It may be possible that the process involves substrate-assisted catalysis. It is known that the transfer of CO$_2$ to biotin only occurs if one of the phosphate oxygen atoms is protonated, which suggests that the phosphate oxygen atom could acts as a catalytic base to abstract the 1’N proton. Interestingly, BC will accept both biotinylated-BCCP and free biotin as substrates.

Furthermore, the bicarbonate dependent ATPase activity displayed by BC is dramatically increased by the presence of both biotin or holo-BCCP (47,49). Thus, the hydrolysis of ATP is synergistic with the binding of another substrate (biotin) - a phenomenon known as substrate-induced synergism (50).
1.3.5 Carboxyltransferase (CT)

1.3.5.1 CT Structure

Carboxyltransferase (CT) catalyses the transfer of CO₂ from carboxybiotin to acetyl-CoA. Little is known about the structure of *E. coli* CT; however, the structure of the CT domain from the yeast ACC has recently been determined to 2.7 Å resolution, and is dimeric (Fig. 1.7) (51). The crystal structure shows that each CT domain molecule comprises two subdomains (N and C domains) that are intimately associated with each other. It is interesting to note that the N- and C- terminal halves show significant sequence homology to the β and α subunits of bacterial CTs. The N and C-domains of yeast CT have little sequence identity yet display similar folds, with a central β–β–α superhelix. The fold is characteristic of the crotonase/ClpP superfamily, many members of which are acyl-CoA-dependant enzymes that catalyse various reactions involved in fatty acid β-oxidation (52-55). Co-crystallisation of the enzyme with CoA has allowed the identification of the active site, which is at the interface of the dimer and contains residues from the N- and C-domains of both monomers. The CoA molecule associates with the N-domain and the N1 and N6 atoms of the adenine base are recognised by hydrogen bonds to the polypeptide chain. The phosphate groups are located near three conserved basic cationic residues (Arg^{1733}, Lys^{2034}, Arg^{2036}). It is postulated that biotin binds at a small β-sheet in the C-domain of the other monomer in the active site.

Further structural studies of yeast CT domain have revealed the molecular details of inhibitor binding. The herbicide haloxyfop has been captured in the active site (56). Unexpectedly, haloxyfop binding requires large conformational changes within the active site, forming a large hydrophobic pocket in the enzyme. Residues
known to affect herbicide sensitivity are located in the affected region. The Pfizer lead compound, CP-640186 (an inhibitor of mammalian ACC) has also been co-crystallised with yeast CT (57). This inhibitor displays a distinct mechanism of inhibition and it is thought that it binds at the putative biotin-binding site.

Figure 1.7. The crystal structure of the CT domain of yeast ACC including bound CoA. The yeast CT is dimeric, with each of the monomers split into two distinct subdomains. One CT monomer is shaded in grey, while the other monomer is shaded to highlight the N-terminal (blue) and C-terminal (gold) subdomains. The binding pocket is highlighted by the presence of CoA in the crystal structure (PDB code 1OD2).
1.3.5.2 CT Mechanism

The chemical mechanism of CT was originally probed by determining the pH rate profile of the reverse reaction (58). Initial velocity patterns and inhibition studies have revealed an ordered kinetic mechanism of CT, with malonyl-CoA binding before carboxybiotin (25). Acid-base catalysis is required in this process because the proton on the N1' of biotin must be removed to allow for carboxyl transfer from malonyl-CoA. This study demonstrated that a single ionising group on the enzyme, with a pKa of 7.5, must be unprotonated for catalysis. It was postulated that this group was a cysteine residue which acts as the base to remove the proton from the N1' of biotin. However, the recent structural evidence shows no cysteine in the active site. Furthermore, mutational analysis of the active site has revealed that no residue acts as a general base and it is now thought that the N1' atom of biotin itself functions as the general base (51). This is in direct contrast with the crotonases, where an acidic sidechain of the enzyme is required for catalysis.

The transfer of the carboxyl from carboxybiotin to acetyl-CoA can occur by either a stepwise mechanism (protein abstraction from the methyl group of acetyl-CoA occurs before carboxyl transfer) or by a concerted mechanism (proton removal is simultaneous to carboxyl transfer). To date, studies to distinguish between the two mechanisms have not been performed with CT. However, results using pyruvate carboxylase suggest that the mechanism is stepwise (59).
1.3.6 A Working Model of an ACC complex

Extensive structural and biochemical studies of the ACC subunits have allowed a working model for the ACC complex to be proposed. Investigation of BC revealed that dimer formation is required for activity, even though the active sites for from the dimer interface (38). This observation suggested a model in which the two subunits of BC are mechanistically linked in a fixed cycle and can not catalyse carboxylation simultaneously. Instead, the two subunits alternate catalytic reactions, with one site performing catalysis while the second is releasing product. This mechanism would require two BCCP molecules to be well separated from each other, as each would interact with a different BC active site. However, a separate study of BCCP implied that at least two biotinyl-domains interact during catalysis (33). It is now believed that the ACC complex contains 2 BC and 4 BCCP subunits, with a BCCP dimer interacting with each BC molecule. Indeed, a recent analysis of the BC-BCCP complex has shown this stoichiometry to be correct (60).

The structural interactions of the BC-BCCP complex are still unknown but have been shown to involve the N-terminal residues of BCCP (60). Future studies will no doubt concentrate on elucidating these interactions and analysing the CT-BCCP complex in the hope of building a clearer picture of the mechanism of the full ACC complex.
1.4 Biotin Protein Ligase (BPL)

1.4.1 The Biotinylation Reaction

The enzyme biotin protein ligase (BPL), also known as holocarboxylase synthase (HCS), is responsible for attaching biotin to the specific lysine residue at the active site of newly synthesised biotin dependant enzymes. This is a post-translational modification of exceptional specificity (61). In *E. coli*, only a single lysine within the BCCP subunit of ACC is recognised and biotinylated by endogenous BPL. Enzymatic biotinylation is an ATP-dependant two-step reaction resulting in the formation of an amide linkage between the carboxyl group of biotin and the ε-amino group of lysine (Fig. 1.8). In the first step, the carboxyl group of biotin is activated by the formation of an adenylate. This is subsequently subjected to nucleophilic attack by the active lysine residue of the biotinyl domain forming the biotinylated species. This reaction is closely related to that of the activation of the amino acyl-tRNA synthetases (62).

*E. coli* BCCP is 156-amino acids in length, with only the C-terminal 82, the biotin binding domain, required for efficient enzymatic biotinylation (63). Similarly sized domains are recognised by BPLs in carboxylases from other bacteria (64), yeast (65), and mammals (66). Primary structure analysis of these biotin binding domains reveals a very high degree of similarity, with the active lysine residue located in a highly conserved (A/V)MKM motif. Furthermore, biotin-accepting enzymes can be recognised and biotinylated by BPL derived from widely divergent species (64,67-69), indicative of a process highly conserved throughout evolution.
Figure 1.8. The biotinylation of BCCP. This two-step reaction is catalysed by BPL. Firstly, the carboxyl group of biotin is activated by the formation of an adenylate. This is subsequently subjected to nucleophilic attack by the active lysine residue of the BCCP biotinyl domain forming the biotinylated-species.
1.4.2 Enzyme Classification

Genetic analyses indicate that each organism has a single BPL-encoding gene, which is responsible for processing all biotin-dependant enzymes. The enzyme is required for health growth and deletion of the bpl gene has been shown to be lethal. BPL/HCSs have been grouped into four classes which all share a highly conserved central catalytic domain but differ at each termini (Fig. 1.9). Class I and II enzymes are derived from bacteria. They both have a C-terminal domain of unknown function but differ at the N-terminus, with class II enzymes displaying an N-terminal DNA binding domain. Eukaryotic BPLs fall into Classes III and IV, they vary in size from the prokaryotic enzymes, often having larger N-terminal domains which have little or no similarity to bacterial enzymes. None of the eukaryotic proteins contains sequences that suggest any DNA binding activity. The role of this N-terminus is unclear, but an N-terminally truncated BPL from the yeast *Saccharomyces cerevisiae* shows a dramatic reduction in enzyme activity (65). It has been speculated recently that this domain may play a part in acceptor-substrate recognition (70).
Chapter 1: Biotinylation

<table>
<thead>
<tr>
<th>Class</th>
<th>Example</th>
</tr>
</thead>
</table>
| I     | A. aeolicus  
|       | M. tuberculosis |
| II    | E. coli |
| III   | A. thaliana |
| IV    | H. sapien |

Figure 1.9. Biotin Protein Ligase Classification. BPL enzymes can be divided into four separate classes based on their tertiary structure. Bacterial BPLs are members of classes I and II, whereas classes III and IV are eukaryotic BPLs.

1.4.3  *E. coli* Biotin Protein Ligase (BirA)

1.4.3.1 The Bifunctional Nature of BirA - Controlling Biotin Biosynthesis

Genetic and biochemical studies have been performed on BPLs from several species, of which the best characterised is the BPL from *E. coli*, BirA (EC 6.3.4.15). This 35.3 kDa class II BPL is bifunctional, acting both as a biotin ligase and as a DNA-binding repressor of the biotin operon (71-73). The biotin operon of *E. coli* (*bio*) contains five genes that are involved in biotin biosynthesis, with the operator located between *bioA* and *bioB* (74). This repressor function is triggered by the formation of the intermediate biotinyl-AMP, which acts as the corepressor, with the BirA-biotinyl-AMP complex binding to the *bio* operator as a dimer. Although biotinyl-AMP is the corepressor, intracellular biotin levels influence gene expression due to
the ordered addition of substrates to BirA, with biotin binding first. The BirA-biotinyl-AMP complex is thermodynamically very stable which results in very low concentrations of both unligated BirA and biotin within the cell (75). The switch from biotin ligase to a repressor reflects competition between apo-BCCP and operator DNA for the BirA-biotinyl-AMP complex. Thus, biotin synthesis is intimately linked to both biotin and apo-BCCP concentrations in vivo. This elegant biochemical control system ensures that when all apo-BCCP has been biotinylated the stable BirA-biotinyl-AMP complex will bind to the bio operator and turn off the transcription of the biotin biosynthetic genes (since no additional biotin biosynthesis is required).

Studies of a second class II BPL from the bacterium *Bacillus subtilis* have revealed a similar role for BPL in regulating biotin biosynthesis (76,77). Furthermore, comparative genomics suggests that the existence of class II BPL within a genome strictly correlates with the presence of putative BPL-binding sites upstream of biotin operons (78).

### 1.4.3.2 Structure of BirA

Despite their importance in all living cells, the structure of only a single BPL, BirA from *E. coli*, has been determined. The crystal structure of BirA has been resolved to 2.3 Å in both the monomeric form and the DNA binding holo-dimeric form (79-81). The monomer consists of three distinct domains and typifies class II BPLs (Fig 1.10). The N-terminal domain has a helix-turn-helix fold characteristic of major groove DNA binding proteins, used in *bio* repression to bind to the operator of the *bio* operon. The central catalytic domain has been shown to contact biotin, with the
Chapter 1: Biotinylation

complete loop encompassing residues 110-128 enclosing the cofactor. Located within this loop is a highly conserved glycine rich sequence $^{115}\text{GRGRXG}^{120}$ which was first proposed as the ATP binding site, however recent characterisation of single-site mutants suggest that the motif functions in biotin and biotinyl-AMP binding (82). In a recent study, a single mutation within this conserved sequence was shown to have a dramatic effect on the activity of BirA. An R118G BirA mutant had non-specific activity, displaying a high level of promiscuous protein biotinylation (83). In effect, the removal of Arg118 led to the loss of the specificity of the enzyme for BCCP. The authors explained this phenomenon by postulating that the mutant releases biotinyl-AMP from the active site; and the increased non-specific activity is non-enzymatic, but results from chemical acylation from the released active intermediate. Thus, Arg118 is linked with the binding of biotinyl-AMP.

Little is known about the function of the small C-terminal domain, although recently residues within this domain have been shown to interact with BCCP. Perhaps more interestingly, an arginine residue in the C-terminus (R317) has been proposed to contribute to the interaction with ATP even though it is some distance from the proposed catalytic domain (84).

Binding of the intermediate by the BirA monomer results in ordering of flexible loops of the central domain, resulting in more favourable energetics of dimerisation (85,86). However, the resulting interactions are only weak and dimer formation may only occurs on the DNA surface (81). Recent studies have revealed the half life of this DNA-protein complex to be approximately 400 seconds (87). In the dimeric structure the $\beta$-sheets in the central domain of each monomer are arranged side by side, forming a single, seamless antiparallel $\beta$-sheet. The protruding
N-terminal DNA binding domains of the holo-dimer are then thought to contact two 12 bp sequences at the termini of the operator sequence (88).

Figure 1.10. The structure of the *E. coli* BirA dimer with biotin bound at each active site. One subunit is highlighted as follows- N-terminal DNA-binding domain (pink), catalytic domain (gold), and C-terminal domain (blue). (PDB codes 1HXD)

1.4.3.3 The BirA-BCCP Complex

The elucidation of the three-dimensional structures *E. coli* BirA (79-81) and *E. coli* BCCP (26-28) has allowed a model to be constructed of the BirA/BCCP complex (89). The key feature of this proposed interaction is that it utilises the same surface
loop in BirA necessary for homo-dimerisation. The proposed model suggests that the complex is formed by hydrogen bonding of β-strands from each protein, creating an extended β-sheet (Fig. 1.11). The difference between the two interactions is the relational orientation of the two domains. In the homodimer, strand Val189-Lys194 from the first BirA subunit forms an antiparallel β-sheet interaction with the same residues in the second subunit. In contrast, in the heterodimer the same stand of BirA participates, but the resulting β-sheet interaction is parallel. This resulting mutually exclusive protein-protein interaction is used to explain the regulation of biotin biosynthesis by in vivo concentrations of apo-BCCP (89).

Figure 1.11. A model of the *E. coli* BirA:BCCP complex. BirA is shown in silver with biotin bound at the active site. BCCP is shown in gold. Notice the parallel β-sheet interactions between BirA and BCCP (PDB code 1K67).
1.5 Aims

To gain further insight into the detailed protein-protein interactions that control biotin transfer we have analysed the reaction between BPL and apo-BCCP from the hyperthermophilic organism *Aquifex aeolicus* (90). This bacterium grows optimally at 95 °C on hydrogen, oxygen, carbon dioxide and mineral salts. Enzymes from extremophiles (extremozymes) are offering new opportunities for biocatalysis as a result of their extreme stability (91-93). Analysis of the *A. aeolicus* genome identified bpl and accB homologues; the predicted BPL is from the group I class (which also includes *M. tuberculosis*) which lack the N-terminal DNA-binding domain found in *E. coli* BirA (94). In *E. coli*, we have expressed active *A. aeolicus* BPL, the biotin-binding domain of BCCP as a six histidine N-terminal fusion (BCCPA67) as well as a series of BCCP mutants. Biotinylation of apo-BCCPA67 by BPL was most efficient at 70 °C and we have carried out kinetic analyses and proteolysis experiments at this temperature. Furthermore, we describe the isolation and characterisation of a chemically crosslinked BPL:BCCPA67 complex for the first time. This study is the first detailed characterisation of a class I BPL and the first investigation of a post-translational modification complex from a hyperthermophilic organism.
Chapter 2: Enzymatic and Chemical Analysis of the BPL from *Aquifex aeolicus*.
2.1 Analysis of the *A. aeolicus* genome

The complete genome sequence of *Aquifex aeolicus* consists of 1512 predicted open reading frames (90). We performed a BLAST search on the complete genome and identified two ORFs of 233aa and 154aa with high sequence homology to *E. coli* BirA (20.9% identity, 35.2% similarity) and BCCP (33.8% identity, 46.9% similarity) respectively. The pairwise sequence alignments generated by CLUSTAL W are shown in figure 2.1 and these enabled us to design PCR primers to clone the *A. aeolicus* BPL and BCCP genes (95). We noted from this initial analysis that the *A. aeolicus* BPL differs from the *E. coli* BirA in that it lacks an N-terminal DNA-binding domain which places it in the group I class of BPLs along with those from *M. tuberculosis* and *T. maritima* (94).

Previous studies on full-length *E. coli* BCCP (156 aa) revealed that the protein forms a tight complex with a second acetyl-CoA carboxylase subunit in solution and complicates biochemical studies (96). In most cases, the biotin carrier domain of biotin-containing enzymes is located at the C-terminal end of the carboxylase, with the biotinyl-lysine about 35 residues from the C-terminus. Deletion studies revealed that a minimum of 75-80 residues of the BCCP is recognised by *E. coli* BirA. The truncated form of the *E. coli* BCCP containing 87aa has been used in biochemical and structural studies (BCCP-87) and here we expressed *A. aeolicus* BCCP lacking 67 residues from the N-terminus (BCCPΔ67, Fig. 2.1) with an N-terminal six histidine-tag (total length 96aa). The homology scores between the *A. aeolicus* BCCPΔ67 and *E. coli* BCCP-87 domains are 51.9% identity and 69.6% similarity (Fig. 2.1).
Figure 2.1. Sequence alignments of BCCP (A) and BPL (B) from *E. coli* and *A. aeolicus*. Pairwise alignment was prepared using CLUSTAL W.

**A.** The (↓) shows the start residue of BCCP-87 domain used in previous studies. The (↑) shows the start codon of the BCCPA67 domain. The biotinylated lysine residue is highlighted in blue. Residues highlighted in red were suspected crosslinking residues (see section 3.2). Secondary structural elements of the BCCP-87 domain are indicated and the (*) indicates the "thumb" region.

**B.** The (--) and (+) indicate the pairs of disordered surface loops which are close in space in the *E. coli* BirA structure. The ($) indicates the trypsin cleavage sites of *A. aeolicus* BPL and the (*) is the site of subtilisin cleavage of *E. coli* BirA.
2.2 Cloning, expression and purification of BPL

The *A. aeolicus* *bpl* gene was amplified by PCR using the primers BPLfor, BPLrev and *A. aeolicus* genomic DNA as a template. The resulting product was cloned into the plasmid pCR2.1 (Invitrogen). DNA sequencing confirmed the previously published gene sequence, with the exception of a single base change at position 325 (T→C), which results in the substitution of a cysteine residue with an arginine. Subsequently the *bpl* gene was cloned into a pET-28a expression vector (Novagen) using the restrictions enzymes NcoI and BamHI; the resulting plasmid was named pET-28a/bpl (Fig. 2.2). Expression was performed in various *E. coli* cells (DE3 lysogens); we found optimum recovery of protein using the BL21(DE3) strain. Cells were grown in shake flasks at 37°C and expression induced with 1mM isopropyl-1-thio-β-D-galactopyranoside (see "Materials and Methods").

![Expression Vector pET28a/bpl. Aquifex bpl gene cloned using NcoI and BamHI restriction sites.](image)

Figure 2.2 - Expression Vector pET28a/bpl. *Aquifex* bpl gene cloned using *NcoI* and *BamHI* restriction sites.
The predicted pI of the *A. aeolicus* BPL is 9.1 and contains a high proportion of positively charged residues and so cation-exchange chromatography was used to purify the enzyme in a single step (See Fig. 2.3). Initially the crude lysate was incubated at 60°C which resulted in the precipitation of a significant quantity of *E. coli* proteins. It was then necessary to dialyse the sample overnight (20°C) against 10mM HEPES (pH 7.5) as immediate loading of an untreated extract onto a ResourceS column resulted in very poor binding (<5%). It is unclear why this step was necessary, but after dialysis binding to the cation exchange column approached 100%. BPL eluted from the column at 200mM NaCl (Fig. 5) and we obtained the enzyme with a purity of greater than 95% (as determined by SDS-PAGE). Electrospray mass spectrometry analysis gave the molecular weight of the protein as 26636.83 ± 2.34 Da, consistent with the post-translational removal of the N-terminal methionine residue, and accurate to within experimental error of the predicted value of 26634.57 Da. The final yield of BPL using this method was >10mg per litre of cell culture and this protein was used for all subsequent kinetic and cross linking analysis.

**Figure 2.3. Purification of *A. aeolicus* BPL**

Protein purification was analysed by SDS-PAGE under reducing conditions. *Lane 1*, low molecular weight marker. *Lane 2*, BPL cell lysate. *Lane 3*, BPL cell lysate after heat purification. *Lane 4*, BPL after ResourceS purification.
2.3 Cloning, expression and purification of BCCPΔ67

We designed primers to clone a truncated domain of the A. aeolicus bccp gene missing the first 201 base pairs which encode the N-terminal 67 amino acids of A. aeolicus BCCP (ACCBΔ67for and ACCBreve). The truncated gene was amplified from genomic DNA using PCR and cloned into the pCR2.1 vector (Invitrogen) using the restriction enzymes Ncol and BamHI. DNA sequencing confirmed the expected gene sequence, and the bccpΔ67 gene was subsequently cloned into a pET-derived expression vector with an N-terminal hexahistidine tag. The final construct was named pET6H/accbΔ67 (Fig 2.4).

Figure 2.4. Expression Vector pET6H/accbΔ67. Aquifex accbΔ67 gene cloned using Ncol and BamHI restriction sites. The resulting construct encodes an N-terminal hexahistidine tag.
E. coli BL21(DE3) competent cells were used for recombinant expression (described under "Materials and Methods") and the BCCPΔ67 cell lysate was first purified by nickel-affinity chromatography (see Fig. 2.5). The protein eluted with 200mM imidazole and, as precipitation had been observed at high concentrations of this eluant, it was immediately diluted 1:1 with 10 mM HEPES (pH 7.5) and dialysed against this buffer. SDS-PAGE analysis indicated BCCPΔ67 to be >90% pure but electrospray mass spectrometry revealed the presence of two distinct species - the first of molecular weight 10740.11 ±1.112 Da, corresponded to the predicted mass of apo-BCCPΔ67 (10739.63 Da); the second corresponding to the holo-form (biotinylated), with a mass increase of 226Da (10965.41 Da; predicted mass 10965.6 Da). This confirmed that the A. aeolicus BCCPΔ67 domain folded correctly, was recognised and biotinylated by the host E. coli BirA. To separate the apo- and holo-forms of BCCPΔ67 we employed anion exchange chromatography in a similar way to that used for E. coli BCCP-87 (97). Fractions from the column were analysed by electrospray mass spectrometry and the apo-protein eluted at a slightly lower salt concentration than the holo-form (160-240 mM NaCl vs. 240-320 mM NaCl). Approximately 80% of the apo-BCCPΔ67 was resolved from the holo-form by collecting only the leading fractions of the protein peak. The final yield of apo-BCCPΔ67 was ~5-10 mg per litre of cell culture and ~1 mg per litre of the holo-form.

![Figure 2.5. Purification of A. aeolicus BCCPΔ67.](image-url)

Protein purification was analysed by SDS-PAGE under reducing conditions. Lane 1, low molecular weight marker. Lane 2, BCCPΔ67 cell lysate. Lane 3, BCCPΔ67 after Ni affinity purification. Lane 4, apo-BCCPΔ67 after MonoQ purification.
2.4 Biological properties of BPL

Activity assays were performed with BPL by measuring the incorporation of [\(^{14}\text{C}\)]biotin into the purified apo-BCCPA67 biotin-accepting domain (98). In initial experiments we observed optimal enzyme activity at pH 8.5 (Fig. 2.6a), and magnesium ions, ATP, biotin and apo-BCCPA67 were all required for activity. The activity of the enzyme was also measured at varying temperatures, with optimal activity at 70°C. Increasing the temperature above 70°C resulted in enzyme precipitation, together with a dramatic loss in activity (Fig 2.6b). BPLs tolerance of other nucleotide sources was measured by replacing ATP with UTP, GTP, or CTP. No BPL activity was detected for any of these three substrates-suggesting the enzyme is completely dependent on ATP for its nucleotide supply (data not shown).

In assays performed with the mutant BCCPA67 K117L as the biotin acceptor no biotinylation was observed, verifying K117 as the active residue and demonstrating the specificity of the BPL catalysed reaction.
Chapter 2: *Aquifex aeolicus* BPL

**Figure 2.6. Biological Properties of** *A. aeolicus* **BPL**. A. Relative activity of BPL with changing pH. B. Relative activity of BPL with changing temperature.

### 2.5 Kinetic Analysis of BPL

The kinetic constants for D-biotin, MgATP, and apo-BCCP\(\Delta67\) were determined using steady-state kinetics (Fig. 2.7) The \(K_m\) for D-biotin was determined to be 439.6 ± 69.2 nM. The \(K_m\)s for BPLs from other species range from low nanomolar to low
micromolar; 67 ± 11nM (S. cerevisiae BPL), 300nM (E. coli BirA), 130nM (A. thaliana HCS) and 3.3mM (chicken liver HCSI) (65,84,99,100). The $K_m$ for MgATP was 15.14 ± 1.52 μM, which is similar to that determined for the S. cerevisiae BPL (20.9 ± 3μM) and A. thaliana HCS (4.4 μM). In contrast, the $K_m$ for MgATP for E. coli BirA is around 300μM. Finally, the $K_m$ for apo-BCCPΔ67 was 160.43 ± 32.43 μM. A range of biotinylation substrates have been used in assays of BPL activity with cross-species reactivity frequently observed e.g. S. cerevisiae BPL has a $K_m$ of 11.1 ± 1mM for E. coli BCCP-87. However, we could not test E. coli BCCP-87 as a substrate for BPL since the rate of biotinylation at 37°C was outwith the lower limit of detection in our assay.

As discussed in chapter 1, the first step in all biotinylation reactions studied thus far involves the synthesis of a biotinyl-5'-AMP intermediate and the release of inorganic pyrophosphate (PPi). This molecule is the substrate for biotin transfer to BCCP and is also the co-repressor of E. coli BirA. To prove that BPL also synthesises biotinyl-5'-AMP we incubated BPL, biotin and [14C] MgATP (the ATP was labelled in the 8 position). We observed the synthesis of [14C] biotinyl-5'-AMP by the use of an avidin binding assay (Clarke, D. J.; MChem research project). Moreover, we carried out inhibition studies with PPi and found that it acted as a competitive inhibitor relative to ATP and as a non-competitive inhibitor relative to biotin (Clarke, D. J.; MChem research project). This implies that MgATP binds to BPL prior to biotin, in contrast to E. coli BirA where the order is reversed (75). Similar orders of binding have been observed in other BPLs such as S. cerevisiae and A. thaliana, which like A. aeolicus BPL, have no DNA-binding domain (65,99,101).
We also noted that biotinylation was inhibited by the addition of NaCl in concentrations above 200mM.

**Figure 2.7. Steady-state kinetic analysis of BPL substrate binding.** The activity of *A. aeolicus* BPL was measured under steady-state conditions at 70 °C. Two substrates were kept at constant saturating levels while the concentration of the third substrate was varied over the ranges shown above in the graphs. From the curves, $K_m$ values for biotin (A), MgATP (B), and apo-BCCPΔ67 (C) were determined (see Materials and Methods).
2.6 Limited Proteolysis of BPL

We subjected BPL to limited proteolysis in the presence and absence of biotin and MgATP (Fig. 2.8). Digestion with both trypsin and chymotrypsin resulted in formation of a fragment of ~21 kDa. Chymotrypsin digestion also produced an array of smaller peptide fragments. Preincubation of BPL with saturating amounts of biotin or MgATP had little effect on the protein’s susceptibility to digestion. However, preincubation with both substrates dramatically increased the resistance of BPL to proteolysis, suggesting that the binding of the substrates and/or the formation of the intermediate, biotinyl-5’-AMP, plays a role in protecting the protease cleavage site. LC-MS analysis of the peptide fragment produced from BPL after treatment with trypsin revealed the presence of two distinct species of mass $215549.54 \pm 2.57$ Da and $21678.62 \pm 5.93$ Da. Primary structure analysis of BPL established these masses corresponded to trypsin cleavage between R44 and K45, and K45 and W46 adjacent to the proposed catalytic centre and biotinyl-5’-AMP binding site.
Figure 2.8. Proteolysis of BPL. *A. aeolicus* BPL was treated with trypsin or chymotrypsin either with or without equilibrating the enzyme with 1mM MgATP and/or 50μM biotin. *Lanes* 1-4, Trypsin digest; *lane* 1, BPL; *lane* 2, BPL + MgATP; *lane* 3, BPL + biotin; *lane* 4, BPL + MgATP and biotin. *Lanes* 5-8 Chymotrypsin digest; *lane* 5, BPL; *lane* 6, BPL + MgATP; *lane* 7, BPL + biotin; *lane* 8, BPL + MgATP and biotin.

2.7 Discussion

Analysis of the complete genome of the hyperthermophile *A. aeolicus* revealed the presence of BPL and BCCP homologues (see Fig. 2.1). The *A. aeolicus* BPL enzyme belongs to the class I group of BPLs since it lacks the DNA-binding domain found in *E. coli* BirA and is the smallest characterised thus far. Eukaryotic BPLs also lack predicted DNA-binding domains but have large N-terminal extensions with unknown functions (102). The full-length *A. aeolicus* BCCP has a C-terminus showing high sequence homology to the biotin domains of biotin-carboxylases and contains the eight amino acid ‘thumb’ motif found in *E. coli* BCCP (33,102,103). The N-terminus has a large proportion of charged residues, and displays little similarity to any other BCCPs.
Using recombinant proteins isolated from *E. coli* we have characterised the full-length BPL and BCCP biotinylation domain BCCPΔ67 from a hyperthermophile, and we have gained insight into this extremely specific post-translational modification reaction at high temperatures. We found *A. aeolicus* BPL to be monomeric and so competing homo-dimerisation interactions found in *E. coli* BirA are not present. We isolated a mixture of apo- and holo-forms of *A. aeolicus* BCCPΔ67 and so conclude that it must be a substrate for *E. coli* BPL *in vivo*. Biotinylation in hyperthermophiles proceeds via the two-step reaction sequence found in other organisms (Fig 1.2). Isolated *A. aeolicus* BPL could biotinylate apo-BCCPΔ67 at temperatures up to 70°C albeit at a slow rate. It is interesting to compare the *A. aeolicus* BPL:BCCPΔ67 biotinylation reaction with that of a mutant *E. coli* BirA lacking the N-terminal DNA binding domain (BirA65-321) and *E. coli* BCCP-87. The BirA65-321 mutant could synthesise biotinyl-5'-AMP and transfer biotin to apo-BCCP-87 at the same rate as wild-type BirA. However, the affinity of BirA65-321 mutant for biotin and biotinyl-5'-AMP was decreased 100-fold and 1000-fold respectively (104). This suggested that in BirA, the N-terminal domain is somehow involved in tight-binding of the two ligands. In future, it would be interesting to study a BPL:BirA chimera by fusing the DNA-binding domain at the N-terminus of *A. aeolicus* BPL.

Substrate $K_m$ values for BPLs from a number of species have been shown to range from the low nanomolar to low millimolar. In steady-state kinetic assays at 70°C, the *A. aeolicus* BPL bound biotin, MgATP and apo-BCCPΔ67 with affinities of 439.6 nM, 15.14 μM and 160.43 μM respectively. The kinetic constant for MgATP suggests that *A. aeolicus* BPL resembles those from eukaryotic biotin.
auxotrophs (low micromolar). In contrast, *E. coli* BirA binds MgATP with a $K_m$ in the low millimolar range which reflects its dual function as both repressor of biotin biosynthesis and biotin ligase. It is interesting to note that *A. aeolicus* contains all the genes required to convert pimelate to biotin (*bioW, bioF, bioA, bioD* and *bioB*) suggesting it can synthesise this vitamin but the *in vivo* concentration of biotin within *A. aeolicus* cells is unknown. The $K_m$ for the apo-BCCPΔ67 domain used in this study is high compared to others but this may reflect the fact that the first 67 amino acid residues, which contain a high number of charged residues, could play an important role in tight binding to BPL. Most biochemical studies use these truncated BCCP domains and future work using full length BCCPs should elucidate the role of the N-terminal interaction with BPL. The calculated $k_{cat}/K_m$ for biotin of $1.7 \pm 0.1 \times 10^4$ M$^{-1}$ s$^{-1}$ is 300-, 100- and 35-fold smaller than the *E. coli* BirA, yeast and *A. thaliana* BPL enzymes respectively but reflects the fact that the *A. aeolicus* BPL $k_{cat}$ is low at 70 °C (cf *A. aeolicus* grows optimally at 95 °C) (65,99,105).

To analyse the domain structure of BPL, we investigated its susceptibility to limited proteolysis and found that treatment with trypsin produced two ~20kDa fragments differing in length by only one residue. Mass spectrometry revealed that cleavage had occurred after residues R44 and K45 which, by comparison with *E. coli* BirA, are predicted to lie near the putative intermediate binding site (Fig. 2.1). Treatment of BPL with trypsin and chymotrypsin in the presence of biotin did not protect the enzyme from either protease and similarly, MgATP alone failed to block proteolysis. However, incubation of the enzyme in the presence of both substrates rendered *A. aeolicus* BPL protease resistant. The same region is protease sensitive in *S. cerevisiae* BPL and is also protected by incubation with biotin and MgATP (65).
Chapter 2: *Aquifex aeolicus* BPL

The *E. coli* BirA structure contains five surface loops, four of which are in the central domain with loop regions [110-128, 212-233] and [140-146, 193-199] close together in three dimensional space (61). The region containing [110-128] in *E. coli* BirA is highly conserved to residues [32-50] in *A. aeolicus* BPL whereas the other loop regions have low pairwise sequence homology. A protease-sensitive site has been reported between residues 217 and 218 of BirA. In contrast, *A. aeolicus* BPL is not cleaved at this site but is cleaved in the adjacent loop region [32-50]. This suggests that this highly-conserved region forms an exposed loop near the biotinyl-5’-AMP binding site (Fig. 2.1). These flexible, unstructured regions are also involved in BCCP binding and are believed to become more rigid upon substrate binding (61,106).
Chapter 3: Characterisation of the BPL:BCCP Complex from *Aquifex aeolicus*
Chapter 3: The *Aquifex aeolicus* BPL:BCCP Complex

3.1 Chemical crosslinking of BPL and BCCP

Although structures of *E. coli* BirA and both apo-and holo-BCCP-87 have been determined, our goal was to isolate a BPL:BCCP complex for biochemical and structural studies. Previous work in our laboratory used the chemical crosslinking agent 1-ethyl-3-(dimethylamino-propyl)-carbodiimide (EDC) to isolate an *E. coli* flavodoxin/flavodoxin reductase complex, so we used this reagent to crosslink BPL and various forms of BCCPΔ67 (107).

Initially we incubated BPL and apo-BCCPΔ67 in the presence of excess EDC at room temperature with and without saturating amounts of biotin and MgATP, but we did not observe any crosslinked species of predicted molecular weight ~36kDa on SDS-PAGE. However, a species was observed when the incubation was carried out at elevated temperatures, with 60°C being the optimum (see Fig. 3.1 A). The presence of the substrates had no observable effect on crosslinking. Interestingly, when BPL was incubated with holo-BCCPΔ67 and EDC the amount of crosslinked species generated was significantly reduced compared to the apo- form (Fig. 3.1 B); suggesting that the active lysine residue (E117) in BCCPΔ67 may be involved in chemical crosslinking.

Purification of the BPL:BCCPΔ67 complex from unreacted proteins was achieved using size exclusion chromatography on Sephadex 75, which resolved the mixture into three peaks (Fig. 3.2). Analysis by SDS-PAGE revealed the BPL:BCCPΔ67 complex had a molecular weight of 37 kDa (Fig. 3.2, inset). Electrospray analysis of the complex gave a molecular weight of 37,221 ± 200 Da which agrees well with the predicted mass of a 1:1 heterodimer.
Figure 3.1. SDS-PAGE analysis of chemical crosslinking assays. Gel A, crosslinking of BPL and apo-BCCPΔ67. Gel B, crosslinking of BPL and holo-BCCPΔ67. Both Gels lanes 1-5, assay after 0, 5, 10, 15, and 30 minutes respectively. Gel A lanes 6 and 7, control assays with BCCPΔ67 alone and BPL alone.

Figure 3.2. Purification of the chemically crosslinked BPL:apo-BCCPΔ67 complex by size exclusion chromatography. The chromatogram above was obtained when the cross-linking reaction was applied to a Superdex 75 column. The three peaks correspond to the crosslinked complex (7-8ml), BPL (10ml) and BCCPΔ67 (11-12ml). Insert. SDS-PAGE analysis of the column fractions. Lane 1, cross-linking reaction before purification. Lane 2-11, 1 ml fractions eluting between 6-15ml.
Chapter 3: The *Aquifex aeolicus* BPL:BCCP Complex

3.2 Identification of Potential Residues Involved in BPL:BCCPΔ67

Crosslinking

The recently proposed model of the *E. coli* BirA:BCCP-87 complex, described in section 1.4.3.3 (Fig. 1.11), was obtained using data from structural studies, sequence analysis, mutagenesis and limited proteolysis experiments (89). The model (PDB code 1K67) was built using the coordinates of the BirA dimer in the presence of biotin (PDB code 1HXD) and BCCP-87 (PDB 1BIA) and it suggests that regions 119-128 and 87-90 of BCCP (also identified by NMR studies) interact directly with BirA. We have mapped the *A. aeolicus* BPL/*E. coli* BirA sequence alignment (Fig. 2.1) onto the *E. coli* BirA:BCCP-87 structure to allow identification of candidate residues that may be cross-linked by EDC in our isolated BPL:BCCPΔ67 complex.

The proposed model suggests that the complex is formed by hydrogen bonding of β-strands from each protein, creating an extended β-sheet (Fig. 1.10). The 3D model has several pairs of charged residues (R or K and D or E) that are between 1.9 and 5.7 Å apart (see Fig. 3.3). The side chains of the closest pair, BirA R119/BCCP-87 E119 are 1.92 Å apart and are located close to the bound biotin and active lysine (K122) residue. Two other pairs, BirA K283/BCCP-87 E147 and BirA K194/BCCP-87 E128, are separated by 5-6 Å in the model. However, computer-aided re-alignment of these side chains reduced the separation to 3.5 and 1.0 Å respectively. It is interesting to note that all these residues are conserved between the *A. aeolicus* and *E. coli* proteins (Fig. 2.1, highlighted in red). The residues L95 on BirA and Q126 on BCCP-87 are 4.3 Å apart but are not conserved. These are substituted by K19 on BPL and E121 in BCCPΔ67 and furthermore, computer-
modelled substitutions of BPL L95K and BCCP-87 Q126E reduce the distance between these two residues to 1.8Å, suggesting they could form a potential ion-pair in the *A. aeolicus* BPL:BCCPΔ67 complex.

These four ions pairs in the *A. aeolicus* proteins were subject to mutational studies to identify what role, if any, they played in mediating protein crosslinking. Single amino-acid mutations of BCCPΔ67 were prepared, selectively removing each glutamate residue, and their ability to crosslink wild-type BPL monitored. A fifth BCCPΔ67 mutant, K117L, was also prepared; this BCCP mutant lacks the active lysine.

Figure 3.3. A model of the *E. coli* BirA:BCCP-87 complex. Four amino-acid pairs at the protein:protein interface are highlighted. Residues in red are situated on BCCP and residues in blue are situated on BPL (BirA). These pairs correspond to potential ion pairs in the *A. aeolicus* complex (see text).

A series of truncated bccpΔ67 mutant genes, with glutamate residues replaced by alanine residues, were produced using the megaprimer method (108). The mutations were confirmed by DNA sequencing and the genes were subsequently cloned into a pET-derived expression vector with an N-terminal hexahistidine tag. The resulting constructs were then transformed into *E. coli* BL21 (DE3) cells for expression (as described in "Materials and Methods"). In addition, a further mutant was constructed, accbΔ67 K117L, which contains a mutation removing the active lysine residue (see Appendix 1).

All of the mutant proteins were purified in a similar manner to the wild type BCCPΔ67 protein (see section 2.3). Nickel-affinity chromatography was employed, followed by anion exchange. The purification procedure was monitored by SDS-PAGE and LC-MS (see Fig. 3.4, Table 3.1). Pure apo-protein was obtained for each mutant expressed.

![Figure 3.4. SDS-PAGE analysis of purified BCCPΔ67 mutants. Lane 1, BCCPΔ67 E114A; lane 2, BCCPΔ67 E121A; lane 3, BCCPΔ67 E123A; lane 4, BCCPΔ67 E142A. All proteins were 1 mg/ml concentration.](image)
Table 3.1. Analysis of BCCPA67 mutants by LC-MS. All mutants were analysed by LC-MS using a C5 reverse-phase HPLC column attached directly to a Platform-MS. The masses of all mutants were consistent with their theoretical masses. Post-translational modifications observed were noted.
Chapter 3: The *Aquifex aeolicus* BPL:BCCP Complex

### 3.2.2 Crosslinking BCCPΔ67 Mutants with BPL

The purified apo-BCCPΔ67 mutants were assayed for their ability to crosslink to wild-type BPL using EDC. The assays were essentially the same as described in "Materials and Methods" and section 3.1. The crosslinking was monitored by SDS-PAGE and the results are displayed in Fig. 3.5.

Under identical conditions, the BCCPΔ67 K117L mutant crosslinked to BPL in comparable amount to wild-type BCCP; indicating that the active lysine residue of
Chapter 3: The *Aquifex aeolicus* BPL:BCCP Complex

BCCP is not directly involved in crosslinking within the BCCPΔ67:BPL complex. One of the four glutamic acid mutants, BCCPΔ67 E121A, showed a significant reduction in crosslinking potential; suggesting that this residue plays an important role in crosslinking BCCPΔ67 to BPL.

### 3.2.3 Cloning, Expression and Purification of a *bpl* K19A mutant.

Analysis of the *E. coli* BCCP:BirA model revealed that the E121 residue of *aquifex* BCCPΔ67 is in close proximity to K19 of *Aquifex* BPL. In order to test if K19 does indeed have a role in crosslinking the *Aquifex* BCCPΔ67:BPL complex a BPL K19A mutant was prepared, and it's ability to crosslink with wild-type BCCPΔ67 was monitored.

A *bpl* K19A mutant gene was produced using the megaprimer method (108). The mutation was confirmed by DNA sequencing and the gene subsequently cloned into the pET-28a expression vector. The resulting construct was then transformed into *E. coli* BL21 (DE3) cells for expression (as described in "Materials and Methods").

The expressed protein was purified in an identical manner to wild-type BPL, using cation exchange chromatography. However, the resulting yield of pure protein was significantly less than wild-type (~4mg / litre of cell culture). Analysis of the pure protein by LC-MS revealed that a single species of mass 26585.98 ± 10.82 Da, which is in agreement with the predicted mass of 26577.5 Da.
3.2.4 Chemical crosslinking BPL K19A with BCCPA67.

The purified BPL K19A was assayed for its ability to crosslink to wild-type BCCPA67. The assays were essentially the same as described in "Materials and Methods" and section 3.1. The crosslinking was monitored by SDS-PAGE and the results are displayed in Figure 3.5.

By direct comparison with wild-type BPL it is clear that the BPL mutant has a reduced ability to crosslink with BCCPA67 (Fig. 3.6). This result suggests that the K19A residue is involved in the formation of the BPL:BCCPA67 complex. However, a small quantity of the crosslinked species is observed after treatment of BPL K19A and BCCPA67 with EDC. This may be due to non-specific crosslinking, i.e. the E121 residue of BCCPA67 may crosslink to another basic residue on BPL K19A which is close in space.

![Figure 3.6. Chemical crosslinking of BPL K19A with BCCPA67. Both gels lane 1, low molecular weight marker; lanes 2-6, assay after 0, 5, 15, 30 and 60 minutes respectively. Gel A, BCCPA67 and wild-type BPL. Gel B, BCCPA67 and BPL K19A.](image)
Chapter 3: The *Aquifex aeolicus* BPL:BCCP Complex

3.3 The Role of the "Thumb" Region of BCCPΔ67 in BPL::BCCPΔ67 Crosslinking.

Extensive X-ray crystallographic and NMR studies have been performed on both the apo- and holo-BCCP biotinylation domain (see section 1.3.3). In the first reported structure, Athappilly and Hendrickson showed that the domain is a remarkably symmetrical structure consisting of two sets of four antiparallel β-sheets (26). However the symmetry of the molecule is disrupted by a protruding eight residue structure known as the "thumb", which has been shown to interact with the biotin moiety in holo-BCCP. This structural motif is not found in all other biotinylated proteins. For example, the *P. shermanii* 1.3S biotinoyl domain lacks the 8 amino acid motif and the structure of this domain has been shown to display a folding pattern essentially identical to *E. coli* BCCP biotinoyl domain lacking the "thumb" (109). However, primary structure alignments suggest that the "thumb" is present in *A. aeolicus* BCCP (see figure 2.1).

Studies by Reche and co-workers showed that an *E.coli* BCCP biotinoyl domain mutants lacking the "thumb" motif was efficiently biotinylated in both *in vitro* and *in vivo* experiments (31). These results suggest that it is unlikely that the "thumb" moiety has a role in biotinylation of the BCCP by cognate BirA. However, the observation that the expression of an *E. coli* BCCP "thumbless" mutant fails to restore both growth and fatty acid biosynthesis to a temperature sensitive mutant strain of *E. coli* lacking a BCCP, suggests the thumb plays a critical role in the ACC reaction (33).
Chapter 3: The *Aquifex aeolicus* BPL:BCCP Complex

Although structural studies have revealed that biotinylation of BCCP has no significant effect on the overall structure of the protein domain, biochemical studies have highlighted significant differences in the susceptibility of the two protein forms to both chemical modification and proteolysis, indicating a difference in protein dynamics (35). More recently, this increased stability of holo-BCCP has been attributed to the interaction of the biotin moiety with the "thumb" region of the BCCP (32). Elegant experiments performed by Solbiati and co-workers demonstrated that a chemically biotinylated BCCP biotinoyl domain "thumbless" mutant displayed sensitivities to trypsin and chemical modification similar to that of wild-type apo-BCCP, and was significantly more susceptible to modification than wild type holo-BCCP.

In our initial experiments chemically crosslinking *A. aeolicus* BPL and BCCPΔ67 we noticed a significant difference in the propensity to crosslink to BPL between the apo- and holo-forms of BCCPΔ67 (section 3.1). Experiments using the holo-BCCPΔ67 produced significantly less of the chemically crosslinked protein:protein complex. This result suggested that the active lysine residue (K117) of BCCPΔ67 may be participating in chemical crosslinking to BPL. However, further experiments revealed that a K117L BCCPΔ67 mutant crosslinked to BPL in comparable amounts to wild type apo-BCCPΔ67 (section 3.2). This latter result demonstrates that our first hypothesis was incorrect; that the active lysine plays no significant role in crosslinking BCCPΔ67 and BPL. To test if the interaction between the biotin moiety and the "thumb" region of *A. aeolicus* BCCPΔ67 has an effect on reducing protein crosslinking both apo and holo- BCCPΔ67 "thumbless" mutants were prepared and chemically analysed.
3.3.1 Cloning, expression and purification of *A. aeolicus AccBΔ67 "thumbless" mutants.

Initially, a "thumbless" BCCPΔ67 mutant, with the replacement of residues 89-SPAPGA-94 with an AAA linker sequence, was produced using the megaprimer method (108). The mutations were confirmed by DNA sequencing and the gene was subsequently cloned into a pET-derived expression vector with an N-terminal hexahistidine tag (see Appendix 1). The resulting construct, named pET6H/AccbΔ67ΔTh, was then transformed into *E. coli* BL21 (DE3) cells for expression (as described in "Materials and Methods").

The mutant protein was purified in a similar manner to the wild-type BCCPΔ67 protein (see section 2.3). Nickel-affinity chromatography was employed, followed by anion exchange. The purification procedure was monitored by SDS-PAGE and LC-MS. Interestingly, the expressed protein was only isolated in the apo-form; suggesting that the mutant was not efficiently biotinylated by BirA *in vivo*. LC-MS analysis of the protein detected a single species of mass 10472.0 Da, which is in agreement with the predicted mass of 10472.4 Da.

To produce the holo-form of the 'thumbless' BCCPΔ67, chemical biotinylation of the protein was attempted using biotinoyl-N-hydroxysuccinimide. Unfortunately, this reaction was not specific for the active lysine (K117), and resulted in multiple sites of biotinylation, presumable at each of the five lysine residues within the protein as well as the amino-terminus. In order to overcome this problem we produced a second 'thumbless' mutant, in which the K117 was the sole lysine residue within the protein.
Chapter 3: The *Aquifex aeolicus* BPL:BCCP Complex

The production of the "sole-lysine thumbless" BCCPΔ67 mutant was undertaken in two stages using the megaprimer method. The procedure involved substituting four lysine residues with arginines (see Appendix 1). Fortunately, three of the lysine residues were close in the primary sequence and could be changed using the same mutant primer. The final lysine was mutated separately. After the introduction of the mutations, the sequence was verified by DNA-sequencing and the resulting gene, named $AccB\Delta67\DeltaTh\ 4KR$, was cloned, expressed, and purified as described previously. Analysis of the resulting protein, named BCCPΔ67ΔTh4KR, by LC-MS revealed a single species of mass 10583.0 Da, consistent with the predicted mass of 10584.4Da.

### 3.3.2 Chemical biotinylation and purification of BCCPΔ67ΔTh 4KR

Chemical biotinylation of the "thumbless" BCCP mutant, BCCPΔ67ΔTh4KR, was achieved used a previously described method (32). The protein was incubated with activated biotin (biotinyol-hydroxy-succinimide) for four hours. After optimisation the final yield of biotinylated protein was approximately 25%. Furthermore, analysis by LC-MS revealed the presence of a small amount of protein (3%) modified with covalent attachment of two biotin moieties. Presumably, the amino terminus of the protein is slightly susceptible to biotinylation under the conditions used.

Separation of the apo- and holo-BCCPΔ67 mutant was achieved by RP-HPLC using a C5 column, and holo-BCCPΔ67ΔTh4KR was isolated in 98% purity.
3.3.3 Chemical crosslinking BPL and apo- and holo- BCCPΔ67ΔTh 4KR

The purified apo- and holo- forms of BCCPΔ67ΔTh 4KR were assayed for their ability to crosslink to wild-type BPL. The assays were essentially the same as described in "Materials and Methods" and section 3.1. The crosslinking was monitored by SDS-PAGE and the results are displayed in Fig. 3.7.

Under identical conditions, both the apo- and holo-BCCP "thumbless" mutants crosslinked to BPL in comparable amounts. This is in stark contrast to the results of crosslinking apo- and holo-BCCPΔ67 to BPL, where there is a significant decrease in the propensity of the holo-form to crosslink to BPL. Therefore, removing the "thumb" region of BCCP has an effect similar to lack of biotinylation and these results suggest that the interaction between the biotin moiety and the protruding thumb region of BCCP is responsible for the decrease in the ability of the BCCP to recognise BPL.

Fig. 3.7. Chemical crosslinking of BPL with apo- and holo-BCCPΔ67ΔTh4KR mutants. All gels lane 1, low molecular weight marker. All gels lanes 2-6, assay after 0, 5, 15, 30 and 60 minutes respectively. Gel A, apo-BCCPΔ67ΔTh4KR. Gel B, holo-BCCPΔ67ΔTh4KR.
Chapter 3: The *Aquifex aeolicus* BPL:BCCP Complex

3.4 Discussion

There is evidence that ion pair networks are important for the recognition of BCCP by BirA in *E. coli*. A recent combined mutagenesis/biological selection approach identified two single glutamate residues E119 and E147 of *E. coli* BCCP-87 that appear to interact with BPL (98). A BCCP-87 E119K mutant was inactive as a substrate for BirA, whereas the E147K protein could be biotinylated, albeit poorly. It is presumed that these acidic BCCP-87 residues interact with basic BirA counterparts and mutation of BirA residues K277 and R317 were found to effect biotinylation and ATP binding respectively. This surprising result suggested that the C-terminal domain of BirA, which had been ascribed no biochemical function, also plays a significant role in apo-BCCP and substrate recognition (84).

It has been shown that ion pair networks are a common feature in heat-resistant proteins and are believed to play important roles in their increased thermal stability (92,110). Since both the *A. aeolicus* BPL and BCCP contain a large number of charged residues, and we observed inhibition of biotinylation at high salt concentrations we presume that ionic interactions are involved in the formation of the hyperthermophilic BPL:BCCPΔ67 complex. To investigate the formation of the BPL:BCCPΔ67 heterodimer we used the chemical cross-linking agent EDC to capture a BPL:BCCP complex for the first time. The zero-length EDC reagent activates acidic residues on one protein to form an unstable urea derivative (111). This derivative then reacts with a nucleophile (such as lysine) on another protein to form an amide link between the two proteins. Incubation of BPL and apo-BCCPΔ67 in the presence of EDC led to the time-dependent appearance of a species of ~37kDa on SDS-PAGE gels (Fig. 3.1), which is in agreement with the predicted mass of a 1:1
complex of BPL and apo-BCCPΔ67. We observed that BPL, BCCPΔ67 and the complex eluted earlier than predicted from the size exclusion column. Nevertheless, the complex was easily separated from the unreacted proteins using this procedure (Fig. 3.2) and allowed us to confirm its mass by electrospray mass spectrometry.

Overlaying the primary sequence of the *A. aeolicus* BCCP and BPL proteins onto the model of the *E. coli* BCCP-87:BirA protein interface assembled by Weaver and co-workers (89), allowed us to identify four ion-pairs which may be involved in crosslinking BCCPΔ67 and BPL. We then recombinantly produced a series of four BCCPΔ67 mutants, with single mutations of the glutamic acid residues thought to be present at the protein-protein interface. Analysis of the ability of each member of the mutant library to crosslink to wild-type BPL allowed us to identify a single glutamic acid residue on BCCPΔ67, E121, which is involved in crosslinking (Fig. 3.4). The proposed ion-pair partner to BCCPΔ67 E121 was BPL K19, and a BPL K19A mutant did indeed show a reduced ability to crosslink to wild-type BCCPD67; verifying it's role in crosslinking.

Interestingly, in our initial crosslinking experiments, we noticed the formation of the complex was significant reduced between BPL and holo-BCCPΔ67 (Fig. 3.1B). Initially, we hypothesised that this was because the active lysine residue of BCCP, K117, was involved in crosslinking to BPL; therefore the amino sidechain of K117 would be effectively blocked from crosslinking upon biotinylation. However, this hypothesis was rejected when a BCCPΔ67 K117L mutant was shown to crosslink to BPL in comparable amount to wild-type apo-BCCPΔ67- demonstrating that the active lysine residue is not required for crosslinked complex formation.
Chapter 3: The *Aquifex aeolicus* BPL:BCCP Complex

Although the published 3D structures of the apo- and holo- forms of BCCP-87 show no major structural differences, some structural studies (both NMR and x-ray) have concluded that the lack of any major differences between them might not be wholly reflected in their behaviour in solution (28). NMR titration experiments were carried out with BirA and apo-BCCP-87 and, in light of our data, it would be interesting to repeat this work with BirA and holo-BCCP-87 to determine if any differences arise. Recent elegant studies by Cronan and Solbiati, described in chapter 3.3, highlight a difference in the stability of apo-BCCP-87 and holo-BCCP-87 to proteolysis and they propose the majority of this increased stability is due to an interaction between the biotin moiety and the "thumb" domain of *E. coli* BCCP (32). The authors attribute this increased stability of the holo-BCCP to a reduction in the dynamics of the protein upon biotinylation.

Primary sequence analysis revealed that the *A. aeolicus* BCCPΔ67 also contains a well-conserved thumb domain located between residues 89-SPAPGA-94. Production of apo- and holo-BCCPΔ67 mutants lacking the "thumb" region allowed us to analysis the effect of the "thumb" upon crosslinking to BPL. We found that apo-BCCPΔ67ΔTh crosslinked to BPL in comparable amounts to wild-type apo-BCCPΔ67- demonstrating that the "thumb" region is not required for the recognition of BCCP by BPL. This result is strengthen by previous research which shows that an *E. coli* BCCP "thumbless" mutant is efficiently biotinylated in vivo and in vitro (31). More interestingly, holo-BCCPΔ67ΔTh is also chemically crosslinked to BPL in significant amounts. It seems that the "thumb" region of BCCP is required for BPL to distinguish between the apo- and holo- forms of its protein substrate. We hypothesise that a biotin-thumb interaction exists in holo-BCCPΔ67, similar to that found in
Chapter 3: The *Aquifex aeolicus* BPL:BCCP Complex

*E. coli* BCCP. If this hypothesis is correct it is easy to imagine that the reduction in the dynamics of the protein, which are a consequence of this interaction, hampers the formation of the ion pair network at the BPL:BCCP heterodimer interface, thus reducing chemical crosslinking. This hypothesis allows us to speculate on a mechanism of biotinylation in which, upon attachment of biotin, the reduction in dynamics of BCCP aid the dissociation of the ligase-holoBCCP complex.

In light of these results, future work should concentrate on measuring the dynamics of the apo- and holo- forms of both wild-type BCCPΔ67 and its "thumbless" mutant. Indirect evidence for any difference in the dynamics of the two forms of the proteins could be established by testing the susceptibility of the proteins to proteolysis, in a similar manner as performed by Cronan and Solbiati (32).
Chapter 4: Defensins - Important Antimicrobial Peptides of Innate Immunity
4.1 The Diversity and Classification of Antimicrobial Peptides

Antimicrobial peptides (AMPs) are gene encoded polypeptides of fewer than 100 amino acids (112,113). They are widespread throughout the plant and animal kingdom and typically display broad-spectrum antimicrobial activity at physiological concentrations. The diversity of AMPs is so great that it is difficult to categorise them except broadly on the basis of their secondary structure (all reported AMPs are catalogued online at www.bbcm.univ.trieste.it/~tossi/amsdb.html). However, the fundamental structural trait shared by the majority of AMPs is the ability of the molecule to adopt a shape in which clusters of hydrophobic and cationic residues are spatially organised in discrete patches- known as amphipathic design. In some classes, such as defensins and bactenecins, this amphipathic nature is constrained by a rigid framework of antiparallel β-sheets held together by disulfide bridges (114,115). In contrast, linear peptides, such as magainin, only adopt an amphipathic structure when they enter a membrane and become predominantly α-helical in nature (116). Members of a third class of AMPs are characterised by the predominance of one or two amino acids within their primary structure. These linear peptides, examples include the tryptophan-rich indolicidin, the histidine-rich histatins, and the proline/arginine-rich PR39, separate hydrophobic and cationic sidechains around an extended peptide scaffold (117-119). Examples of the various classes of AMPs are given in table 4.1.

It is interesting to note that anionic AMPs have been discovered and characterised. However, these peptides are restricted to a few animal species and
Chapter 4: Defensins

tissues (120). In recent years there has been growing interest in these fascinating molecules for development of novel therapeutics (121-125).

<table>
<thead>
<tr>
<th>Class</th>
<th>Sequence</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear, α-helical</td>
<td>Cathelicidin (LL-37)</td>
<td>Human</td>
</tr>
<tr>
<td></td>
<td>Buforin II</td>
<td>Vertebrate</td>
</tr>
<tr>
<td></td>
<td>Magainin 2</td>
<td>Frog</td>
</tr>
<tr>
<td>One Disulfide Bond</td>
<td>Bactenecin 1</td>
<td>Cow</td>
</tr>
<tr>
<td>Two Disulfide Bonds</td>
<td>Protegrin 1</td>
<td>Pig</td>
</tr>
<tr>
<td>Three Disulfide Bonds</td>
<td>α-defensin (HNP3)</td>
<td>Human</td>
</tr>
<tr>
<td></td>
<td>β-defensin (HBD2)</td>
<td>Human</td>
</tr>
<tr>
<td></td>
<td>θ-defensin (Resistin)</td>
<td>Monkey</td>
</tr>
<tr>
<td>Four Disulfide Bonds</td>
<td>Hepcidin</td>
<td>Human</td>
</tr>
<tr>
<td>Linear, not α-helical</td>
<td>PR-39</td>
<td>Pig</td>
</tr>
<tr>
<td></td>
<td>Histatin 5</td>
<td>Human</td>
</tr>
</tbody>
</table>

Table 4.1. Overview of antimicrobial peptides - Representatives from each class. Cysteines paired in disulfide bonds are noted. C-terminal amides are noted by a. For a comprehensive inventory see www.bbcm.univ.trieste.it/~tossi/amsdb.html.

4.2 Vertebrate Defensins

4.2.1 Classification

The defensins are a diverse class of vertebrate antimicrobial peptides, which are characterised by a common triple β-sheet fold held together by six disulfide linked...
Chapter 4: Defensins

cysteine residues. These evolutionary related molecules were first classified in 1985 (114,126), and have recently been the focus of several review papers (127-129). The two main defensin subfamilies, α- and β-defensins, differ in cysteine spacing and connectivity of the disulfide bonds (Fig. 4.1). More recently, a third defensin subfamily of θ-defensins has been isolated from the rhesus macaque monkey (130).

The θ-defensins are cyclic peptides, formed by splicing two nine amino acid peptide precursors. This subfamily is thought to have evolved in primates but has been inactivated in humans by mutations which incorporate premature stop codons (131).

Antimicrobial peptides found in invertebrates and plants which contain six (or eight) cysteines have also been classified as defensins (132-136). However, their evolutionary relationship to vertebrate defensins is unclear (137), and this introduction will concentrate on the latter.

![Diagrams of α-, β-, and θ-defensins](image)

**Figure 4.1.** The disulfide bridging connectivity in α, β, and θ-defensins. α-defensins are characterised by C1-C6, C2-C4 and C3-C5 disulfide bridging. β-defensins are characterised by C1-C5, C2-C4 and C3-C6 disulfide connectivity. θ-defensins are characterised by an amide-linked backbone that forms a cyclic structure with three disulfide bonds.
4.2.2 Biological Role of Mammalian Antimicrobial Defensins - The Innate Immune System.

Although mammals are constantly exposed to a myriad of microbes, very few succeed in causing systemic infection. This is predominantly due to the barrier function of the skin and mucous membranes which form a mechanical obstruction, and the presence of antimicrobial substances. If this barrier is breached, invading microbes are contained and eliminated by the host immune system. Innate immunity is the first line of constitutively pre-existing host defence that is rapidly mobilised following pathogen detection, and consists of two major components (138,139).

Firstly, invading microbes are challenged with resident macrophages and various humoral protectants, including ß-defensins, lectins and complement. In humans, four ß-defensins (HBD1-HBD4) have been described in detail, and many more ß-defensin genes have been identified within the genome (140-145). The expression of these ß-defensin genes in barrier and secretory epithelial cells is constitutive in some cases (HBD1), whereas in others the gene is expressed in response to infection, such as HBD2-4 (146,147). Similar epithelial ß-defensins derived from other mammals have been characterised and are reviewed in ref. (148).

The second arm of innate immunity involves the production of chemokines for the recruitment and activation of polymorphonuclear leukocytes (PMNs). Half a century ago, a crude protein fraction of PMNs displaying antimicrobial activity was isolated and named “phagocytin” (149-151). Subsequent investigation of "phagocytin" revealed a complex mixture of antimicrobial proteins and peptides. Antimicrobial proteins found in PMNs include: lactoferrin, lysozyme, cathepsin G, azurocidin, calprotectin and bactericidal/permeability-increasing protein (BPI) (152).
Chapter 4: Defensins

The peptide fraction of "phagocytin" consists of cationic antimicrobial peptides that belong to the α-defensin (126-129) and cathelicidin (153-156) families. The α-defensins are stored in primary (azurophil) granules (157,158). When PMNs ingest microorganisms into phagocytic vacuoles, these granules fuse to the vacuoles and deliver their contents onto the microorganism. Consequently, the local concentration of α-defensin can be very high (>10 mg/ml). In contrast, cathelicidins are stored in secondary (specific) granules of the PMNs and are mainly delivered extracellularly (159).

Human PMNs contain four α-defensins (HNP1-4) and one cathelicidin (LL-37/hCAP-18) (126,159-161). However, the type and number of AMPs in mammalian PMNs varies between species. Bovine PMN do not contain α-defensins, but express many β-defensins and cathelicidins (156,162,163). Porcine and ovine PMNs contain no defensins and only express cathelicidins (155,156,164), whereas murine PMNs contain only a single cathelicidin (165). Interestingly, the PMNs of Old World monkeys have been shown to contain θ-defensins (circular peptides, known as lectins, with effective antiviral activity) in addition to their α-defensins and cathelicidin (131,166-170). The antimicrobial arsenal of PMNs is further enhanced by the production of a variety of toxic oxidants. These are formed by reactions between superoxide anions produced by NADPH oxidase and myeloperoxidase. Additionally, the PMNs from some species contain oxidants that are derived from a type II inducible nitric oxide synthase (iNOS) (171).

In addition to expression in PMNs, α-defensins are also found in Paneth cells in humans (HD5-6), mice and rhesus monkeys (172-176). These secretory cells are located in the basal regions of the Lieberkuhn crypts, tiny pits throughout the small
intestine. The release of α-defensins and other antimicrobial substances from these cells is stimulated by the LPS of invading intraluminal bacteria. Paneth cell α-defensins, also known as cryptdins, are reviewed in refs. (177,178).

The complex nature and number of host defence molecules indicate that no single mechanism is adequate to control and fight the threat of microbial infection. Furthermore, it is clear that defensins play a critical role in the innate immune response.

4.2.3 Defensin Genes, Biosynthetic Pathways, and Posttranslational Processing

In humans, the genes encoding α- and β-defensins are located in a cluster on chromosome 8p23 (179-183). This defensin cluster is incompletely mapped, and is highly polymorphic- the copy-number of specific defensin genes are variable between individuals (184). Recently, four additional defensin gene clusters have been discovered, which contain transcribed genes. However, their protein products have not yet been fully characterised (145). Analogous defensin gene clusters have been identified in mice (145,185-187).

Defensin genes are initially expressed as a pre-propeptide, containing an N-terminal endoplasmic reticulum/golgi targeting sequence (signal pre-sequence), the adjacent anionic precursor sequence (propeptide sequence) and the cationic mature peptide at the C-terminus (Fig. 4.2). In α-defensins, the charge on the propeptide and mature peptide balance (188). In 1999, a paper by Wilson et al. reported the cellular processing of a murine Paneth cell α-defensin (189). The metalloprotease matrilysin (MMP7) was shown to cleave the propeptide sequence to produce the
mature peptide, and the interaction of these two molecules has been the subject of a further study (190). More recently, the protease trypsin was shown to be the processing enzyme of human α-defensin HP-5 (191,192). To date, no detailed analysis regarding the processing of β-defensins has been published.

Figure 4.2. Structure of prototypical defensin gene and peptide. The gene product is a prepropeptide. The pre-sequence encodes a signal sequence. The pro-sequence is removed by a specific protease either before or after secretion (see text). The mature peptide is antimicrobial.

4.2.4 Genetic Regulation of Defensins

Defensin biosynthesis can be regulated by microbial signals, development signals, or cytokine response and has been described in detail for human, bovine and murine molecules. Human PMN α-defensins are synthesised constitutively by the bone-marrow precursors of PMNs during specific differentiation stages of the cells
The defensins are processed into the mature form before packaging into primary (azurophil) granules. Once matured, PMNs do not express α-defensins and cease granular synthesis. They are then released into the bloodstream and are free to enter tissues. Paneth cell α-defensins are also thought to be synthesised constitutively. However, they are only released into the lumen after exposure to bacteria or cholinergic stimuli (194). This release has recently been shown to be moderated by a Ca^{2+}-activated K(+) channel mIKCa1 (195). Once released, the α-defensins are processed by their specific proteases, such as matrilysin or trypsin.

The expression of HBD-1, which is produced predominantly in the urinary tract, has been found to be constitutive, as has its murine homologue mBD-1 (196). In contrast, HBD-2 expression is upregulated by infection and inflammatory stimuli in human studies (197). In cell culture experiments, HBD-2 expression was stimulated by interleukin (IL)-1α, IL-1β, tumour necrosis factor-α (TNF-α), microorganisms, and lipopolysaccharide (LPS) (196). This upregulation mechanism is dependant on a specific nuclear factor-κB (NF-κB) binding site upstream of the HBD-2 gene. The upregulation of several murine and bovine β-defensins by the NF-κB pathway has also been demonstrated (185,198). Both HBD-3 and HBD-4 are upregulated during infection (143,199). In vitro, the production of HBD-3 by keratinocytes and airway epithelial cells is stimulated by TNF-α or bacteria. However, the genetic control is mediated by an NF-κB independent pathway which is not fully understood. There is also evidence of upregulation of HBD-2 and HBD-3 in epithelia by Rhinovirus-16 and HIV-1 (200,201).
Chapter 4: Defensins

In some cases, the synthesis of defensins is regulated developmentally. For example, rabbit alveolar macrophages have been shown to synthesise and accumulate α-defensins during the first postnatal month (202).

4.2.5 Defensin Structure

Although purified in 1985, the first structural characterisation of a defensin was not published until 1991 (203). This crystallographic study elucidated the structure of α-defensin HNP-3 to a resolution of 1.9 Å. HNP-3 folded to form a triple stranded β-sheet held together by 3 disulfide bonds. This novel fold has also been shown to be adopted by human HNP-1 and rabbit NP-2 and NP-5, despite very little sequence identity between the molecules (204-207). Recently, the structures of human β-defensins have been extensively investigated and both crystallographic and NMR structures of HBD-1, -2, and -3 have been determined (208-211). In addition NMR structures of several murine and bovine defensins have also been published (212). The secondary structures of these β-defensins are similar; and essentially the same as the α-defensins (Fig 4.3 and 4.4). Although the two subfamilies display different disulfide connectivity and spacing both display the same triple-stranded β-sheet fold. There is no hydrophobic core in these small molecules, which suggests the fold is stabilized by the disulfide bonds. The presence of a small N-terminal α-helical segment has been observed in several β-defensins (HBD-2, mBD-8), however it is unclear what, if any, significant role this feature has (210,212). The quaternary structure of defensins seems to vary between molecules. For example HNP-1 was dimeric in its crystalline form, whereas other α-defensins are predominantly monomeric. Furthermore, native gel electrophoresis suggests that HBD-1 is
monomeric, while HBD-2 and HBD-3 are dimeric. However, NMR analysis of HBD-2 and HBD-3 show no indication of dimerisation, while crystallographic data suggests oligomerisation of HBD-2. It maybe that the more dynamic environment of NMR is responsible for the lack of higher order structure observed using this method. It has been postulated that higher order oligomerisation is essential for antimicrobial activity and further analysis of defensin oligomerisation, both in solution and in the presence of membranes, is required.

Figure 4.3. Sequence alignment of human \( \beta \)-defensins HBD1-3. Notice the conserved cysteine residues which form disulfides between Cys 1 and 5, 2 and 4, and 3 and 6. Apart from the cysteine residues and two conserved glycine residue, there is very little sequence homology between the peptides.
Chapter 4: Defensins

Figure 4.4. The crystal structures of representatives of $\alpha$- and $\beta$-defensins. 

**A.** The structure of $\alpha$-defensin HNP-3. The molecule is dimeric, one monomer is coloured gold and the disulfide bonds are highlighted in red. Notice the characteristic triple stranded $\beta$-sheet fold.

**B.** A space-fill model of HNP-3, residues are highlighted by charge. Notice the aliphatic nature of the molecule.

**C.** The structure of $\beta$-defensin HBD-1. Again, the molecule is dimeric within the unit cell and one monomer is highlighted. The characteristic triple stranded $\beta$-sheet is adopted; also notice the small N-terminal $\alpha$-helix, which is displayed by some $\beta$-defensins.

**D.** A space-fill representation of HBD-1 which again highlights the aliphatic nature of defensins.
4.2.6 Antimicrobial Activity of Defensins

Most defensins exhibit broad spectrum antimicrobial activity against both bacteria and fungi (199,213-217). Furthermore, many show impressive in vitro activity against microorganisms resistant to conventional antibiotics. However, antimicrobial activity is sensitive to the ionic strength and the concentration of divalent cations. Under optimal conditions (low salt concentration), activity is observed at concentrations as low as 1-10 µg/ml (nM to µM range). Increasing the concentrations of salts and plasma proteins competitively inhibits defensin activity. There is evidence that some enveloping viruses, including HIV, are also inactivated by defensins (218-220). Indeed, human α-defensins have been implicated among the molecules that are responsible for the antiviral activity secreted by CD8+ T cells of HIV-non-progressors; although these results have recently been questioned (221,222).

At higher concentrations, some defensins have cytotoxic effects against mammalian cells (223-225). Furthermore, cells exposed to high concentrations of defensins generate pro-inflammatory signals suggesting that defensins could be responsible for some tissue injury- especially in the lung (226,227).

A recent seminal study by Salzman et al. has described the potent in vivo activity of a human α-defensin (HD-5) in protecting mice from oral administration of virulent Salmonella typhimurium (228). Using transgenic mice expressing HD-5, the authors demonstrated that this defensin protected animals from an infection that was fatal for wild-type mice. Protection was achieved quickly upon administration of bacteria, and lower counts of S. typhimurium were observed in the intestinal lumen, with less dissemination to other organs. Interestingly, protection was only observed
Chapter 4: Defensins

if the dose of *S. typhimurium* was given orally, not intraperitoneally, supporting the hypothesis that defensins act as “locally secreted antibiotics”.

From early studies of defensin peptides, it was thought that the strictly conserved cysteines residues, and the resulting characteristic disulfide bridges, were essential for antimicrobial activity. Indeed, the disulfide bridging had been shown to be integral for maintaining the triple-stranded \( \beta \)-sheet tertiary structure, and it was assumed this tertiary structure governed antimicrobial activity. However, in 2002 Nagaraj and co-workers showed peptide fragments of bovine neutrophil \( \beta \)-defensin BNBD-2 had antibacterial activities which were independent of the number and location of the disulfide bridges (229). To my knowledge, this is the first study to analyse the direct effects of the disulfide bonds on activity. Recently, this has been followed by studies showing that the antimicrobial activity of HBD-3 is independent of disulfide bridging (230). Furthermore, analysis of the murine \( \alpha \)-defensin, Cryp4, and disulfide variants revealed a similar relationship (231).

So, what biological role does this invariant structural feature perform? It has been demonstrated that the chemotactic properties (see section 4.2.9) of HBD-3 are dependant on disulfide-bond formation (230). There is also evidence that disulfide crosslinking protects defensins from proteolysis by proteases (231).

### 4.2.7 Mechanism of Antimicrobial Activity

Defensins, like all cationic AMPs, exploit a fundamental difference between the microbial and host's cell membrane. Bacterial membranes tend to be rich in hydroxylated phospholipids (such as phosphatidylglycerol and cardiolipin) and contain lipoteichoic acid or lipopolysaccharide (LPS) moieties in their outer
membrane (Fig. 4.6). These constituents are anionic and consequently they generate a negative charge on the outer membrane of the bacterium. In contrast, eukaryotic membranes are principally composed of zwitterionic phospholipids, such as phosphatidylcholine and phosphatidylethanolamine. Thus, the outer leaflet of eukaryotic membrane has very little net charge. In the model proposed by Shai, Matsuzaki and Huang, the first step of cationic AMP action involves an electrostatic interaction between the peptide and the anionic bacterial membrane (232,233). This theory is supported by experimental data, using artificial membranes, which clearly implies electrostatic interactions are crucial for targeting defensins to the bacterial cell wall (234-237). What happens once the defensin molecule has located the bacterial membrane is less certain. The theories proposed to date can be split into two camps- those that propose that higher order structures are formed inside the membrane and those that do not (238).

In the first case, the ‘barrel-stave’ model, there is evidence for the formation of defensin multimers following initial interaction with the membrane. The observed leakage of dye markers from liposomes implies that pores are formed, at least transiently, within the membrane (236). It is proposed that after the initial interaction with the membrane, and upon reaching a threshold concentration, the defensins start to self-aggregate and insert deeper into the membrane core. Aggregation allows for a minimal exposure of the peptide hydrophilic residues to the hydrophobic membrane interior, as the peptides adopt a transmembrane configuration. Lipids may also intercalate with the peptides within these channels. These pores then allow the leakage of intracellular components out of the cell and/or defensins into the cell.
Chapter 4: Defensins

In the second case, 'the carpet model', the defensin molecules act against microorganisms though a relatively diffuse manner, often equated with detergents (238). A high density of peptides accumulates on the membrane surface. Phospholipid displacement, changes in membrane fluidity and/or reduction in membrane barrier properties subsequently lead to membrane disruption. This theory requires the defensin to adopt no specific quaternary structure. Thus, when a threshold concentration of peptide is reached, the membrane is subject to unfavourable energetics and membrane integrity is lost. In this theory defensin molecules do not need to enter the hydrophobic environment of the membrane core (Fig. 4.5).

In both theories, it seems that disruption of the cell wall and rapid membrane depolarisation is the primary action resulting in cell death. However, there is evidence that defensins may have intracellular targets, since they have been shown to bind to both glycoproteins, and DNA (166,239). Therefore, secondary defensin targets may exist. Early studies by Lehrer et al. demonstrated that defensins act on both the outer and inner membranes of Gram-negative bacteria independently, and the lethal consequences of defensin exposure are correlated specifically to the disruption of the inner membrane (240).
Figure 4.5. The proposed mechanisms of the antibacterial activity of defensins. Hydrophilic regions of the peptides are coloured red, hydrophobic regions of the peptides are coloured blue. A. The ‘barrel-stave’ model. The attached peptides aggregate and insert into the membrane bilayer so that the hydrophobic peptide regions align with the lipid core region and the hydrophilic peptide regions form the interior region of the pore. B. The ‘carpet’ model. The peptides disrupt the membrane by orienting parallel to the surface of the lipid bilayer, carpeting the membrane.

4.2.8 AMP Resistance - Bacterial Countermeasures

Although AMPs display broad spectrum activity, both constitutive and inducible resistance has been documented in bacteria (238). These 'bacterial countermeasures' include, reducing the net negative charge of the bacterial cell envelope; expelling AMPs through energy dependant pumps; and digestion of AMPs with proteases (241). These resistance mechanisms have been noted as key virulence factors in many pathogenic bacteria.
Chapter 4: Defensins

Figure 4.6. Examples of lipid A from the Gram-negative bacterium *Salmonella enteritica* (A) and Lipoteichoic acid from the Gram-positive bacterium *Staphylococcus aureus* (B). Modifications of the bacterial cell envelope components involved in defensin resistance. The cationic modifications are highlighted in red. In A, addition of both amino-arabinose and C₁₆ fatty acids have been documented. In B, addition of aniline.

Reducing the negative charge of the cell wall appears to be the most common mechanism of resistance, and has been achieved by both Gram-positive and Gram-negative bacteria (Fig. 4.6). Gram-positive bacteria have a characteristic thick cell wall composed of peptidoglycan and teichoic acid polymers. Teichoic acid consists of alternating phosphate and alditol groups, and the phosphate groups confer an overall negative charge to the molecule. Partial neutralisation of teichoic acid, by
modifying the phosphate group with D-alanine, has been achieved in many Gram-positive defensin resistant bacteria (staphylococci, streptococci, enterococci, listeriae and bacilli) (242). Addition of D-alanine requires four proteins (DltA, -B, -C and -D) which are encoded in an operon. Experimental evidence has demonstrated that disruption of this dlt operon in S. aureus abolishes defensin resistance (243).

Furthermore, it is also believed that another strain of S. aureus modifies phosphatidylglycerol with L-lysine, producing lysylphosphatidylglycerol (LPG), as a means of resisting AMPs (244). The two free amino groups of the lysine moiety of lysylphosphatidylglycerol result in a net positive charge on the molecule. The gene mprF encodes a lysine transferase (using lysine-t-RNA as a substrate) and is required for the biosynthesis of this unusual phospholipid and ΔmprF S. aureus mutants are more susceptible to defensins (244).

In Gram-negative bacteria a similar resistance mechanism exists (Fig. 4.4). One of the major components of the outer membranes of these bacteria is lipopolysaccharide (LPS), and reduction in the anionic charge of LPS has been well documented. The anionic component of LPS, lipid A, is composed of a glucosamine dimer flanked by phosphate groups and linked to five or more fatty acids (245). Modification of the lipid A moiety with aminoarabinose has been documented in Salmonella enterica, Pseudomonas aeruginosa and Burkholderia cepacia (246-249). The free amino group of aminoarabinose leads to the reduction in the net charge of lipid A, and confers resistance to defensins and other cationic AMPs (Fig. 4.4). The pmrE and pmrHFIJKL genes are responsible for the biosynthesis of aminoarabinose and its addition to lipid A, and S. enterica pmrF mutants have been shown to be virulence attenuated in mice (250). Modification of lipid A with
phosphoethanolamine is also found in cationic AMP resistant *S. enterica* and is thought to alter the net charge of lipid A in a similar way (251).

These genes which confer resistance to defensins and other cationic AMPs in Gram-positive bacteria are under the control of the PhoP-PhoQ regulon (246). This two-component 'master-regulatory' system controls many virulence factors in response to changes in magnesium and calcium ion concentrations (252).

Bacterial efflux pumps, which are proton-motive-force dependent, have long been known to confer resistance to organic dyes and disinfectants. Recently, however, they have been linked with increased resistance to cationic AMPs. The MtrCDE exporter has been shown to increase resistance to the mammalian AMPs protegrin and cathelicidin in *Neisseria gonorrhoeae* and similar mechanisms have been demonstrated in some clinical strains of *S. aureus* (253,254).

Finally, in a recent publication by Islam *et al.* a novel concept of resistance to HBD-1 and LL-37 was documented in the *Shigella* species (255). The mechanism involves inhibiting the production of these peptides in human rectal epithelial cells. The event involves *Shigella* plasmid DNA but is not fully understood.

**4.2.9 Other Biological Roles of Defensins**

In addition to their broad spectrum antimicrobial properties, there is growing evidence that defensins have additional biological activity and act as important signal molecules in the immunological repertoire; see review ref (129,256). Recent reports suggest that defensins act as chemokines, and thus serve as a bridge between the innate and adaptive immune systems. Over fifteen years ago, HNP-1 and -2 were shown to be chemotactic for human monocytes, and more recently they have been
shown to induce migration of human T lymphocytes and dendritic cells (DCs) (257,258). Similar activity has been observed for β-defensins for example, HBD-2 is chemotactic for immature DCs, T-cells and neutrophils treated with TNF-α (259,260). In each case, HBD-2 acts by binding to the CCR6, a member of a large family of G-protein coupled receptors (GPCR). Mouse β-defensin-2 also triggers maturation of DCs though an interaction with Toll-like receptor 4 (TLR4) (261,262). Finally, HBD-2 also induces mast-cell migration via chemotaxis through an, as yet, unknown pathway (263).

Neutrophil derived α-defensins from humans, rats, guinea pigs and rabbits cause peritoneal mast cells to release histamine (264). This process is triggered at nM concentrations of defensin, and acts through a G-protein-dependant process. A similar activity has also been observed using HBD-2, but not HBD-1 (265). Histamine, as well as being a mediator of inflammation, is an important bioactive marker for mast-cell degranulation. As the granules of human and mouse mast-cells contain both β-defensins and cathelicidins (some of which are chemotactic for neutrophils), as well as a range of serine proteases, this process can be viewed as a communication circuit between the bodies neutrophils and mast-cells (266,267).

In 1994, human α-defensins were shown to interact with complement (268). Purified HNP-1-3 from human neutrophils were immobilised and subsequently shown to bind C1q and trigger the classic pathway of complement activation (269). However, more recent work has cast doubt on these early findings, as van den Berg et al. demonstrated that HNP-1 in solution inhibited activation of the classical pathway (270). To date the exact effect of defensin/complement interactions remain unclear. In addition, there is evidence that some rabbit neutrophil defensins are
opsonins, aiding the ingestion of *Klebsiella pneumoniae* by macrophages (271); α-defensins have also shown to have both cytotoxic and growth promoting effects under controlled conditions (225,272-276).

Finally, recent evidence suggests that many defensins are lectins - they recognise and bind carbohydrates (166,169,170). The antiviral activity of some defensins has been attributed to this property. HNP-1-3 and retrocyclin (a θ-defensin) have been shown to bind to glycoprotein B2 (gB2) of herpes simplex virus type 2 (HSV-2) - gB2 has an essential role in allowing HSV-2 to attach and enter the target cell (169,277,278).

### 4.3 Defensins in human health

Current and rapidly expanding information on defensin expression, structure, and biological activity has established a firm foundation for further study of the role that these peptides may have in human health and disease (279). However, to date, only limited work has been performed exploring the direct relationship between defensin/AMP function and disease pathogenesis.

#### 4.3.1 Lung Disease - Cystic Fibrosis

One possible example of inactivation of defensin and other AMPs causing human disease has been proposed for cystic fibrosis (CF). CF is the most common autosomal recessive genetic disease caused by the mutation in the cystic fibrosis transmembrane regulator (*CFTR*) gene. The *CFTR* gene encodes a regulated chloride channel that also functions as a regulator of other ion channels, and mutation of this
gene leads to an imbalance in the airway surface liquid. The main cause of morbidity and mortality in CF is respiratory failure due to the progressive destruction of the airway and lungs by a cycle of recurring infection and inflammation. These infections are typified in early childhood by *S. aureus* and *Hemophilus influenzae*, but eventually these are succeeded by *Pseudomonas aeruginosa* and this infection indicates the onset of progressive lung disease.

These events indicate a local defect in epithelial host defence, as the infection is strictly limited to the lungs and does not spread elsewhere (280). In 1996, Smith *et al.* proposed that this impairment was due to inhibition of defensin and cathelicidin activity by an abnormally high ionic environment in the CF lung (281). This theory is supported by evidence that HBD-1 is inactivated in CF, and that overexpression of cathelicidin in a CF xenograft model restores bacterial killing (217,282). According to this theory, the thin layer of fluid that coats the respiratory epithelia, known as the airway surface liquid (ASL), normally has a low salt concentration relative to blood plasma- conditions which favour the activity of cationic AMPs. In CF, the ability to absorb salt from the epithelial cell fluid is compromised which leads to an increased salt concentration in the ASL, and consequently a loss in AMP activity. However, measuring the salt concentration of ASL has proved problematic and very little direct evidence supports this theory. Alternative hypotheses have been proposed, and the role of defensins in CF pathogenesis remains speculative (283,284).

### 4.3.2 Dermatological Diseases - Psoriasis and Atopic Dermatitis

As outlined in section 4.2.2-4.2.4, defensin genes in epithelial cells are upregulated during infection and injury. However, a similar but much longer-lasting increase in
defensin and cathelicidin expression has been observed in some inflammatory conditions, such as psoriasis (141,147,285,286). Evidence suggests that patients with psoriasis have a decreased incidence of several types of infection (287). This resistance has been proposed to be due increased epithelial AMP expression in these patients.

In contrast, recent studies have extended these association by demonstrating that patients with atopic dermatitis have insufficient amounts of LL-37 and HBD-2 expressed at the skin surface (288). These patients are frequently subjected to skin infections by several microorganisms.

Finally, another dermatological example of the effect of defensins is seen in burn wounds. The concentration of defensins and cathelicidins is reduced in burn wound fluid in comparison with incision wounds (289). Bacterial infection of burn wound is a common problem, particularly with *P. aeruginosa*, a bacterium particularly sensitive to the AMPs down-regulated. Thus, a relative deficiency of AMPs in the burn environment might contribute to local and systemic immunosuppression.

### 4.3.3 Inflammatory Bowel Diseases

Increased AMP expression has been observed in association with several diseases of the gut (290). HD-5 is expressed from Paneth cells in the colon in inflammatory bowel disease but is not present in normal colonic mucosa, HD-6 and HBD-2 has also been linked to this disorder (291-293).

Surface intestinal epithelial cells induce the expression of five defensins in active ulcerative colitis but not in normal or inactive disease (294). In contrast,
Chapter 4: Defensins

Crohn's disease is characterised by the down-regulation of HBD-2 and -3 in the colon (295). This disease, which involves inflammation and frequent infection of the bowel, may be a 'defensin deficiency syndrome' (296).

Gastrointestinal infections are also associated with defensin expression. *H. pylori* infection has been shown to induce HBD-2 expression in the human gastric epithelium (297). Finally, in a startling example, the expression of both LL-37 and HBD-1 is turned off in early *Shigella* infections in humans (255). This down-regulation has been proposed as a virulence factor in the development of the disease.
4.4 Defensin-Related Peptide - Defr1

The dogma that all defensins contain the conserved six-cysteine motif was recently challenged by a report by our collaborators, who recently described a murine β-defensin gene in C57Bl/6 mice that encoded a 34 amino acid peptide with only five cysteine residues (Defr1, accession no. AJ344114, Fig. 4.5) (298). This gene is a variant allele of Defb8, a classical defensin which encodes six cysteines and is found in all other inbred murine strains tested (299). Defr1 was shown to be expressed mainly in the heart and testis, and a synthetic Defr1 peptide displayed potent antimicrobial activity against both Gram-positive and Gram-negative bacteria.

In the work reported here we have performed both structural and functional analysis of Defr1 and its six-cysteine analogue (Defr1 Y5C) and report a novel relationship between the tertiary structure and antimicrobial activity of this unique defensin.

![Sequence of Defb8, Defr1 and Defr1 Y5C](image)

**Figure 4.7. Sequence of Defb8, Defr1 and Defr1 Y5C.** Numbering corresponds to the six conserved cysteines found in β-defensins. Notice the loss of Cys1 in Defr1. The overall charge of each sequence is +5.
Chapter 5: Structural and functional characterisation of defensin related peptide-1 (Defr1).
5.1 Antimicrobial Properties of Defr1 and Defr1 Y5C

5.1.1 Antimicrobial Activity of Defr1 and Defr1 Y5C

In a previous study, Dorin and coworkers demonstrated that, despite having only five cysteine residues, murine Defr1 displayed antimicrobial activity at nanomolar concentrations (298). Here we analysed synthetic Defr1 and its six cysteine analogue, Defr1 Y5C, and have determined the minimum bactericidal concentrations (MBCs) of the oxidised and reduced forms of both peptides against a diverse panel of clinically relevant microbes (Table 5.1). Numerous methods have been used in various laboratories to determine the antimicrobial activities of defensins (229,231,300). Typically, the test organism is exposed to the peptide (0 - 24 hours), followed by growth on liquid or solid media and subsequent determination of optical densities and/or counting of surviving colonies relative to a control. Antimicrobial activities (μg/ml) are quoted as MICs, MBCs, LD₅₀ or LD₉₀ (dose required to kill 50% or 90% bacteria).

At the outset we incubated *P. aeruginosa* PAO1 with peptides for 0, 1, 4 and 24 hours and noted that the MBCs (99.99% bacteria killed) were the same for each time point indicating that the peptides killed within an hour of administration (Fig. 5.1). We therefore carried all the reported broth dilution assays with a 1 hour incubation time (see Materials and Methods).

The test peptides were assayed against a panel of multiresistant, clinically relevant pathogens which included highly transmissible *P. aeruginosa* strains responsible for outbreaks in CF clinics in Manchester (C3425), Liverpool (H183), Brisbane (C4269), and Melbourne (C3781); *B. cepacia* complex isolates (ATCC
Chapter 5: Defensin related peptide 1

25416 and J2315); Sten. maltophilia isolates (C1980, C3625, C3626, C3627); Ralstonia pickettii (C3079) and Ralstonia eutropha (C3081); Bordetella bronchiseptica J3083; MRSA S113; and Candida albicans J2922. Other bacteria investigated were laboratory reference stains of E. coli (ATCC 25922), P. aeruginosa (PAO1), S. aureus (ATCC 25923), and E. faecalis (ATCC 29212).

Figure 5.1. Time-Kill Assay. A defensin killing assay (including PAO1 and Defr1 at 5 x MBC: 25 μg/ml) was prepared and a bacterial sample was removed at 30-minute intervals, and plated to determine the viable count (CFU/ml). The data shows that bacterial viability was reduced significantly within the first 30 minutes, and completely abolished by 60 minutes.

Defr1 exhibited broad spectrum antimicrobial activity, with MBCs ranging from 3-12.5 μg/ml against all organisms tested, with the exception of the two B. cepacia complex isolates. Defr1 also displayed antifungal activity against Candida
The high-resistance to Defr1 displayed by the B. cepacia complex isolates (MBC>100μg/ml) is in agreement with well documented evidence that these bacteria are highly resistant to cationic AMPs (301,302).

In contrast, the six cysteine analogue Defr1 Y5C displayed MBCs in the range 50-100 μg/ml against Gram-negative bacteria, ≥100 μg/ml against Gram-positive bacteria, and 25 μg/ml antifungal activity. The increased MBCs of Defr1 Y5C compared with Defr1 highlight a significant increase in potency for the 5 cysteine variant. Recent work by Wu et al. has demonstrated that the activity of human β-defensin 3 (HBD3) is independent of the number and connectivity of disulfide bridges within the molecule (230). To test this hypothesis we repeated the antimicrobial assays with both peptides after complete reduction with excess DTT. Reduced Defr1 Y5C had similar MBCs to its oxidised form, supporting the hypothesis that its activity is independent of the presence of disulfide bonds. However, the MBCs of reduced Defr1 were significantly higher than its oxidised form and similar to the MBCs for Defr1 Y5C. Thus, in this instance, it appears that the presence of disulfide bonds does influence the antimicrobial activity of the 5 cysteine containing peptide.
### Chapter 5: Defensin related peptide 1

<table>
<thead>
<tr>
<th>Organism</th>
<th>Strain</th>
<th>Description</th>
<th>MBC&lt;sup&gt;a&lt;/sup&gt; (µg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Defr1</td>
</tr>
<tr>
<td>Gram-negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. aeruginosa</em></td>
<td>PAO1</td>
<td>Lab Strain</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>C3425</td>
<td>Manchester</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>H183</td>
<td>Liverpool</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>C4269</td>
<td>Brisbane</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>C3781</td>
<td>Melbourne</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>P235</td>
<td>LPS Rough Mutant</td>
<td>6</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>ATCC</td>
<td>Culture Collection</td>
<td>8</td>
</tr>
<tr>
<td><em>B. cepacia</em></td>
<td>ATCC</td>
<td>Culture Collection - G1</td>
<td>&gt;100</td>
</tr>
<tr>
<td></td>
<td>J2315</td>
<td>Clinical Isolate - GIII</td>
<td>&gt;100</td>
</tr>
<tr>
<td><em>S. maltophilia</em></td>
<td>C1980</td>
<td>Clinical Isolate</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>C3625</td>
<td>Clinical Isolate</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>C3626</td>
<td>Clinical Isolate</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>C3627</td>
<td>Clinical Isolate</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>C3642</td>
<td>Clinical Isolate</td>
<td>12.5</td>
</tr>
<tr>
<td><em>B. bronchoseptica</em></td>
<td>J3083</td>
<td>Clinical Isolate</td>
<td>6</td>
</tr>
<tr>
<td><em>R. picketti</em></td>
<td>C3079</td>
<td>Clinical Isolate</td>
<td>3</td>
</tr>
<tr>
<td><em>R. eutropha</em></td>
<td>C3081</td>
<td>Clinical Isolate</td>
<td>3</td>
</tr>
<tr>
<td>Gram-positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. faecalis</em></td>
<td>ATCC</td>
<td>Culture Collection</td>
<td>6</td>
</tr>
<tr>
<td>S. aureus</td>
<td>ATCC</td>
<td>Culture Collection</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>25923</td>
<td>Clinical Isolate - MRSA</td>
<td>6</td>
</tr>
<tr>
<td>Fungi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>C. albicans</em></td>
<td>J2922</td>
<td>Clinical Isolate</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5.1 Antimicrobial activity of Defr1 and Defr1 Y5C. <sup>a</sup>MBC, values are the minimum concentration required to kill 99.99% of initial inoculum. <sup>b</sup>ND, not determined.
5.1.2 Salt Sensitivity of Defr1 and Defr1 Y5C

For further comparison we also analyzed the effect of high salt concentrations on the antimicrobial activities of Defr1 and Defr1 Y5C against *P. aeruginosa* (Fig. 5.2). The activity of Defr1 Y5C was extremely salt sensitive, killing only 10% of 1 x 10^5 bacteria in the presence of 25 mM NaCl and being completely inactive at 50 mM NaCl. In contrast, Defr1 displayed 100% killing at 25 and 50 mM NaCl, 84% killing at 150 mM NaCl and 28% killing even at 300 mM NaCl. These salt sensitivity results further highlight the significant differences between the antimicrobial activity of the 6 cysteine and the 5 cysteine Defr1 peptides.

![Graph showing percentage killed vs NaCl concentration](image)

**Figure 5.2. The effect of NaCl concentration on the antimicrobial activity of Defr1 and Defr1 Y5C.** *P. aeruginosa* PAO1 cells (1 x 10^5 CFU in 100 μl) were resuspended in 10 mM sodium phosphate buffer, 0.1% ISB, pH 7.4 containing 0, 25, 50, 150 and 300 mM NaCl. Bacteria were challenged with defensin at 4 x MBC (24 μg/ml for Defr1 and 200 μg/ml Defr1 Y5C) for 1 hour at 37 °C before serial dilutions of the assay mixture were plated on ISA plates and grown overnight at 37 °C. CFUs were counted and the percentage bacteria killed determined.
5.2 Structural properties of Defr1 and Defr1 Y5C

To rationalise the striking differences observed in antimicrobial activity between Defr1 and Defr1 Y5C, the structure of the two peptides was investigated. This was undertaken using a combination of native gel electrophoresis, reverse-phase chromatography, and high-resolution mass spectrometry using Fourier transform ion cyclotron resonance (FT-ICR-MS). In contrast to both NMR and x-ray crystallography, FT-ICR-MS requires only microgram amounts of each peptide for analysis.

5.2.1 Native Gel Electrophoresis

To analyse the structure of Defr1 and Defr1 Y5C they were subjected to non-denaturing electrophoresis on 16% Tricine gels (Fig. 5.3). In their oxidised states, the predominant forms of both peptides migrated with apparent molecular weights of ~7kDa which suggests that both defensins can dimerise under these conditions. Reduction of both peptides with excess DTT increased their apparent mobility to ~4kDa which indicates they are present as monomers and that disulfide bonds are involved in maintaining their tertiary structure. It is interesting to note the observation of significant amounts of Defr1 dimer even after incubation with DTT which suggests increased stability of the Defr1 dimer over the Defr1 Y5C dimer under reducing conditions.

Previous studies on various antimicrobial peptides have revealed that some, but not all, can dimerise. For example, HBD3 is a potent antimicrobial peptide isolated from human skin which has been the subject of intensive structural and functional studies (142). By native gel electrophoresis dimeric HBD3 was observed under oxidising conditions whereas monomeric HBD3 was detected after reduction
Chapter 5: Defensin related peptide 1

(211). Modeling studies predict that the dimer is formed through a combination of electrostatic salt bridges and H-bonding between amino side chains. Octomers, formed by amide backbone interactions, have also been observed in the crystal lattice of HBD2 (209). The formation of higher order aggregates of antimicrobial peptides has been proposed as one factor which contributes to their ability to disrupt bacterial membranes (303). Since oxidised Defr1 and Defr1 Y5C behave in a similar manner by gel electrophoresis, dimerisation alone cannot explain the significant differences observed in their antimicrobial activity.

![Figure 5.3. Colloidal Coomassie-stained non-denaturing 16% Tricine gel of oxidised and reduced Defr1 and Defr1 Y5C. lane 1, Defr1 reduced; lane 2, Defr1 oxidised; lane 3, Defr1 Y5C reduced; lane 4, Defr1 Y5C oxidised. Mass markers in kDa are indicated.](image)

**5.2.2 Characterization of the nature of dimerisation of Defr1 and Defr1 Y5C using FT-ICR mass spectrometry**

To identify the structural differences between Defr1 and Defr1 Y5C, which could account for the difference in antimicrobial activity, high-resolution FT-ICR mass spectrometry was employed. The resolving power and accuracy of this instrument
Chapter 5: Defensin related peptide 1

allows the determination of the mass and isotopic distribution of large biomolecules. A comparison of experimentally observed values to those predicted based on the empirical formula for the oxidised and reduced forms of both defensins can then be used to determine the number of disulfide bonds in each species.

Figure 5.4. Defr1 Y5C FT-ICR isotopic envelope. The deconvoluted isotopic envelope from FT-ICR nanospray analysis of oxidised Defr1 Y5C. (A), Triangles correspond to the isotopic envelope expected from a Defr1 Y5C monomer with all cysteines oxidised i.e. contains 3 disulfides. (B), shows the equivalent spectra for the dimeric form of Defr1 Y5C. Again the triangles represent the isotopic envelope expected from a dimer with all cysteines oxidised i.e. 6 disulfide bonds. The elemental composition of oxidised Defr1 Y5C monomer with 3 disulfides is C_{157}H_{324}N_{50}O_{43}S_{6}, average mass 3722.4490 Da and a Defr1 Y5C dimer with 6 disulfide bonds is C_{314}H_{508}N_{100}O_{86}S_{12}, average mass 7444.898 Da.
Analysis of the ion envelope and deconvolution for Defr1 Y5C suggested the presence of two species (Fig. 5.4 A and B). The isotopic distributions and masses fit very well to those predicted for a Defr1 Y5C monomer containing 3 disulfide bonds (elemental composition $C_{157}H_{254}N_{50}O_{43}S_{6}$, average mass 3722.4490 Da, Fig. 5.4A) and to a dimeric Defr1 Y5C with 6 disulfide bonds ($C_{314}H_{508}N_{100}O_{86}S_{12}$, average mass 7444.898 Da, Fig. 5.4B). Artificial dimerisation and formation of higher order aggregates has been observed for proteins under the conditions used in electrospray ionisation (ESI). However, we do not believe the dimer to be an artifact since it was also observed by native-gel electrophoresis in a 6-fold excess over the monomer. Nevertheless, since the dimer peak displays 16% relative abundance compared to the monomer, it is clear that the dimer is not stable to the electrospray process suggesting dimer formation by weak, non-covalent interactions.

![5 disulfide bonds](image)

**Figure 5.5. Defr1 FT-ICR isotopic envelope.** High resolution mass spectrum of oxidised Defr1 dimeric isoforms. The triangles correspond to the isotopic envelope calculated from the Defr1 amino acid sequence containing five disulfide bonds. The elemental composition for Defr1 dimer is $C_{326}H_{518}N_{100}O_{88}S_{10}$, average mass 7566.9761 Da.
Chapter 5: Defensin related peptide 1

FT-ICR analysis on oxidised Defr1 produced the mass spectrum which upon
denconvolution gave rise to one major species (Figure 5.5) whose isotopic
distribution matched that expected for a fully oxidised dimer containing 5 disulfide
bonds (elemental composition for Defr1 dimer C_{326}H_{518}N_{100}O_{88}S_{10}, average mass
7566.9761 Da). Interestingly, there was no peak corresponding to Defr1 monomer. In
contrast to that observed for Defr1 Y5C, the Defr1 dimer remains intact under
electrospray conditions, which implies a strong interaction between monomers. Since
a Defr1 dimer contains 10 cysteine residues, and a fully oxidised isoform has 5
disulfide bonds, dimerisation can only occur through formation of at least one
intermolecular disulfide bridge.

5.2.3 Dissociation of defensin dimers

Collision induced dissociation (CID) and electron capture dissociation (ECD) are
powerful mass spectrometry techniques used for analysing protein structure which
complement traditional methods (304,305). These fragmentation techniques were
employed to characterize the protein-protein interactions mediating dimerisation in
Defr1 Y5C and Defr1.

For CID analysis of the Defr1 Y5C dimer, a peak corresponding exclusively
to the +5 charge state of the dimer (m/z 1490) was isolated and subjected to
dissociation (Figure 5.6). The +5 ion readily dissociates into two monomers with +3
(m/z 1242) and +2 (m/z 1862) charge states (Fig. 5.6 – inset). Dissociation of the
dimer occurs without fragmentation of the peptide backbone demonstrating that the
Defr1 Y5C dimer is unstable thereby supporting our hypothesis that dimerisation is
mediated through non-covalent interactions.
Chapter 5: Defensin related peptide

Figure 5.6. Collision-induced dissociation (CID) spectra of Defr1 Y5C. The ion envelope for nanospray FT-ICR analysis of Defr1 Y5C is shown with each ion annotated as monomeric [M], or dimeric [D]. Inset - the dimer D^5+ ion with m/z 1488.905 (monoisotopic mass) was subjected to CID and gave rise to monomeric fragment ions with +3 and +2 charge states.

In stark contrast, the Defr1 dimer was stable to the same CID conditions used for Defr1 Y5C. When the stable dimeric ion corresponding to the +7 charge state (m/z 1082) was isolated and subjected to dissociation no significant monomeric ‘daughter’ fragments were observed (Fig. 5.7). Increasing the amount of gas into the collision cell still did not dissociate the dimer, but did allow us to partially sequence the peptide since it gave rise to a distinct b-type fragment series from the N-termini up to the location of the first cysteine (b2-b10) and a y-type fragment resulting from the loss of a C-terminal lysine (Fig. 5.7 - inset). Such stability under CID conditions indicates the Defr1 dimer is held together by covalent bonding.

100
Chapter 5: Defensin related peptide 1

Figure 5.7. Collision-induced dissociation (CID) spectra of Defri. The +7 charge state (m/z 1081.401, monoisotopic mass) generated by nanospray FT-ICR MS was isolated and subjected to collisional activation with argon for 40ms prior to detection. Fragments arising from activation are labeled. The Defri species remains intact but CID produces b and y ions arising from the cleavage of the amide backbone. Inset, magnification of the y ions at m/z 1200-1300.

These observations are complimented by ECD, where cleavage of Cys-Cys disulfide bridges is known to be a favored process (306). When the isolated +7 charge state of the Defri dimer was subjected to ECD, the molecule readily dissociated into monomers with charge states +2, +3 and +4 (m/z 1893, 1262, and 949 respectively (Fig. 5.8). On closer inspection (Fig. 5.8, inset), we observed low intensity species with masses +/- 16 Da either side of the +2 monomer peak, which indicate the gain/loss of a sulfur atom. We can explain the appearance of these species only if the Defri dimer is formed by a covalent intermolecular disulfide bond. Here, the ECD
process has cleaved the dimer into monomers by two mechanisms. Firstly, symmetric cleavage of the S-S bond to give a monomer signal with \( m/z \) 1892.5. The alternative pathway involves asymmetric cleavage of the C-S bond of the intermolecular disulfide bridge to give rise to two monomers; one with a persulfide SH at \( m/z \) 1909 and its corresponding partner having lost S at \( m/z \) 1876.5. The mechanisms of the cleavage of the S-S and S-C bonds are analogous to those previously observed by CID (307).

![Figure 5.8. Electron capture dissociation (ECD) of Defr1. Defr1 was subjected to nanospray FT-ICR MS and the \([D]^7+\) dimer ion (\( m/z \) 1081.401) was isolated and irradiated with electrons for 50ms. This caused the +7 ion to fragment into +5, +4, +3 and +2 monomers. The magnification of the +2 monomer ion series is, which results from both symmetric and asymmetric cleavage of the intermolecular S-S bond, is shown.](image)

The use of high-resolution FT-ICR, coupled with two dissociation techniques has revealed structural differences between Defr1 Y5C and Defr1 and for the first time identified a defensin dimer containing an intermolecular disulfide bond.
5.2.4 Reverse-phase high-pressure liquid chromatography

High-pressure liquid chromatography (HPLC) on a C$_{18}$ reverse-phase column was used to investigate structural differences between the oxidised and reduced forms of both peptides. Oxidised Defrl Y5C eluted as a single, sharp peak at 26.8 mins on a 20-35% acetonitrile gradient. This species had a mass of 3722.0 Da by ESI-MS in good agreement with the predicted value of 3722.3 Da for Defrl Y5C with the loss of six hydrogens. Upon incubation with excess DTT the single peak eluted earlier in the gradient at 23.6 mins and this shift in retention time suggested that reduction of the disulfide bonds caused a change in the structure of the defensin (the mass of this species is 3728.1 Da). It is thus apparent that Defrl Y5C folds to give a single species. Upon refolding from a denatured state we would normally expect a shift to shorter retention time on a reverse-phase column resulting from burial of hydrophobic residues. The refolding mechanism of defensins is not well understood but appears to be governed by the formation of the disulfide bonds. Since defensins are amphipathic and contain surface-exposed hydrophobic and cationic patches, these properties can be postulated to explain this unusual chromatographic behavior, which has also been observed for the oxidative folding of HBD2 (230).

In contrast, oxidised Defrl eluted as a broad peak between 22 and 24 mins which we were unable to resolve despite trying a range of conditions (under ESI conditions Defrl 7566.0 Da in agreement with a dimer with the loss of 10 protons). Upon reduction with excess DTT Defrl eluted as a single, sharp peak at 21.8 mins (mass 3788.0 Da, predicted mass for fully-reduced 3788.2 Da). This data suggests that during oxidative refolding Defrl has formed a number of different species all of which are dimeric. This phenomenon has been observed for HBD3 which was shown
Chapter 5: Defensin related peptide 1

to form several different oxidised isoforms from a single reduced precursor (230). The differences in the chromatographic behavior of Defr1 and Defr1 Y5C again suggest that their structures are dissimilar.

Figure 5.9. HPLC analysis of the oxidised and reduced forms of Defr1 and Defr1 Y5C. Analysis was performed on a C18 reverse phase HPLC column using a 15-40% acetonitrile gradient over 40 minutes. (A). Analysis of Defr1 Y5C, the red trace corresponds to oxidised Defr1 Y5C and the blue trace corresponds to reduced Defr1 Y5C. (B). Analysis of Defr1, the red trace corresponds to oxidised Defr1 and the blue trace corresponds to reduced Defr1.
5.2.5 Determination of Disulfide Connectivity of Defr1 and Defr1 Y5C

To determine their S-S connectivities, both defensins were subjected to protease cleavage and analysis of the resulting peptide fragments by MALDI-TOF and Q-TOF mass spectrometry. Trypsin cleavage of Defr1 Y5C produced two major fragments which were easily separated by HPLC (Table 5.2). Fragment 1 had a mass of 1711.87 Da and was unmodified by treatment with chymotrypsin, consistent with the species (D1-R7) + (C16-R20) + (C32-K34) containing 2 disulfide bonds (calculated mass 1711.84 Da). Because (C32-K34) contains two adjacent cysteines (Cys5 and Cys6) this data alone cannot be used to assign the absolute connectivity. However, from the mass of the first fragment we can assign the connectivity as either Cys1-Cys5 and Cys3-Cys6 or Cys1-Cys6 and Cys3-Cys5. The second tryptic fragment had a mass of 1817.54 Da, and was susceptible to chymotrypsin producing a species with mass 1532.70 Da. These results are consistent with the second tryptic fragment being (N8-R15) + (I23-K31) containing 1 disulfide bond (calculated mass 1816.86 Da), and its chymotrypsin derivative being (N8-Y14) + (I23-F30) (calculated mass 1532.66 Da), allowing the assignment of a disulfide bond between Cys2-Cys4 (Fig. 5.10).
Chapter 5: Defensin related peptide 1

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Calc. Mass (±1)(^a)</th>
<th>Obs. Mass(^a)</th>
<th>Calc. Mass Chymo (±1)(^b)</th>
<th>Obs. Mass Chymo(^b)</th>
<th>Disulfide Bridge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment 1</td>
<td>(D1-R7) + (C16-R20) + (C32-K34)</td>
<td>1711.84</td>
<td>1711.87</td>
<td>1711.84</td>
<td>1711.55</td>
</tr>
<tr>
<td>Fragment 2</td>
<td>(N8-R15) + (I23-K31)</td>
<td>1816.86</td>
<td>1817.54</td>
<td>1532.66</td>
<td>1532.70</td>
</tr>
</tbody>
</table>

Table 5.2. Determination of the disulfide bond connectivity of oxidised Defr1 Y5C. Trypsin digest of the peptide produced two fragments. The purified peptides (Fragment 1 and Fragment 2) were analyzed by MALDI-Tof. Each fragment was also digested separately with chymotrypsin before assignment.

\(^a\)Calculated and observed masses of tryptic fragments (Da)

\(^b\)Calculated and observed masses of fragments after treatment with chymotrypsin (Da)

\(^c\)Assignment of S-S connectivity for fragment 1 was determined to be C1-C5 C3-C6 using CID (see text).
Chapter 5: Defensin related peptide 1

Figure 5.10. Proteolysis of Def1 Y5C. Def1 was first digested with trypsin which produced two fragments, consistent with β-defensin topology. A chymotrypsin digestion of each fragment was used to verify assignment.

To distinguish between Cys5 and Cys6 in fragment 1 we used a Q-TOF tandem mass spectrometer and CID to sequence the peptides. Dissociation of the +3 charge state of fragment 1 resulted predominantly in the sequential loss of residues from the N-terminus of the fragment (D1-R7). However, a +2 species was observed with m/z 405.48 Da (deconvoluted mass 808.96 Da), which we attribute to (C16-R20) + (C33-K34) (calculated mass 808.41 Da), which allows the assignment of the disulfide bridge between Cys3-Cys6 and thus, by elimination, Cys1-Cys5. In summary, oxidised Def1 Y5C monomers contain the typical β-defensin S-S connectivity (Cys1-Cys5, Cys2-Cys4, Cys3-Cys6). It is worth noting that the NMR
structure of mBD-8 (NEPVSCIRNGGICQYRCIGLRHKIGTCGSPFKCCK) has been determined (PDB accession code 1E4R) (212). This defensin is very similar to Defr1 Y5C, the only differences being at the N-terminus (NEPVS of mBD-8 replaces DPVT of Defr1 Y5C). Since the sequences of mBD-8 and Defr1 Y5C are highly similar and they both display β-defensin connectivities we suggest that mBD-8 is a good model for the 3D structure of Defr1 Y5C.

Figure 5.11. CID Spectrum of Defr1 Y5C tryptic fragment 1. The 3+ charge state (m/z 572.22) was isolated and subject to CID. This led predominantly to sequential loss of N-terminal residues, yielding a series of y-ions (Main diagram, loss of residues highlighted. Inset B, cleavages shown in red). However, a 2+ species with m/z 405.48 was observed which was attributed to cleavage between Cys5 and Cys6 (Inset A and Inset B, highlighted in purple). This species allowed the assignment of Cys3-Cys6 and Cys1-Cys5).

In contrast to its six cysteine analogue, cleavage of oxidised Defr1 with a combination of trypsin and chymotrypsin produced a complex mixture of peptide products which proved impractical to separate by HPLC. Instead, the unfraccionated
mixture was analyzed by MALDI-TOF MS. The mass spectra obtained from analysis of the reduced peptide products revealed that proteolysis had been incomplete, and many peptides were present which contained uncleaved, internal proteolysis sites (see table 5.3A). Analysis of the oxidised tryptic digest peptide products enabled the assignment of disulfide connectivities (see table 5.3B). Peaks were observed with masses consistent with peptide fragments with Cys2-Cys3, Cys2-Cys4, Cys3-Cys4, and Cys4-Cys4 disulfide bonds. To confirm the initial assignment of these masses and identify potential disulfide bonds we again carried out N-terminal sequencing of selected peptides by CID. This allowed us to confirm the identity of the N-terminal fragment (D1-R7), and peptides containing disulfide bridges between Cys2-Cys3 and Cys2-Cys4.

Our analysis revealed Defr1 has S-S connectivities not typical of either α or β defensins. It appears from our proteolysis/mass mapping data that certain cysteine residues (e.g. Cys2) can form disulfide bonds with more than one other cysteine. This can only be explained if Defr1 is a mixture of topologically different dimeric isoforms, rather than a single species with a defined connectivity. The HPLC analysis also supports the unusual nature of Defr1. The broad peak resulting from RP-HPLC analysis of oxidised Defr1 again suggests that Defr1 is heterogeneous. The high-resolution FT-ICR data prove that all of these isoforms are dimeric connected by an intermolecular disulfide bridge. However, it appears that the intermolecular disulfide bond is not formed exclusively between two equivalent cysteine residues from each monomer e.g. Cys5'-Cys5 or Cys4'-Cys4. Rather, our proteolysis and chromatography data suggest that the Defr1 dimer can be formed via
any two cysteine residues from each monomer from a possible 15 combinations e.g. Cys2'-Cys3.

Recent work on synthetic HBD3 led to the surprising discovery that it did not fold into a single isoform with typical β-defensin disulfide connectivity in vitro (230). Instead, at least 7 topologically distinct isoforms were isolated with unusual intramolecular connectivities, all of which displayed similar antimicrobial activity. It is noteworthy that the S-S connectivity of native, HBD3 purified from human skin has not been determined (142). Interestingly, none of these isoforms contained an intermolecular disulfide bond. The authors were unable to determine the individual connectivity of these isoforms, so devised a synthetic protocol which allowed the control of disulfide bond formation. We are currently exploring whether this, or a similar, strategy can be used to control the formation of a specific intermolecular disulfide bond in Defr1.
### Chapter 5: Defensin related peptide 1

#### (A)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Calc. Mass (+1)</th>
<th>Obs. Mass</th>
<th>Relative Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D1-R7)</td>
<td>863.46</td>
<td>863.71</td>
<td>72</td>
</tr>
<tr>
<td>(I6-Y14)</td>
<td>1023.50</td>
<td>1023.98</td>
<td>91</td>
</tr>
<tr>
<td>(I6-R15)</td>
<td>1179.60</td>
<td>1179.92</td>
<td>11</td>
</tr>
<tr>
<td>(H21-F30)</td>
<td>1046.50</td>
<td>1046.95</td>
<td>67</td>
</tr>
<tr>
<td>(N8-R20)</td>
<td>1452.71</td>
<td>1452.72</td>
<td>26</td>
</tr>
<tr>
<td>(C16-F30)</td>
<td>1588.80</td>
<td>1588.67</td>
<td>20</td>
</tr>
<tr>
<td>(R15-F30)</td>
<td>1744.90</td>
<td>1744.54</td>
<td>100</td>
</tr>
<tr>
<td>(C16-K34)</td>
<td>2051.01</td>
<td>2051.21</td>
<td>13</td>
</tr>
</tbody>
</table>

#### (B)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Calc. Mass (+1)</th>
<th>Obs. Mass</th>
<th>Relative Abundance</th>
<th>N-terminal Sequencing</th>
<th>Disulfide Bridge</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D1-R7)</td>
<td>863.46</td>
<td>863.36</td>
<td>72</td>
<td>DPVTY</td>
<td></td>
</tr>
<tr>
<td>(I6-R20)</td>
<td>1450.70</td>
<td>1450.83</td>
<td>100</td>
<td>NGGI</td>
<td>C2-C3</td>
</tr>
<tr>
<td>(H21-F30)</td>
<td>1719.88</td>
<td>1719.66</td>
<td>58</td>
<td>HKI</td>
<td>C2-C3</td>
</tr>
<tr>
<td>(R15-R20) + (H21-F30)</td>
<td>1760.90</td>
<td>1760.61</td>
<td>87</td>
<td>HKI</td>
<td>C3-C4</td>
</tr>
<tr>
<td>(I6-Y14) + (I23-F30)</td>
<td>1801.83</td>
<td>1801.50</td>
<td>47</td>
<td>HKI</td>
<td>C2-C4</td>
</tr>
<tr>
<td>(I6-Y14) + (H21-F30)</td>
<td>2066.98</td>
<td>2067.31</td>
<td>48</td>
<td>HKI</td>
<td>C2-C4</td>
</tr>
<tr>
<td>(H21-F30) + (H21-F30)</td>
<td>2089.99</td>
<td>2090.28</td>
<td>6</td>
<td>HKI</td>
<td>C4-C4</td>
</tr>
<tr>
<td>(I6-Y14) + (H21-K34)</td>
<td>2527.18</td>
<td>2526.82</td>
<td>11</td>
<td>HKI</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3 Determination of the disulfide bond connectivity of oxidised Defr1. Oxidised Defr1 was digested with trypsin and chymotrypsin and the peptide fragments analyzed by MALDI-TOF. (A) The peptide mixture from the digest was reduced with TCEP and masses assigned accordingly. (B) MALDI-TOF analysis of the peptide digest without reduction and assignment of disulfide bonds. Assignments were confirmed by N-terminal sequencing by CID peptide mass mapping where indicated.

\(^a\) N-terminal sequencing data was collected using CID to resolve peptides which could not be unambiguously assigned (see text).
5.3 Discussion

β-Defensins are characterized by having six conserved cysteines which are oxidised to form three disulfide bonds. It was assumed that these were required for antimicrobial activity but several recent studies have begun to explore their structural and functional roles. Analogues of bovine β-defensin BNBD-12 containing one, two and three disulfide bridges, synthetic peptides corresponding to the carboxy-terminal segment of bovine β-defensin BNBD-2 and a cysteine-free, full-length analogue of HBD3 all displayed antimicrobial activity suggesting that the three disulfide bridges are not absolutely required (229,230,308). Moreover, to explore whether high cationic charge is the only requirement for activity, small, linear peptides corresponding to regions of HBD3 were analyzed for their ability to kill different pathogens (309). Somewhat surprisingly, the least positively-charged peptides proved to be the most active, which suggests that defensins contain inherent subtle structural features which come into play upon interaction with the membrane target.

Previously we reported the discovery of a defensin (Defrl) with five cysteine residues expressed in murine heart and testis which had potent antimicrobial activity (298). This defensin lacks the first of the conserved six cysteines found in other members of the family. Here we explored the effect of this loss on the structural and antimicrobial properties of by comparison with its six cysteine analogue Defri Y5C. Defrl Y5C has modest, salt-sensitive, antimicrobial activity against a panel of microbial pathogens and analysis by electrophoresis and mass spectrometry suggest it forms a non-covalently bound dimer in solution. Peptide mass mapping revealed that this defensin has Cys1-Cys5, Cys2-Cys4, Cys3-Cys6 disulfide connectivity.
typical of a β-defensin. In contrast, Defr1 has significantly higher antimicrobial activity against *P. aeruginosa* even under raised salt concentrations. High resolution mass spectrometry combined with two complementary dissociation techniques revealed a dimer containing an intermolecular disulfide bond, a novel structural feature not previously observed within the defensin family. Proteolysis and HPLC analysis suggest that oxidised Defr1 is a complex mixture of dimer isoforms varying in their intra and intermolecular disulfide connectivities. Recent studies of defensins HBD3 and BNB2 have concluded that short linear peptide fragments as well as those containing only one intramolecular disulfide could have potential as therapeutic agents. Our results demonstrate that defensins containing intermolecular disulfide bonds have novel structural and antimicrobial properties and suggest that covalently-bound dimeric cationic peptides are promising targets for future study.
Chapter 6: Recombinant production of Defr1 and a Defr1 derivative.
Chapter 6: Recombinant production of Defr1 and derivatives

6.1 Cloning the *Defr1* gene and its derivatives

One of the factors which hamper studies of defensins has been the lack of large quantities of recombinant peptides. To address this issue we undertook a study to develop an expression system for defensin production in *E. coli*.

For this investigation we overexpressed and purified murine Defr1. Furthermore, we also overexpressed a derivative of Defr1, Defr1 cys, which contained only a single cysteine residue in the Cys 5 position (the other four cysteines were substituted by serine residues, See Appendix I). Upon refolding this peptide formed a dimeric molecule containing an intermolecular disulfide bond. Functional studies of this peptide have again highlighted that a significant increase in antimicrobial activity is linked with the presence of an intermolecular disulfide bond.

6.1.1 Cloning the *Defr1* gene

Oligonucleotide primers were designed in order to clone the *Defr1* gene from murine cDNA. The primers, Defr1 AlwNI For and Defr1 AlwNI Rev, incorporated an AlwNI restriction site and both the 3'- and 5'-end of the gene. After amplification by PCR, the product was first cloned into the PCR2.1 vector before digestion with AlwNI and ligation into the commercial plasmid pET31b. The resulting construct, pET31b/Defr1, encoded a 125aa ketosteroid isomerase protein tag (KSI) at the 5'-end of the defensin gene (Figure 6.1). The DNA sequence between the KSI tag and defensin gene contained a unique methionine residue, which would facilitate cleavage of the expressed fusion protein by cyanogen bromide.
6.1.2 Designing the Defr1 1cys gene, a one-cysteine analogue of Defr1

Earlier work using synthetic Defr1 highlighted the presence of an intermolecular disulfide bond (see chapter 5). This unique structural feature had dramatic effects on both the structure and antimicrobial activity of the peptide.

In order to investigate these effects further we planned a strategy for the expression of a derivative of Defr1 which contained only a single cysteine, located in the Cys 5 position. Within this analogue of Defr1, named Defr1 1cys, the remaining cysteine were substituted for serine residues. The construction of the Defr1 1cys gene was achieved by recursive PCR using the four overlapping primers- named Defr1...
Chapter 6: Recombinant production of Defr1 and derivatives

1Cys P1-P4 (310). The gene incorporated *Alw*NI sites at the 3' and 5'-ends and was designed using codons which were optimised for expression in *E. coli*.

The *Defr1* 1cys gene was subsequently cloned into pET31b as described in section 6.1.1 and the resulting construct was named pET31b/Defr1 1cys.

### 6.2 Overexpression and purification of KSI-defensin fusion protein

*E. coli* BL21 Rosetta(DE3)pLysS were used to express both recombinant defensins. After induction with IPTG, these competent cells facilitated high level overexpression of both constructs as insoluble inclusion bodies. Expression was monitored by SDS-PAGE (Figure 6.2).

The cells were lysed by treatment with lysozyme and freeze-thawing and the inclusion bodies were purified by a series of washes as outlined in *Materials and Methods*. Typically, after purification inclusion bodies were 80% pure and the yield was in excess of 200 mg/l of cell culture.

![Figure 6.2](image)

**Figure 6.2. Expression, purification and cleavage of KSI-defensin fusion peptide.** **A.** Overexpression of KSI-Defr1. **Lane 1,** cell extract before induction. **Lane 2,** cell extract after 8 hour induction. **B.** Cleavage of KSI-Defr1. **Lane 1,** purified KSI-Defr1 fusion protein. **Lane 2,** KSI-Defr1 after cyanogen bromide cleavage.
6.3 Cyanogen bromide cleavage of KSI-defensin fusion protein

After purification, the KSI defensin fusion protein was subject to sulfitolysis. This procedure protected the free cysteine residues by capping each cysteine thiol with a sulfite group. This protection strategy has been successfully used in the expression of recombinant insulin (311). The protected fusion protein was then subject to cleavage by excess cyanogen bromide. Great care was taken to ensure that the cleavage reaction was performed under nitrogen and in the absence of light. The reaction was allowed to proceed for 16 hours. Generally, cleavage was achieved in good yield (>80%, see figure 6.2B).

Separation of the cleaved defensin from the KSI tag after cleavage proved more problematic. For Defr1, this was finally achieved by freeze-drying the cleaved mixture and extracting with a minimum amount of 40% acetonitrile/60% water/0.01% TFA. The highly hydrophobic KSI is insoluble in this solvent, and the protected Defr1 was extracted in good yield (Figure 6.3). For the cleaved Defr1 Y5C extraction was easily achieved using 10mM sodium phosphate buffer (pH 7.5).
6.4 Deprotection and refolding recombinant defensins

6.4.1 Deprotection and refolding Defr1

Deprotection of Defr1 was monitored by LC-MS. Analysis of the protected-Defr1 revealed a mass of 4188.0 Da, consistent with the mass of Defr1 with the addition of 5 sulfite groups (calc. mass 4183.55 Da). After reduction with TCEP full deprotection was observed and the mass of the deprotected recombinant Defr1 was 3788.0 Da, consistent with the mass of fully reduced Defr1 (3788.2 Da).

Refolding Defr1 proved problematic. Initial attempts at reducing the concentration of the denaturing agent guanidine (GuHCl) by rapid dilution and dialysis resulted in precipitation of the peptide. The most successful results were achieved by dialysing against refolding buffer A at 4°C. However, this procedure still resulted in significant precipitation of the peptide (Table 6.1).

Addition of a small quantity of oxidised and reduced thiols (e.g. cysteine/cystine) to refolding buffers has been known to dramatically increase yield of refolding proteins containing disulfide bonds- as the thiol additives catalyse disulfide bond shuffling within the protein. Indeed, on addition of both oxidised and reduced glutathione and cysteine/cystine yields of refolded defensin increased dramatically (refolding buffers B and C). However, LC-MS analysis of the resulting oxidised Defr1 showed no intermolecular-disulfide bound dimer species (Table 6.1). Instead the majority of the peptide seemed to have formed two intramolecular disulfides, while the fifth cysteine residue was capped with the thiol additive (either glutathione or cysteine).
Chapter 6: Recombinant production of Defr1 and derivatives

<table>
<thead>
<tr>
<th>Refolding Buffer</th>
<th>Yield</th>
<th>Obs. Mass (Da)</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15%</td>
<td>3784.7</td>
<td>2 intramolecular disulfides, 1 reduced cysteine? (Calc. 3784.2 Da)</td>
</tr>
<tr>
<td>B</td>
<td>80%</td>
<td>4090.0</td>
<td>2 intramolecular disulfides, addition of glutathione (Calc. 4088.3 Da)</td>
</tr>
<tr>
<td>C</td>
<td>80%</td>
<td>3903.0</td>
<td>2 intramolecular disulfides, addition of cysteine (Calc. 3902.4 Da)</td>
</tr>
</tbody>
</table>

Table 6.1. Refolding of recombinant Defr1. Refolding Defr1 was attempted under three different conditions. Unfortunately, no oxidised dimer species was detected. \(^{a}\) Yield has been approximated.

6.4.2 Deprotection and refolding Defr1 1cys

In contrast to Defr1, Defr1 1cys underwent oxidation to form an intermolecular disulfide bound dimer in high yield. The mass of the protected peptide was found to be 3804.0 Da, in agreement with the calculated value of 3802.9 Da. Defr1 1cys was deprotected by reduction with excess TCEP and LC-MS analysis revealed the reduced mass to be 3724.3 Da (Calculated 3723.9 Da). After exhaustive dialysis to remove the reducing agent an intermolecular disulfide was formed in high yield (>90%) and the resulting Defr1 1cys covalent dimer was found to have a mass of 7446.0 Da (calculated 7445.9 Da). As this Defr1 derivative has only one cysteine residue there is only one possible isoform of the dimeric species, and the intermolecular disulfide bridge can be assigned as Cys5-Cys5.

Finally, dimeric Defr1 1cys was purified by cation exchange chromatography on a Tricorn S column using a linear gradient of 0 to 1M NaCl over 25 column volumes. The recombinant protein eluted between 300-350mM NaCl.
Chapter 6: Recombinant production of Defr1 and derivatives

6.5 Antimicrobial activity of recombinant defensins

6.5.1 Antimicrobial activity of recombinant Defr1

Although recombinant Defr1 could not be folded to form a covalent dimer, enough monomeric Defr1 to perform activity assays was isolated and purified. Defr1, which had been refolded using refolding buffer B and consequently had one cysteine capped with glutathione, was assayed against \textit{P. aeruginosa} PAO1. The recombinant defensin was dialysed into 0.01% acetic acid and the assay was performed as described in sections 5.1 and 7.6.

Recombinant Defr1 did exhibit antimicrobial activity and displayed an MBC of 50 µg/ml against PAO1 in both oxidised and reduced forms. In comparison with synthetic Defr1, we noted that the activity of the reduced forms of the peptides displayed identical activities. However, the oxidised form of recombinant Defr1, which we know lacks an intermolecular disulfide, displayed no increased activity. Monomeric recombinant Defr1 seems to display antimicrobial properties similar to Defr1 Y5C, with oxidation state having little effect on activity. These results again highlight the importance of the intermolecular S-S bond within dimeric Defr1.

6.5.2 Antimicrobial activity of recombinant Defr1 1cys

In contrast to recombinant Defr1, the Defr1 1cys mutant was successfully purified in a dimeric state, containing a single intermolecular disulfide bond. The peptide was assayed against both Gram negative \textit{(P. aeruginosa} PAO1 and \textit{E. coli} ATCC 25922) and one Gram positive bacterium (MRSA S113) as outlined in sections 5.1 and 7.6. The activity of oxidised Defr1 1cys was similar to synthetic Defr1 for all the bacteria tested (See Table 6.2). Furthermore, the antimicrobial activity of the recombinant
defensin analogue was reduced after reduction of the intermolecular disulfide bond with excess DDT. It is also interesting to note that reduced Defr1 1cys seems to be significantly more active than reduced Defr1. This may be explained by the change in the hydrophobic nature of the peptide by the substitution of four cysteine residues with serines. These substitutions would have an effect on the hydrophobicity of the molecule- a factor which is known to effect antimicrobial activity. This effect remains to be investigated in future work.

<table>
<thead>
<tr>
<th>Organism</th>
<th>MBC (µg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Def 1Cys Ox</td>
</tr>
<tr>
<td>Gram Negative</td>
<td></td>
</tr>
<tr>
<td><em>P. aeruginosa</em></td>
<td>6</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>&gt;3</td>
</tr>
<tr>
<td>Gram Positive</td>
<td></td>
</tr>
<tr>
<td><em>S. aureus</em></td>
<td>6-12</td>
</tr>
</tbody>
</table>

Table 6.2. Activity of recombinant Defr1 1cys. The activity of Defr1 1cys was determined against three species of bacteria. The peptide was tested in both oxidised and reduced forms. The data is displayed with the comparative data from synthetic Defr1.
Chapter 7: Materials & Methods
Chapter 7: Materials and Methods

7.1 General materials

7.1.1 General reagents

Chemicals and solvents were of the appropriate quality and were purchased from Amersham Bioscience, Promega, Sigma, Pierce, Biorad, Gibco BRL, Vivascience, New England Biolabs, Water Micromass, or Invitrogen unless otherwise stated.

_Aquifex aeolicus_ genomic DNA was a kind gift from Ronald V. Swanson (Diversa), Robert Huber and Karl Stetter (University of Regensburg). Synthetic defensin peptides were produced using standard solid phase synthetic procedures (Albachem).

7.1.2 Solutions and buffers

_TAE buffer_ — 40 mM Tris-HCl, 20 mM Acetic acid, 1 mM EDTA (pH 8.3).

**Buffer A** — 10 mM HEPES (pH 7.5).

**Buffer B** — 10 mM HEPES, 1 M NaCl (pH 7.5).

**6His-Binding Buffer** — 20 mM Tris-HCl, 0.5 M NaCl, 5 mM imidazole (pH 7.5).

**6His-Charge Buffer** — 100 mM NiSO₄.

**6His-Elution Buffer** — 20 mM Tris-HCl, 0.5 M NaCl, 1 M imidazole (pH 7.5).
Chapter 7: Materials and Methods

KSI-Resuspension Buffer – 50 mM Tris-HCl, 25% Sucrose, 1 mM EDTA, 0.1% NaAzide, 10 mM DTT (pH 8.0).

KSI-Lysis Buffer – 50 mM Tris-HCl, 1% Triton X-100, 100 mM NaCl, 0.1% NaAzide, 10 mM DTT (pH 8.0).

KSI-Wash Buffer (with Triton) – 50 mM Tris-HCl, 1% Triton X-100, 100 mM NaCl, 1 mM EDTA, 0.1% NaAzide, 1 mM DTT (pH 8.0).

KSI-1 M Guanidine Buffer – 50 mM Tris-HCl, 1% Triton X-100, 100 mM NaCl, 1 M Guanidine-HCl, 1 mM EDTA, 0.1% NaAzide, 1 mM DTT (pH 8.0).

KSI-6 M Guanidine Buffer – 50 mM Tris-HCl, 1% Triton X-100, 100 mM NaCl, 6 M Guanidine-HCl, 1 mM EDTA, 0.1% NaAzide, 1 mM DTT (pH 8.0).

Buffer C – 50 mM Tris-HCl, 8 M Guanidine (pH 8.0).

Buffer D – 50 mM Tris-HCl, 1 mM EDTA, 6 M guanidine (pH 6.2).

Buffer E – 50 mM Tris-HCl (pH 8.2).

Buffer F – 50 mM Tris-HCl, 1 mM EDTA (pH 6.2).

Buffer G – 50 mM Tris-HCl, 1 mM EDTA, 1 M NaCl (pH 6.2).
Refolding Buffer A- 50 mM Tris-HCl, 20% DMSO (v/v) (pH 7.8).

Refolding Buffer B- 50 mM Tris-HCl, 1 mM EDTA, 3 mM [R] Glutathione, 0.3 mM [O] Glutathione (pH 7.8).

Refolding Buffer C- 50 mM Tris-HCl, 1 mM EDTA, 3 mM [R] Cysteine, 0.3 mM [O] Cysteine (pH 7.8).

Chymotrypsin digestion Buffer- 50 mM Tris-HCl, 20 mM CaCl₂ (pH 8.2)

Trypsin digestion Buffer- 50 mM Tris-HCl, 0.1% sodium azide, 5 mM EDTA (pH 6.2).

7.1.3 Media

Sterilisation of media – All media was autoclaved at 121 °C for 15 mins at 15 psi prior to use.

Agar plates – bacto-agar (15 g/l) was added to the specified media to prepare agar plates.

Luria Bertani (LB) – tryptone (10 g/l), yeast extract (5 g/l), sodium chloride (10 g/l); pH adjusted to 7.5 with sodium hydroxide.
Chapter 7: Materials and Methods

2YT - tryptone (16 g/l), yeast extract (10 g/l), sodium chloride (5 g/l); pH adjusted to 7.5 with sodium hydroxide.

SOC - tryptone (20 g/l), yeast extract (5 g/l), sodium chloride (0.5 g/l), magnesium sulfate (5 g/l), glucose (3.2 g/l); pH adjusted to 7.5 with sodium hydroxide.

Iso-sensitest (IS) - 23.4 g/l (Oxoid Ltd.).

7.2 Molecular Biology – DNA manipulation

7.2.1 Bacterial cell lines

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Description/ Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10\textsuperscript{TM}</td>
<td>F\textsuperscript{+}mcrA D(mrr-hsdRMS-mcrBC) \Phi80lacZAM15 ΔlacX74 deoR recA1 araD139 Δ(ara-leu)7697 gal/U gal/K rpsL endA1 nupG</td>
<td>Used For transforming DNA ligations.</td>
</tr>
<tr>
<td>BL21(DE3)</td>
<td>F\textsuperscript{+}omp Thsds\textsubscript{B} (r\textsubscript{B} \textsuperscript{-} m\textsubscript{B} \textsuperscript{-}) gal dcm (DE3)</td>
<td>General purpose expression host.</td>
</tr>
<tr>
<td>Rosetta(DE3)pLysS</td>
<td>F\textsuperscript{+}omp Thsds\textsubscript{B} (r\textsubscript{B} \textsuperscript{-} m\textsubscript{B} \textsuperscript{-}) gal dcm (DE3) pLysSRARE2 (Cam\textsuperscript{R})</td>
<td>Expression host for expression of mammalian genes in E. coli.</td>
</tr>
</tbody>
</table>
Chapter 7: Materials and Methods

7.2.2 Oligonucleotide primers

The following oligonucleotide primers were using in this study. Restriction sites are highlighted.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPL For</td>
<td>TTCTTAACCATGGGCTTCAAAAACCTGATCTGG</td>
</tr>
<tr>
<td>BPL Rev</td>
<td>TTAAGGATCCTAGAACCAGACAGGCTGAACTCTCC</td>
</tr>
<tr>
<td>AccB A67</td>
<td>GTAACCAGGTTGAACAGGAAGAA</td>
</tr>
<tr>
<td>AccB Rev</td>
<td>GGAATCTCTAAGCCTGTGCTATAAAGG</td>
</tr>
<tr>
<td>AccB E114A</td>
<td>CTAAGGTATTAGAACAGCAGCTCTAAGGT</td>
</tr>
<tr>
<td>AccB E121A</td>
<td>ATGAACGAGATAAGCGAGGAGAGCGAGTGAGG</td>
</tr>
<tr>
<td>AccB E123A</td>
<td>GGTAGACGCGGTGCTGACGCAACCG</td>
</tr>
<tr>
<td>AccB E142A</td>
<td>AAAGCTCTAGCTGTTATGAC</td>
</tr>
<tr>
<td>AccB E141L</td>
<td>ATCTAGCTACCTAGGTGTTGTTAGAA</td>
</tr>
<tr>
<td>AccB E143A</td>
<td>GGAAGAGTTGAGCAGATCCCTCTGAGG</td>
</tr>
<tr>
<td>AccB K17L</td>
<td>CAGATGCTGATGGATCCAGTAACCTAC</td>
</tr>
<tr>
<td>DefrI AlwNI For</td>
<td>CAGATGCTGATGGATCCAGTAACCTAC</td>
</tr>
<tr>
<td>DefrI AlwNI Rev</td>
<td>CAGACATCTGTCACTTGCGATCCAGTTGAAAAAG</td>
</tr>
<tr>
<td>DefrI 1cys P1</td>
<td>CAGATGCTGATGGATCCAGTAACCTACAGTCGAAACGGCAGATAAGC</td>
</tr>
<tr>
<td>DefrI 1cys P2</td>
<td>ATCTTTATGCTTAGCAGCAAGGCTTGCTCGTACTGGTGAATGTCCCGTTT</td>
</tr>
<tr>
<td>DefrI 1cys P3</td>
<td>CTAGGCTAGAAGATTGGAACTTCTGAGATCTGTTTCAAAATGCAAG</td>
</tr>
<tr>
<td>DefrI 1cys P4</td>
<td>CAGCATCTGTCAGCTTGCAATTTGAAAGGATGCC</td>
</tr>
</tbody>
</table>

7.2.3 Storage of bacterial stocks

LB or 2YT medium containing the appropriate antibiotics were used for the short term storage of *E. coli*. Colonies of bacteria were stored on inverted agar plates at 4 °C for up to 28 days. For long term storage strains were frozen (-80 °C) in LB or 2YT media containing 20% glycerol.
7.2.4 Transformation of *E. coli* competent cells with recombinant plasmid DNA

Competent cells were transformed according to the manufacturers’ instructions. DNA (up to 40 ng) was added to an aliquot of competent cells and gently mixed. This was left on ice for 5 or 30 minutes before the cells were heat shocked (42 °C, 30 sec). The cells were then grown in 80 or 250 μl SOC medium at 37 °C for 1 hr. Finally, the cells were spread to dryness on selective agar plates and incubated at 37 °C overnight.

7.2.5 Preparation of plasmid DNA

Plasmid DNA was prepared using QiAprep® Spin Miniprep Kit (Qiagen) following the manufacturer’s instructions.

7.2.6 Digestion of DNA with restriction endonucleases

The required amount of DNA (0.5-1 μg) was treated with the appropriate amount of endonuclease and buffer and incubated for at least 2 hours at 37 °C. Blue/orange loading dye (Promega) was added and the sample was subjected to electrophoresis on agarose. The gels were viewed and photographed under UV-light.

7.2.7 Electrophoresis of DNA

The required amount of agarose was added to 1xTAE buffer (typically 1.2g / 100ml) and heated at 100 °C until it dissolved. The solution was allowed to cool to 55 °C and ethidium bromide was added to a final concentration of 0.5 μg/ml. The gel was then poured into the casting mould and allowed to set at room temperature. DNA was
loaded and electrophersed using a potential difference of 100 V for an adequate time to achieve separation.

7.2.8 Purification of DNA

DNA was purified from agarose using QIAquick® Gel Extraction Kit (Qiagen) following the manufacturer’s instructions.

7.2.9 Amplification and megaprimer mutagenesis of DNA

Polymerase chain reactions (PCRs) and sequencing reactions were performed using a Perkin Elmer 480 thermal cycler. A typical reaction contained – 2 Ready to Go PCR™ beads (Amersham Biosciences), DNA template (1µl), primer 1 (forward) (5µl, 10µM), primer 2 (reverse) (5µl, 10µM), and distilled water (39 µl). Reactions were overlayed with oil and cycled 30 times at 95 °C for 1 min, 54 °C for 1 min and 72 °C for 2 min. Heating at 72 °C for 10 min terminated the reaction. The PCR product was then subjected to agarose gel electrophoresis and the required band was excised and the DNA purified as described in section 9.11. In mutagenic PCR reactions the resulting PCR product (the megaprimer) was used as primer 2 in the second reaction.

7.2.10 Recursive PCR

The gene assembly technique used was based on the recursive PCR protocol as described by Prodromou and Pearl (310). A gene sequence encoding the Defr1 1Cys peptide was designed using codons optimised for high level expression in a bacterial host. Four oligonucleotide primers were then synthesised spanning the entire
sequence (Defr1 1Cys P1-4, see section 9.5), these were 40-50 bp in length and consecutive primers overlapped by 10-20 bp. *A/wNI* restrictions sites were incorporated at the beginning and end of the sequence to facilitate subsequent incorporation into the pET31b expression vector. A codon encoding a methionine residue was also introduced directly before the peptide sequence.

The recursive PCR (30 cycles; 2 min at 95 °C, 2 min at 56 °C, and 1 min 72 °C) was performed using a Perkin Elmer 480 thermal cycler. The reaction contained 20-30 pmol of the outermost primers, 2-3 pmol of the internal primers, 10 µl of 10 x Herculase buffer, 5 units of Herculase® polymerase (Stratagene), 4 µl of Dimethyl-sulfoxide (DMSO), 4 µl of 100 µM of each dinucleotide triphosphates (dNTPs: dATP, dCTP, dGTP, dTTP), and dH2O was added to a final volume of 100 µl. The PCR product was then subjected to agarose gel electrophoresis and the required band was excised and the DNA purified as described in section 9.11.

### 7.2.11 Direct cloning of PCR products

All PCR products were cloned into the pCR 2.1-TOPO vector using the TOPO TA Cloning® kit (Version J) using the manufacturer’s guidelines.

### 7.2.12 DNA sequencing

Automated DNA sequencing was performed on an ABI prism 377 DNA sequencer using the Sanger dideoxy chain termination method. Each sequencing reaction typically contained – DNA template (~5 pmol) (2 µl), primer (10 µM, 1 µl), distilled water (13 µl) and Big Dye 4.1 (4µl; PE Biosystems, UK). The reaction was overlayed with oil and cycled 30 times at 96 °C for 30 sec, 45 °C for 15 sec and 60
Chapter 7: Materials and Methods

°C for 4 min. Sequence data was analysed using Contig Express within the Vector NTI Advance™ V9 software package.

7.2.13 Cloning into plasmid vectors

The DNA fragment cut with suitable restriction enzymes (13 µl), the host vector cut with suitable restriction enzymes (4 µl), 10x T4 DNA ligase buffer (2 µl), and T4 DNA ligase (2 µl) were gently mixed and incubated overnight at room temperature. 2 µl of the reaction mixture was used to transform suitable competent cells.

7.3 Molecular Biology - Protein Expression & Purification

7.3.1 Polyacrylamide gel electrophoresis (PAGE)

SDS-PAGE was used to analyse proteins on the basis of their molecular mass. The technique used was the discontinuous buffer system of Laemmli. Alternatively, proteins were analysed on precast 12% Bis-Tris Nu-PAGE gels (Invitrogen) according to manufacture’s instructions.

Proteins were analysed in their native form by native gel electrophoresis (NGE) using 16% Tricine Nu-PAGE gels (Invitrogen). For NGE, samples were mixed with equal volumes of non-reducing Novex Tricine SDS sample buffer. Gels were visualised using Comassie blue or GelCode Blue Stain reagent (Pierce).

7.3.2 Large Scale Expression of A. aeolicus BPL and mutant

The pET28a/ aaBPL vector was transformed into E. coli BL21(DE3) cells (Novagen). A single colony was added to 200 ml LB supplemented with kanamycin (30 µg / ml) and grown overnight at 37 °C and 250 rpm. This seed culture was then
used to inoculate 4 litres of fresh growth medium and grown subject to the same
conditions to $A_{600} = 1.0$ before induction with isopropyl-1-thio-β-D-galactopyranoside (ITPG) to a final concentration of 1.0 mM. After a further 3 hours
the cells were harvested by centrifugation (5000 rpm for 15 minutes at 4 °C), washed
buffer A, and stored at -20 °C.

7.3.3 Purification of *A. aeolicus* BPL and mutant

Cells overexpressing BPL were resuspended in buffer A and disrupted by sonication
(15 pulses of 30 seconds at 30 second intervals) at 4 °C. The cell debris was removed
by centrifugation at 15,000 rpm for 20 minutes at 4 °C. One tablet of Complete™
Proteinase Inhibitor Cocktail (Roche) was added to the cell lysate before it was
heated for 20 minutes at 60 °C. The resulting precipitated cellular protein was
removed by centrifugation at 15,000 rpm for 20 minutes at 4 °C. The supernatant
was filtered through a 0.45 μM membrane before it was loaded onto a Resource-S
cation exchange column (Amersham Biosciences, 20 °C) equilibrated with buffer A.
The BPL protein was eluted with a linear salt gradient (0-100% buffer B) over 20
column volumes. Fractions containing BPL were run on either SDS-PAGE or Nu-
PAGE gels and those fractions judged to be 95% pure were pooled and stored in
buffer A containing 20% glycerol (v/v) at -20°C. Protein concentration was
determined using the Bio-Rad protein assay kit. The aaBPL K19A mutant was
expressed and purified in a similar manner.
7.3.4 Large scale expression of His6-tagged *A. aeolicus* BCCPΔ67 and mutants

Overexpression of *A. aeolicus* BCCPΔ67 was achieved by transforming *E. coli* BL21(DE3) cells with the plasmid pET6H/aaAccBA67. A single colony was added to 200ml 2YT supplemented with ampicillin (100 μg / ml) and grown overnight at 37 °C and 250 rpm. This seed culture was then used to inoculate 4 litres of fresh growth medium and grown subject to the same conditions to $A_{600} = 1.0$ before induction with IPTG (1.0 mM final concentration). After a further 3 hours the cells were harvested by centrifugation (5000 rpm for 15 minutes at 4 °C), washed in binding buffer, and stored at -20 °C.

7.3.5 Purification of His6-tagged *A. aeolicus* BCCPΔ67 and mutants

Cells overexpressing His$_6$-BCCPΔ67 were resuspended in binding buffer (5 ml per gram of wet cell paste) and disrupted by sonication (15 pulses of 30 seconds at 30 second intervals) at 4 °C. The cell debris was removed by centrifugation at 15,000 rpm for 20 minutes at 4 °C, after which the supernatant was filtered through a 0.45μm membrane prior to chromatography.

The cell lysate was loaded onto a Hitrap® chelating affinity column (Amersham Biosciences, 20°C) previously charged with 6His-charge buffer and equilibrated with 6His-binding buffer. The column was then washed with 5 column volumes of 6His-binding buffer before bound material was eluted using a linear gradient of 0-100% 6His-elution buffer. Fractions were analysed by UV spectroscopy and SDS-PAGE, those containing BCCPΔ67 were pooled and dialysed overnight against 4 litres of buffer A at 20°C.
Chapter 7: Materials and Methods

Apo- and holo BCCPΔ67 were separated by applying the BCCPΔ67-containing fractions eluted from the nickel column onto a Mono-Q column (Amersham Biosciences, 20 °C) pre-equilibrated in buffer A. The column was then washed with 20 column volumes of Buffer A, before the protein was eluted with a salt gradient (0-100% buffer B) over 25 column volumes. The elution was monitored by UV spectroscopy and fractions containing apo- and holo-BCCPΔ67, screened by LC-MS, were pooled and stored in buffer A containing 20% glycerol (v/v) at −20°C. Due to the low proportion of aromatic residues in BCCP Δ67, protein concentration was evaluated by measuring the absorbance at 280nm and using the molar extinction coefficient 3960 M⁻¹ cm⁻¹, calculated using the Vector NTI Advance™ V9 software.

The expression and purification of the BCCPΔ67 mutants were performed essentially as above and the resulting protein was stored at the same conditions.

7.3.6 Overexpression of Defr1-KSI and Defr1 1Cys-KSI fusion protein as inclusion bodies

The pET31b/Defr1 vector was transformed into *E. coli* Rosetta(DE3)pLysS cells (Novagen). A single colony was added to 200ml 2YT media supplemented with ampicillin (100 μg/ml) and chloramphenicol (10 μg/ml) and grown overnight at 37 °C and 250 rpm. This seed culture was then used to inoculate 4 litres of fresh growth medium and grown subject to the same conditions to A600 = 0.3 before induction with ITPG (1.0 mM final concentration). After a further 5 hours the cells were harvested by centrifugation (5000 rpm for 15 minutes at 4 °C).
7.3.7 Purification of Defl1-KSI and Defr1 1Cys-KSI fusion protein

Cells overexpressing KSI-fusion protein were resuspended in KSI-resuspension buffer (5 ml per gram of wet cell paste) and disrupted by sonication (4 pulses of 30 seconds at 30 second intervals, 50% power) at 4 °C. Following sonication 100 µl lysozyme (50 mg/ml), 250 µl DNaseI (10 mg/ml), and 50 µl MgCl₂ (500 mM) were added and the sample mixed for 10 seconds using a vortexer. Lysis buffer (5 ml per gram of wet cell paste) was then added, the sample mixed, and incubated at room temperature for 45 minutes. After the addition of 700 µl of EDTA (500 mM, PH 8.0), the mixture was flash-frozen in liquid nitrogen and then thawed at 37 °C for 45 minutes. 500 µl MgCl₂ (500 mM) was subsequently added before the sample was incubated at room temperature. After the viscosity of the sample decreased, typically 30-60 minutes, 700 µl EDTA (500 mM) was added and the sample placed on ice. Crude inclusion bodies were collected by centrifugation (11,000 g, 20 minutes, 4 °C).

The inclusion bodies were first washed by resuspending in KSI-Wash Buffer (10 ml/g original cell pellet), disrupting by sonication (4 pulses of 30 seconds at 30 second intervals, 50% power) at 4 °C, and harvesting by centrifugation (11,000 g, 20 minutes, 4 °C). This step was repeated three times. Finally, the inclusion bodies were washed twice in KSI-1 M Guanidine Buffer (10 ml/g original cell pellet) and collected by centrifugation (11,000 g, 20 minutes, 4 °C). The purity of the inclusion bodies after each wash step was monitored by SDS-PAGE.
Chapter 7: Materials and Methods

7.3.8 Protection of KSI fusion proteins by sulfite addition

Washed inclusion bodies were dissolved in KSI-6 M Guanidine Buffer (10ml/ g original cell pellet) and any insoluble debris was removed by centrifugation (30,000 g, 20 minutes, 4 °C). Sulfitolysis was performed by incubating the protein with 300 mM sodium sulfite and 60 mM sodium tetrathionate for 16 hours at 20 °C. The sample was subsequently precipitated by either addition of an equal volume of 100mM ZnCl₂ or by dialysis against 4 litres of water for 2 hours at 20 °C. Finally the protected fusion protein was collected by centrifugation (11,000 g, 20 minutes, 4 °C).

7.3.9 Cleavage and separation of KSI fusion protein

The protected fusion protein was dissolved in a minimum amount of 80% formic acid under nitrogen, and cyanogen bromide was added to a final concentration of 100 mM. After incubation for 16 hours in the absence of light and under nitrogen the sample was freeze dried.

Defr1 was separated from the KSI tag by extraction with a minimum amount 40% acetonitrile/ 60% water/ 0.01% TFA for 1 hour at 20 °C. In contrast the Defr1 1Cys mutant was extracted using 10 ml 10mM sodium phosphate buffer (pH 7.5). The resulting suspensions were centrifuged (30,000 g, 20 minutes, 4 °C), and the soluble fraction containing the cleaved peptide was further clarified by filtration using a 0.22 μm filter. The purity and the cleaved peptide was analysed by SDS-PAGE and LC-MS. The organic solvent was removed from Defr1 by freeze-drying overnight.
7.3.10 Deprotection and refolding Defr1

Defr1 peptide, dissolved in buffer C, was deprotected by treatment with 10 mM TCEP for 4 hours at 20 °C, and monitored by LC-MS. The excess TCEP was removed by precipitating the reduced peptide, which was achieved by dialysing against water at 20 °C for a further 4 hours. Reduced Defr1 was finally dissolved in buffer D.

Refolding of Defr1 was attempted by dialysis against either refolding buffer A, B or C. The time and temperature of each dialysis was varied in an attempt to improve yield (See section 6.4.1).

7.3.11 Deprotection and oxidation of Defr1 1Cys

The Defr1 1Cys peptide was deprotected by treatment with 10 mM TCEP for 4 hours at 20 °C and monitored by LC-MS. The sample was oxidised by removal of excess TCEP by exhaustive dialysis against buffer E. The formation of the intermolecular disulfide bond was monitored by LC-MS.

7.3.12 Final purification of Defr1 and Defr1 1Cys

Both Defr1 and Defr1 1Cys were dialysed into buffer F and loaded onto a Tricon-S cation exchange column (Amersham Biosciences) equilibrated with buffer F. The column was washed with 20 column volumes of buffer F, before the protein was eluted with a 0-100% gradient of buffer G over 25 column volumes. The elution was monitored using UV spectroscopy, SDS-PAGE and LC-MS. Fractions containing pure recombinant defensin were stored at -20°C.
Chapter 7: Materials and Methods

7.4 Protein Chemistry

7.4.1 Assay of A. aeolicus BPL

BPL activity was assayed by measuring the incorporation of $[^{14}\text{C}]$biotin into purified BCCPΔ67, in a similar way to that previously described (98). Except where stated otherwise, the reaction contained 10 mM HEPES (pH 8.5), 100 μM ATP, 200 μM MgCl₂, 10 μM biotin, 1 μM $[^{14}\text{C}]$biotin (specific activity 54mCi/mmol), 0.1 mg/ml bovine serum albumin, and 400 μM apo-BCCPΔ67. The reaction was initiated by the addition of purified BPL to a final concentration of 1 μM, and incubated at 70°C for 30 min. The reaction was terminated by the addition of ice-cold trichloroacetic acid (final concentration 25% w/v), and incubation on ice for 30 min. The resulting protein precipitate was removed by centrifugation (15,000 rpm for 10 minutes). Aliquots of the supernatant were added to 5 ml of scintillation fluid (ICN biomedicals), and radioactivity was measured using a Tri-carb 210 OTR liquid scintillation counter (Packard).

For kinetic analysis each of the substrate concentrations (biotin, ATP, BCCP) were varied accordingly. Values for $K_m$ and $V_{max}$ were determined by Michaelis-Menten analysis on SigmaPlot 2001 software. In some assays, to obtain sufficiently high levels of activity for accurate detection, it was necessary to continue until greater than 10% of the limiting substrate had been used. In these instances the data was transformed using the method of Lee and Wilson and plotted as transformed values $s'$ and $v'$ (312).
7.4.2 Limited proteolysis of *A. aeolicus* BPL

Proteolysis of apo-BPL and substrate-bound-BPL were investigated using the proteases trypsin (Sigma) and chymotrypsin (Promega). Substrate-bound BPL was prepared by incubating BPL (15 µM) for 20 min at 60 °C with saturating amounts of biotin (40 µM), MgATP (2 mM), or both. The samples were then cooled for 10 min before treatment with protease (with a final protease:substrate ratio of 1:20 (w/w)) and incubation at 37 °C for 30 min. Digestion was terminated by the addition of SDS sample buffer and boiling for 5 minutes, and the extent of proteolysis was analysed by SDS-PAGE.

7.4.3 Chemical crosslinking of *A. aeolicus* BPL and BCCPΔ67 mutants

Purified BPL (15 µM) and BCCPΔ67 mutants (45 µM) were covalently cross-linked using 1-ethyl-3-(dimethylamino-propyl)-carbodiimide (EDC, 10 mM) at 60 °C for 60 mins. Aliquots were withdrawn at various time intervals, quenched with ammonium acetate (40 mM), and analysed by SDS-PAGE.

The cross-linked complex was prepared on a larger scale and separated from BPL and BCCPΔ67 by gel filtration. To prepare the complex we incubated 5 mg each of BPL and BCCP, EDC (10 mM) in a final volume of 5 ml 10 mM HEPES (pH 8.5) for 60 min at 60 °C. After incubation the reaction was quenched with 40mM hydroxylamine and analysed by SDS-PAGE. The mixture was passed through a Superdex 75 column (AP-Biotech) equilibrated in 10 mM HEPES (pH 8.5) and 100 mM NaCl and the purified protein was stored at −20 °C.
7.4.4 Chemical Biotinylation

Chemical biotinylation was performed essentially as previously described (32). Briefly, the purified BCCP mutants, BCCPΔ67ΔTh and BCCPΔ67ΔTh4KR (90 μM, 10mM HEPES pH 8.0), were incubated with biotinyol-hydroxy-succinimide (250 μM, in DMSO) at room temperature for four hours. The solution was then dialysed against 10mM HEPES pH 8.0 overnight with one buffer change; the reaction was monitored by LC-MS. The biotinylated product was separated from unreacted apo-protein by HPLC using a C5 reverse phase column and a gradient of 50-65% acetonitrile over 20 minutes.

7.5 Protein Characterisation

7.5.1 High performance liquid chromatography (HPLC)

HPLC was performed using a Beckman System Gold analytical instrument fitted with an autosampler and a diode array. Typically, proteins were monitored at 214 and 280 nm and data was analysed using Beckman Gold software. The instrument was a kind gift from Glaxo-Smith-Kline, and was specially customised by Parcel-Force.

7.5.2 Liquid chromatography-mass spectrometry (LC-MS)

LC-MS was performed on a MicroMass Platform II quadrupole mass spectrometer equipped with an electrospray ion source. The spectrometer cone voltage was ramped from 40 to 70 V and the source temperature set to 140 °C. Protein samples were separated with a Waters HPLC 2690 with a Jupiter C5 reverse phase column (5 μm, 250 x 4.6 mm, Phenomenex) directly connected to the spectrometer. Proteins
were eluted from the column with a 5-95% acetonitrile (containing 0.01% TFA) gradient at a flow rate of 0.2 ml/min. Peptide samples were separated with a Jupiter C18 reverse phase column (5 μm, 250 x 4.6 mm, Phenomenex). Typically, separation was achieved using a 20-50% acidified acetonitrile gradient over 60 min, and a flowrate of 1 ml/min. The total ion count in the range 500-2000 m/z was scanned at 0.1 s intervals. The scans were accumulated, spectra combined and the molecular mass determined by the MaxEnt and Transform algorithms of the Mass Lynx software (Micromass, U. K.).

7.5.3 FT-ICR Analysis

Characterization of the oxidation state of the defensin peptides Defr1, and Defr1 Y5C, was performed using the accurate mass capabilities of a 9.4 T FT-ICR mass spectrometer (Bruker Daltonics). Using this instrument, two dissociation methods were applied to the isolated peptides. For Sustained Off Resonance Collision Induced Dissociation (SORI-CID) a given charge state was isolated by sweep excitation and subjected to CID with Argon as the collision gas for 500 ms. To perform ECD a given charge state was isolated by sweep excitation and subjected to electron irradiation for 50 ms using a barium oxide coated high-surface area (5mm diameter) dispenser cathode (HeatWave) An experimental pulse sequence which combined ECD and SORI-CID was also employed. This work was performed under the supervision of Dr. Pedita Barran and with the help of Dr. Nick Polfer.
Chapter 7: Materials and Methods

7.5.4 Protease digestion and peptide mass mapping of defensins

The peptide mass mapping for Defrl and Defrl Y5C was conducted using a MALDI-TOF (Tof-Spec 2E, Micromass, U.K.). Enzymatic digestion of Defrl (500 µg/ml) with trypsin (100 µg/ml) and chymotrypsin (100 µg/ml) was performed at 37 °C in chymotrypsin digestion buffer and allowed to proceed for four hours, before termination by addition of 0.05% TFA. For analysis of the reduced proteolytic digest 1 µl of the peptide mixture was removed and reduced with 100 mM TCEP. For MALDI-MS analysis oxidised and reduced peptide digests were desalted and concentrated using C18 ZipTips (Millipore Corp.). 1 µl of analyte and 1 µl of matrix solution (a saturated solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile with 0.1% TFA) were mixed, left to air-dry on a MALDI plate and MALDI-MS results were obtained in the positive mode.

For analysis of Defrl Y5C, peptide of concentration 500 µg/ml was digested with trypsin (100 µg/ml). The reaction was performed in trypsin digestion buffer and was allowed to proceed for four hours at 37 °C before termination by addition of 0.05% TFA. Cleavage products were separated by HPLC using a Phenomenex C18 Jupiter column (5 µm, 250 x 4.6 mm) and a 10-35% acetonitrile gradient and the two major species were collected and freeze-dried. The resulting peptides were reconstituted in chymotrypsin digestion buffer and each split into three aliquots. One aliquot was analyzed in its oxidised state; the second was reduced with 20 mM TCEP; and the third was digested with chymotrypsin (100 µg/ml) for 4 hrs at 37 °C. All samples were purified and concentrated using C18 ZipTips and analyzed by MALDI-TOF MS as described above.
Chapter 7: Materials and Methods

7.5.5 Collision induced dissociation (CID) of peptides

CID analysis performed using a Q-TOF tandem mass spectrometer (Micromass, U.K.) equipped with a nanospray source. Samples were typically analysed in 50% acetonitrile, 0.1% TFA and sprayed from gold/palladium coated tips (Proxeon Biosystems). Specific ions were mass selected by the quadrupole MS, subjected to CID using argon and the resulting fragments were analyzed in the TOF MS.
7.6 Microbiology

7.6.1 Bacterial isolates

All bacterial isolates were provided by Prof. John Govan from the Edinburgh Cystic Fibrosis Microbiology Laboratory and Strain Repository (ECFML).

<table>
<thead>
<tr>
<th>Organism</th>
<th>Strain</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gram-negative</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. aeruginosa</em></td>
<td>PAO1</td>
<td>Type strain</td>
</tr>
<tr>
<td></td>
<td>C3425</td>
<td>Manchester Epidemic Strain</td>
</tr>
<tr>
<td></td>
<td>H183</td>
<td>Liverpool Epidemic Strain</td>
</tr>
<tr>
<td></td>
<td>C4269</td>
<td>Brisbane Epidemic Strain</td>
</tr>
<tr>
<td></td>
<td>C3781</td>
<td>Melbourne Epidemic Strain</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>ATCC 25922</td>
<td>Type strain</td>
</tr>
<tr>
<td><em>B. cepacia</em></td>
<td>ATCC25416</td>
<td>Type strain</td>
</tr>
<tr>
<td><em>B. cenocepecia</em></td>
<td>J2315/LMG16656</td>
<td>Type strain</td>
</tr>
<tr>
<td><em>S. maltophilia</em></td>
<td>C1980</td>
<td>Clinical Isolate</td>
</tr>
<tr>
<td></td>
<td>C3625</td>
<td>Clinical Isolate</td>
</tr>
<tr>
<td></td>
<td>C3626</td>
<td>Clinical Isolate</td>
</tr>
<tr>
<td></td>
<td>C3627</td>
<td>Clinical Isolate</td>
</tr>
<tr>
<td><em>B. bronchoseptica</em></td>
<td>J3083</td>
<td>Clinical Isolate</td>
</tr>
<tr>
<td><em>R. picketti</em></td>
<td>C3079</td>
<td>Clinical Isolate</td>
</tr>
<tr>
<td><em>R. eutropha</em></td>
<td>C3081/LMG1199</td>
<td>Clinical Isolate</td>
</tr>
<tr>
<td><strong>Gram-positive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. faecalis</em></td>
<td>ATCC 29212</td>
<td>Type strain</td>
</tr>
<tr>
<td><em>S. aureus</em></td>
<td>ATCC 25923</td>
<td>Type strain</td>
</tr>
<tr>
<td></td>
<td>S113/ J2918</td>
<td>Clinical Isolate - MRSA</td>
</tr>
<tr>
<td><strong>Fungi</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>C. albicans</em></td>
<td>J2922</td>
<td>Clinical Isolate</td>
</tr>
</tbody>
</table>
7.6.2 Preparation of defensin for antimicrobial assays

The synthetic peptides (Albachem, Edinburgh) were lyophilised (in 250 μg aliquots) and kept at −70 °C. Prior to use, the peptide was dissolved in 250 μl of 0.01% acetic acid to give a stock solution with a final concentration of 1 mg/ml. The peptide preparations were held at 4°C during use and after which at −70 °C.

Reduction of the peptides was performed by adding 10 mM dithiothreitol (DTT) and incubating at room temperature overnight. The oxidation state of each peptide prior to performing antimicrobial assays was determined by LC-MS.

7.6.3 Broth dilution antimicrobial assay of defensins

Test bacteria were grown to mid-logarithmic phase in ISB (Oxoid) and then diluted to 2×10⁶ CFU/ml in 10 mM sodium phosphate containing 1% (v/v) ISB, pH 7.4. The bacterial samples were washed three times in buffer to ensure removal of salt. Varying concentrations of test peptide (peptide range 1.5 μg/ml to 100 μg/ml) were incubated in 100 μl of the bacterial suspension at 37°C for 1 hour. 10-fold serial dilutions of the incubation mixture were plated on ISA plates, incubated at 37°C, and the number of colony forming units (CFU) determined the following day. The minimum bactericidal concentration (MBC) was determined as the concentration of peptide where we observed >99.99% killing of the initial inoculum. All assays were performed in duplicate and repeated on two independent occasions. The MBC was obtained by taking the mean of all the results, and experimental errors were within one doubling dilution. Dose-response curves could be determined from surviving CFU.
7.6.4 Time-kill assay of defensins

A defensin killing assay using *P. aeruginosa* PAO1 and Defr1 at 5 x MBC (25 μg/ml) was performed as described above. Bacterial sample was removed at 30-minute intervals, and plated on agar to determine the viable count (CFU/ml).

7.6.5 Salt sensitivity of defensins

The effect of salt on antimicrobial activity was tested by incubating 100 μl of 1-5 x 10^5 CFU of bacteria in 10 mM potassium phosphate, 1% (v/v) Iso-Sensitest, pH 7.4 which contained various concentrations of NaCl (0-300 mM). The bacteria were then challenged with peptide at a concentration four times the MBC.
Chapter 8: References


Chapter 8: References


151
Chapter 8: References


Chapter 8: References


156
Chapter 8: References


Chapter 8: References


Chapter 8: References


Chapter 8: References


163
Chapter 8: References


Chapter 8: References


Chapter 8: References


Chapter 8: References


Appendix 1: Recombinant proteins and synthetic peptides used in this study.
Defr1 and Mutants

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>10</th>
<th>20</th>
<th>34</th>
</tr>
</thead>
<tbody>
<tr>
<td>defb8</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
</tr>
<tr>
<td>defr1</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
</tr>
<tr>
<td>defr1 Y5C</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
</tr>
<tr>
<td>defr1 1cys</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
<td>IRNGGIQYRIGLRHKIGTGSFPCK</td>
</tr>
</tbody>
</table>
Appendix 2: Publications.
Biotinylation in the hyperthermophile *Aquifex aeolicus*

Isolation of a cross-linked BPL:BCCP complex

David J. Clarke, Joseph Coulson, Ranald Baillie and Dominic J. Campopiano

School of Chemistry, University of Edinburgh, UK

Biotin protein ligase (BPL) catalyses the biotinylation of the biotin carboxyl carrier protein (BCCP) subunit of acetyl CoA carboxylase and this post-translational modification of a single lysine residue is exceptionally specific. The exact details of the protein–protein interactions involved are unclear as a BPL:BCCP complex has not yet been isolated. Moreover, detailed information is lacking on the composition, biosynthesis and role of fatty acids in hyperthermophilic organisms. We have cloned, overexpressed and purified recombinant BPL and the biotinyl domain of BCCP (BCCPA67) from the extreme hyperthermophile *Aquifex aeolicus*. In vitro assays have demonstrated that BPL catalyses biotinylation of lysine 117 on BCCPA67 at temperatures of up to 70 °C. Limited proteolysis of BPL with trypsin and chymotrypsin revealed a single protease-sensitive site located 44 residues from the N-terminus. This site is adjacent to the predicted substrate-binding site and proteolysis of BPL is significantly reduced in the presence of MgATP and biotin. Chemical crosslinking with 1-ethyl-3-(dimethylamino-propyl)-carbodiimide (EDC) allowed the isolation of a BPL:apo-BCCPA67 complex. Furthermore, this complex was also formed between BPL and a BCCPA67 mutant lacking the lysine residue (BCCPA67 K117L) however, complex formation was considerably reduced using holo-BCCPA67. These observations provide evidence that addition of the biotin prosthetic group reduces the ability of BCCPA67 to heterodimerize with BPL, and emphasizes that a network of interactions between residues on both proteins mediates protein recognition.

**Keywords:** biotin protein ligase; *Aquifex aeolicus*; biotinylation; protein recognition; chemical crosslinking.

The enzymes of bacterial fatty acid biosynthesis have been suggested as good targets for the development of novel antibacterial agents since several natural product and synthetic inhibitors of this pathway are already known [1]. Moreover, significant differences in fatty acid biosynthesis between bacteria and mammals should allow selective inhibition of the microbial enzymes. The first committed step of bacterial fatty acid biosynthesis is catalysed by a multisubunit acetyl-CoA carboxylase [2]. This biotin-dependent complex is composed of biotin carboxylase, carboxytransferase and biotin carboxyl carrier protein (BCCP) subunits, the exact composition of which is species-specific. The *Escherichia coli* acetyl-CoA carboxylase has been intensively studied, because the subunits can be separated or expressed individually in an active form. Biotin is covalently bound to a specific lysine residue in the BCCP subunit [3,4]. Biotinylated enzymes transfer carbon dioxide from bicarbonate to organic acids to form cellular metabolites, using the biotin prosthetic group as a mobile carboxyl carrier [5]. Biotin protein ligase (BPL), also known as holocarboxylase synthase (HCS, EC 6.3.4.10) catalyses this post-translational attachment via a two-step reaction (Scheme 1 [6]).

\[
\text{Biotin} + \text{MgATP} \rightarrow \text{Biotinyl-5'-AMP} \rightarrow \text{PPi} \\
\text{Biotinyl-5'-AMP} + \text{apo-BCCP} \rightarrow \text{Biotinylated-BCCP} + \text{AMP}
\]

Scheme 1.

Genes encoding BPLs have been identified in a number of organisms, but the best-characterized BPL is the 35.3 kDa BirA protein from *E. coli* [7,8]. BirA is a bifunctional protein that can act as both an enzyme and a DNA-binding protein; it catalyses protein biotinylation when *in vivo* biotin concentrations are low, but becomes a repressor of the expression of biotin biosynthetic enzymes when biotin concentrations are increased. The crystal structure of the biotin-bound protein, determined at 2.3 Å resolution, shows the enzyme has three domains [9,10]: an N-terminal domain that contains a helix-turn-helix fold for DNA binding; a central catalytic domain, which contains a highly conserved GRGRRG motif shown to be involved in biotin binding [11]; and a small C-terminal domain which has been postulated to mediate dimerization with apo-BCCP [12]. The recent determination of the structure of a BirA dimer in
the absence of DNA provides insight into how the N-terminal DNA-binding domain interacts with the 40 bp biotin operator sequence [12]. The structure of the apo- and holo-forms of the biotinylation domain of \textit{E. coli} BCCP (known as BCCP-87) have been determined by X-ray crystallography and NMR [13-15]. The BCCP domain is a barrel consisting of two antiparallel \( \beta \)-sheets each containing four strands. The N- and C-termini are close together at one end, and the biotinylated lysine is exposed on a tight \( \beta \)-turn at the opposite face of the molecule. Surprisingly, the structures of the apo- and holo- forms are remarkably similar suggesting that biotinylation causes few significant changes in the domain tertiary fold.

To gain further insight into the detailed protein–protein interactions that control biotin transfer we have analysed the reaction between BPL and apo-BCCP from the hyperthermophilic organism \textit{Aquifex aeolicus} [16]. This bacteria grows optimally at 95 °C on hydrogen, oxygen, carbon dioxide and mineral salts. Enzymes from extremophiles (extremozymes) are offering new opportunities for biocatalysis as a result of their extreme stability [17,18].

Analysis of the \textit{A. aeolicus} genome identified BirA and BCCP homologues; the predicted BPL is from the group I class (which also includes \textit{M. tuberculosis}) which lack the N-terminal DNA-binding domain found in \textit{E. coli} BirA [19]. In \textit{E. coli}, we have expressed active \textit{A. aeolicus} BPL, the biotin-binding domain of \textit{A. aeolicus} BCCP as a His\(_6\) N-terminal fusion (BCCPA67) as well as an \textit{A. aeolicus} BCCP mutant lacking the active lysine residue (K117L). Biotinylation of apo-BCCPA67 by BPL was most efficient at 70 °C and we have carried out kinetic analyses and proteolysis experiments at this temperature. Furthermore, we describe the isolation of a chemically crosslinked BPL:BCCPA67 complex for the first time. This study is the first characterization of post-translational modification complex from a hyperthermophilic organism.

**Experimental procedures**

**Materials**

All chemicals used in the preparation of buffers were at least of reagent grade. Nu-PAGE gels were obtained from Invitrogen; restriction endonucleases were purchased from New England Biolabs; [\( ^{14} \text{C} \)]biotin (54 mCi mmol\(^{-1}\)) was from Amersham Biosciences; and 1-ethyl-3-(dimethylaminopropyl)-carbodiimide (EDC) was from Sigma. PCR was performed using Ready To Go PCRTM beads (Amersham Biosciences), and the isolated 723 bp fragment containing the \textit{A. aeolicus} BPL gene was cloned in \textit{NcoI} and \textit{BamHI} digested pET28a (Novagen), producing the expression vector pET28a/BPL. An \textit{NcoI}/\textit{BamHI} digest was performed on plasmid pCR2.1/BPL, and the resulting 259 bp fragment containing the truncated \textit{Bcp} gene was ligated in a \textit{NcoI}/\textit{BamHI}-digested pET28a (Novagen), producing the expression vector pET28a/BPL. An \textit{NcoI}/\textit{BamHI} digest was performed on plasmid pCR2.1/BccpA67 and the resulting 259 bp fragment containing the truncated \textit{Bcp} gene was ligated in a \textit{NcoI}/\textit{BamHI}-digested pET derivative (Novagen). The resulting expression vector, pET6H/BccpA67, produced a His\(_6\) fusion at the N-terminus of bccpA67.

A \textit{bcpA67} mutant gene encoding a mutation of the active site lysine to a leucine residue (K117L) was produced by the PCR megaprimer method [21]. The primers used were BCCPA67, BCCP-rev and BCCPA67 restriction digest was performed on the genomic DNA template by polymerase chain reaction using primers BPL-for and BPL-rev; and BCCPA67 and BCCP-rev, respectively. The PCR products were cloned into plasmid pCR2.1 (Invitrogen) using standard TOPO cloning procedures, yielding the plasmids pCR2.1/BPL and pCR2.1/BccpA67. Positive clones were sequenced to confirm the fidelity of the insert and a restriction digest was performed on the pCR2.1/BPL plasmid using the restriction endonucleases \textit{NcoI} and \textit{BamHI}. The isolated 723 bp fragment containing the \textit{A. aeolicus} BPL gene was cloned in \textit{NcoI}/\textit{BamHI}-digested pET28a (Novagen), producing the expression vector pET28a/BPL.

**Expression and purification of \textit{A. aeolicus} BPL**

The \textit{pET}28a/BPL vector was used to transform \textit{E. coli} BL21(DE3) cells (Novagen). A single colony was used to inoculate 200 mL LB broth supplemented with kanamycin (30 \( \mu \text{g/mL} \)) and grown overnight at 37 °C and 250 r.p.m. This seed culture was then used to inoculate 4 L of fresh broth supplemented with kanamycin (30 \( \mu \text{g/mL} \)) and grown overnight at 37 °C to \( D_{600} = 1.0 \) before induction with 1.0 \( \text{mM} \) isopropyl thiogalactoside (IPTG). After a further 3 h growth the cells were harvested by centrifugation (4000 \( \text{g} \) for 15 min at 4 °C) and washed with 10 mM Hepes (pH 7.5). The cells were resuspended in 10 mM Hepes buffer (pH 7.5) and disrupted by sonication (15 pulses of 30 s at 30-second intervals) at 4 °C. The cell
debris was removed by centrifugation at 27 000 g for 20 min at 4 °C.

One tablet of Complete™ Protease Inhibitor Cocktail (Roche) was added to the cell lysate before it was incubated at 60 °C for 20 min. Precipitated cellular debris was removed by centrifugation at 27 000 g for 20 min at 4 °C. The supernatant was filtered through a 0.45-μm membrane before it was loaded onto a 6-mL Resource-S cation exchange column (Amersham Biosciences) equilibrated with 10 mM Hepes (pH 7.5) at room temperature. The BPL protein was eluted with a linear salt gradient (0-100% 10 mM NaCl in 10 mM Hepes, pH 7.5) over 20 column volumes (120 mL). Fractions containing BPL (eluting at ~200 mM NaCl) were analysed by SDS/PAGE and those fractions judged to be 95% pure were pooled and stored in 10 mM Hepes (pH 7.5) containing 20% glycerol (v/v) at −20 °C. Protein concentration was determined using the Bio-Rad protein assay kit.

Expression and purification of Apo-BCCPΔ67 and BCCPΔ67 K117L from A. aeolicus

Overexpression of A. aeolicus BCCPΔ67 was achieved by transforming E. coli BL21(DE3) cells with the plasmid pET6H/BccpA67. A single colony was used to inoculate 200 mL 2YT supplemented with ampicillin (100 μg·mL⁻¹) and grown overnight at 37 °C and 250 r.p.m. This seed culture was then used to inoculate 4 L of fresh growth medium and grown at 37 °C to D₆₀₀ = 1.0 before induction with IPTG (1.0 mM final concentration). After a further 3 h the cells were harvested by centrifugation (4000 g for 15 min at 4 °C) and washed in binding buffer (20 mM Tris/HCl, pH 7.5, 0.5 mM NaCl, 5 mM imidazole). The cells were resuspended in binding buffer (5 mL per gram of wet cell paste) and disrupted by sonication (15 pulses of 30 s at 30-second intervals) at 4 °C. The cell debris was removed by centrifugation at 27 000 g for 20 min at 4 °C, after which the supernatant was filtered through a 0.45-μm membrane prior to chromatography.

The cell lysate was applied to a Hitrap® chelating affinity column (Amersham Biosciences) previously loaded with charge buffer (100 mM NiSO₄) and equilibrated with binding buffer at room temperature. The column was then washed with 5 column volumes of binding buffer before bound material was eluted using a linear gradient of 0-100% elution buffer (20 mM Tris/HCl, pH 7.5, 0.5 mM NaCl, 1 mM imidazole). Fractions were analysed by SDS/PAGE and those containing BCCPΔ67 were pooled and dialysed overnight against 4 L of 10 mM Hepes (pH 7.5) at 20 °C.

Apo-BCCPΔ67 and holo-BCCPΔ67 were separated by applying the BCCPΔ67-containing fractions eluted from the nickel column onto a 1-mL Mono-Q column (Amersham Biosciences) pre-equilibrated with 10 mM Hepes (pH 7.5) at room temperature. The column was then washed with 20 column volumes of 10 mM Hepes (pH 7.5), before the protein was eluted with a salt gradient (0–100% 10 mM Hepes, 1 mM NaCl, pH 7.5) over 25 column volumes. Fractions containing apo-BCCPΔ67 (confirmed by LC-MS analysis) were pooled and stored in 10 mM Hepes (pH 7.5) containing 20% glycerol (v/v) at −20 °C. Due to the low proportion of aromatic residues in BCCPΔ67, protein concentration was evaluated by measuring the absorbance at 280 nm and using the conversion factor calculated using vector ntr5 software.

The expression and purification of the BCCPΔ67 K117L mutant was performed in a similar way to the wild type protein. Elution from the Mono-Q column produced a single, apo-form peak.

Mass spectrometry characterization of proteins

Mass spectrometry was performed on a MicroMass Platform II quadrupole mass spectrometer equipped with an electrospray ion source. The spectrometer cone voltage was ramped from 40 to 70 V and the source temperature set to 140 °C. Protein samples were separated with a Waters HPLC 2690 with a Phenomenex C5 reverse phase column directly connected to the spectrometer. The proteins were eluted from the column with a 5-95% acetoniitrile (containing 0.01% trifluoroacetic acid) gradient at a flow rate of 0.4 mL·min⁻¹. The total ion count in the range 500–2000 m/z was scanned at 0.1 s intervals. The scans were accumulated and spectra combined and the molecular mass determined by the MAXENT AND TRANSFORM algorithms of the MASS LYNX software (MicroMass).

Assay of A. aeolicus BPL

BPL activity was assayed by measuring the incorporation of [14C]-biotin into purified BCCPΔ67, in a similar way to that described previously [22]. Except where stated otherwise, the reaction contained 10 mM Hepes (pH 8.5), 100 μM ATP, 200 μM MgCl₂, 10 μM biotin, 1 μM [14C]-biotin (specific activity 54 mCi·mmol⁻¹), 0.1 mg·mL⁻¹ bovine serum albumin, and 400 μM apo-BCCPΔ67. The reaction was initiated by the addition of purified BPL to a final concentration of 1 μM, and incubated at 70 °C for 30 min. The reaction was terminated by the addition of ice-cold trichloroacetic acid (final concentration 25% w/v), and incubation on ice for 30 min. The resulting protein precipitate was removed by centrifugation (27 000 g for 10 min). Aliquots of the supernatant were added to 5 mL of scintillation fluid (ICN biomedicals), and radioactivity was measured using a Tri-carb 210 OTR liquid scintillation counter (Packard). The extent of BCCPΔ67 biotinylation was deduced from the decrease in [14C]-biotin in the supernatant.

For kinetic analysis each of the substrate concentrations (biotin, ATP, BCCP) was varied accordingly. Values for Kₘ and Vₘₐₓ were determined by Michaelis–Menten analysis using SIGMAPLOT 2001 software. In some assays, to obtain sufficiently high levels of activity for accurate detection, it was necessary to continue until more than 10% of the limiting substrate had been used. In these instances the data was transformed using the method of Lee and Wilson and plotted as transformed values s' and v' [23].

To demonstrate the formation of the reaction intermediate, biotinyl-5'-AMP, we employed a streptavidin-binding assay. Briefly, the reaction contained 10 mM Hepes (pH 8.5), 10 μM biotin, 100 μM [8-14C]-ATP (specific activity 60–62 mCi·mmol⁻¹), 200 μM MgCl₂ and 0.1 mg·mL⁻¹ bovine serum albumin. The reaction was initiated by the addition of purified BPL to a final concentration of 5 μM, and incubated at 70 °C for 30 min. Ice-cold trichloroacetic acid (final concentration 10% w/v) was used to terminate
the reaction and the resulting precipitate of BPL was removed by centrifugation. Aliquots of the assay were then spotted onto a single SAM® Biotin Capture Membrane (Promega). Unreacted [14C]ATP was removed by washing each membrane four times in 2 M NaCl, four times in 2 M NaNO3 in 1% H2PO4, and twice in water. Finally the membrane was added to 5 mL of scintillation fluid (ICN biomedicals), and the radioactivity of the retained, bound biotinyl-[14C]AMP was measured using a Tri-carb 210 OTR liquid scintillation counter (Packard).

**Limited proteolysis of BPL**

Proteolysis of apo-BPL and substrate-bound-BPL were investigated using the proteases trypsin (Sigma) and chymotrypsin (Promega). Substrate-bound BPL was prepared by incubating BPL (15 µm) for 20 min at 60 °C with saturating amounts of biotin (40 µm), MgATP (2 mm), or both. The samples were then cooled for 10 min before treatment with protease, with a final protease/substrate ratio of 1:20 (w/w), and incubation at 37 °C for 30 min. Digestion was terminated by the addition of SDS sample buffer and boiling for 5 min. The extent of proteolysis was analysed by SDS/PAGE and densitometry analysis of the gel spots was performed using IMAGEMASTER TOTAL LABORATORY Software (Amersham Biosciences).

**Chemical crosslinking of A. aeolicus BPL and Apo-BCCPA67**

Purified BPL (15 µm) and either apo-BCCPA67, holo-BCCPA67 or BCCPA67 K117L (45 mm) were covalently cross-linked using 1-ethyl-3-(dimethylamino-propyl)-carbodiimide (EDC, 10 mm) at 60 °C for 60 mins. Aliquots were withdrawn at various time intervals, quenched with ammonium acetate (100 mm), and analysed by SDS/PAGE.

The cross-linked complex was prepared on a larger scale and separated from BPL and BCCPA67 by gel filtration. To prepare the complex we incubated 5 mg each of BPL and BCCP, EDC (10 mm) in a final volume of 5 mL 10 mm Hepes (pH 8.5) for 60 mins at 60 °C. The mixture was concentrated to 1 mL and then passed through a Superdex 75 column (Amersham Biosciences) equilibrated in 10 mM Hepes (pH 8.5) and 100 mM NaCl. The purified protein was stored at −20 °C.

**Results**

**Analysis of the A. aeolicus genome**

The complete genome sequence of A. aeolicus consists of 1512 predicted open reading frames [16]. We performed a BLAST search on the complete genome and identified two ORFs of 233 aa and 154 aa with high sequence homology to E. coli BirA (20.9% identity, 35.2% similarity) and BCCP (33.8% identity, 46.9% similarity), respectively. The pairwise sequence alignments generated by CLUSTAL W [24] are shown in Fig. 1 and these enabled us to design PCR primers to clone the A. aeolicus BPL and BCCP genes. We noted from this initial analysis that the A. aeolicus BPL differs from the E. coli BirA in that it lacks an N-terminal DNA-binding domain which places it in the group I class of BPLs along with those from Mycobacterium tuberculosis and Thermotoga maritima [19].

Previous studies on full-length E. coli BCCP (156 aa) revealed that the protein forms a tight complex with the biotin carboxylase (BC) subunit in solution, which complicates biochemical studies [25]. In most cases, the biotin carrier domain of biotin-containing enzymes is located at the C-terminal end of the carboxylase, with the biotinyllysine about 35 residues from the C-terminus. Structural studies revealed that a 65-70 amino acid fragment of BCCP, previously suggested by deletion mutagenesis, is required to form a minimal structured biotin domain [26]. Various truncated forms of the E. coli BCCP have been used in biochemical and structural studies, containing between 80 and 87 residues from the C-terminus of the protein. Here we
expressed *A. aeolicus* BCCP lacking 67 residues from the N-terminus (BCCPA67, Fig. 1) with an N-terminal His<sub>6</sub>-tag (total length 96 aa). The homology scores between *A. aeolicus* BCCPA67 and *E. coli* BCCP-87 (a domain containing 87 C-terminal amino acids) are 51.9% identity and 69.6% similarity (Fig. 1).

**Cloning, expression and purification of BPL**

The *A. aeolicus* bpl gene was amplified by PCR using *A. aeolicus* genomic DNA as a template and cloned into plasmid pCR2.1. DNA sequencing confirmed the previously published gene sequence, with the exception of a single base change at position 325 (T→C), which results in the substitution of a cysteine residue with an arginine. Subsequently the bpl gene was cloned into a pET expression vector for expression in various *E. coli* cells (DE3 lysogens); we found optimum recovery of protein using the BL21(DE3) strain. Cells were grown in shake flasks at 37 °C and expression induced with 1 mm IPTG (see Experimental procedures).

The predicted pl of the *A. aeolicus* BPL is 9.1 and as the enzyme contains a high proportion of positively charged residues, cation-exchange chromatography was used to purify it in a single step (Fig. 2, lanes 2–4). Initially the crude lysate was incubated at 60 °C which resulted in the precipitation of a significant quantity of *E. coli* proteins. It was then necessary to dialyse the sample overnight (20 °C) against 10 mm Hepes (pH 7.5) as immediate loading of an untreated extract onto a ResourceS column resulted in very poor binding (< 5%). It is unclear why this step was necessary, but after dialysis binding to the cation-exchange column approached 100%. BPL eluted from the column at 200 mm NaCl and we obtained the enzyme with a purity of greater than 95% (as determined by SDS/PAGE). Electrospray mass spectrometry analysis gave the molecular mass of the protein as 26636.8 ± 2.3 Da, consistent with the post-translational removal of the N-terminal methionine residue, and accurate to within experimental error of the predicted value of 26634.6 Da. The final yield of BPL using this method was > 10 mg per litre of cell culture and this protein was used for all subsequent kinetic and cross linking analysis.

**Cloning, expression and purification of BCCPA67**

We designed primers to clone a truncated domain of the *A. aeolicus* bccpA67 gene missing the first 201 bp, which encode the N-terminal 67 amino acids of *A. aeolicus* BCCP (Fig. 1). The truncated gene was amplified from genomic DNA using PCR and cloned into the pCR2.1 vector. DNA sequencing confirmed the expected gene sequence, and the bccpA67 gene was subsequently cloned into a pET-derived expression vector with an N-terminal His<sub>6</sub>-tag. *E. coli* BL21(DE3) competent cells were used for recombinant expression (described under Experimental procedures) and the BCCPA67 cell lysate was first purified by nickel-affinity chromatography (Fig. 2, lanes 6–8). The protein eluted with 200 mm imidazole and, as precipitation had been observed at high concentrations of this eluant, it was immediately diluted 1:1 with 10 mm Hepes (pH 7.5) and dialysed against this buffer. SDS/PAGE analysis indicated BCCPA67 to be > 90% pure but electrospray mass spectrometry revealed the presence of two distinct species. The first, of molecular mass 10739.6 Da, corresponded to the predicted mass of apo-BCCPA67 (10739.6 Da) while the second corresponding to the holo-form (biotinylated), with a mass increase of 226.1 Da (10965.7 Da). This confirmed that the *A. aeolicus* BCCPA67 domain folded correctly, and was recognized and biotinylated by the host *E. coli* BirA. To separate the apo- and holo-forms of BCCPA67 we employed anion exchange chromatography in a similar way to that used for *E. coli* BCCP-87 [27]. Fractions from the column were analysed by electrospray mass spectrometry and the apo-protein eluted at a slightly lower salt concentration than the holo-form (160–240 mm NaCl vs. 240–320 mm NaCl). Approximately 80% of the apo-BCCPA67 was resolved from the holo-form by collecting only the leading fractions of the protein peak. The final yield of apo-BCCPA67 was 5–10 mg per litre of cell culture and ~1 mg per litre of the holo-form.

**Cloning, expression and purification of BCCPA67 K117L mutant**

A mutant of the truncated bccpA67 gene, with the active lysine residue (K117) replaced by a leucine residue, was produced using the megaprimer method [21]. The mutation was confirmed by DNA sequencing before the gene was cloned into a pET-derived expression vector with an N-terminal His<sub>6</sub>-tag and the resulting construct was then transformed into *E. coli* BL21(DE3) cells for expression (as described in Experimental procedures). The BCCPA67 K117L protein was purified using nickel-affinity chromatography and the protein eluted with 200 mm imidazole (Fig. 2, lanes 10–12). Protein-containing fractions were immediately dialysed against 10 mm Hepes (pH 7.5). Further purification on anion-exchange chromatography gave a
single species with a mass of 10724.8 ± 1.1 Da, consistent with the predicted mass of apo-BCCPA67 K117L of 10724.6 Da. A species was not present at +226 Da, an indication that in vivo biotinylation had not occurred. The yield of the apo-BCCPA67 K117L mutant was ≈15 mg of protein per litre of cell culture.

Biochemical properties of BPL

Activity assays were performed with BPL by measuring the incorporation of [14C]biotin into the purified apo-BCCPA67 biotin-accepting domain [22]. In initial experiments we observed optimal enzyme activity at pH 8.5, and magnesium ions, ATP, biotin and apo-BCCPA67 were all required for activity. The activity of the enzyme was also measured at varying temperatures, with optimal activity at 70 °C. Activity was seen to decrease by roughly 50% for every 10 °C drop in temperature, and increasing the temperature above 70 °C resulted in enzyme precipitation, together with a dramatic loss in activity (data not shown). The tolerance of BPL for other nucleotide sources was measured by replacing ATP with UTP, GTP or CTP. No BPL activity was detected for any of these three substrates, suggesting that the enzyme is completely dependent on ATP for its nucleotide supply (data not shown).

In assays performed with BCCPA67 K117L as the biotin acceptor no biotinylation was observed, verifying K117 as the active residue and demonstrating the specificity of the BPL catalysed reaction.

Kinetic analysis of BPL

The kinetic constants for d-biotin, MgATP and apo-BCCPA67 were determined using steady-state kinetics (Fig. 3). The $K_m$ for d-biotin was determined to be 440 ± 70 nM. The $K_m$ values for BPLs from other species range from low nanomolar to low micromolar; 67 ± 11 nM (Saccharomyces cerevisiae BPL), 300 nM (E. coli BirA), 130 nM (Arabidopsis thaliana HCS) and 3.3 mm (chicken liver HCS) [28-31]. The $K_m$ for MgATP was 15.1 ± 1.5 μM, which is similar to that determined for the S. cerevisiae BPL (20.9 ± 3 μM) and A. thaliana HCS (4.4 μM). In contrast, the $K_m$ for MgATP for E. coli BirA is around 500 μM. It should be noted that the kinetic analyses for each BPL were performed under slightly different reactions conditions, for example an elevated temperature was used in the study presented here. Finally, the $K_m$ for apo-BCCPA67 was 160 ± 32 μM. A range of biotinylation substrates have been used in assays of BPL activity with cross-species reactivity frequently observed, e.g. S. cerevisiae BPL has a $K_m$ of 11.1 ± 1 mm for E. coli BCCP-87. However, we could not test E. coli BCCP-87 as a substrate for BPL because the rate of biotinylation at 37 °C was outside the lower limit of detection in our assay.

As shown in Scheme 1 the first step in all biotinylation reactions studied thus far involves the synthesis of a biotinyl-5'-AMP intermediate and the release of PF. This molecule is the substrate for biotin transfer to BCCP and is also the corepressor of E. coli BirA. To prove that A. aeolicus BPL synthesises biotinyl-5'-AMP we incubated BPL with biotin and [14C]MgATP at 70 °C and used streptavidin-coated membranes to capture radioactive biotinyl-5'-[14C]AMP (data not shown). Furthermore, we noted that biotinylation was inhibited by the addition of NaCl in concentrations above 200 mM.

Proteolysis of BPL

We subjected BPL to limited proteolysis in the presence and absence of biotin and MgATP (Fig. 4). Digestion with both trypsin and chymotrypsin resulted in formation of a fragment of ≈21 kDa. Chymotrypsin digestion also produced an array of smaller peptide fragments. We found that only 34% of total BPL remained after trypsin cleavage in the absence of substrates. However, preincubation of BPL with saturating amounts of biotin or MgATP separately increased its resistance to digestion (50% and 63% remained, respectively). Moreover, preincubation with both substrates dramatically increased the resistance of BPL to proteolysis with trypsin (98.9% remained). Comparative analysis with chymotrypsin showed that 11% of BPL remained intact after digestion. Preincubation of the enzyme with MgATP afforded little protection (13% of BPL remaining), whereas 34% and 92% BPL remained after preincubation with biotin and biotin and ATP. Taken together these results suggest that the binding of the substrates and/or the formation of the intermediate, biotinyl-5'-AMP, plays a role in protecting BPL from protease cleavage.

![Fig. 3. Steady-state kinetic analysis of BPL substrate binding. The activity of A. aeolicus BPL was measured under steady-state conditions at 70 °C. Two substrates were kept at constant saturating levels while the concentration of the third substrate was varied over the ranges shown above in the graphs. From the curves, $K_m$ values for biotin (A), MgATP (B) and apo-BCCPA67 (C) were determined (see Experimental procedures).](image-url)
Fig. 4. Proteolysis of *A. aeolicus* BPL. *A. aeolicus* BPL was treated with trypsin or chymotrypsin either with or without equilibrating the enzyme with 1 mM MgATP and/or 50 μM biotin. Lanes 1-4, Trypsin digest; lane 1, BPL; lane 2, BPL + MgATP; lane 3, BPL + biotin; lane 4, BPL + MgATP and biotin. Lanes 5-8 Chymotrypsin digest; lane 5, BPL; lane 6, BPL + MgATP; lane 7, BPL + biotin; lane 8, BPL + MgATP and biotin.

LC-MS analysis of the peptide fragment produced from BPL after treatment with trypsin revealed the presence of two distinct species of mass 215549.5 ± 2.6 Da and 21678.6 ± 5.9 Da. Primary structure analysis of BPL established these masses corresponded to trypsin cleavage between R44 and K45, and K45 and W46 adjacent to the proposed catalytic centre and biotinyl-5'-AMP binding site.

Chemical crosslinking of BPL and BCCP

Although structures of *E. coli* BirA and both apo- and holo-BCCP-87 have been determined, our goal was to isolate a BPL:BCCP complex for biochemical and structural studies. Previous work in our laboratory used the chemical crosslinking agent EDC to isolate an *E. coli* flavodoxin—flavodoxin reductase complex, so we used this reagent to crosslink BPL and various forms of BCCPA67 [32]. Initially we incubated BPL and apo-BCCPA67 in the presence of excess EDC at room temperature with and without saturating amounts of biotin and MgATP, but we did not observe any crosslinked species of predicted molecular mass ≈ 36 kDa on SDS/PAGE (data not shown). However, a species was observed when the incubation was carried out at elevated temperatures, with 60 °C being the optimum (Fig. 5A). The presence of the substrates had no observable effect on crosslinking. Interestingly, when BPL was incubated with holo-BCCPA67 and EDC the amount of crosslinked species generated was significantly reduced compared to the apo form (Fig. 5B). Moreover, the incubation of BPL with the BCCPA67 K117L mutant led to the formation of crosslinked complex in comparable amounts to that using apo-BCCPA67 (Fig. 5C). Purification of the BPL: BCCPA67 complex from unreacted proteins was achieved using size exclusion chromatography, which resolved the mixture into three peaks (Fig. 6). We noted that both BPL, BCCPA67 and the complex eluted from the size exclusion column at retention volumes different to that predicted by their molecular masses (45, 35 and 70 kDa, respectively). However, analysis by SDS/PAGE revealed that the BPL:BCCPA67 complex eluted from the column first and had a molecular mass of 37 kDa (Fig. 6, inset). Electrospray analysis of the complex gave a molecular mass of 37 200 ± 200 Da which agrees well with the predicted mass of a 1:1 heterodimer.

Discussion

The attachment of biotin to the specific lysine residue of the apo-forms of biotin-requiring enzymes is a complex, multistep reaction. The BPL enzyme (also known as holocarboxylase synthetase, HCS) catalysing this process first activates biotin as biotinyl-5'-AMP then transfers the biotin to a specific lysine of the BCCP domain. The BPLs and BCCPs from a diverse range of organisms including *E. coli* (BirA), yeast, human and plant have been isolated and it has been shown that the BPL from one organism can biotinylate the BCCP domain from another [28]. This suggests some degree of structural homology between these proteins and primary structure analysis reveals there is a high degree of amino acid sequence similarity throughout the catalytic domain of the BPL family and the biotinyl domain of BCCPs [33]. An understanding of the protein-
protein interactions that mediate this highly specific reaction requires three dimensional structures of each of the components. The structure of the *E. coli* BirA monomer in complex with biotinyl-lysine revealed details of the protein-substrate interactions but several loops within the active site were disordered [9]. More recently, the structure of the BirA dimer has provided insights into how the ligase also acts as a transcriptional repressor by binding to the *E. coli* biotin operon operator [12]. The structures of the apo- and holo-forms of *E. coli* BCCP-87, determined by X-ray and NMR, are virtually identical and showed that the biotinyl-lysine residue is located at an exposed β-turn, flanked by highly conserved methionine residues [13,15]. A more recent NMR study, combined with results from site-directed and random mutagenesis [29,34,35], allowed modelling of the elusive *E. coli* BPL:BCCP-87 complex and it appears that its formation is dependent on subtle, competing protein–protein interactions [36].

Analysis of the complete genome of the hyperthermophile *A. aeolicus* revealed the presence of BPL and BCCP homologues (Fig. 1). The *A. aeolicus* BPL enzyme belongs to the class I group of BPLs since it lacks the DNA-binding domain found in BirA and is the smallest characterized thus far. Eukaryotic BPLs also lack predicted DNA-binding domains but have large N-terminal extensions with unknown functions [33]. The full-length *A. aeolicus* BCCP has a C-terminus showing high sequence homology to the biotin domains of biotin-carboxylases and contains the eight amino acid ‘thumb’ motif found in *E. coli* BCCP [33,37,38]. The N-terminus has a large proportion of charged residues, and displays little similarity to any other BCCPs.

Using recombinant proteins isolated from *E. coli* we have characterized the full-length BPL and BCCP biotinylating domain BCCPA67 (with a His6 N-terminal tag) from a hyperthermophile. We have gained insight into this extremely specific post-translational modification reaction at high temperatures and used features of the two *A. aeolicus* proteins to capture a BPL:BCCP complex. We found *A. aeolicus* BPL to be monomeric, and thus competing homodimerization interactions found in *E. coli* BirA are not present. We isolated a mixture of apo- and holo-forms of *A. aeolicus* BCCPA67 and so conclude that it must be a substrate for *E. coli* BPL in vivo. Biotinylation in hyperthermophiles proceeds via the two-step reaction sequence found in other organisms (Scheme 1). Isolated *A. aeolicus* BPL could biotinylate apo-BCCPA67 at temperatures up to 70 °C albeit at a slow rate. It is interesting to compare the *A. aeolicus* BPL:BCCPA67 biotinylation reaction with that of a mutant *E. coli* BirA lacking the N-terminal DNA binding domain (BirA65-321) and *E. coli* BCCP-87. The BirA65-321 mutant could synthesize biotinyl-5′-AMP and transfer biotin to apo-BCCP-87 at the same rate as wild-type BirA. However, the affinity of BirA65-321 mutant for biotin and biotinyl-5′-AMP was decreased 100-fold and 1000-fold, respectively [39]. This suggested that in BirA, the N-terminal domain is somehow involved in tight-binding of the two ligands. In future, it would be interesting to study a BPL:BirA chimera by fusing the DNA-binding domain at the N-terminus of *A. aeolicus* BPL.

Substrate *Kₘ* values for BPLs from a number of species have been shown to range from the low nanomolar to low millimolar. In steady-state kinetic assays at 70 °C, the *A. aeolicus* BPL bound biotin, MgATP and apo-BCCPA67 with affinities of 440 nM, 15.1 μM and 160 μM, respectively. The kinetic constant for MgATP suggests that *A. aeolicus* BPL resembles those from eukaryotic biotin auxotrophs (low micromolar). In contrast, *E. coli* BirA binds MgATP with a *Kₘ* in the low millimolar range which reflects its dual function as both repressor of biotin biosynthesis and biotin ligase. It is interesting to note that *A. aeolicus* contains all the genes required to convert pimelate to biotin (*bioW*, *bioF*, *bioA*, *bioD* and *bioB*) suggesting it can synthesize this vitamin but the in vivo concentration within *A. aeolicus* cells is unknown. The *Kₘ* for the apo-BCCPA67 domain used in this study is high compared to others but this may reflect the fact that the first 67 amino acid residues, which contain a high number of charged residues, could play an important role in determining the *Kₘ* of the full-length BPL:BCCP complex.

**Fig. 6.** Purification of the chemically crosslinked BPL:apo-BCCPA67 complex by size-exclusion chromatography. The chromatogram above was obtained when the cross-linking reaction was applied to a Superdex 75 column. The three peaks correspond to the crosslinked complex (7–8 mL), BPL (10 mL) and BCCPA67 (11–12 mL). Insert: SDS/PAGE analysis of the column fractions. Lane 1, cross-linking reaction before purification. Lane 2–11, 1 mL fractions eluting between 6 and 15 mL.
role in tight binding to BPL. Most biochemical studies use these truncated BCCP domains and future work using full length BCCPs should elucidate the role of the N-terminal interaction with BPL. It is also possible that the addition of the His$_6$-tag to the protein has altered its kinetic properties and may contribute to the abnormally high $K_m$ for BCCP-A67. The calculated $k_{cat}/K_m$ for biotin of 1.7 ± 0.1 x 10$^3$ M$^{-1}$s$^{-1}$ is 300-, 100- and 35-fold smaller than the $E. coli$ BirA, yeast and $A. thaliana$ BPL enzymes, respectively [28,30,40] but reflects the fact that the B. aeolicus BPL $k_{cat}$ is low at 70°C (cf. B. aeolicus grows optimally at 95°C).

Limited proteolysis with trypsin produced two fragments of ≈20 kDa, differing in length by only one residue (Fig. 4). Mass spectrometry revealed that cleavage had occurred after residues R44 and K45 which, by comparison with E. coli BirA, are predicted to lie near the putative intermediate binding site (Fig. 1). Treatment of BPL with trypsin and chymotrypsin in the presence of biotin or MgATP decreased the susceptibility to cleavage by a small but noticeable amount. However, incubation of the enzyme in the presence of both substrates rendered a B. aeolicus BPL protease-resistant. The same region is protease-sensitive in S. cerevisiae BPL and is also protected by incubation with both biotin and MgATP [28]. E. coli BirA structure contains five surface loops, four of which are in the central domain with loop regions (110–128, 212–233) and (140–146, 193–199) close together in three-dimensional space [6]. The region containing 110–128 in E. coli BirA is highly analogous to residues 32–50 in B. aeolicus BPL whereas the other loop regions have low pairwise sequence homology. A protease-sensitive site has been reported between residues 217 and 218 of BirA. In contrast, B. aeolicus BPL is not cleaved at this site but is cleaved in the adjacent loop region (32–50). This suggests that this highly conserved region forms an exposed loop near the biotinyl-5'-AMP binding site (Fig. 1). These flexible, unstructured regions are also involved in BCCP binding and are believed to become more rigid upon substrate-binding [6,34].

A recent combined mutagenesis/biological selection approach identified two single glutamate residues E119 and E147 of E. coli BCCP-87 that appear to interact with BPL [28]. E119 is a highly conserved residue in the BPL:apo-BCCP complex and may contribute to the abnormally high $K_m$ for BPL:apo-BCCP complex for the first time. The zero-length EDC reagent activates acidic residues on one protein to form an unstable urea derivative [42]. This derivative then reacts with a nucleophile (such as lysine) on another protein to form an amide link between the two proteins. Incubation of BPL and apo-BCCPΔ67 in the presence of EDC led to the time-dependent appearance of a species of ≈37 kDa on SDS/PAGE gels (Fig. 5A), which is in agreement with the predicted mass of a 1:1 complex of BPL and apo-BCCPΔ67. We noticed that BPL, BCCPΔ67 and the complex eluted earlier than predicted from the size-exclusion column. Future studies will analyse the proteins by equilibrium sedimentation experiments in a similar way to that described for the BCCP-87 and BCCP [25]. Nevertheless, the complex was easily separated from the unreacted proteins using this procedure (Fig. 6) and allowed us to confirm its mass by electrospray mass spectrometry. Interestingly, the complex was not formed between BPL and holo-BCCPΔ67 (Fig. 5B) suggesting that biotinylation had either caused a conformational change in BCCPΔ67 such that it no longer bound to BPL or that the biotin moiety had somehow blocked residues that react with the EDC reagent. Furthermore, a complex was formed between the BCCPΔ67 K117L mutant and BPL both in the absence and presence of saturating amounts of biotin and MgATP (Fig. 5C). This demonstrates that the active lysine residue does not take part in the cross-linking reaction and saturating amounts of both substrates do not inhibit complex formation.

Although the published 3D structures of the apo- and holo- forms of BCCP-87 show no major structural differences, some structural studies (both NMR and X-ray) have concluded that the lack of any major differences between them might not be wholly reflected in their behaviour in solution [15]. NMR titration experiments were carried out with BirA and apo-BCCP-87 and, in light of our data, it would be interesting to repeat this work with BirA and holo-BCCP-87 to determine if any differences arise. Recent elegant studies by Cronan and Solbiati et al. highlight a difference in the stability of apo-BCCP-87 and holo-BCCP-87 to proteolysis and stress the importance of the essential so-called 'thumb' domain of BCCP-87 (residues 91–100) which had previously been shown to interact with the ureido ring of the attached biotin moiety [37,43]. Studies using chemically biotinylated BCCP-87 recently confirmed that this increased stability is an inherent property of holo-BCCP-87 and not due to a conformational change imparted by BPL. Furthermore, thumbless holo-BCCP-87 mutants exhibit little increased stability over their apo-counterparts, implying the majority of this increased stability is due to the thumb–biotin interaction. The authors conclude that the more protease sensitive apo-BCCP has a more dynamic form than the holo-protein. The B. aeolicus BCCPΔ67 also contains a well-conserved thumb domain (Fig. 1) and we are currently producing thumbless BCCPΔ67 mutants for analysis by EDC cross-linking with BPL (D. Clarke and D. Campopiano, unpublished results).

A recent study suggested that the C-terminal domain of BirA, which had been ascribed no biochemical function, also plays a significant role in apo-BCCP and substrate recognition [29]. It has been shown that ion pair networks are a common feature in heat-resistant proteins and are believed to play important roles in their increased thermal stability [17,41]. As both the B. aeolicus BPL and BCCP contain a large number of charged residues and we observed inhibition of biotinylation at high salt concentrations, we presume that ionic interactions are involved in the formation of the hyperthermophilic BPL:BCCPΔ67 complex. To investigate the formation of the BPL:BCCPΔ67 heterodimer we used the chemical cross-linking agent EDC to capture a
model (PDB code 1K67) was built using the coordinates of the BirA dimer in the presence of biotin (PDB code 1HXD) and holo-BCCP-87 (PDB 1BIA). Residues in both E. coli proteins thought to be responsible for BirA:BCCP-87 complex formation are conserved in A. aeolicus BPL and BCCPΔ67 (Fig. 1). A current goal is to identify the charged residues taking part in the EDC-mediated crosslinking reaction and A. aeolicus BPL and BCCPΔ67 mutants are currently being studied using high-temperature in vitro biotinylation and chemical crosslinking assays.

Acknowledgements

We wish to thank Profs. K. Steuer and R. Huber (University of Regensburg) for the gift of A. aeolicus chromosomal DNA. The Nuffield Foundation Bursary Scheme is acknowledged for its support of (J. C.). This work was supported by the Biotechnology and Biological Sciences Research Council, UK, and the University of Edinburgh.

References

Biotinylation in *Aquifex aeolicus* (Eur. J. Biochem. 270) 1287

Defensins are cationic antimicrobial peptides that have a characteristic six-cysteine motif and are important components of the innate immune system. We recently described a ß-defensin-related peptide (Defrl) that had potent antimicrobial activity despite having only five cysteines. Here we report a relationship between the structure and activity of Defrl through a comparative study with its six-cysteine-containing analogue (Defr1 Y5C). Against a panel of pathogens, we found that oxidized Defrl had significantly higher activity than its reduced form and the oxidized and reduced forms of Defr1 Y5C. Furthermore, Defrl displayed activity against Pseudomonas aeruginosa in the presence of 150 mM NaCl, whereas Defr1 Y5C was inactive. By using non-denaturing gel electrophoresis and Fourier transform ion cyclotron resonance mass spectrometry, we observed Defrl and Defr1 Y5C dimers. Two complementary fragmentation techniques (collision-induced dissociation and electron capture dissociation) revealed that Defr1 Y5C dimers form by noncovalent, weak association, as expected for a non-native disulfide bond. In contrast, Defrl dimers are resistant to collision-induced dissociation and are only disassociated into monomers by reduction using dithiothreitol. This is indicative of Defrl dimerization being mediated by an intramolecular disulfide bond. Proteolysis and peptide mass mapping revealed that Defrl Y5C monomers have ß-defensin disulfide bond connectivity, whereas oxidized Defrl is a complex mixture of dimeric isoforms with as yet unknown inter- and intramolecular connectivities. Each isoform contains one intermolecular and four intramolecular disulfide bonds, but because we were unable to resolve the isoforms by reverse phase chromatography, we could not assign each isoform with a specific antimicrobial activity. We conclude that the enhanced activity and stability of this mixture of Defrl dimeric isoforms are due to the presence of an intramolecular disulfide bond. This first description of a covalently cross-linked member of the defensin family provides further evidence that the antimicrobial activity of a defensin is linked to its ability to form stable higher order structures.

© 2004 by The American Society for Biochemistry and Molecular Biology, Inc.

Published, JBC Papers in Press, August 17, 2004, DOI 10.1074/jbc.M404690200

Dominie J. Campopiano†, David J. Clarke*, Nick C. Polfer†, Perdita E. Barran‡‘, Ross J. Langley†, John R. W. Gova§, Alison Maxwell** and Julia R. Dorin**

†From the School of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, the Department of Medical Microbiology, University Medical School, University of Edinburgh, Edinburgh EH8 9AG, and **Medical Research Council Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, Scotland, United Kingdom

ß-Defensins are members of the antimicrobial peptide family that are important components of the mammalian innate immune response (1, 2). In addition to potent bactericidal activity, they can also act on T lymphocytes and immature dendritic cells thus playing key roles in adaptive immunity (3). They are produced as propeptides and are processed to a mature secreted peptide, which have six canonical cysteine residues with spacing and intramolecular disulfide bridge connectivity (Cys1–Cys3, Cys2–Cys4, and Cys5–Cys6) distinct from the similar ß-defensins (4). A recent bioinformatic analysis of the human and mouse genomes revealed a significantly higher number of ß-defensin genes than previously thought, although the function of each individual defensin has still to be determined in vivo (5).

Two strong lines of evidence have demonstrated the significance of the antimicrobial activity of defensins in vivo. First, it was demonstrated that deleting the gene that encodes the enzyme matrilysin, which processes ß-defensins in gut paneth cells, resulted in mice in which orally administered bacteria survived in greater numbers and were more virulent (6). Second, a human ß-defensin gene expressed in mice resulted in transgenic animals that were markedly resistant to oral challenge with the virulent Salmonella typhimurium (7). In addition, mice that are deficient in the ß-defensin 1 gene have a phenotype consistent with a defect in microbial resistance (8). The mature peptides of ß-defensins are 30–45 amino acids in length and are amphipathic, i.e. they have discrete cationic and hydrophobic patches. Their net positive surface charge implies an initial electrostatic interaction between the peptide and negatively charged components of the bacterial cell wall, e.g. lipopolysaccharide or teichoic acid (9). Indeed, reduction of the negative charge of these molecules by enzymatic covalent modification has been observed in antimicrobial peptide-resistant bacterial strains (10). The exact mechanism of how defensins bind and disrupt the bacteria membrane is still the subject of intensive study, but it appears that active peptides must display an appropriate balance of hydrophobicity and net positive charge (11, 12). It has been proposed that the conserved disulfide bridges impart a structural core, and conserved residues on the surface are under selective pressure against rapidly evolving bacteria (13, 14). In direct contrast to this perceived structure-function relationship, Wu et al. (15) have recently demonstrated that the antimicrobial activity of human ß-defensin 3 (HBD3) is independent of disulfide bridging. How...
were incubated in 100 μl of cells (1-5 × 10^5 CFU) at 37 °C for 1 h. The concentration of peptide where we observed >99.99% killing of forming units (CFU)/ml in 10 mm potassium phosphate containing 1% (v/v) Iso-Sensitest broth, pH 7.4, which contained various concentrations of NaCl (0-300 mM). The bacteria were then digested with trypsin at a concentration four times the MBC.

Native Gel Electrophoresis—Electrophoresis of Defrl and Defrl Y5C was performed initially by using the accurate mass capabilities of a 9.4 Tesla Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (Bruker Daltonics, Billerica, MA) equipped with a nanospray source. Any minor partially reduced species was removed by acetylation and separation. Using this instrument, two dissociation methods were applied to isolated peptides. For sustained off resonance irradiation dissociation-induced dissociation (SORI-CID), a given charge state was isolated by sweep excitation and subjected to CID with argon as the collision gas for 500 ms. For electron capture dissociation (ECD), a given charge state was isolated by sweep excitation and subjected to electron irradiation for 50 ms using a barium fluoride-coated high surface area (5 mm diameter) dispense cathode (HeatWave, Watsonville, CA) This instrument was also utilized to determine the time scale necessary for full reduction of the peptides.

Electrospray solution and source conditions were the same for accurate mass, SORI-CID, and ECD. All peptide samples were made up to a concentration of 50 μM with MeOH/H₂O/CH₃COOH (50:40:1 v/v). Solutions were ionized by nano-electrospray from gold/palladium-coated tips (Proxeon Biosystems). Ions were accumulated in a hexapole and transferred to the detection cell of the FT-ICR instrument. The predicted isotopic abundance was determined using published methods within the XMASS software (28).

Prior to antimicrobial assays, the oxidation state of the peptides was confirmed using a Micromass Platform II single quadrupole mass spectrometer equipped with an electrospray source. Following our initial experiments on the FT-ICR, the platform had sufficient resolution to confirm that the peptides were fully oxidized or fully reduced. The spectrometer cone voltage was ramped from 40 to 70 V, the source temperature was set to 140 °C. Protein samples were separated with a Waters HPLC 2690 with a Phenomenex CS reverse phase column (5 μm, 250 × 4.6 mm) directly connected to the spectrometer. The proteins were eluted from the column with a 5-95% acetonitrile (containing 0.01% trifluoroacetic acid) gradient at a flow rate of 0.2 ml/min. The total ion count in the range 500-2000 m/z was scanned at 0.1-s inter-

EXPERIMENTAL PROCEDURES

Materials—Defrl and Defrl Y5C were chemically synthesized by the standard solid phase methodology (Abelcrom Ltd.) and were refolded and oxidized in air as described previously (26).

Antimicrobial Activity Assays—The strains of microbes used in this study are as follows: Escherichia coli ATCC 25922, Pseudomonas aeruginosa PA01, Staphylococcus aureus NCTC 25953, Enterococcus faecalis ATCC 29212, and Candida albicans J2922.

Test organisms were grown to mid-logarithmic phase in Iso-Sensitest broth (Oxoid) growth media and then diluted to 1-5 × 10^6 colony-forming units (CFU)/ml in 10 mm potassium phosphate containing 1% (v/v) Iso-Sensitest broth, pH 7.4. Different concentrations of test peptide were incubated in 100 μl of cells (1-5 × 10^6 CFU) at 37 °C for 1 h. 10-Fold serial dilutions of the incubation mixture were plated on Iso-Sensitest plates, incubated at 37 °C, and the CFU was determined the following day. The minimum bactericidal concentration (MBC) is the concentration of peptide where we observed >99.99% killing of the initial inoculum. All assays were performed in duplicate and repeated on two independent occasions. The MBC was obtained by taking the mean of all the results, and experimental errors were within one doubling dilution.

Reduction of the peptides was performed by adding 10 mM dithiothreitol (DTT) and incubating at room temperature overnight. The oxidation state of each peptide prior to performing antimicrobial assays was determined by mass spectrometry.

The effect of salt on antimicrobial activity was tested by incubating 100 μl of 1-5 × 10^6 CFU of bacteria in 10 mm potassium phosphate, 1% (v/v) Iso-Sensitest, pH 7.4, which contained various concentrations of NaCl (0-300 mM). The bacteria were then digested with trypsin at a concentration four times the MBC.

Sequence of Defrl and Defrl Y5C. <table><thead><tr><th>Sequence</th></tr></thead><tr><td>DPVTY1 RGGGQCYQRCGLHDKCQGSPFCOKC</td></tr><tr><td>DPVTY1 RGGGQCYQRCGLHDKCQGSPFCOKC</td></tr></table>
Y5C, and have determined the minimum bactericidal concentration (MBC) at over 40 μg/ml. Peptides were detected at 215 and 280 nm. A gradient, and the two major species were collected and freeze-dried. The samples were subjected to MS and subjected to CID using argon, and the resulting fragments oxidized state; the 2nd aliquot was reduced with 20 mM TCEP; and the 3rd aliquot was digested with chymotrypsin (100 μg/ml) for 4 h at 37 °C. All samples were purified and concentrated using C18 ZipTips and analyzed by MALDI-TOF MS as described above.

CID sequence analysis of the digested peptides was performed by using a Q-TOF tandem mass spectrometer (Micromass, UK) equipped with a nanospray source. Specific ions were mass selected by the quadrupole MS and subjected to CID using argon, and the resulting fragments were analyzed in the TOF MS.

HPLC—Analytical HPLC was performed on a Beckman System Gold HPLC equipped with a Phenomenex C18 Jupiter column (5 μm, 250 × 4.6 mm) and a 10–35% acetonitrile gradient, and the two major species were collected and freeze-dried. The resulting peptides were reconstituted in 50 mM Tris-HCl, 20 mM CaCl₂, pH 8.2, and each was split into 3 aliquots. The 1st aliquot was analyzed in its oxidized state; the 2nd aliquot was reduced with 20 mM TCEP; and the 3rd aliquot was digested with chymotrypsin (100 μg/ml) for 4 h at 37 °C. All samples were purified and concentrated using C18 ZipTips and analyzed by MALDI-TOF MS as described above.

RESULTS AND DISCUSSION

Antimicrobial Properties—In a previous study we showed that despite having only five cysteine residues, murine Defr₁ displayed antimicrobial activity at nanomolar concentrations (26). Here we synthesized the six-cysteine analogue, Defr₁ Y5C, and have determined the minimum bactericidal concentrations (MBCs) of the oxidized and reduced forms of both peptides against a diverse panel of clinically relevant microbes (Table I). Numerous methods have been used in various laboratories to determine the antimicrobial activities of defensins (for example see Refs. 9, 15, and 16). Typically, the test organism is exposed to the peptide (0–24 h), followed by growth on liquid or solid media and subsequent determination of optical densities and/or counting of surviving colonies relative to a control. Antimicrobial activities (μg/ml) are quoted as minimal inhibitory concentrations, MBCs, and LD₅₀ or LD₉₀ values (dose required to kill 50 or 90% bacteria). At the outset we incubated P. aeruginosa PA01 with peptides for 0, 1, 4 and 24 h and noted that the MBCs (99.99% bacteria killed) were the same for each time point, indicating that the peptides killed within an hour of administration. We therefore carried all the reported assays with a 1-h incubation time.

Defr₁ exhibited broad spectrum antimicrobial activity, with MBCs ranging from 3 to 10 μg/ml against all organisms tested. These included the Gram-negative bacteria P. aeruginosa and E. coli as well as the Gram-positives S. aureus and E. faecalis. Defr₁ also displayed antifungal activity against C. albicans. In contrast, the six-cysteine analogue Defr₁ Y5C displayed MBCs in the range 50–100 μg/ml against Gram-negative bacteria, ≥100 μg/ml against Gram-positive bacteria, and 25 μg/ml antifungal activity. The increased MBCs of Defr₁ Y5C compared with Defr₁ highlight a significant increase in potency for the five-cysteine variant. Recent work by Wu et al. (15) has demonstrated that the activity of HBD₃ is independent of the number and connectivity of disulfide bridges within the molecule. To test this hypothesis we repeated the antimicrobial assays with both peptides after complete reduction with excess DTT. Reduced Defr₁ Y5C had similar MBCs to its oxidized form, supporting the hypothesis that its activity is independent of the presence of disulfide bonds. However, the MBCs of reduced Defr₁ were significantly higher than its oxidized form and similar to the MBCs for Defr₁ Y5C. Thus, in this instance, it appears that the presence of disulfide bonds does influence the antimicrobial activity of the five-cysteine containing peptide.

For further comparison we also analyzed the effect of high salt concentrations on the antimicrobial activities of Defr₁ and Defr₁ Y5C against P. aeruginosa (Fig. 2). The activity of Defr₁ Y5C was extremely salt-sensitive, killing only 10% of 1 × 10⁵ bacteria in the presence of 25 mM NaCl and being completely inactive at 50 mM NaCl. In contrast, Defr₁ displayed 100% killing at 25 and 50 mM NaCl, 84% killing at 150 mM NaCl, and 28% killing even at 300 mM NaCl. These salt sensitivity results further highlight the significant differences between the antimicrobial activity of the six-cysteine and the five-cysteine Defr₁ peptides.

Structural Analysis—To rationalize the striking differences that we observed in antimicrobial activity between Defr₁ and Defr₁ Y5C, we investigated the structure of the two peptides by a combination of natural gel electrophoresis, high resolution mass spectrometry, and reverse phase chromatography. One advantage of using FT-ICR MS for studying defensin structure is that only microgram amounts of each peptide are required for analysis in contrast to both NMR and x-ray crystallography.

Native Gel Electrophoresis—To analyze the structure of Defr₁ and Defr₁ Y5C, they were subjected to nondenaturing electrophoresis on 16% Tricine gels (Fig. 3). In their oxidized forms, the predominant forms of both peptides migrated with apparent molecular masses of ~7 kDa, which suggests that they can dimerize under these conditions. Reduction of both peptides with excess DTT increased their apparent mobility to
Structure-Activity Relationships in Defensin Dimers

Defensin dimers are present as monomers and that disulfide bonds are involved in maintaining their tertiary structure. Densitometry analysis of the stained gel gave the following monomer/dimer ratios for each of the peptides: Defrl-reduced (Fig. 3, lane 1), monomer/dimer (2:1); Defrl-oxidized (lane 2), monomer/dimer (1:5); Defrl Y5C reduced (Fig. 3, lane 3), monomer/dimer (>100:1); Defrl Y5C-oxidized (Fig. 3, lane 4), monomer/dimer (1:6). It is interesting to note the observation of significant amounts of Defrl dimer even after incubation with DTT, which suggests increased stability of the Defrl dimer over the Defrl Y5C dimer under reducing conditions.

Previous studies on various antimicrobial peptides have revealed that some, but not all, can dimerize. For example, HBD3 is a potent antimicrobial peptide isolated from human skin, which has been the subject of intensive structural and functional studies (29). By native gel electrophoresis, dimeric HBD3 was observed under oxidizing conditions, whereas monomeric HBD3 was detected after reduction (25). Modeling studies predict that the dimer is formed through a combination of electrostatic salt bridges and H-bonding between amino side chains. Octamers, formed by amide backbone interactions, have also been observed in the crystal lattice of HBD2 (20). The formation of higher order aggregates of antimicrobial peptides has been proposed as one factor that contributes to their ability to disrupt bacterial membranes (11). Because oxidized Defrl and Defrl Y5C behave in a similar manner by gel electrophoresis, dimerization alone cannot explain the significant differences we observed in their antimicrobial activity.

Characterization of the Nature of Dimerization by Mass Spectrometry—To identify the structural differences between Defrl and Defrl Y5C, which could account for the increased antimicrobial activity of the five cysteine-containing defensin, we used high resolution FT-ICR mass spectrometry. The resolving power and accuracy of this instrument allows the determination of the mass and isotopic distribution of large biomolecules. A comparison of experimentally observed values to those predicted based on the empirical formula for the oxidized and reduced forms of both defensins can then be used to determine the number of disulfide bonds in each species.

Analysis of the ion envelope and deconvolution for Defrl Y5C suggested the presence of two species (Fig. 4, A and B). The isotopic distributions and masses fit very well to those predicted for a Defrl Y5C monomer containing three disulfide bonds (elemental composition, C_{157}H_{254}N_{50}O_{43}S_{6}; average mass, 3722.4490 Da; see Fig. 4A) and to a dimeric Defrl Y5C with six disulfide bonds (C_{314}H_{508}N_{100}O_{86}S_{12} and average mass 7444.889 Da, Fig. 4B). Artificial dimerization and formation of higher order aggregates have been observed for proteins under the conditions used in ESI (30). However, we do not believe the dimer to be an artifact because it was also observed by native gel electrophoresis in a 6-fold excess over the monomer. Nevertheless, because the dimer peak displays 16% relative abundance compared with the monomer, it is clear that the dimer is not stable to the electrospray process suggesting dimer formation by weak, noncovalent interactions.

FT-ICR analysis on oxidized Defrl produced the mass spectrum, which upon deconvolution gave rise to one major species (Fig. 5) whose isotopic distribution matched that expected for a fully oxidized dimer containing five disulfide bonds (elemental composition for Defrl dimer, C_{326}H_{518}N_{100}O_{86}S_{10}; average mass, 7566.9761 Da). Most interestingly, there was no peak corresponding to Defrl monomer. In contrast to that observed for Defrl Y5C, the Defrl dimer remains intact under electrospray conditions, which implies a strong interaction between monomers. Because a Defrl dimer contains 10 cysteine residues, and a fully oxidized isoform has five disulfide bonds, dimerization can only occur through formation of at least one intermolecular disulfide bridge.

Dissociation of Defensin Dimers—CID and ECD are powerful mass spectrometry techniques used for analyzing protein struc-
ture, which complement traditional methods (31–33). These fragmentation techniques were employed to characterize the protein-protein interactions mediating dimerization in Defr1 Y5C and Defr1.

For CID analysis of the Defr1 Y5C dimer, a peak corresponding exclusively to the +5 charge state of the dimer (m/z 1490) was isolated and subjected to dissociation (Fig. 6). The +5 ion readily dissociates into two monomers with +3 (m/z 1242) and +2 (m/z 1862) charge states (Fig. 6, inset). Dissociation of the dimer occurs without fragmentation of the peptide backbone, demonstrating that the Defr1 Y5C dimer is unstable and thereby supporting our hypothesis that dimerization is mediated through noncovalent interactions.

In stark contrast, the Defr1 dimer was stable to the same CID conditions used for Defr1 Y5C and gave rise to the most abundant ion corresponding to the +7 charge state of the Defr1 dimer (m/z 1082) (Fig. 7). When this stable dimeric ion was isolated and subjected to dissociation, no significant monomeric “daughter” fragments were observed. Increasing the amount of gas into the collision cell still did not dissociate the dimer, but it did allow us to partially sequence the peptide because it gave rise to a distinct b-type fragment series from the N termini up to the location of the first cysteine (b2–b10) and a y-type fragment resulting from the loss of a C-terminal lysine (Fig. 7, inset). Such stability under CID conditions indicates the Defr1 dimer is held together by covalent bonding.

These observations are complimented by ECD, where cleavage of Cys–Cys disulfide bridges is known to be a favored process (34). When the isolated +7 charge state of the Defr1 dimer was subjected to ECD, the molecule readily dissociated into monomers with charge states +2, +3, and +4 (m/z 1893, 1262, and 949, respectively) (Fig. 8). On closer inspection (Fig. 8, inset), we observed low intensity species with masses ±16 Da on either side of the +2 monomer peak, which indicate the gain/loss of a sulfur atom. We can explain the appearance of these species only if the Defr1 dimer is formed by a covalent intermolecular disulfide bond. Here the ECD process has cleaved the dimer into monomers by two mechanisms. First, symmetric cleavage of the S–S bond gave a monomer signal with m/z 1892.5. The second pathway involves asymmetric cleavage of the C–S bond of the intermolecular disulfide bridge to give rise to two monomers: one with a persulfide SH at m/z 1909 and its corresponding partner having lost S at m/z 1876.5.

The mechanisms of the cleavage of the S–S and S–C bonds are
analagous to those observed previously by CID (35). We are currently using SORI-CID and ECD to carry out a full "top-down" sequence analysis of various defensins, which is out with the scope of this present study. The use of high resolution FT-ICR, coupled with two dissociation techniques, has revealed the difference between Defr1 Y5C and Defr1 and for the first time has identified a defensin dimer containing an intermolecular disulfide bond.

Reverse Phase Liquid Chromatography—HPLC on a C18 reverse phase column was used to investigate structural differ-
These results are consistent with the second tryptic fragment of Defrl Y5C produced two major fragments that were easily cleavage and analysis of the resulting peptide fragments by MALDI-TOF mass spectrometry. Trypsin cleavage was shown to form several different oxidized isoforms from a first fragment we can assign the connectivity as either Cys' - Cys 1 and Cys 3 - Cys 6 or Cys'-Cys 1 and Cys 3 - Cys 6 . The second fragment we can assign the absolute connectivity. However, from the mass of the first fragment we can assign the connectivity as either Cys' - Cys 1 and Cys 3 - Cys 6 or Cysi-Cyse 1 and Cys 3 - Cys 6 . The second fragment we can assign the absolute connectivity. However, from the mass of the first fragment we can assign the connectivity as either Cys' - Cys 1 and Cys 3 - Cys 6 or Cysi-Cyse 1 and Cys 3 - Cys 6 . The second fragment we can assign the absolute connectivity. However, from the mass of the first fragment we can assign the connectivity as either Cys' - Cys 1 and Cys 3 - Cys 6 or Cysi-Cyse 1 and Cys 3 - Cys 6 . The second fragment we can assign the absolute connectivity. However, from the mass of the first fragment we can assign the connectivity as either Cys' - Cys 1 and Cys 3 - Cys 6 or Cysi-Cyse 1 and Cys 3 - Cys 6 .

**Table II**

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Calculated mass (Da)</th>
<th>Observed mass (Da)</th>
<th>Calculated mass (Da)</th>
<th>Observed mass (Da)</th>
<th>Disulfide bridge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment 1</td>
<td>(Asp 1'-Arg 2') + (Cys 16'-Arg 20') + (Cys 22'-Lys 24)</td>
<td>1711.84</td>
<td>1711.87</td>
<td>1711.84</td>
<td>1711.55</td>
</tr>
<tr>
<td>Fragment 2</td>
<td>(Asn 1'-Arg 2') + (Ile 23'-Lys 24)</td>
<td>1816.86</td>
<td>1817.54</td>
<td>1532.66</td>
<td>1532.70</td>
</tr>
</tbody>
</table>

* Calculated and observed masses of tryptic fragments (in Da).

**Structure-Activity Relationships in Defensin Dimers**

Oxidized Defrl Y5C eluted as a single sharp peak at 26.8 min on a 20% acetonitrile gradient. This species had a mass of 3722.0 Da by ESI-MS in good agreement with the predicted value of 3722.3 Da for Defrl Y5C with the loss of six hydrogens. Upon incubation with excess DTT, the single peak eluted earlier in the gradient at 23.6 min, and this shift in retention time suggested that reduction of the disulfide bonds caused a change in the structure of the defensin (the mass of this species is 3728.1 Da). It is thus apparent that Defrl Y5C folds to give a single species. Upon refolding from a denatured state, we would normally expect a shift to shorter retention time on a reverse phase column resulting from burial of hydrophobic residues. The refolding mechanism of defensins is not well understood but appears to be governed by the formation of the disulfide bonds. Because defensins are amphipathic and contain surface-exposed hydrophobic and cationic patches, these properties can be postulated to explain this unusual chromatographic behavior, which has also been observed for the oxidative folding of HBD2 (15).

In contrast, oxidized Defrl eluted as a broad peak between 22 and 24 min, which we were unable to resolve despite trying a range of conditions (under ESI conditions Defrl 7566.0 Da in agreement with a dimer with the loss of 10 protons). Upon reduction with excess DTT, Defrl eluted as a single sharp peak at 21.9 min (mass 3789.0 Da, predicted mass for fully reduced 3788.2 Da). These data suggest that during oxidative refolding Defrl has formed a number of different species all of which are dimeric. This phenomenon has been observed for HBD3, which was shown to form several different oxidized isoforms from a single reduced precursor (15). The differences in the chromatographic behavior of Defrl and Defrl Y5C again suggest that their structures are dissimilar.

**Determination of Disulfide Connectivity**—To determine their S-S connectivities, both defensins were subjected to protease cleavage and analysis of the resulting peptide fragments by MALDI-TOF and Q-TOF mass spectrometry. Trypsin cleavage of Defrl Y5C produced two major fragments that were easily separated by HPLC (Table II). Fragment 1 had a mass of 1711.87 Da and was unmodified by treatment with chymotrypsin, consistent with the species (Asp 1'-Arg 2') + (Cys 16'-Arg 20') + (Cys 22'-Lys 24) containing two disulfide bonds (calculated mass 1711.84 Da). Because (Cys 22'-Lys 24) contains two adjacent cysteines (Cys 2' and Cys 6), these data alone cannot be used to assign the absolute connectivity. However, from the mass of the first fragment we can assign the connectivity as either Cys 1' - Cys 6 and Cys 2' - Cys 5 or Cys 1'-Cys 6 and Cys 2' - Cys 5. The second tryptic fragment had a mass of 1817.54 Da and was susceptible to chymotrypsin producing a species with mass 1532.70 Da. These results are consistent with the second tryptic fragment being (Asn 1'-Arg 2') + (Ile 23'-Lys 24) containing 1 disulfide bond (calculated mass 1816.86 Da), and its chymotrypsin derivative being (Asn 1'-Tryr 14') + (Ile 23'-Phes 20) (calculated mass 1532.66 Da), allowing the assignment of a disulfide bond between Cys 2' - Cys 6. To distinguish between Cys 5 and Cys 6 in fragment 1, we used a Q-TOF tandem mass spectrometer and CID to sequence the peptides. Dissociation of the +3 charge state of fragment 1 resulted predominantly in the sequential loss of residues from the N terminus of the fragment (Asp 1'-Arg 2'). However, a +2 species was observed with m/z 404.96 Da (deconvoluted mass 408.92 Da), which we attribute to (Cys 16'-Arg 20') + (Cys 22'-Lys 24) (calculated mass 807.41 Da), which allows the assignment of the disulfide bridge between Cys 2' and Cys 5 and thus, by elimination, Cys 1'-Cys 6. In summary, oxidized Defrl Y5C monomers contain the typical β-defensin S-S connectivity (Cys 1'-Cys 6 and Cys 2' - Cys 5). It is worth noting that the NMR structure of mBD-8 (NEPVSICRRGGIGCYRCIGL-RHKIGTGCSPPFKCK) has been determined (Protein Data Bank accession code 1E4R) (29). This defensin is very similar to Defrl Y5C, the only differences being at the N terminus (NEPVS of mBD-8 replaces DPVT of Defrl Y5C). Because the sequences of mBD-8 and Defrl Y5C are highly similar and they both display β-defensin connectivities, we suggest that mBD-8 is a good model for the three-dimensional structure of Defrl Y5C.

In contrast to its six-cysteine analogue, cleavage of oxidized Defrl Y5C with a combination of trypsin and chymotrypsin produced a complex mixture of peptide products that proved impractical to separate by HPLC. Instead, the unfraccionated mixture was analyzed by MALDI-TOF MS. The mass spectra obtained from analysis of the reduced peptide products revealed that proteolysis had been incomplete, and many peptides were present that contained uncleaved, internal proteolysis sites (see Table III). Analysis of the oxidized tryptic digest peptide products enabled the assignment of disulfide connectivities (see Table IV). Peaks were observed with masses consistent with peptide fragments with Cys 5-Cys 6 and Cys 2'-Cys 5 or Cys 1'-Cys 6 and Cys 3'-Cys 5. The second tryptic fragment had a mass of 1817.54 Da and was susceptible to chymotrypsin producing a species with mass 1532.70 Da. These results are consistent with the second tryptic fragment being (Asn 1'-Arg 2') + (Ile 23'-Lys 24) containing 1 disulfide bond (calculated mass 1816.86 Da), and its chymotrypsin derivative being (Asn 1'-Tryr 14') + (Ile 23'-Phes 20) (calculated mass 1532.66 Da), allowing the assignment of a disulfide bond between Cys 2' - Cys 6. To distinguish between Cys 5 and Cys 6 in fragment 1, we used a Q-TOF tandem mass spectrometer and CID to sequence the peptides. Dissociation of the +3 charge state of fragment 1 resulted predominantly in the sequential loss of residues from the N terminus of the fragment (Asp 1'-Arg 2'). However, a +2 species was observed with m/z 404.96 Da (deconvoluted mass 408.92 Da), which we attribute to (Cys 16'-Arg 20') + (Cys 22'-Lys 24) (calculated mass 807.41 Da), which allows the assignment of the disulfide bridge between Cys 2' and Cys 5 and thus, by elimination, Cys 1'-Cys 6. In summary, oxidized Defrl Y5C monomers contain the typical β-defensin S-S connectivity (Cys 1'-Cys 6 and Cys 2' - Cys 5). It is worth noting that the NMR structure of mBD-8 (NEPVSICRRGGIGCYRCIGL-RHKIGTGCSPPFKCK) has been determined (Protein Data Bank accession code 1E4R) (29). This defensin is very similar to Defrl Y5C, the only differences being at the N terminus (NEPVS of mBD-8 replaces DPVT of Defrl Y5C). Because the sequences of mBD-8 and Defrl Y5C are highly similar and they both display β-defensin connectivities, we suggest that mBD-8 is a good model for the three-dimensional structure of Defrl Y5C.

Our analysis revealed Defrl has S-S connectivities not typical of either α- or β-defensins. It appears from our proteolysis/mass mapping data that certain cysteine residues (e.g. Cys 5) can form disulfide bonds with more than one other cysteine. This can only be explained if Defrl is a mixture of topologically different dimeric isoforms, rather than a single species with a defined connectivity. The HPLC analysis also supports the unusual nature of Defrl. The broad peak resulting from reverse phase-HPLC analysis of oxidized Defrl again suggests that Defrl is heterogeneous. The high resolution FT-ICR data prove that all of these isoforms are dimeric connected by an
intermolecular disulfide bridge. However, it appears that the intermolecular disulfide bond is not formed exclusively between two equivalent cysteine residues from each monomer, e.g., Cys$^7$-Cys$^8$ or Cys$^4$-Cys$^4$. Rather, our proteolysis and chromatography data suggest that the Defr1 dimer can be formed via any two cysteine residues from each monomer from 15 possible combinations, e.g., Cys$^2$-Cys$^2$. Recent work on synthetic HBD3 led to the surprising discovery that it did not fold into a single isoform with typical β-defensin disulfide connectivity in vitro (15). Instead, at least seven topologically distinct isoforms were isolated with unusual intramolecular connectivities, all of which displayed similar antimicrobial activity. It is noteworthy that the S-S connectivity of native HBD3 purified from human skin has not been determined (29). Most interestingly, none of these isoforms contained an intermolecular disulfide bond. The authors were unable to determine the individual connectivity of these isoforms, so they devised a synthetic protocol that allowed the control of disulfide bond formation. We are currently exploring whether this strategy can be used to control the formation of a specific intermolecular disulfide bond in Defr1.

**Conclusions—**β-Defensins are characterized by having six conserved cysteines that are oxidized to form three disulfide bonds. It was assumed that these were required for antimicrobial activity, but several recent studies have begun to explore their structural and functional roles. Analogues of bovine β-defensin BNBD-12 containing one, two, and three disulfide bridges, synthetic peptides corresponding to the C-terminal segment of bovine β-defensin BNBD-2, and a cysteine-free, full-length analogue of HBD3 all displayed antimicrobial activity suggesting that the three disulfide bridges are not absolutely required (15, 16, 36). Moreover, to explore whether high cationic charge is the only requirement for activity, small linear peptides corresponding to regions of HBD3 were analyzed for their ability to kill different pathogens (37). Somewhat surprisingly, the least positively charged peptides proved to be the most active, which suggests that defensins contain inherent subtle structural features that come into play upon interaction with the membrane target.

Previously we reported the discovery of a defensin (Defr1) with five cysteine residues expressed in murine heart and testis that had potent antimicrobial activity (26). This defensin lacks the first of the conserved six cysteines found in other members of the defensin family. Here we explored the effect of this loss on the structural and antimicrobial properties by comparison with its six-cysteine analogue Defr1 Y5C. Defr1 Y5C has modest, salt-sensitive antimicrobial activity against a panel of microbial pathogens, and analysis by electrophoresis and mass spectrometry suggest it forms a noncovalently bound dimer in solution. Peptide mass mapping revealed that this defensin has a novel structural feature not previously observed within the defensin family. Proteolysis and HPLC analysis suggest that oxidized Defr1 is a complex mixture of dimer isoforms varying in their intra- and intermolecular disulfide connectivities. Recent studies of defensins HBD3 and BN2B have concluded that short linear peptide fragments as well as those containing only one intramolecular disulfide could have potential as therapeutic agents. Our results demonstrate that defensins containing intermolecular disulfide bonds have novel structural and antimicrobial properties and suggest that covalently bound dimeric cationic peptides are promising targets for future study.

**Acknowledgments—**We thank Robert Smith, Bryan McCulloch, Nick Tomczyk and Dr. Jim Creanor for their help with mass spectrometry analysis. We also thank Professor Nick Hastie of the MRC Human Genetics Unit and Dr. Pat Langridge-Smith (Director, Scottish Instrumentation and Resource Centre for Advanced Mass Spectrometry (SIR-CAMS)) for their enthusiastic support of this work.
REFERENCES

Is it biologically relevant to measure the structures of small peptides in the gas-phase?


a School of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, UK
b Department of Medical Microbiology, University Medical School, University of Edinburgh, Edinburgh EH8 9AG, UK
c MRC Human Reproductive Sciences Unit, Centre for Reproductive Biology, The Chancellors Building, Edinburgh EH16 4SB, UK
d Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106-9510, USA

Received 15 September 2004; accepted 27 September 2004
Available online 10 November 2004

Abstract

Recent developments in sample introduction of biologically relevant molecules have heralded a new era for gas-phase methods of structural determination. One of the biggest challenges is to relate gas-phase structures, often measured in the absence of water and counter ions, with in vivo biologically active structures. An advantage of gas-phase based techniques is that a given peptide can be analysed in a variety of different forms, for example, as a function of charge state, or with additional water molecules. Molecular modelling can provide insight into experimental findings and help elucidate the differences between structural forms. Combining experiment and theory provides a thorough interrogation of candidate conformations. Here two important naturally occurring peptide systems have been examined in detail and results are assessed in terms of their biological significance.

The first of these is gonadotropin-releasing hormone (GnRH), a decapeptide which is the central regulator of the reproductive system in vertebrates. We have examined several naturally occurring variants of this peptide using Ion Mobility Mass Spectrometry and Electron Capture Dissociation (ECD) in conjunction with Fourier Transform Ion Cyclotron Mass Spectrometry (FT-ICR-MS). Candidate conformations are modelled using the AMBER force field. Single amino acid changes, for example Gly6 → Ala6, or Ala6 → D-Ala6, have observable effects on the gas phase structure of GnRH. It has been shown that evolutionary primary sequence variations are key to the biological activity of GnRH, and it is thought that this is due to different binding affinities at target receptors. This work provides strong evidence that this activity is structurally based. The second system examined is the relationship between the quaternary structure and activity of two novel β-defensins. FT-ICR mass spectrometry has been employed to characterize di-sulphide bridging and dissociation based experiments utilised to investigate their structural core. Defr1, with five cysteines, exists as a covalently bound disulphide linked dimer; Defr1 Y5C with six cysteines also is observed as a dimer, but non-covalently bound, suggesting that this defensin has a tendency to aggregate. The activity of Defr1 is 10 times higher than that of Defr1 Y5C when tested against the pathogen Pseudomonas aeruginosa. The results from these studies could inform future design of novel GnRH type ligands and anti-microbial agents, and illustrate the power of gas-phase based techniques for solving peptide structures.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Ion mobility; Peptides; ECD; FT-ICR

1. Introduction

Advances in proteomic technologies, coupled with the sequencing of the human genome have resulted in vast numbers of peptide ions being ‘measured’ daily in mass
s stirred extensively, and the clear solution used for the experiments. The peptide concentration was determined using UV-visible spectrometry at a wavelength of 280 nm.

The results indicated that the folding of the peptide is influenced by the presence of crowding agents, such as polyethylene glycol (PEG) and D-sorbitol. The folding process was found to be dependent on the peptide concentration and the type and concentration of the crowding agent. The data suggest that the crowding agents affect the peptide folding by perturbing the solvent structure and thus altering the free energy landscape of the folding reaction.

In summary, the results presented in this study provide insights into the folding of the studied peptide in the presence of crowding agents. The findings can be useful for designing strategies to control the folding of similar peptides in biotechnological and medical applications.
2. Experimental

2.1. Mass spectrometry

Mammalian GnRH used in this study was obtained from Sigma-Aldrich. Ciona1 variants were synthesised by conventional solid phase methodology and purified by HPLC. All GnRH peptides were examined without further purification and made up at a concentration of 20 μM (H2O:MeOH:CH3COOH, 0.49:0.49:0.02) for ECD and at 100 μM under the same solution conditions for ion mobility. β-Defensins were obtained from Albachem (UK) and used without further purification. Electrospray solution and source conditions were the same for accurate mass, SORI-CID and ECD. Defensins were sprayed at a concentration of 50 μM (MeOH:H2O:CH3COOH, 50:49:1). Dimeric signal was still observed for Defr1 Y5C peptide on ten fold dilution. Solutions were ionised by electro spray and nano-electrospray, from gold/palladium coated tips (Proxeon Biosystems).

Analysis was conducted on two instruments. Ion mobility experiments were conducted on an instrument in the laboratory at UCSB which has been described in detail elsewhere [26]. Ions drift through a cell pressured to 5 Torr with helium at room temperature and are mass selected prior to detection. Arrival time distributions (ATDs) were collected at room temperature over no more than 5 min for each GnRH variant. ATDs were converted to collision cross sections according to transport theory [27] for comparison with values obtained from modelled peptides structures.

Characterization of the oxidation state of the defensin peptides Defr1, and Defr1 Y5C, was performed using the accurate mass capabilities of a 9.4 T FT-ICR mass spectrometer (Bruker Daltonics) at the University of Edinburgh. Using this instrument, two dissociation methods were applied to the isolated peptides. For Sustained Off Resonance Collision Induced Dissociation (SORI-CID) a given charge state was isolated by sweep excitation and subjected to CID with Argon as the collision gas for 500 ms. To perform ECD on both the GnRH peptides and the defensins, a given charge state was isolated by sweep excitation and subjected to electron irradiation for 1 ns at 300 K, to determine the conformational flexibility of the minimised structures. Representative snapshots of co-ordinates obtained during this run are those discussed within the text. These structures were viewed and analysed using VMD [30]. An orientation averaged projection cross section is calculated from each candidate structure geometry, which is scaled according to the temperature dependent helium interaction potential [31].

3. Results and discussion

3.1. GnRH—gonadotropin releasing hormone

3.1.1. Ion mobility studies on mammalian GnRH and Ciona1 structural variants

Results of the measurements made and comparison with model geometries are shown in Table 1. The four peptides here are mammalian GnRH and three structural variants, one of which is the naturally occurring form for the lower non-vertebrate Ciona1 (a type of sea squirt). The other two are mutants of Ciona1 where the sixth residue (shown in bold) has been substituted Ala → Gly and Ala → D-Ala. Primary sequences for the two naturally occurring peptides are shown below:

Mammalian GnRH: pGlu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH2
Ciona1: pGlu-His-Trp-Ser-Tyr-Ala-Leu-Ser-Pro-Gly-NH2

There is good agreement between experiment and theory for all peptides examined with differences within expected experimental error. The number in brackets indicates that a
range of cross sections are found for low energy structures, however, the mean value is close to that found experimentally in each case.

Two major trends are apparent in this data: the first is that the Ciona singly charged peptides have slightly larger cross sections than their doubly charged equivalents, but for the mammalian form the reverse is true. The second is that the size order of cross sections obtained for all the forms of GnRH both singly and doubly charged is found to be:

Mammalian GnRH < Ciona Gly6 Cional DAla6 < Ciona1. These trends are best discussed with reference to the molecular models obtained.

3.1.2. Molecular modelling—the structures

For gas phase peptide ions the amino acid which has the highest gas phase basicity is presumed the first to protonate [32]. Considering the peptides studied here, it is most likely to be Arg 8 for the mammalian form and His2 for the Ciona1 variants. To assign a second proton to the \([M+2H]^{2+}\) ions we assume that His2 will also protonate in mammalian GnRH. In the absence of a second basic amino acid for the Ciona1 variants the next most basic site is the C terminal amide group. These assignments are used in modelling to obtain low energy structures whose calculated cross sections appear in the second and fourth columns of Table 1. In all force field calculations sites of protonation are fixed at the start of measurements however experimentally, and particularly on collisional activation, protons on less basic sites are potentially mobile [33]. The relevance to this study will be discussed later, although the good correlation between experimental and calculated cross sections, suggests that it is not a significant problem with the ion mobility measurements of these structures.

Figs. 1 and 2 illustrate representative structures obtained for structures of the \([M+H]^+\) and \([M+2H]^{2+}\) GnRH peptide ions from molecular dynamics runs at 300 K. In all cases the backbone forms a loop which is stabilised by the presence of hydrogen bonds predominantly from side chain groups to backbone carbonyls. With mammalian GnRH Arg8 provides an excellent ‘cap’ to the peptide backbone (Fig. 1A). In the Ciona1 variants, this role is taken by His2 (Fig. 1B–D). The presence of the second protonation site at the amidated C terminal glycine allows for additional H bond formation, with carbonyl groups along the polypeptide chain. This bonding is analogous to an effect which occurs from C terminal lysine groups [34]. The effect of these cross chain hydrogen bonds is to tighten the structure, which accounts for the decrease in size from \([M+H]^+\) to \([M+2H]^{2+}\). The capping of the polypeptide backbone can be seen in Fig. 2 for all the Ciona1 peptides. The protonated histidine residue in particular with Ciona Gly6 (C) and DAla6 (D) is able to interact with groups situated on the opposite face of the polypeptide chain. There is a play off here with cumblic repulsion that would occur between the two charged groups, and might be expected to increase the cross section of the doubly protonated ions. However, the Ciona1 decapeptide systems orientate themselves such that coulombic repulsion is not dominant, consistent with the slight decrease in cross section from single to doubly protonated ions.

The difference in the cross sections obtained for the single and doubly protonated forms of the mammalian form follows an opposite trend. Here the protonated guanidinium group forms hydrogen bonds to backbone carbonyl groups as can be seen in Fig. 1A. This bonding fixes the structure and subsequent protonation of the bulky histidine side chain results in charge crowding, angling the imadozole ring away from
the peptide, providing a slight increase in the cross section for the \([M+2H]^2+\) ion.

As stated above, for the Cional peptides conformations obtained via simulated annealing are similar. The natural Cional GnRH is somewhat larger than its chiral analogue CionalDAla. Cross sections for this DAla6 variant in both charge states are similar to those for the Gly6 variant. This implies that the presence of the D amino acid at residue 6 allows for a slightly more compact structure which can form a \(\alpha\) helix II type turn, as with the achiral glycine residue. It appears that the LAla isoform introduces steric hindrance which yields a more open polypeptide than that found in either the DAla or the Gly form. This is supported by molecular dynamics calculations. Fig. 3 illustrates this for Cional and CionalGly6. It can be seen that the distance between residues 3 and 7 (Trp and Leu) is significantly greater in the Cional peptide over the course of a 1 ns dynamics run than that for CionalGly6. Across this defined co-ordinate, which effectively is the diameter of the quasi circular form taken by these peptides, the Cional1Gly6 \([M+H]^+\) peptide exhibits somewhat higher conformational flexibility than Cional. The cross-peptide distance obtained for Cional1DAla averages at 6.3 Å which falls midway between that seen here for Cional1Gly6 and Cional.

### 3.1.3. ECD of mammalian GnRH

The use of relative fragmentation intensities to determine gas-phase stability has been employed by many researchers. When examining small molecular systems insights obtained from collision or photo induced dissociation are used with electronic structure calculations, in order to confirm proposed ionic conformations [35,36]. Such approaches rely on accurate knowledge of the process by which fragmentation proceeds. ECD is a relatively new technique, which has principally been applied to determination of primary structures of peptide and proteins [37,38]. Since the precise mechanism of the electron capture and subsequent dissociation process is not yet known, it is difficult to interpret results in terms of the secondary structures of the ions involved. However, we believe that this technique can provide information on gas phase conformations. Polfer [39] has obtained exceptionally reproducible fragmentation patterns from performing ECD.
A similar effect is seen for the production of the c ion. As might be expected from peptides where 9 out of 10 amino acids are the same, there are also some comparable features. For example, the c_{2}^{+} and c_{5}^{+} product channels appearing at similarly low abundances for each GnRH variant.

Quantitative analysis of this data with reference to gas phase conformation(s) of these peptides is not conclusive. If the ECD mechanism proceeds via electron capture at a proton which is hydrogen bonded to a backbone carbonyl, then it appears that the presence of L-Ala reduces the formation of such a bond, although two out of the 10 low energy structures examined did show such a bond. Since there is still some fragmentation here for the Ciona1 peptide, it may simply be that the opening of the peptide backbone lessens the likelihood for hydrogen bond formation at this point. The use of force fields to provide gas phase structures is limited by the absence of a parameterised mobile proton. Our calculations have all employed the amidated N terminus as the second protonation site. This choice introduces some flexibility and as can be seen in Fig. 2B-D, where the protonated amidated glycine forms hydrogen bonds to several backbone carbonyls. Whilst our ion mobility measurements support this structure, it is possible that on collisional activation (for example in the transfer hexapole of the FT-ICR) such a proton would be mobile prior to ECD. A proper comparison of ECD data to gas phase structures must consider this.

3.1.4. Biological significance

Many thousand synthetic analogues of GnRH have been made and so the significance of each amino acid in this peptide towards binding at the receptors in the anterior pituitary has been well established in this intense research field [6]. What is clear from earlier work is that the presence of a chiral amino acid at residue 6 significantly reduces binding at the mammalian GnRH I receptor [6,42]. This is presumed to be due to the inability of peptides to form the tight type II'β turn which preconfigures the ligand prior to receptor binding and activation. The structures shown in Fig. 1B and C taken in parallel with the ion mobility results illustrate that the Ciona1 peptide has a looser conformation than the DA1a6 or Gly6 structural variants. Some low energy structures for both the mammalian GnRH and the DA1a6 and Gly6 Ciona1 peptides do exhibit such a turn at or around residue 6. Data obtained from the [M+2H]^{2+} ions is less relevant here, since the C terminal glycine is probably not protonated in vivo, however our preliminary ECD results also imply a different gas phase conformation for these peptide ions, which also reflects the influence that single amino acid changes can have on structure. The binding affinity of the Ciona1 peptides will be the subject of a future communication [43].

3.2. B-Defensins

3.2.1. High resolution mass spectrometry of defensins

The sequences of the two defensins examined here are below. They differ by just one amino acid, the substitution of Tyr → Cys forming Defr1-Y5C which contains six cysteines,
the deconvoluted signal seen from the Defrl dimer fits the extremely closely responding to the monomer of Derfi (Fig. 6A), and supports the accurate mass assignment we have made. Interestingly there was no peak corresponding to the monomer of Defrl + iodoacetamide (region not shown), suggesting that no monomer is present in solution prior to mass spectrometric analysis. This strongly implies that Defrl exists as a dimer linked by an intermolecular disulfide bridge, and that if there is some partially reduced defensin in solution it will be at cysteines which are not involved in the intermolecular disulfide bridge. In short this peptide is dimeric with five disulfide bridges. To further characterize this we employed dissociation techniques including reduction of disulfide by ECD to probe the stability of the defensin peptides in the gas phase. For Defrl Y5C the isotopic distribution obtained fits well to that predicted for a fully oxidized monomeric peptide, so we can conclude that this peptide contains three intramolecular disulfide bonds (Fig. 6B). We also observed a Defrl Y5C dimer which appears to be composed of two non-covalently associated monomers, with all six disulfide bonds formed (Figs. 6C and 6B). Since this mass spectrum was acquired under denaturing acidic nano-spray conditions, and in the harsh desolvation conditions of the Bruker Apex III source, we presume that Defrl Y5C must form a dimer due to favourable non-covalent interactions, which is the conclusion of others on observation of oligomeric peptides via MS [44]. On lowering the concentration to 5 μM we still observe the dimeric form of this peptide. This tendency to aggregate is also born out by gel electrophoresis, which has also revealed a dimeric form of Defrl Y5C [25].

3.2.2. Determination of the structural core via CID and ECD

CID of the Defrl dimer gives rise to b-type fragments from the N terminus up to the location of the first cysteine (b2-b10) and a y-type fragment resulting from the loss of the C-terminal lysine, indicating a distinct structural core (Fig. 6A). Some very low intensity species are apparent with masses ±32 arising from monomer units formed by cleavage of the C-S bond of the intermolecular cysteine bridge. The lack of any significant monomeric ‘daughter’ fragments, demonstrates that the dimer is covalently bound. These findings are further supported by ECD where cleavage of Cys–Cys disulfide bridges is a favoured process [41]. Here this appears to occur preferentially at the intermolecular disulfide bridge leading directly to two monomers (Fig. 7). We can conclude from this, that the intermolecular bridge is accessible for cleavage, at least when this dimer exists as a gas phase ion.

We also performed CID on the Defrl Y5C dimer and in contrast to Defrl it readily dissociated into monomers (Fig. 6B and C), supporting the conclusion of the accurate mass work above that this is a non-covalently bound dimer. b2-4 fragment ions were observed, corresponding to loss of the shorter N terminal section of this peptide, but as for Defrl1, the disulfide bridged core resisted collision induced dissociation.

Fig. 5. (A) High resolution mass spectrum of oxidized Defrl dimeric isoforms. The triangles correspond to the isotopic envelope calculated from the Defrl amino acid sequence containing five disulfide bonds [46]. (B) Deconvoluted isotopic envelope from FT-ICR nanospray analysis of oxidized Defrl Y5C. The triangles correspond to the isotopic envelope expected from a Defrl Y5C monomer with all cysteines oxidized, i.e containing three disulfides. (C) Equivalent spectra for the dimeric form of Defrl Y5C, here the triangles represent the isotopic envelope expected from a dimer with all cysteines oxidized, i.e., six disulfide bonds [46].

from the five-cystine containing murine defensin Defrl.

Defrl DPVTYIRNGGICQYRCIGLRHKIGTCGSFKCK
Defrl Y5C DPVTClRNGGICQYRCIGLRHKIGTCGSFKCK

Using FT-ICR mass spectrometry we have been able to unambiguously define the number of cross-linked cysteine residues in each peptide, by comparing the measured isotopic distribution to that expected for a fully or partially reduced species. Since it is possible to reduce cysteine in the electrospray process, accurate mass was performed on defrl pre-incubated with iodoacetamide. After this treatment, the deconvoluted signal seen from the Defrl dimer fits exactly to that expected for fully oxidized Defrl, ruling out anomalous gain or loss of hydrogen’s during the electrospray process, and supports the accurate mass assignment we have made. Interestingly there was no peak corresponding to the monomer of Defrl + iodoacetamide (region
3.2.3. Determination of disulphide connectivity's via 'top-down' sequencing and peptide mass mapping

A combination of two dissociation techniques were employed to attempt to discern the disulphide bridging of Defr1: ECD to cleave (selectively) the cysteine bridges and SORI-CID to dissociate the peptide. If sustained ECD alone is used, it is possible to sequence almost the entire peptide however connectivity information is not always forthcoming [45]. The low energy nature of the ECD process employed here, means that for low charge states, even on backbone
Fig. 7. ECD of Def1, here the [D]3+ dimer ion (m/z = 1081.401) was isolated and irradiated with electrons for 50 ms. The +7 ion underwent fragmentation into +5, +4, +3 and +2 monomers. Shown is an expansion of the +2 monomer ion series which results from both symmetric and asymmetric cleavage of the intermolecular S–S bond.

Fig. 8. Strategy for assigning disulphide connectivity using ECD and SORI-CAD.
cleavage, dissociation into measurable fragment ions does not occur, since non-covalent interactions preserve the gas-phase complex [45]. As stated above, with CID alone cleavage of disulphide bridges is unfavourable and N terminal sequencing to the first cystine only limits the analysis. The strategy which combines ECD and collisional activation is illustrated in Fig. 8.

On isolating the dimeric ion at 1082.797 (z = 7) and subjecting it to ECD followed by SORI-CID several fragmentation channels are apparent (Fig. 9). A series of y-type ions ensue from the loss of N terminal amino acids, which are accompanied by water loss peaks. A dominant channel is reduction of the dimer via cleavage of the intermolecular-disulphide bridge, forming monomer ions as described above. Although this also proceeds via cleavage at the thiol ether bond, resulting in signature satellite peaks —32 Da and +33 Da around the dimer. This provides a useful diagnostic of peaks which arise from the dissociation of the dimeric peptide or, of internal disulphide bridges. Table 2 illustrates some of the internal fragment ions which were observed. Isolation and fragmentation of a dimeric ion gives rise to two strong y-type ions which correspond to the loss of CK+ S and to CCK from a monomer defensin mass. These peaks are not present if a monomer ion is isolated post ECD and subjected to CID. This indicates that they arise from ECD fragmentation of the dimeric ion, into a monomer followed by CID possibly assisted by the location of the Lysine side chain, such that both CID and ECD are directed by the interaction of a protonated site with the peptide backbone [33,37]. The fact that we observe this loss of CK+ S only from the dimer, suggests that this arises from a metastable precursor which contains a cleaved thiol–ether bond, and indeed we see such ions (Figs. 7 and 9). The fact that this loss channel is opened to produce the b32 + S ion with greater intensity that the b32 — S ion indicates a greater stability for the monomer-S species and that is certainly what we see (Fig. 7). Subsequent

Table 2
Internal fragment ions arising from ECD followed by CID, the numbering scheme corresponds to peaks in Fig. 9

<table>
<thead>
<tr>
<th>Peak number in Fig. 5</th>
<th>ECD/CID fragment residue numbers</th>
<th>Disulphide bridge</th>
<th>Satellite S species</th>
<th>Calculated mass (M+H)⁺ (Da)</th>
<th>Observed mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 23-31* and/or 15-1831-34</td>
<td>4-5 and/or 3-6</td>
<td>Yes</td>
<td>892.44</td>
<td>892.49</td>
<td></td>
</tr>
<tr>
<td>2 6-1223-31 and 14-2532-33</td>
<td>2-4 and/or 3-6</td>
<td>Yes</td>
<td>1637.80</td>
<td>819.9217 (2+)</td>
<td></td>
</tr>
<tr>
<td>3 15-2033-34</td>
<td>3-6</td>
<td>No</td>
<td>791.41</td>
<td>791.4440</td>
<td></td>
</tr>
</tbody>
</table>

* Unlikely to be this internal fragment due to the existence of the satellite S peaks which point to the inclusion of an additional cleaved Cys.
fragmentation gives the $b_{13}$ ion shown by Fig. 5 to be a dominant product. These findings suggest that the intermolecular disulphide bridge is formed between $\text{Cys}^1$-$\text{Cys}^3$. This bond is cleaved in some cases asymmetrically which can lead to CID fragment ions which sequence from the C terminus. This process is also dependent on prior cleavage of an internal $\text{Cys}^5$-$\text{Cys}^6$ bond.

The peaks listed in Table 2 provide evidence for the existence of the internal disulphide bonds 2$\text{Cys}^2$-$\text{Cys}^4$ (peaks 1 and 2) and of the internal 3$\text{Cys}^3$-$\text{Cys}^4$ bond (peak 3) where no evidence for asymmetric cleavage is found. Elegant though this technique is, we cannot be certain that these are the only connectivities available. An alternative route was taken whereby the two defensins were subjected to proteolytic cleavage, and subsequent peptide mass mapping and sequencing [25]. Data obtained from Defrl Y5C proved that this six cysteine containing peptide possesses typical $\beta$-defensin $S$-$S$ connectivity ($\text{Cys}^1$-$\text{Cys}^5$, $\text{Cys}^2$-$\text{Cys}^4$, $\text{Cys}^3$-$\text{Cys}^6$). In contrast on digestion, oxidized Defrl yielded a complex mixture of peptide products. Ions were obtained with masses correlating to $\text{Cys}^2$-$\text{Cys}^3$, $\text{Cys}^2$-$\text{Cys}^4$, $\text{Cys}^3$-$\text{Cys}^4$ and $\text{Cys}^4$-$\text{Cys}^4$ disulphide bonds, several of which were confirmed by tandem MS sequencing. In short, disulphide connectivity's present in Defrl are not soley typical of either $\alpha$ or $\beta$ defensins. $\text{Cys}^2$ in particular appears able to form $S$-$S$ bonds with other cysteine than $\text{Cys}^4$. Our conclusion is that Defrl presents a mixture of topologically different isoforms, rather than a single species with a defined connectivity.

3.2.4. Biological relevance

The activity of these $\beta$-defensins against *Pseudomonas aeruginosa* PAO1 was assessed to determine their minimum inhibitory concentrations (MICs) [25]. The MICs were also determined in the presence of the reducing agent dithiothreitol (DTT, 10 mM). The results from our killing assays when taken in tandem with the MS data are remarkable. Defrl Y5C has identical activity in its oxidized and reduced states supporting the recent conclusions of Wu et al. [21] with HBD3 that disulphide-bonds are not required for $\beta$-defensin antimicrobial activity. However, Defrl is an order of magnitude more active in its oxidized form. This demonstrates that this 10 cysteine containing intermolecular, covalently-linked dimer is a significantly more potent form of this peptide. Defensins are found in high concentrations at cell surfaces [18], and here we have demonstrated the activity of a preformed dimeric species. It is possible that this provides a more effective seed for the formation of higher order aggregates than a monomeric or even a reduced 6 cysteine defensin. Alternatively the presence of the untethered N terminal section of Defrl may be assisting initial interactions with cell membranes. These observations have implications for the future design of novel anti-microbial agents.

4. Conclusions

The work presented here demonstrates the power of gas-phase techniques in determining peptide structures. Several complimentary methodologies have been applied in order to probe conformational variations in two important peptide systems which arise from single amino acid changes in primary sequences. We have shown how structural variants of GnRH can be elucidated by both ion mobility mass spectrometry and the use of ECD. Molecular modelling has generated structures with cross sections that correspond to those found experimentally. The presence of achiral glycine at residue 6 enables the formation of a tighter structure which assists the binding affinity of this peptide to the mammalian GnRH 1 receptor. The fact that the gas phase structure can be related to a biologically active form is not entirely surprising, given the lower dielectric medium experienced by a ligand as it binds to a membrane bound protein.

Employing the high resolution capabilities of FT-ICR we have determined the oxidation state of the covalently bound $\beta$-defensin dimer Defrl, and developed a strategy which combines ECD and SORI-CAD to map out its disulphide bridging pattern. We have linked the oligomerisation state of a defensin with its antimicrobial activity and these findings imply that there is a critical concentration of this naturally occurring peptide that is necessary to kill bacteria, which is enhanced tenfold by the formation of an intra-molecular disulphide bridge. These findings which have been inspired by collaborations with bio-medical researchers illustrate the potential for solving real biological problems in the gas phase.

Acknowledgements

We would like to thank the Scottish Higher Education Funding Council (SHEFC), Scottish Enterprise (SE), Bruker Daltonics and the School of Chemistry (University of Edinburgh) for funding the world-class facilities of the Scottish Instrumentation and Research Centre for Advanced Mass Spectrometry (SIRCAMS). The Edinburgh Protein Interaction Centre (EPIC) funded in part by the Welcome Foundation, is thanked for use of its computational facilities. In addition we thank the EPSRC (particularly for the award of an Advanced Research Fellowship to PEB) the MRC, the Cystic Fibrosis Trust, the Royal Society and the NSF for funding these studies.

References

The elemental composition for DefrI dimer is C_{326}H_{518}N_{100}O_{88}S_{10}, average mass 7566.9761 Da. The elemental composition of oxidized DefrI Y5C monomer with three disulfides is C_{157}H_{254}N_{50}O_{43}S_{6}, average mass 3722.4490 Da and a DefrI Y5C dimer with six disulfide bonds is C_{314}H_{508}N_{100}O_{86}S_{12}, average mass 7444.898 Da.
Cloning, expression, purification, crystallization and preliminary X-ray characterization of the full-length single-stranded DNA-binding protein from the hyperthermophilic bacterium Aquifex aeolicus

David J. Clarke, Christopher G. Northey, Lynsey A. Mack, Iain W. McNae, Dmitriy Alexeev, Lindsay Sawyer and Dominic J. Campopiano

Cloning, expression, purification, crystallization and preliminary X-ray characterization of the full-length single-stranded DNA-binding protein from the hyperthermophilic bacterium Aquifex aeolicus

Single-stranded DNA-binding (SSB) proteins stabilize single-stranded DNA, which is exposed by separation of the duplex during DNA replication, recombination and repair. The SSB protein from the hyperthermophile Aquifex aeolicus has been overexpressed in Escherichia coli, purified and characterized and crystals of the full-length protein (147 amino acids; M, 17131.20) have been grown by vapour diffusion from ammonium sulfate pH 7.5 in both the absence and presence of ssDNA [dT(pT)₅₅]. All crystals diffract to around 2.9 A resolution and those without bound DNA (native) belong to space group P2₁, with two tetramers in the asymmetric unit and unit-cell parameters a = 80.97, b = 73.40, c = 109.76 A, β = 95.1°. Crystals containing DNA have unit-cell parameters a = 108.65, b = 108.51, c = 113.24 A and could belong to three closely related space groups (P2₁2₁2₁ or I₄₁) with one tetramer in the asymmetric unit. Electrospray mass spectrometry of the crystals confirmed that the protein was intact. Molecular replacement with a truncated E. coli SSB structure has revealed the position of the molecules in the unit cell and refinement of both native and DNA-bound forms is under way.

1. Introduction

Single-stranded DNA-binding (SSB) proteins have been shown to play an essential role in many aspects of DNA metabolism (Chase & Williams, 1986). They preferentially bind and protect vulnerable single-stranded DNA (ssDNA), which is formed transiently during DNA replication, recombination and repair. SSB proteins are characterized by the presence of a conserved OB-fold motif (oligonucleotide/oligosaccharide/oligopeptide-binding fold), which is typically 100 amino acids in length (Murzin, 1993).

SSB proteins can be divided into two distinct groups based on their quaternary structure. Eukaryotic SSB proteins, known as replication protein A (RPA), are exemplified by human RPA, which has a heterotrimeric structure comprising three subunits: RPA70, RPA32 and RPA14 (of molecular weights 70, 32 and 14 kDa, respectively; Wold, 1997). The RPA complex contains six OB folds, four of which bind DNA: three on RPA70 and one on RPA32 (Bastin-Shanower & Brill, 2001). An N-terminal domain on RPA70 has also been shown to be involved in protein–protein interactions (Jacobs et al., 1999). In contrast, bacterial SSB proteins form homotetramers, with each subunit containing one DNA-binding domain (Raghunathan et al., 1997). These DNA-binding domains are located at the N-termini of the individual SSB protein subunits and form the characteristic OB folds. While the N-terminus of each subunit binds ssDNA and contains the homotetramer interface, it is thought that the C-terminal domain is involved in interactions with other protein components of DNA metabolism. The C-terminal domain of bacterial SSB proteins exhibits low sequence homology across species, with the exception of the terminal six residues, which form a highly conserved negatively charged DDDDPF motif. This motif is essential for the function of Escherichia coli SSB protein in vivo (Curth et al., 1996) and has been shown to interact directly with the 3'-5' ssDNA-degrading exonuclease I (Genschel et al., 2000). The tails of both the E. coli and the Sulfolobus solfataricus SSB proteins are not involved in DNA binding, but are thought to play roles in mediating protein–protein interactions with other subunits within the DNA polymerase complex (Bruck et al., 2002). There is also evidence that a mutually exclusive interaction between the C-terminal domain of E. coli SSB protein, DNA polymerase and primase is utilized as a three-point switch to initiate the exchange of places of these two proteins on DNA (Yuzhakov et al., 1999). Furthermore, a recent report suggests that the interaction between the DNA polymerase and SSB from RB69 (a T4-like bacteriophage) results in an increase in the overall affinity of the SSB protein for ssDNA (Sun & Shamoo, 2003). Finally, Gulbis and coworkers have recently
crystallization papers

proposed a positively charged patch on the χ subunit of Pol III holoenzyme which may interact with the C-terminal acidic region of SSB (Gulbis et al., 2004).

The DNA-binding domains and OB folds from SSB proteins have been well studied and structural information is available from a variety of organisms spanning all three kingdoms of life (Bochkarev et al., 2001; Bochkarev et al., 1997, 1999; Webster et al., 1997; Raghunathan et al., 2000; Yang et al., 1997; Kerr et al., 2003). However, crystallization of a full-length bacterial protein has proved problematic and most studies have used proteolytic N-terminal fragments of SSB proteins; consequently, little is known about the structure of its C-terminal domain.

Efforts to crystallize the intact E. coli SSB tetramer resulted in autolysis during crystallization and the structure determination omitted 30 amino acids from the C-terminus (Matsumoto et al., 2000). It has been postulated that the C-terminal domain of the E. coli SSB is cleaved to decrease unfavourable interactions for crystallization which result from its high glutamine content. The most recent crystallographic study by Kerr and coworkers presents a 1.2 Å structure of a trypsin-cleaved fragment of the SSB from the crenarcheote S. solfataricus, missing some 28 amino-acid residues from the C-terminal tail (Fig. 1).

Extensive investigations of the binding mode for ssDNA to SSB have revealed a complex series of protein–protein and protein–DNA interactions (Lohman & Ferrari, 1994; Raghunathan et al., 2000). Different binding modes [referred to as (SSB)$_{35}$ and (SSB)$_{65}$] and cooperativities have been observed that are dependent upon oligonucleotide length, salt and protein concentration. For example, E. coli (SSB)$_{35}$ binds about 35 nucleotides and in this case only two of the four subunits in the tetramer bind to the DNA. In contrast, in the (SSB)$_{65}$ binding mode all four subunits of the tetramer are involved in DNA binding, although it appears that there is a ‘limited’ type of intertetramer cooperativity. Thus, using dT(pT)$_{59}$ various combinations with ratios of one SSB subunit binding to one or two DNA oligomers or two SSB subunits to one DNA are possible depending on the conditions. However, the use of a high (>0.2 M NaCl) salt concentration appears to favour formation of one SSB tetramer binding to one dT(pT)$_{59}$.

To investigate the structure of a full-length SSB, we report here the cloning, overexpression, crystallization and initial data collection for crystals of the SSB protein (147 amino acids; $M_r$ 17 131.20; Fig. 1) from the hyperthermophilic bacterium A. aeolicus (SSB Aae) in both the free and the DNA-bound forms. In contrast to E. coli SSB, primary structure analysis of the SSB Aae reveals a polyglutamic acid region at its C-terminus and a EDEIPF motif (Fig. 1). We hope that the crystal structure of the A. aeolicus SSB protein will facilitate the study of the complex protein–protein interactions mediated through the C-terminus of bacterial SSB proteins and the data may also reveal the structural basis for the increased stability of this SSB at elevated temperatures. Further, the DNA-bound structure may reveal details of the (SSB)$_{65}$ binding mode.

2. Cloning, expression and purification

The ssg gene was identified from the complete A. aeolicus genome sequence (Deckert et al., 1998), amplified by polymerase chain reaction and the resulting 451 bp fragment was subsequently inserted into the pET-23a expression vector (Novagen) using NdeI/HindIII restriction sites. The fidelity of the construct, pET23a-ssg, was verified by DNA sequencing before transformation of E. coli BL21(DE3)/pLysS (Novagen). Cells were grown in 2YT growth media supplemented with ampicillin (100 μg ml$^{-1}$) in shake flasks at 310 K and 250 rev min$^{-1}$ and OD$_{600}$ = 0.8 prior to induction with 1.0 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). After a further 4 h of growth, cells were collected by centrifugation and stored at 253 K.

The cell pellet was resuspended in 10 ml buffer A (50 mM Tris–HCl pH 7.0) per gram of cell paste, disrupted by sonication and centrifugation and stored at 253 K.

Table 1. \[\begin{array}{|c|c|}
\hline
SSB Aae & SSB Aae (126) EIEKLEGE - - - EKFPETDEEDIEIPF (147) \\
SSB Eco & SSB Eco (150) QAGSPQQPQASPAPA8NEPMDFDCDPFP (147) \\
SSB Eco & SSB Eco (128) GRYRGGRRG - - - RQQENEEEEE (148) \\
\hline
\end{array}\]

Figure 1. Sequence alignment of the SSB proteins from A. aeolicus (SSB Aae), E. coli (SSB Eco) and S. solfataricus (SSB Sso). The asterisks (\*) under residue L112 of both SSB Aae and SSB Eco denotes the limit of the resolution of chymotryptic fragment (SSB; residues 1-135, represented by a $ sign under W135) used to determine the structure of the native and DNA-bound SSB complex described by Raghunathan et al. (1997, 2000) (PDB codes Ikw and leyg, respectively). The hash symbol (#) under residue R119 indicates the C-terminus of the tryptic fragment of S. solfataricus SSB (SSB Sso) crystallized by Kerr et al. (2003) (residues 1-119; PDB code I07). The addition sign (+) under N145 of SSB Eco denotes the limit of structure determination from the autolytic fragment crystallized by Matsumoto et al. (2000) (residues 1-145; PDB code Iqve). Notice the glutamate-rich C-terminus of SSB Aae in comparison to the glutamine-rich tail of SSB Eco.

theoretical weight of full-length *A. aeolicus* SSB protein (17 131.2 Da).

### 3. Crystallization

Initial crystals were obtained using Molecular Dimensions Structure Screens 1 and 2 and the sitting-drop vapour-diffusion method at 290 K. The drop consisted of 5 µl protein solution (7 mg ml⁻¹ in buffer C) and 5 µl precipitant. Over two weeks, small crystals of native protein were observed under three different conditions, with the best quality obtained using 100 mM HEPES pH 7.5, 2% (v/v) PEG 400, 2.0 M (NH₄)₂SO₄ pH 7.5 as the precipitant. After refining the crystallization conditions, larger crystals were obtained after four weeks using the same precipitant at a pH of 7.0. Co-crystallization of the DNA-bound protein was achieved by mixing 7.5 mg ml⁻¹ protein in 50 mM Tris pH 7.0, 0.1 M NaCl and 1.5 hI precipitant. Over two weeks, small crystals of similar dimensions were obtained after four weeks using the same precipitant at a pH of 7.0. Analysis of the diffraction data under three different conditions, with the best quality obtained using 100 mM HEPES pH 7.5 as the precipitant. After refining the crystallization conditions, larger crystals were then used to aid location of the second tetramer.

### 4. Data collection and processing

Crystals of native protein of approximate dimensions 0.1 x 0.2 x 0.2 mm were flash-cooled in a 20% glycerol well solution and X-ray data for the native SSB were collected at 100 K (Cryostream cooler; Oxford Cryosystems, Oxford, England) on a MAR Research 345 imaging plate mounted on an Enraf–Nonius FR591 rotating-anode generator, λ = 1.5418 Å, fitted with Osmic mirrors and operating at 40 kV, 110 mA. Crystals of similar dimensions were obtained for the DNA-bound form and data were also collected at 100 K on station 14.2 (λ = 0.978 Å) at the SRS, CLRC Daresbury Laboratory. Analysis of the diffraction data for both crystals using MOSFLM/SCALEA (Leslie, 1992; Collaborative Computational Project Number 4, 1994) produced the data shown in Table 1 and allowed the assignment of the native crystals to space group *P₂₁.*

For the DNA-bound data, similar processing statistics were obtained with space groups *I₂22, I₂22, I₂2₂₂,* and *I₄₁.*

For the native SSB data set, a model of the SSB from *Escherichia coli* (PDB code 1qv; Matsumoto et al., 2000) was used to search for an initial solution using MOLREP (Vagin & Teplyakov, 1997). The search molecule was trimmed of its flexible loops and amino-acid side chains to produce a tetrameric polyaniline structure. The top-rotation-function solution produced a satisfactory translation-function solution that was then used to aid location of the second tetramer. No solution was obtained using the *S. solfataricus* structure as a search model. For the DNA-bound SSB, a multi-copry search with MOLREP using the partially refined SSB *Sae* dimer (*R* = 0.272, *R*_ave = 0.308) provided solutions, the best of which contained two dimers per asymmetric unit in each of the three space groups. 20 cycles of rigid-body refinement were followed by ten cycles of restrained refinement. The statistics for this process are also shown in Table 1. Refinement of both crystal forms is currently in progress while attempts are being made to improve the diffraction quality of the crystals.

To ensure that no autolysis of the protein had occurred, a single crystal of DNA-bound SSB was dissolved in 10 mM HEPES pH 8.1 and analysed by SDS-PAGE and ESI-MS, which revealed no significant degradation of the protein. SDS-PAGE analysis produced a single protein band running at an anomalous weight of ~23 kDa.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Data-collection and reduction statistics.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>Native</td>
<td>DNA-bound</td>
</tr>
<tr>
<td>Data collection</td>
<td>Home source</td>
<td>SRS Daresbury</td>
</tr>
<tr>
<td>Oscillation range (°)</td>
<td>180 x 1</td>
<td>120 x 1</td>
</tr>
<tr>
<td>Space group</td>
<td><em>P₂₁</em></td>
<td><em>I₂₂₂</em></td>
</tr>
<tr>
<td>Unit-cell parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a (Å)</td>
<td>80.97</td>
<td>80.36</td>
</tr>
<tr>
<td>b (Å)</td>
<td>73.40</td>
<td>73.40</td>
</tr>
<tr>
<td>c (Å)</td>
<td>109.76</td>
<td>113.20</td>
</tr>
<tr>
<td>β (°)</td>
<td>95.11</td>
<td>90.90</td>
</tr>
<tr>
<td>Tetramers per AU</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Vₐ (Å³ Da⁻³)</td>
<td>2.37</td>
<td>1.89</td>
</tr>
<tr>
<td>Resolution limits (Å)</td>
<td>70-2.92 (1.40-2.92)</td>
<td>78.28-2.80 (2.95-2.80)</td>
</tr>
<tr>
<td>No. observations</td>
<td>10085 (12811)</td>
<td>119053 (12240)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.9 (98.9)</td>
<td>99.9 (99.9)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.057 (0.212)</td>
<td>0.060 (0.255)</td>
</tr>
<tr>
<td>MOLREP R factor</td>
<td>0.567</td>
<td>0.560</td>
</tr>
<tr>
<td>MOLREP correlation coefficient</td>
<td>0.359</td>
<td>0.316</td>
</tr>
<tr>
<td>Restrained refinement R factor</td>
<td>0.425</td>
<td>0.401</td>
</tr>
<tr>
<td>Restrained refinement Rmerge</td>
<td>0.487</td>
<td>0.507</td>
</tr>
</tbody>
</table>

*For one SSB tetramer (Mr = 68 400) and one DNA 69-mer (Mr = 20 927). † Rmerge = √(Σ(Ih)-|Ih|)/ΣΣ Ih, where (Ih) is the mean intensity of the h symmetry-equivalent reflections.*

**Figure 2** Analysis of the crystallized DNA-bound SSB protein by ESI-MS. The main figure shows the deconvoluted mass of 17 127.3 (obtained using Transform software; Micromass UK), consistent with the theoretical value of 17 131.2; right inset, ion envelope of crystallized SSB protein; left inset, SDS-PAGE analysis of crystallized SSB protein results in a single band running at an anomalous weight of ~23 kDa.
single band around 23 kDa in keeping with the observed anomalous mobility of the native SSB. Only one major species was observed by ESI–MS with a mass of 17 127.3 ± 2.7 (Fig. 2), in good agreement with the predicted weight. No LCMS data could be obtained from the dissolved native SSB crystal.

5. Discussion
It is clear from the sequence alignment of the SSB proteins that the *A. aeolicus* and the *E. coli* proteins are more closely related to each other than either is to *S. solfataricus* SSB (Fig. 1). This is borne out by the fact that a molecular-replacement solution using the *E. coli* structure was obtained relatively easily, whilst no satisfactory solution could be obtained with the *S. solfataricus* structure. The very high resolution of *S. solfataricus* SSB reveals why this should be so in that the actual molecular structure is much more closely related to the eukaryotic SSB fold than that of *E. coli* SSB (Kerr et al., 2003). Consequently, despite a modest sequence identity, the structures are distinct.

For the DNA-bound crystals reported here, there is an ambiguity as regards the space group. Given an SSB tetramer in the asymmetric unit, using the monomer *M* of 17 100 and that of the DNA as 20 927, the current state of the refinement, such a screw dyad axis in 1222. It is impossible for sequence identity, the structures are distinct. The expected VM  can be calculated to be 2.43, 2.11, 1.86 or 1.65 ÅDa⁻¹ for zero, 0.5, one or two bound DNA 69-mers per tetramer. The expected 1:1 complex requiring one DNA oligomer per asymmetric unit corresponds to a VM of 1.86 ÅDa⁻¹, which is within the range found by Matthews (1968), albeit quite close to the lower limit. The physiological tetramer as observed in the native structure sits on a crystallographic dyad in both *I*₄₁, and *D*₂2₂, whereas the tetramer sits on a screw dyad axis in *I*₂₂. It is impossible for there to be exact twofold symmetry for the SSB tetramer with a single DNA oligomer bound, although a pseudo-twofold arrangement is possible. Given the limited resolution of the present X-ray data and the current state of the refinement, such a situation cannot yet be ruled out. However, if the DNA oligomer is shared between two tetramers in some fashion, this could permit the DNA-bound tetramer to lie upon a crystallographic dyad, while maintaining four subunits and a single DNA molecule in the asymmetric unit (Ferrari et al., 1994).

The initial electron-density maps in each of the three space groups all show extra electron density near regions of the protein expected to bind DNA (Raghunathan et al., 2000). Examination of the maps together with the statistics shown in Table 1 leads us to prefer *I*₂₂₂ as the space group, but we are continuing to refine all three possibilities. These refinements should clarify this uncertainty and also allow us to estimate the occupancy of the DNA.

In summary, our expression and purification strategy has produced full-length SSB from the hyperthermophile *A. aeolicus* with no autolysis observed by mass spectrometry and SDS–PAGE. The flexible C-terminal tail is present in the crystals reported here, unlike the truncated SSB used in both the *E. coli* and *S. solfataricus* structure determinations. Our initial refinement of the structures of both forms has allowed the clear assignment of the electron density to residues 1–38 and 41–108 and we are currently refining the models in an effort to distinguish the C-terminal residues.

Note added in proof: During the processing of this manuscript a report has been published describing the crystallization of *S. solfataricus* SSB (Savvides et al., 2004).

We wish to thank Professors Karl Stetter and Robert Huber (University of Regensburg) for the kind gift of *A. aeolicus* chromosomal DNA and Professor Jim Naismith and Dr Iain Kerr (University of St Andrews) for kindly providing the coordinates of the SSB fragment from *S. solfataricus* prior to publication. The Biotechnology and Biological Sciences Research Council UK and the University of Edinburgh supported this work (DfC, LAM).

References