CHEMICAL SYNTHESIS, RECOMBINANT EXPRESSION AND
STRUCTURAL CHARACTERISATION OF COMPLEMENT PROTEIN
MODULES

Joanne Marie O'Leary

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh
September 1999
Unless otherwise stated, the work described in this thesis is my own work and has not been submitted in whole or in part for a degree or other qualification at this or any other University.

University of Edinburgh
September 1999
ACKNOWLEDGEMENTS

I would like to thank my supervisor, Paul Barlow, for his help and supervision throughout this project. I also wish to acknowledge Professor Ramage for the opportunity to work in his laboratory.

The help, encouragement and friendship I have received from everyone in the Barlow group is greatly appreciated. In particular I would like to thank Nick Mullin for his help and advice in the lab; Dusan Uhrin, Brian Smith and Stanka Uhrinova for their NMR expertise; Marina Kirkitadze for her help with CD and calorimetry and John Parkinson and Robert Wakefield for their assistance with all things computer related. From Lab. 120 I would like to thank everyone for making it an enjoyable place to work and in particular Gideon Grogan and Gareth Roberts for their advice and support. I also wish to acknowledge Kevin Shaw for his technical assistance with peptide synthesis. Many thanks to everyone who helped with proof-reading.

Finally, I would like to thank Jonathan for his support, encouragement and patience, Jim for being a good friend and my family.
# Table of Contents

Table of Contents

Abbreviations

## Chapter One: Introduction

1.1 **The Complement System**

1.1.1 Activation of the classical pathway  
1.1.2 The lectin pathway  
1.1.3 Activation of the alternative pathway  
1.1.4 Terminal lytic pathway  
1.1.5 Effects of complement activation  
1.1.6 Control of the complement system  
1.1.7 The RCA proteins and CP-modules  
1.1.8 CP-module structure

1.2 **Membrane Cofactor Protein**

1.2.1 Structure  
1.2.2 Biosynthesis and tissue distribution  
1.2.3 Function as a complement inhibitor  
1.2.3.1 Role of MCP in reproduction  

15
16
20
20
21
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2.3.2 Role of MCP in xenotransplantation</td>
<td>21</td>
</tr>
<tr>
<td>1.2.4 MCP as a receptor for pathogens</td>
<td>22</td>
</tr>
<tr>
<td>1.2.4.1 MCP as the measles virus receptor</td>
<td>22</td>
</tr>
<tr>
<td>1.2.4.2 MCP is a receptor for pathogenic <em>Neisseria</em></td>
<td>24</td>
</tr>
<tr>
<td>1.2.4.3 Role of MCP in Streptococcal infection</td>
<td>24</td>
</tr>
<tr>
<td><strong>1.3</strong> Factor H</td>
<td>25</td>
</tr>
<tr>
<td>1.3.1 Structure</td>
<td>25</td>
</tr>
<tr>
<td>1.3.2 Function as a complement regulator</td>
<td>26</td>
</tr>
<tr>
<td>1.3.3 Non-complement ligands for factor H</td>
<td>26</td>
</tr>
<tr>
<td><strong>1.4</strong> Strategies for the production of protein for structural studies</td>
<td>27</td>
</tr>
<tr>
<td>1.4.1 Chemical synthesis of proteins</td>
<td>28</td>
</tr>
<tr>
<td>1.4.1.1 Synthetic strategies</td>
<td>28</td>
</tr>
<tr>
<td>(a) Solid support</td>
<td>33</td>
</tr>
<tr>
<td>(b) Protecting groups</td>
<td>34</td>
</tr>
<tr>
<td>(c) Carboxylate activation</td>
<td>35</td>
</tr>
<tr>
<td>1.4.1.2 Chemical ligation</td>
<td>38</td>
</tr>
<tr>
<td>1.4.2 Recombinant protein expression</td>
<td>40</td>
</tr>
<tr>
<td>1.4.2.1 Protein expression in <em>E. coli</em></td>
<td>40</td>
</tr>
<tr>
<td>1.4.2.2 Protein expression in yeast</td>
<td>41</td>
</tr>
<tr>
<td><strong>1.5</strong> Summary</td>
<td>42</td>
</tr>
</tbody>
</table>
CHAPTER TWO: MATERIALS AND METHODS

2.1 General methods

2.2 Chemical synthesis
2.2.1 Materials
2.2.2 Solid phase synthesis
2.2.3 Purification of MCP~1
2.2.4 Purification of fH~15

2.3 Recombinant expression of MCP~1
2.3.1 Materials
2.3.2 Prokaryotic expression of MCP~1
2.3.3 Eukaryotic expression of MCP~1
2.3.4 Eukaryotic expression of MCP~1(T85A)

2.4 Structural characterisation of MCP~1
2.4.1 Differential scanning calorimetry
2.4.2 Circular dichromism
2.4.3 NMR spectroscopy
2.4.3.1 Water suppression
2.4.3.2 Multi-dimensional NMR
2.4.3.3 Homonuclear NMR experiments
   (a) Correlated spectroscopy (COSY)
   (b) Total correlation spectroscopy (TOCSY)
### Contents

- Nuclear Overhauser effect spectroscopy 77
  - 2.4.3.4 Heteronuclear experiments 78
    - (a) Heteronuclear spin quantum correlation (HSQC) spectroscopy 78
    - (b) 3D-Heteronuclear NMR spectroscopy 78
  - 2.4.3.5 Band selective 2D experiments for resonance assignment of proline residues 79

### CHAPTER THREE: CHEMICAL SYNTHESIS OF CP-MODULES

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>81</td>
</tr>
<tr>
<td>3.2</td>
<td>Chemical synthesis of MCP-1</td>
<td>82</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Solid phase synthesis</td>
<td>82</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Purification of MCP-1</td>
<td>86</td>
</tr>
<tr>
<td>3.2.2.1</td>
<td>Size exclusion chromatography</td>
<td>88</td>
</tr>
<tr>
<td>3.2.2.2</td>
<td>Renaturation of MCP-1</td>
<td>90</td>
</tr>
<tr>
<td>3.2.2.3</td>
<td>Anion exchange chromatography</td>
<td>94</td>
</tr>
<tr>
<td>3.2.2.4</td>
<td>Further purification of MCP-1</td>
<td>96</td>
</tr>
<tr>
<td>3.3</td>
<td>Chemical Synthesis of fH-15</td>
<td>99</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Solid phase synthesis</td>
<td>99</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Purification of fH-15</td>
<td>101</td>
</tr>
<tr>
<td>3.3.3</td>
<td>N-terminal TbFmoc derivatisation of fH-15</td>
<td>105</td>
</tr>
</tbody>
</table>
CHAPTER FOUR: RECOMBINANT EXPRESSION OF MCP-1

4.1 Expression of MCP-1 in *E.coli* 128

4.1.1 Expression in *E. coli* JM109 129

4.1.2 Minimising formation of inclusion bodies 134

4.1.2.1 Expression under sub-optimal conditions 134

4.1.2.2 Employment of alternative strains of *E. coli* 135

4.1.3 Solubilisation and renaturation of inclusion bodies 139

4.2 Expression of MCP-1 in *Pichia pastoris* 144

4.2.1 Expression and purification of MCP-1 146

4.2.2 Expression of MCP-1(T85A) 152

4.3 Discussion 158
## CHAPTER FIVE: STRUCTURAL CHARACTERISATION OF MCP-1

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>161</td>
</tr>
<tr>
<td>5.2</td>
<td>Differential Scanning calorimetry studies</td>
<td>161</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Study of MCP-1 expressed in <em>E. coli</em></td>
<td>162</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Studies of MCP-1 expressed in <em>P. pastoris</em></td>
<td>163</td>
</tr>
<tr>
<td>5.3</td>
<td>Circular dichromism studies</td>
<td>165</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Study of MCP-1 expressed in <em>E. coli</em></td>
<td>166</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Studies of MCP-1 expressed in <em>P. pastoris</em></td>
<td>168</td>
</tr>
<tr>
<td>5.4</td>
<td>Analysis of MCP-1 by NMR spectroscopy</td>
<td>168</td>
</tr>
<tr>
<td>5.4.1</td>
<td>1D $^1$H-NMR studies</td>
<td>169</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Comparison of MCP-1 and MCP-1(T85A) by 2D $^1$H-$^1$N heteronuclear spectroscopy</td>
<td>169</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Analysis of MCP-1 by 2D- and 3D-NMR spectroscopy</td>
<td>173</td>
</tr>
<tr>
<td>5.5</td>
<td>Resonance Assignment of MCP-1</td>
<td>173</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Sequential assignment of MCP-1</td>
<td>174</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Assignment of prolines</td>
<td>181</td>
</tr>
<tr>
<td>5.5.3</td>
<td>Chemical shift index for MCP-1</td>
<td>186</td>
</tr>
<tr>
<td>5.5.4</td>
<td>Regular secondary structures from spin-spin couplings $^3J_{HN,\alpha}$</td>
<td>187</td>
</tr>
<tr>
<td>5.5.5</td>
<td>Hydrogen bonds from D$_2$O experiments</td>
<td>189</td>
</tr>
<tr>
<td>5.5.6</td>
<td>NMR structure determination of MCP-1</td>
<td>192</td>
</tr>
</tbody>
</table>
CHAPTER SIX: CONCLUSIONS

6.1 Chemical Synthesis and Recombinant Expression of CP-modules

6.2 Structural Characterisation of MCP-1

6.3 Summary

References

Lectures and Courses Attended
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBP</td>
<td>Mannose binding protein</td>
</tr>
<tr>
<td>MASP</td>
<td>Mannose binding protein-associated serine protease</td>
</tr>
<tr>
<td>fD</td>
<td>Factor D</td>
</tr>
<tr>
<td>fB</td>
<td>Factor B</td>
</tr>
<tr>
<td>MAC</td>
<td>Membrane attack complex</td>
</tr>
<tr>
<td>RCA</td>
<td>Regulators of complement activation</td>
</tr>
<tr>
<td>CR1</td>
<td>Complement receptor type 1</td>
</tr>
<tr>
<td>CR2</td>
<td>Complement receptor type 2</td>
</tr>
<tr>
<td>MCP</td>
<td>Membrane cofactor protein</td>
</tr>
<tr>
<td>DAF</td>
<td>Decay accelerating factor</td>
</tr>
<tr>
<td>fH</td>
<td>Factor H</td>
</tr>
<tr>
<td>CP-module</td>
<td>Complement protein module</td>
</tr>
<tr>
<td>VCP</td>
<td>Vaccinia virus complement control protein</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>COSY</td>
<td>Correlated spectroscopy</td>
</tr>
<tr>
<td>TOCSY</td>
<td>Total correlation spectroscopy</td>
</tr>
<tr>
<td>NOESY</td>
<td>Nuclear Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser effect</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear single quantum correlation spectroscopy</td>
</tr>
<tr>
<td>DPFGSE</td>
<td>Double pulsed field gradient spin echo</td>
</tr>
<tr>
<td>ESMS</td>
<td>Electrospray mass spectrometry</td>
</tr>
<tr>
<td>MALDI-TOF MS</td>
<td>Matrix assisted laser desorption time-of flight mass spectrometry</td>
</tr>
<tr>
<td>DSC</td>
<td>Differential scanning calorimetry</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichromism</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>Reverse-phase high performance liquid chromatography</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>AIEX</td>
<td>Anion exchange chromatography</td>
</tr>
<tr>
<td>HIC</td>
<td>Hydrophobic interaction chromatography</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecylsulphate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SPPS</td>
<td>Solid phase peptide synthesis</td>
</tr>
<tr>
<td>Boc</td>
<td>t-Butoxycarbonyl</td>
</tr>
<tr>
<td>Fmoc</td>
<td>9-Fluorenlymethylcarbonyl</td>
</tr>
<tr>
<td>TbFmocCl</td>
<td>17-tetrazeno(a,c,g,i)fluorenlymethylchloroformate</td>
</tr>
<tr>
<td>DCM</td>
<td>Dichloromethane</td>
</tr>
<tr>
<td>DMF</td>
<td>Dimethylformamide</td>
</tr>
<tr>
<td>DCC</td>
<td>Dicyclohexylcarbodiimide</td>
</tr>
<tr>
<td>DIC</td>
<td>Diisopropylcarbodiimide</td>
</tr>
<tr>
<td>HOBt</td>
<td>1-Hydroxybenzotriazole</td>
</tr>
<tr>
<td>HOCT</td>
<td>Ethyl 1-hydroxy-1H-1,2,3-triazole-4-carboxylate</td>
</tr>
<tr>
<td>DIEA</td>
<td>N,N-diisopropylethylamine</td>
</tr>
<tr>
<td>TFA</td>
<td>Trifluoroacetic acid</td>
</tr>
<tr>
<td>βME</td>
<td>β-mercaptoethanol</td>
</tr>
<tr>
<td>GSH</td>
<td>Reduced glutathione</td>
</tr>
<tr>
<td>GSSG</td>
<td>Oxidised glutathione</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-transferase</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>BMG</td>
<td>Buffered minimal glycerol</td>
</tr>
<tr>
<td>BMM</td>
<td>Buffered minimal methanol</td>
</tr>
<tr>
<td>EndoHf</td>
<td>Endoglycosidase Hf</td>
</tr>
<tr>
<td>Name</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Alanine</td>
<td>Ala</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>Asp</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>Glu</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
</tr>
</tbody>
</table>
To Jonathan
INTRODUCTION

1.1 The Complement System

The immune system serves to protect the host against infection. The innate immune system provides the first line of defence against invading pathogens, which are destroyed either by phagocytosis or with bactericidal substances such as lysozyme. The adaptive immune system provides a more specific response which includes the production of antibodies to particular antigens. Whilst the innate immune system provides a rapid response to infection, the adaptive immune response requires longer to take effect but offers increased protection against re-infection by the same pathogen.

The complement system serves as a major defence and clearance system in the bloodstream, by identifying and removing foreign substances and immune complexes. It consists of over thirty soluble and membrane-bound proteins. Complement activation occurs via two major routes, called the classical and alternative pathways. The classical pathway is activated by immunoglobins after the identification of a foreign body by an antibody. Antibody-dependent activation of complement also occurs via the lectin pathway. Activation of the alternative pathway is antibody independent and occurs if the pathogen provides a suitable site for activation of early-acting complement components. An enzymatic cascade mechanism is initiated upon activation of the complement system resulting in opsonisation, formation of the membrane attack complex, which can cause cell
death, and the release of anaphylatoxins, which are involved in the inflammatory response.

The cascade-like nature of the complement system results in the amplification of a small initiating stimulus to cause a large response. This could potentially damage host tissue both at the site of activation and, since activated molecules may be transported around the body in the bloodstream, elsewhere. To prevent damage to host tissue a number of modulatory proteins exist that rigorously control activation. A schematic of the complement system is illustrated in Figure 1.1.

1.1.1 Activation of the classical pathway

The classical pathway is primarily activated by the interaction of C1 with antibody/antigen complexes although it can also interact with other molecules including polyanions, certain viral proteins and nucleic acids. C1 is a calcium ion-dependent serum protein complex consisting of one molecule of Clq, two ofClr and two of CIs. Although Clq is involved in antibody recognition it does not directly activate complement as it does not exhibit enzymatic activity. It has an unusual conformation being composed of six globular heads connected by a collagen triple helix to a central core. Each of the heads contain three non-identical domains that can bind immunoglobulins.Clr and CIs are serine protease proenzymes that bind to the collageneous portion of Clq. Multiple interactions between Clq and the activator are required for complement activation as the binding of Clq to the \( F_c \) domains of monomeric IgG is weak. Clq therefore binds to clustered IgG in, for example, immune complexes which provide multiple binding sites. This low affinity for \( F_c \) domains prevents Clq binding to isolated antibodies in solution. A conformational change in Clq occurs upon the binding of two or more of the Clq heads to the activator molecule. This
results in the autoactivation of C1r which cleaves and then activates C1s. Activated C1s cleaves C4, a plasma protein, to form two fragments; C4b and a smaller anaphylatoxic fragment C4a. Although C4b does not possess an enzymatic site its presence facilitates the cleavage of C2 by C1s to yield the classical pathway C3 convertase C4b2a. This convertase cleaves C3, a homologue of C4, to form C3b in an analogous manner to the cleavage of C4.

C3b and C4b each contain a thioester which reacts with any available nucleophiles such as sugar hydroxyls or amino groups close to the point of activation. Thus C3b and C4b can covalently bind to both foreign and host cells via ester or amide bonds. Thioesters are readily hydrolysed and inactivated by water limiting the range of action of C3b/C4b. They are therefore restricted to binding to the cell surface at which they are activated. This process, known as opsonisation, coats foreign cells with C3b/C4b to facilitate their uptake by phagocytes.

Covalent binding of a C3b molecule to C4b2a forms the classical pathway C5 convertase, C4b2a3b. In an analogous manner to C3 and C4, C5 is cleaved to yield C5a and C5b. The C5b fragment initiates the formation of the membrane attack complex (MAC) which ultimately results in cell death.

1.1.2 The lectin pathway

The lectin pathway activates via the classical pathway but bypasses the C1 complex. Mannose binding protein (MBP) is a member of the collectins, a group of carbohydrate binding proteins with a common carbohydrate recognition domain.
Figure 1.1 The complement system. The major reactions in the complement activation pathways, resulting in the formation of the membrane attack complex (MAC), are illustrated. Complement regulatory proteins are shown in shaded circles. Abbreviations as in the text.
The collectins are structurally and functionally related to C1q. However, MBP is activated by binding to carbohydrate ligands on bacteria or viral cell surfaces or to a glycosylated variant of IgG rather than to the Fc domain of antibodies as in the case of C1q. As in the classical pathway the cleavage of C4 and C2 is mediated by the presence of a serine protease, MBP-associated serine protease (MASP). MASP is structurally related to Cls and forms a complex with MBP to initiate the formation of the C3 convertase.

1.1.3 Activation of the alternative pathway

In contrast to the classical pathway, there is no initiating factor for activation of the alternative pathway. It operates continuously at a low level. In the absence of an initiating factor, the formation of the C3 convertase enzyme is possible because C3 is continuously activated in the fluid phase either by serum proteases or by nucleophilic attack on its thioester bond. Additionally, perturbation of the structure of C3 by the activating molecule may expose the thioester making it susceptible to hydrolysis.

The low level of spontaneous hydrolysis of the C3 thioester bond produces C3i which can associate with Factor B (fB), a serine protease, to form C3iB. Factor D (fD) then cleaves the bound fB to yield the alternative pathway C3 convertase C3iBb. C3iBb then activates C3 to form C3b, which subsequently results in the formation of the C3 convertase C3bBb in an analogous manner to the formation of C3iBb. In this way a positive feedback loop can be generated, see Figure 1.2. The alternative pathway C5 convertase, C3bBb3b, is generated by the binding of C3b to the C3 convertase. Activation of the alternative pathway also occurs if the classical pathway convertase C4b2a is the originator of the C3b on the
target cell. Thus the alternative pathway can serve to amplify the effects of classical path activation.

Figure 1.2 C3 amplification cycle. The spontaneous hydrolysis of the thioester bond of C3 results in the formation of the C3 convertase C3iBb which subsequently converts C3 to C3b to produce more C3 convertase (C3bBb). Adapted from Liszewski et al. 6

1.1.4 Terminal lytic pathway

The terminal lytic pathway involves complement components C5-C9 which generate the membrane attack complex (MAC). Its first step is the formation of the C5-convertases, C4b2a3b and C3bBb3b, in the classical and alternative pathways respectively. The only proteolytic event in this pathway is the cleavage of C5 to form C5a and C5b. C5b binds to C6 and the MAC is formed by the binding of C7, C8 and between one and eighteen molecules of C9. The C-terminus of C9 is relatively hydrophobic and intercalates into the membrane of the complement activator forming large pores in the membrane. The number
of molecules of C9 bound determines the type of lesion. The resulting change in osmotic pressure results in death of the pathogen.

1.1.5 Effects of complement activation

Activation of the complement system results in the formation of many biologically important mediators. The principal role of complement is the identification and clearance of foreign material. The opsonisation of pathogens by C3b results in attack by phagocytic cells and the terminal lytic components of the activation cascade.

The release of the anaphylatoxins C3a, C4a and C5a upon activation of the complement system forms part of the inflammatory response. Inflammation is a localised response which results in an influx of white cells, complement and antibody to a site of injury or infection. The anaphylatoxins cause smooth muscle contraction, vasodilation and increased vascular permeability resulting in swelling and redness due to an increase in local blood supply. They also induce the release of histamine by mast cells which serves to amplify the inflammatory effects.

Complement is also important for the clearance of immune complexes which can form when an antigen is exposed to an antibody. The binding of C3b to these complexes enables them to bind to erythrocytes which have a C3b receptor on their surface. The immune complexes are then transported to the liver for clearance.
The multiple effects of complement, as described above, mean that any disruption to the correct functioning of the cascade has important consequences. Over-activation of complement can cause damage to host tissue in autoimmune disorders such as rheumatoid arthritis. Diminished complement activity increases susceptibility to infection and can result in disorders such as systemic lupus erythematosus due to ineffective clearance of immune complexes. The correct functioning of complement, to provide a mechanism for host defence against infection without causing host damage, is realised by a series of control proteins that operate throughout the complement cascade.

1.1.6 Control of the complement system

Since complement activation proceeds by a cascade mechanism, a small initial stimulus can rapidly result in the formation of large quantities of active components. To prevent host-damage the system is controlled stringently by numerous regulatory proteins and enzymes. These regulatory proteins are, in general, not present on foreign micro-organisms, thus providing a mechanism for distinguishing self from non-self and preventing complement attack on autologous tissue \(^8\). Furthermore, the C3 and C5 convertases decay by dissociation of C2a and factor Bb in the absence of the control proteins to limit the effects of complement activation.

The proteins which regulate the complement system do so at every stage of the cascade. For example, Cl inhibitor, a serine protease inhibitor, prevents over-activation of the classical pathway by binding to activated Clr and Cls \(^9\). These then dissociate from the Clq activator complex leaving Clq bound to the antibody. Formation of the membrane attack
complex on cells is inhibited by CD59 and homologous restriction factor (HRF). CD59 probably blocks the binding of multiple copies of C9, whereas HRF binds C8 and prevents further assembly of the MAC.

Although the complement system is tightly controlled at each stage of the cascade, the majority of the regulatory proteins act at the critical C3 activation stage. C3b is a key component of the complement system as it is the point at which the three activation pathways meet. The formation of C3b is essential for the opsonisation and subsequent lysis of microbial targets. Control of C3b also restricts the release of C5a, the most potent anaphylatoxin, by limiting the formation of the C5 convertase C4b2a3b. A family of control proteins, the regulators of complement activation (RCA proteins), serve to prevent over-activation at this stage and to protect the host from complement mediated damage.

1.1.7 The RCA proteins and CP-modules

The RCA proteins are central to the regulation of the complement system as they bind the activated components C3b/C4b and therefore control C3 and C5 convertase formation. The regulatory proteins serve either as cofactors for the factor I-mediated cleavage of C3b/C4b or to accelerate the decay of the C3 and C5 convertases. The RCA proteins: complement receptor 1 (CR1); membrane cofactor protein (MCP); decay accelerating factor (DAF); C4 binding protein (C4bp); factor H (fH) and complement receptor type 2 (CR2), are related functionally, genetically and structurally. The genes for these regulatory proteins are closely related and, in humans, are located on the same chromosome. This locus is
Introduction

known as the RCA gene cluster \(^{13}\). In combination with their strong sequence similarities this suggests they may have evolved from a common ancestor \(^{14}\).

The RCA proteins share a common structural motif known as the complement control protein module (CP-module), which is composed of approximately sixty amino acids with a number of conserved residues at specific positions \(^{15}\). This consensus sequence includes one tryptophan, two prolines, four cysteines and several hydrophobic residues and glycines, see \(\text{Figure 1.3}\). The cysteines form disulphide bonds in a 1-3, 2-4 pattern \(^{16}\).

\[
\begin{array}{cccccccccccc}
C & P & I & N & G & F & G & I & F & C & G & F & I & G & C & W & P & C \\
V & H & Y & V & Y & Y & V & L & L \\
\end{array}
\]

\(\text{Figure 1.3} \quad \text{The CP-module consensus sequence.} \quad \text{The residues are spaced to approximately reflect their positions in the consensus sequence.}\)

CP-modules frequently occur in contiguous arrays although isolated modules have been identified in, for example, \(C. \text{elegans} \) \(^{17}\). The complement control proteins contain varying numbers of CP-modules, see \(\text{Figure 1.4}\). For example, \(fH\) is a single chain of twenty CP-modules whereas MCP and DAF, the most structurally similar RCA proteins, contain only four. The binding sites for C3b/C4b are contained within the CP-modules of the RCA proteins with at least two modules being required for binding \(^{18-29}\).

The RCA proteins have overlapping roles in regulating complement on cell surfaces and in the fluid phase, see \(\text{Table 1.1}\). Their combined activity provides control in the classical and alternative pathways. The membrane-bound regulators, complement receptor type 1 (CR1),
membrane cofactor protein (MCP) and decay accelerating factor (DAF), function in both the classical and alternative pathways. CR1 is a cofactor for the factor I mediated cleavage of C3b and C4b and has decay accelerating activity for the dissociation of the C3 and C5 convertases. The comprehensive regulatory activity of CR1 has made it the first candidate for complement regulatory therapeutics for inflammatory conditions. DAF and MCP, as their names imply, possess only decay accelerating and cofactor activity respectively. The fluid phase regulators, C4bp and factor H, possess both decay accelerating and cofactor activity. However, they each function in only one of the two activation pathways.

Figure 1.4 Schematic representation of the six RCA proteins
<table>
<thead>
<tr>
<th>Decay accelerating activity</th>
<th>Cofactor activity</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical pathway</td>
<td>Alternative pathway</td>
<td>Classical pathway</td>
</tr>
<tr>
<td>CR1</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MCP</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DAF</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>C4bp</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Factor H</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 1.1 *Inhibitory activities of the RCA proteins.* Adapted from Weisman *et al.*

CP-modules have been identified in over 40 different mammalian proteins in addition to the RCA proteins. These include other complement proteins, the interleukin 2 receptor, factor XIIIb, and GABA<sub>a</sub> receptor 1a. CP-modules bind a diverse range of ligands including the activated complement components C3b/C4b; viruses such as measles virus, the M protein of *Streptococcus pyogenes*, Epstein-Barr virus, echovirus and other ligands including heparin and sialic acid. The occurrence of the CP-module in numerous proteins and its ability to bind a variety of ligands suggests that the modules provide a structural framework for a number of specific protein:protein interactions.

### 1.1.8 CP-module structure

CP-modules are autonomously folding domains separated by short linker sequences. There are no high-resolution structures of complete CP-module-containing proteins, however their modular nature permits a ‘dissect and rebuild’ strategy to obtain structural information. The strategy enables the study of modules from large, often membrane bound and...
glycosylated, proteins that would otherwise be unsuitable for structural elucidation by NMR.

High-field NMR spectroscopy has been utilised to determine the three-dimensional solution structure of two individual CP-modules and a pair of CP-modules from factor H. The structure of the C-terminal pair of CP-modules from vaccinia virus complement control protein (VCP) has also been solved. Similarly, the structure of the N-terminal pair of modules from MCP (MCP-1,2) has been solved by X-ray crystallography. Crystals have also been obtained of the C-terminal pair of modules of DAF (DAF-3,4) although the structure has not yet been determined. In addition to CP-modules, the ‘dissect and rebuild’ method has been employed to study other module-containing proteins including, intracellular signalling proteins, nucleic acid binding proteins, cell surface receptors, and extracellular matrix proteins such as epidermal growth factor-like, fibronectin-type I, -type II and -type III, immunoglobulin-type, C-type lectin and SH2 domains.

The secondary and tertiary structure of all CP-modules studied is largely conserved, indicating that the consensus sequence confers a conserved structural motif on the CP-module. CP-modules essentially consist of a hydrophobic core containing the side-chains of the highly conserved residues wrapped in β-strands. The β-strands form two antiparallel β-sheets, one containing three strands and the other containing two, which run approximately parallel to the long axis of the module. Two disulphide bridges are formed across the module at opposite ends of the long axis. The N- and C- termini are also located at opposite ends of the module. The loop regions show least sequence conservation and the variation within these regions may account for the functional diversity of CP-modules as
different binding sites can be introduced without affecting the overall structure of the module. The solution structures of a fragment of factor H composed of the fifteenth and sixteenth CP-modules and of the C-terminal pair of CP-modules from VCP are illustrated in *Figure 1.5.*

![Figure 1.5](image)

*Figure 1.5 NMR solution structures of fH–15,16 and VCP–3,4. The tryptophan residues are shown as a CPK surface.*
The orientation of the modules with respect to one another may be critical for function as typical binding sites require between two and four CP-modules. Although it is possible to model the structure of individual modules, the lack of sequence conservation in the linker region does not permit reliable modelling of the relative orientation of modules. The solution structures of module pairs currently available do not show a consensus orientation for the modules with respect to each other. In order to fully understand the nature of binding to CP-modules it is necessary to perform further experimental determinations of structure.

1.2 Membrane Cofactor Protein

Membrane cofactor protein (MCP) is a widely distributed cell surface glycoprotein found on all cells examined except erythrocytes. As a regulator of complement activation it is related genetically, functionally and structurally to the other RCA proteins. MCP regulates the complement system by binding the activated complement components C3b and C4b and serving as a cofactor for their cleavage by factor I. The role of MCP as a complement regulator has implications in reproductive biology and for preventing rejection after xenotransplantation. In addition to this role, MCP has been identified as the measles virus receptor and is of interest to tumour immunologists due to its high level of expression on malignant cells.
1.2.1 Structure

The extracellular region of MCP is comprised of four CP-modules (see Figure 1.6), the protein domains common to all RCA proteins, a region rich in serines, prolines and threonines (STP domain) and a region of unknown function. Within the CP-modules there are three N-linked glycosylation sites (in modules 1, 2 and 4) whilst the STP region is heavily O-glycosylated. The carboxyl terminus of the protein is composed of a hydrophobic membrane-spanning domain with a basic cytoplasmic anchor and cytoplasmic tail. MCP is structurally heterogeneous exhibiting two broad bands at 59-68 and 51-58 kDa when analysed by SDS-PAGE. The two forms are a result of alternative splicing in the STP region, with the upper form containing more sialic acid and probably being more heavily O-glycosylated. Further alternative splicing in the cytoplasmic region results in four major isoforms of MCP. Two distinct STP regions and two distinct cytoplasmic tails are generated by the alternative splicing, while the CP-modules are common to all four forms, see Figure 1.7.

**MCP-1**

{\text{CEEPPTFEAMELIGKPKP YYEIGERVYKCKK GYFY IPPLATHTIC DRNHT WLPVSDDACY}}

**MCP-2**

{\text{RETC P YIRDPL MGQAVPANGT YEFGYQMFICHE GY YY LIGEEL YCEILKGVSVA IGKPKP PICE}}

**MCP-3**

{\text{KVLCPPPPKIK NGKHTFSEVEVFYLDATVYSDPAPGPDFFS LIGESTY CDNSV WSRAA PECK}}

**MCP-4**

{\text{VVKCRFP VVE NGKQIGFGKMFYYKATVMECCK G FY LDGSST IV C DSNST WDPKV PKCL}}

**Figure 1.6** Sequence alignment of the four CP-modules of MCP-1. The consensus residues are shown in bold. Adapted from Lublin et al. 80.
The functions of the different isoforms and the roles of the STP region and cytoplasmic tail are not fully understood. The four major isoforms are expressed in the same proportions by most cells \(^7\). It has been observed that there is preferential expression of some of the isoforms in the brain, kidney and salivary gland \(^8\) and isoforms with a larger STP-rich region have been shown to preferentially protect against classical pathway activation of complement \(^8^2\). Rarer isoforms of MCP with further variation in the STP region exist \(^7^9\).

![Schematic of the four common isoforms of MCP. Adapted from Liszewski et al.](image)

The heavy O-glycosylation of the STP-rich region may protect the protein from proteolysis, a role that has been suggested for the carbohydrates of other membrane-bound proteins \(^8^3,8^4\). Both the N- and O-glycans are important for the complement regulatory function of MCP. The N-glycans of CP-modules 2 and 4 (MCP-2 and MCP-4) are required for cytoprotection which is enhanced by the STP region \(^8^5\). The carbohydrate moieties are also important in the exploitation of MCP by the measles virus. Although the N-glycan of MCP-2 is necessary
for measles virus binding and infection \(^{86}\), removal of the O-glycans in the STP-region does not affect binding \(^{87}\).

The structure determination of MCP may assist in understanding the function of the protein as a regulator of the complement system and as a viral receptor. The X-ray crystal structure of the N-terminal CP-modules (MCP-1,2) \(^{49}\) provides the only high-resolution structural data currently available on the protein. The overall structure is in good agreement with those structures of CP-modules solved by NMR \(^{44-48}\). As described previously, the consensus sequence appears to confer a consensus structure on the modules. However, the orientation of the modules in MCP-1,2 is different to those previously observed for fH-15,16 and VCP-3,4, see Figure 1.8. Despite the conserved tertiary structures of the individual modules this variability in the inter-domain interface prevents accurate modelling of CP-module-containing proteins.

The high-resolution NMR structure determination of MCP-1,2 remains a valid target, despite the available crystal structure, due to the inter-modular flexibility of the protein in solution. In the crystal structure, the inter-domain interface of MCP-1,2 was found to be highly flexible. Six copies of the molecule were found in the unit cell showing approximately 15\(^\circ\) variation in the inter-domain angle between the copies. However, it is postulated that movement could be restricted both by the arrangement of the molecules within the crystal and by a calcium ion in the interdomain interface. The introduction of calcium was necessary for crystallisation suggesting that it does have a stabilising effect on the structure.
Figure 1.8 Superimposition of the structures of fH-15,16 (blue), VCP-3,4 (red) and MCP-1,2 (yellow). The structures were overlaid using the program Insight II on the disulphides of the C-terminal module.
1.2.2 **Biosynthesis and tissue distribution**

The broad tissue distribution of MCP is essential for its critical role in protecting against host cell-damage. As an intrinsic regulator of complement, MCP can only protect the cell on which it is expressed \(^{88}\), which is in contrast to the extrinsic regulators of complement such as factor H and CR1 which have limited tissue distribution but are able to regulate globally. Small amounts of soluble MCP are found in plasma, tears and seminal fluid \(^{89}, 90\).

1.2.3 **Function as a complement inhibitor**

MCP regulates complement in both the classical and alternative pathways at the C3 activation stage, binding C3b/C4b and functioning as the cofactor for the factor I-mediated cleavage of the C3 convertases. The binding of C3b/C4b to MCP prevents further deposition of the activated components on the cell and protects the host cell from phagocytosis or lysis.

The binding sites in MCP for the complement ligands C3b and C4b have been identified by constructing deletion mutants and found to be distinct \(^{20}, 91\). The binding site for C3b has been localised to the C-terminal pair of CP-modules (MCP-3,4) although without the second module (MCP-2) the protein does not possess any cofactor activity. The middle pair of modules, MCP-2,3, are essential for the factor I-mediated cleavage of C4b \(^{91}\). Although the N-terminal CP-module (MCP-1) is not involved in binding C3b/C4b, deletion of this module decreases the efficiency of C4b-binding \(^{91}\). The distinct binding sites for the
two complement components illustrate the functional diversity that is contained within the consensus CP-module structure.

The regulation of complement activation by MCP is illustrated by its roles in reproductive biology and in the prevention of hyperacute rejection of xenografts.

1.2.3.1 Role of MCP in reproduction

The foetus could be considered as an allograft, a graft between members of the same species which have different genetic constitutions, and as such could be rejected by the maternal immune response. MCP and other regulatory proteins, such as DAF and CD59, at the maternal-foetal interface may protect the foetus from the maternal complement system. Changes in the function or expression of MCP may result in the spontaneous abortion of the foetus and infertility. MCP may also have a role in the mechanism of egg/sperm attachment and serve to prevent rejection of seminal fluid.

1.2.3.2 Role of MCP in xenotransplantation

Organ transplantation is restricted by the severe shortage of donor organs, a problem which could be overcome by the use of animal organs. Pigs have been selected as potential donors for xenotransplantation as their organs are of similar size to those of humans. However, hyperacute rejection of xenografts occurs as the presence of a foreign organ activates the classical pathway of complement via xenoactive antibodies. Since the complement regulators MCP, DAF and CD59 are species-specific, their presence on the pig organ does not offer any protection against attack by human complement. The development of
transgenic pigs which express human MCP and/or DAF on the graft endothelium could prevent rejection of the organ. In vitro studies have shown that expression of human DAF and MCP on swine endothelial cells provides some protection from human complement\(^8\). A recombinant soluble form of MCP has been shown to inhibit complement activation and to delay rejection of a xenograft in vivo\(^9\), although it is not yet known if this will prove clinically useful.

1.2.4 MCP as a receptor for pathogens

RCA proteins are exploited by a number of micro-organisms for attachment to host cells, for example CR2 is the receptor for Epstein-Barr virus\(^{37,38}\) and DAF is the receptor for Echovirus\(^{39,40}\). In addition to its complement control function, MCP has been identified as a receptor for the measles virus\(^{34,36}\), pathogenic Neisseria\(^9\) and the M protein of Streptococcus pyogenes\(^{19}\).

1.2.4.1 MCP as the measles virus receptor

The measles virus is a major killer of young children in developing nations\(^97\). The measles virus envelope comprises two glycoproteins: the hemagglutinin which binds to MCP and the fusion protein which initiates the fusion between the virus and the cell membrane\(^98\). The binding site for the virus is distinct from those for C3b/C4b and has been localised to the first and second CP-modules, MCP−1,2, by constructing deletion mutants and chimeras between MCP and the structurally similar protein DAF, which does not bind measles virus\(^{20,21}\). The N-glycans and disulphide bonds of MCP−1,2 are necessary for measles virus
binding and infection, implying that the interaction is conformationally dependent \[^86\].

However, whilst the sugar on module 2 is essential for MCP to function as the measles virus receptor the glycosylation of MCP\(~1\) is of only minor importance \[^87\]. Additionally, the type of sugar is unimportant \[^86, 99\]. Since the measles virus binding site is located within the CP-module region, all four major isoforms of MCP bind the virus \[^36, 100, 101\]. The STP region and cytoplasmic tail have no effect on virus binding \[^21\].

Extensive studies have been performed to determine the nature of the interaction between MCP and the measles virus hemagglutinin. A combination of peptide inhibition and mutagenesis identified two distinct virus binding domains in modules 1 and 2 \[^102-105\]. The regions, Gly 67-Trp 86 in MCP\(~1\) and Phe 119-Ile 139 in MCP\(~2\), were identified by peptide inhibition studies \[^102\]. In the crystal structure of MCP\(~1,2\), these regions are located away from the N-glycans on the same side of the molecule which provides a large site for protein:protein interactions. Since the N-glycan on MCP\(~2\) is known to be required for measles virus binding, on the basis of this structural data it has been postulated that the carbohydrate stabilises the conformation of the protein \[^49\]. Other studies have used site-specific mutagenesis and inhibitory monoclonal antibodies to determine the virus binding sites. One study replaced most of the hydrophilic residues in MCP\(~1,2\) with alanine and found that the mutations E58A and R59A in MCP\(~1\) abolished hemagglutinin binding \[^103\]. Another study showed that the mutations E45A, K63A, P73A and D104A reduced binding to hemagglutinin \[^104\]. In a third mutagenesis study, selected residues in MCP\(~1,2\) were replaced with serine \[^105\]. This study indicated that E45, Y54, E58, R59, Y101, I102, R103, D104 and Y117 are critical for the binding of measles virus hemagglutinin to MCP. The results of these mutagenesis studies are summarised in Figure 1.9. Analysis of the crystal
structure of MCP-1,2 indicates that these residues are on the surface of the molecule and form an extended virus binding surface. 

Figure 1.9 Residues in MCP-1,2 implicated in measles virus binding. Residues determined by (a) peptide inhibition studies are in blue, (b) mutagenesis to Ala are underlined and (c) mutagenesis to Ser in red. The positions of β-sheets, as determined from the crystal structure, are indicated by arrows.

1.2.4.2 MCP is a receptor for pathogenic Neisseria

Bacterial meningitis and gonorrhea/inflammatory disease of the pelvis are caused by the human specific pathogens Neisseria meningitidis and Neisseria gonorrhoeae respectively. MCP has been identified as a cellular receptor for the piliated Neisseria bacteria. It is proposed that the binding site is located within the C-terminal pair of CP-modules as binding of the bacteria was inhibited by monoclonal antibodies to MCP-3,4. Since Neisseria also binds to MCP expressed in E. coli it is proposed that the piliated bacteria do not bind to the carbohydrate moiety on MCP-4.

1.2.4.3 Role of MCP in Streptococcal infection

The binding of Streptococcus pyogenes provides a further example of viral exploitation of RCA proteins. This bacterium is a powerful activator of complement and causes pharyngitis and a number of inflammatory skin infections including impetigo and
necrotizing fasciitis. The presence of MCP on the cell surface provides a binding site for the bacterium to the keratinocytes, the most abundant cells in the epidermis. *S. pyogenes* bind another RCA protein, factor H, which is believed to inhibit complement activation on the bacterial cell surface \(^{106,107}\).

### 1.3 Factor H

Factor H is a soluble glycoprotein of 150 kDa that controls the alternative pathway activation of the complement system. The gene for Factor H is located close to the RCA gene cluster and the protein is related structurally and functionally to the other members of this family. Factor H is a key complement regulator as it reduces the positive feedback cycle of C3 in the alternative pathway to a ‘tickover’ mechanism. In the absence of factor H the positive feedback cycle would result in the depletion of native C3.

### 1.3.1 Structure

Factor H is entirely composed of 20 contiguous CP-modules. The single polypeptide chain has an elongated structure as determined by electron microscopy and hydrodynamics \(^{108}\). NMR structure determination of modules from factor H provided the first high-resolution structures of CP-modules \(^{44-47}\).
1.3.2 Function as a complement regulator

Factor H operates by three mechanisms to regulate the alternative pathway and prevent depletion of C3 in plasma. Firstly, factor H binds C3b or C3i to prevent binding to factor B, and therefore inhibits the generation of the C3 convertases C3bBb or C3iBb. The second mechanism is a consequence of the affinity of factor H for the C3 convertase C3bBb (or C3iBb) being similar to that for binding to C3b. This binding destroys the convertase by displacing the Bb component. On the cell surface this decay accelerating function is performed by CR1 and DAF. The third function of factor H is to serve as a cofactor for the factor I-mediated cleavage of C3b in the fluid phase. The related proteins MCP and CR1 serve as membrane bound cofactors for factor I. The inactivation of C3b prevents the formation of the C3 convertase and prevents formation of the membrane attack complex as C5 is unable to bind. There are three binding sites for complement components on factor H. These are located within modules 1-4, 6-10 and 16-20. Cofactor activity is exhibited solely by modules 1-4 and decay accelerating activity is also mediated by these N-terminal modules.

1.3.3 Non-complement ligands for factor H

Factor H regulates the alternative pathway of complement activation by inhibiting the deposition of C3b on sialic acid-rich surfaces. The binding sites for sialic acid are located in modules 6-10 and 13. Factor H also binds other polyanionic surfaces such as heparin to prevent them from autologous complement attack. The heparin-binding sites have been located to CP-modules 7 and 20.
The binding of factor H to \textit{S. pyogenes} enables this pathogen to evade attack by the complement system. By constructing a series of deletion and truncated mutants the binding site for the bacteria has been located to module 7 of factor H \textsuperscript{114}. This module also contains a heparin binding site and the presence of heparin has been shown to inhibit the binding of the M protein of \textit{S. pyogenes} to factor H \textsuperscript{115}.

\subsection*{1.4 Strategies for the Production of Protein for Structural Studies}

The production of large amounts of protein for structural studies has often proved problematic. Although proteolysis of a protein to yield a smaller fragment has been successfully used in some cases \textsuperscript{116,117} it is a restricted method of protein production. The amino and carboxyl termini of the fragment produced are determined by the position of the available proteolytic sites and there is no possibility of isotopic enrichment for structure elucidation by NMR, nor is it possible to perform site-directed mutagenesis \textit{via} this route. Chemical synthesis and recombinant protein expression provide more versatile approaches, permitting modifications to the protein and the potential for large scale production of labelled material.
Chemical synthesis of proteins offers some distinct advantages over recombinant protein expression. It is the only method of introducing NMR active isotopic labels into a protein site specifically. On the other hand, global labelling with $^{15}$N or $^{13}$C to assist structure elucidation is not economically viable. The introduction of unnatural amino acids permits investigation into the functional and structural significance of selected residues. There are several examples of the successful use of chemically synthesised proteins to study protein structure-function relationships. Despite recent advances in methodology, the routine chemical synthesis of proteins containing 60-200 amino acids still presents a challenge. Large proteins have been synthesised by standard solid phase techniques but the synthesis of milligram quantities of protein has generally been restricted to smaller molecules. In addition to the possible synthetic and purification difficulties, the refolding of the polypeptide chain into the correct 3-dimensional structure presents further challenges.

The first peptide synthesis, of the dipeptide glycylglycine, initiated the development of techniques for condensing amino acids to form peptide bonds. Advances in the methodology resulted in the solution phase synthesis of a number of biologically active peptides including oxytocin and secretin. The introduction of solid phase peptide synthesis (SPPS) revolutionised peptide synthesis by circumventing many of the difficulties associated with solution phase chemistry and enabled the synthesis of larger peptides and proteins. Solid phase peptide synthesis has since become a widely used method for: the production of natural products such as hormones and neuropeptides for structural elucidation; studying structure-function relationships by the synthesis of peptide analogues; and in developing therapeutic agents.
Peptide synthesis is essentially the sequential coupling of amino acids to form a polypeptide chain. However, simply reacting two amino acids together does not result in the formation of a dipeptide. The reaction is slow and reversible, the zwitterionic nature of amino acids does not facilitate the formation of a peptide bond as the amino and carboxyl groups are insufficiently reactive. Moreover, since each amino acid has both a carboxyl and amino group, self-condensation and polymerisation will also occur. Thus for the reaction of A and B, four products are possible: A-A, A-B, B-A and B-B. This problem is exacerbated by the reactivity of amino acid side-chains resulting in further unwanted reactions. To unambiguously form a peptide bond, the rate of formation of the required peptide bond must be increased whilst the likelihood of competing reactions is simultaneously reduced. Numerous methods exist to increase the electrophilicity of the carbonyl carbon to facilitate nucleophilic attack by the amino group of the required amino acid to form the peptide bond. A variety of protecting groups is available for N-α-amino groups and chain functionalities to prevent unwanted side reactions. Orthogonal protecting group strategies enable extension of the polypeptide chain by removing the appropriate terminal protecting group whilst maintaining the integrity of side chain protecting groups.

**1.4.1.1 Synthetic strategies**

Although solution phase synthesis has yielded a number of biologically active peptides the methodology does not facilitate the synthesis of larger polypeptides. The technique is both time- and skill-demanding and consequently is not amenable to automation. Excess reagents must be removed after each coupling reaction and all intermediates purified before the synthesis can be continued. The poor solubility of protected polypeptides in organic
solvents presents further problems as protected peptide chains tend to aggregate rather than interact with solvent.

The introduction of solid phase peptide synthesis has, however, enabled the routine production of polypeptides. Solid phase peptide synthesis is the sequential addition of side chain and α-amino protected amino acids to an insoluble polymeric support. By attaching the developing polypeptide to a solid support several distinct advantages over solution phase chemistry are gained. An excess of reagents is used to drive reactions to completion and any unreacted material can be washed away after isolating the resin by filtration. The crystallisation and purification of synthetic intermediates as in solution phase synthesis is rendered unnecessary permitting automation of the process. The solid support aids the solubility of the protected polypeptide allowing the production of longer peptide chains than is possible in solution. A potential disadvantage of solid phase synthesis is that since all intermediates are attached to the solid support, purification of the target molecule can be complex. These intermediates can be either deletion peptides, missing one or more internal residues, or truncated peptides which result from incomplete coupling reactions. The basic methodology is illustrated in Figure 1.10.

The α-amino and side-chain protected C-terminal residue is attached to the solid support resulting in synthesis from the C- to the N-terminus. The α-amino protecting group is removed and, after activation of the carboxyl group of the next protected amino acid, coupling occurs to form the required peptide bond. Any unreacted N-termini are capped to prevent participation in further reactions and excess reagents are washed away. This process is repeated as necessary for each amino acid to complete the synthesis.
There are two strategies for solid phase peptide synthesis defined by the choice of N-terminal protecting group, known as the Boc- and Fmoc-amino acid approaches. Boc chemistry has been used since the first solid phase synthesis. Consequently, the chemistry is mature and well characterised although there are some drawbacks. The methodology is not strictly orthogonal as it relies on differential acid labilities of the transient α-amino protecting group and the permanent side chain protecting groups. The
chemistry has been used since the first solid phase synthesis\textsuperscript{135}. Consequently, the chemistry is mature and well-characterised although there are some drawbacks. The methodology is not strictly orthogonal as it relies on differential acid labilities of the transient $\alpha$-amino protecting group and the permanent side chain protecting groups. The cleavage of the side chain protecting groups at the end of the synthesis requires the use of HF which is difficult to handle, corrosive and can cause side reactions such as N-O acyl migration at serine residues (see Figure 1.11).

The difficulties associated with Boc chemistry resulted in the development of the alternative 9-fluorenylmethoxycarbonyl (Fmoc) methodology\textsuperscript{136}. The protecting group strategy is truly orthogonal; the Fmoc group is cleaved using mild base (see Figure 1.12) enabling the use of mildly acid labile side chain protecting groups. This not only ensures maintenance of side chain protecting groups throughout the synthesis but also avoids exposing the peptide to harsh acid conditions.
Figure 1.12 Piperidine mediated Fmoc deprotection. The Fmoc-piperdine adduct [A] is chromophoric at 302 nm.

(a) Solid support

The introduction of the solid support simplifies peptide synthesis by avoiding the need to isolate and purify intermediates. The type of solid support used depends not only on the N-terminal protecting group strategy but also on the required functionality of the C-terminus. In Boc chemistry PAM (phenylacetamidomethyl functionalised polystyrene) resin is used as the solid support for the synthesis of peptide acids. The resin and side-chain protecting groups are simultaneously cleaved from the peptide on completion of the synthesis with hydrogen fluoride.

The Wang resin \textsuperscript{136, 137} provides the solid support for the synthesis of peptide acids by the Fmoc methodology. Again the resin is polystyrene based and is functionalised with an acid
labile $p$-hydroxybenzyl alcohol linker. The greater acid lability of this linker allows the use of TFA for the cleavage reaction.

(b) Protecting groups

To prevent unwanted side reactions, self-condensation and polymerisation it is essential that all amino acid functional groups not involved in peptide bond formation are protected. The protecting groups must be reversible and their removal accomplished quantitatively and without damage to the polypeptide. At least two types of protection are necessary: side chain protection that is cleaved only on completion of the synthesis and transient N-terminal protection.

Transient N-$\alpha$-amino protection is generally provided by the Boc or Fmoc group. The benzyloxy carbonyl (Z) group was the first successful use of amino protection by urethane formation. However, since this group is removed under harsh acidic conditions or by hydrogenolysis, the need for a protecting group removable under milder conditions remained. The tertiary butyloxy carbonyl (Boc) group is readily cleaved in mild acid due to the stability of the tertiary carbocation. Thus Boc can be cleaved in the presence of appropriate side chain protecting groups such as the Z group. More recently (2-(4-biphenyl)propyl(2)oxycarbonyl) (Bpoc) derivatives have been utilised due to their increased acid lability.

The Fmoc group provides a completely orthogonal protection strategy. The Fmoc group is cleaved by treatment with a secondary amine whilst the side chain protecting groups are cleaved under mildly acid conditions. Therefore it is unnecessary to expose the peptide to
harsh acids such as HF as in the Boc strategy.

The choice of side chain protecting groups to avoid branching of the growing peptide chain depends on the type of transient α-amino group protection. Side chain protection is essential. There are a number of side chain protecting groups available and their use is well documented.141

(c) Carboxylate activation

The efficient formation of peptide bonds requires the activation of the carboxyl group of the N-α-protected amino acid. Increasing the electrophilicity of the carbonyl carbon is generally used as a means to enhance nucleophilic attack by the amino group. However, it is important to avoid over-activation as side reactions may result.

The formation of the acid chloride is perhaps the most obvious synthetic strategy to increase the electrophilicity of the carbonyl group and indeed was used in early attempts at peptide synthesis.142 However, these highly reactive amino acid derivatives are generally not used in peptide synthesis as chiral integrity can be lost due to a process involving the urethane-derived amino protecting groups such as Boc. Other early attempts at carboxyl activation included the use of acid azides.143 This method is limited by the tendency of acid azides to rearrange to isocyanates at elevated temperatures, see Figure 1.13. Since the reaction is slow the method is not widely used except in cases where there is a high risk of racemisation, for example in the coupling of peptide fragments.
Peptide bonds are usually formed using dicyclohexylcarbodiimide (DCC) or diisopropylcarbodiimide (DIC), see Figure 1.14. These coupling agents can be used in situ to activate the carboxyl group in the presence of the amino component or to form an anhydride or active ester prior to the introduction of the amino component. Since these groups are highly activating it is essential that all amino and carboxyl groups not involved in peptide bond formation are protected. Symmetrical anhydrides of α-amino protected amino acids can be used for peptide synthesis although only one of the two molecules of carboxylic acid in the anhydride molecule is incorporated into the product. Unsymmetrical or mixed anhydrides can be used to eliminate this wastage. The anhydride is designed so that nucleophilic attack on the non-amino acid portion is favoured, either by steric or electronic effects.

Activated esters are widely used in peptide synthesis, see Figure 1.15. Although esters are usually only weakly reactive to nucleophiles their reactivity can be greatly enhanced by the addition of an electron-withdrawing group on the alkoxy oxygen. 4-nitrophenylesters are often used as they are fairly mild activating groups, therefore the risk of racemisation is reduced but coupling may be slow or incomplete. Active esters are now commonly formed.
in situ by the DIC/DCC condensation of 1-hydroxybenzotriazole (HOBt) with the protected amino acid. The development of ethyl 1-hydroxy-1H-1,2,3-triazole-4-carboxylate (HOCT) provides an alternative coupling reagent which is reported to result in high coupling efficiencies. Reactions with in situ activating agents tend to be fast even between sterically hindered amino acids and consequently activated esters are widely used in solid phase peptide synthesis.

![Diagram of symmetrical anhydride formation](image)

Figure 1.14 *Symmetrical anhydride formation.* Peptide coupling is frequently achieved using amino acid anhydrides [C] which are prepared using DCC [A] or DIC [B].

![Diagram of alcohols used in esterification](image)

Figure 1.15 *Alcohols used in esterification.* Active esters are obtained from DIC/DCC mediated condensation with alcohols such as HOBt [A] and HOCT [B].
Chemical ligation is the most significant recent development in protein synthesis for the production of larger proteins 149-151. The technique relies on the chemoselective ligation of two or more unprotected peptide fragments in aqueous solution at neutral pH. The first examples of chemical ligation resulted in an unnatural moiety, frequently a thioester at the site of ligation, see Figure 1.16 149. Developments in the methodology enabled a native peptide bond to be formed at the site of ligation by the reaction of an unprotected peptide fragment with a C-terminal α-thioester group and a peptide with an N-terminal cysteine, see Figure 1.17 150. Chemical ligation has been successfully used to synthesise a number of proteins including HIV type-I protease and microbial ribonuclease barnase 149,152.

![Figure 1.16 Thioester-forming chemical ligation. Adapted from Schnölzer and Kent 149.](image)

The development of expressed protein ligation 153 provides a potentially powerful new protein engineering technique. The method utilises native chemical ligation to attach a chemically synthesised peptide to a recombinant protein. The expression vector selected introduces a thioester linkage between the recombinant protein and an intein-chitin binding domain. Cleavage of the intein-chitin binding domain is usually achieved with a thiol reagent. A thioester is generated in the recombinant protein by reaction with thiophenol prior to ligation with the synthetic peptide via an N-terminal cysteine residue. The
methodology exploits the advantages of both recombinant expression and chemical synthesis to alter protein structure and function by introducing unnatural amino acids, post-translational modifications and biophysical probes. The application of chemically synthesised peptides in this manner enables the advantages of synthesis to be applied to proteins far greater in size than could be produced by total synthesis.

Figure 1.17 *Native chemical ligation*. Adapted from Dawson *et al*.

1.4.2 Recombinant protein expression

Recombinant protein technology permits the production of large amounts of protein and provides the most flexibility for the production of large proteins. A number of systems are available allowing expression in bacteria, yeast and insect and mammalian cells. The
exploitation of these systems can result in the production of large amounts of soluble, folded protein. Inherent to the methodology is the ability to define the termini of the protein and to produce mutants. Additionally, since most organisms require only simple nutrients the expressed protein can be isotopically enriched to assist in structural elucidation by NMR spectroscopy.

1.4.2.1 Protein expression in *E. coli*

*Escherichia coli* is a frequently used organism for the high-level production of heterologous proteins. The advantages of this prokaryotic expression system are well documented. In general, large amounts of protein are produced in rapid, inexpensive processes. There is a wide range of cloning vectors available with a variety of promoters to control expression of the gene of interest. The use of gene fusion vectors can assist both the expression and the purification of heterologous proteins. The presence of a fusion partner may increase the solubility of the protein and can be used to used to target the protein. If the fusion protein is produced in a soluble form the fusion partner can serve as an 'affinity tag' enabling the development of generic purification protocols.

The major disadvantages of expression in *E. coli* include the inability to introduce the post-translational modifications of eukaryotic proteins and the restricted ability to form disulphide bonds. Furthermore, over-expressed proteins can form insoluble aggregates known as inclusion bodies which require the protein to be refolded to regain biological activity.
1.4.2.2 Protein expression in yeast

Expression of recombinant proteins from higher eukaryotes in yeast offers several advantages over prokaryotic expression. The expression systems offer a number of the benefits of bacterial systems: they are easy to use, offer the possibility to manipulate foreign genes and require only simple nutrients for growth. Eukaryotic expression systems have the additional advantage of possessing the appropriate environment for post-translational processing and secretion into the media. Thus the recombinant proteins produced are similar to the native proteins; they are folded, disulphide bonds are formed and, if naturally secreted or membrane-bound, are glycosylated. By incorporating a cleavable signal peptide the recombinant protein is secreted and is generally the major species in the medium since few native yeast proteins are secreted. A potential disadvantage is that although N-linked glycosylation occurs at the correct position in the protein the carbohydrate pattern differs from the native form and can be excessively large $^{155,156}$.

The use of methylotrophic yeasts such as *Pichia pastoris* for the production of recombinant proteins is rapidly increasing. *P. pastoris* grows on a simple medium that utilises methanol as its sole carbon source. This enables the proteins to be readily $^{15}$N and $^{13}$C labelled to assist in structure determination. Isotopic enrichment of proteins expressed in *P. pastoris* has facilitated the elucidation of a number of solution structures by NMR $^{48,157,158}$. 
1.5 Summary

This thesis describes a comparison of methods to produce CP-modules for structural analysis by NMR. The chemical synthesis of two CP-modules, the N-terminal module of MCP (MCP-1) and the fifteenth module of factor H (fH-15), and the recombinant expression of MCP-1 in prokaryotic and eukaryotic systems are described. The characterisation of these modules by NMR and other techniques forms part of an ongoing study into CP-module structure and the nature of the inter-modular interface. It is anticipated that structural studies of these modules will aid understanding of CP-module function and their ability to bind a diverse range of ligands.
2.1 General Methods

All buffers were prepared with Milli-Q grade water. Buffer components and many other reagents were purchased from Sigma. SDS-polyacrylamide gel electrophoresis (PAGE) was carried out using a BioRad Minigel system, acrylamide was purchased from National Diagnostics. All SDS-PAGE gels were 17.5% acrylamide. Centrifugation was carried out using the Sorvall RC 5C Plus centrifuge. Protein concentration was achieved using hollow fibre (Millipore Prep/Scale-TFF cartridge) and high-pressure stirred cell ultrafiltration (Amicon). Nitrocellulose membranes for protein concentration, with molecular weight cut-off (MWCO) as described in the text, were purchased from Millipore. Dialysis tubing (Spectra/por 6, cellulose membrane) was purchased from Pierce and Warriner with MWCO as described in the text. Ultraviolet (UV) spectra were recorded on a Hewlett Packard UV/Vis 8543 spectrophotometer. Scintillation counting was performed using a Packard Tri-Carb 2100TR liquid scintillation counter with Packard Ultima scintillation fluid.

Chromatography - Size exclusion, ion exchange and hydrophobic interaction chromatography were carried out using either a Pharmacia Gradifrac system with single wavelength UV detector, a Pharmacia Äkta Explorer system or a Waters protein purification system with 996 photodiode array detector.
High performance liquid chromatography (HPLC) was carried out on a Gilson 715 system or a Waters system comprising a 600 pump and a 486 UV/Vis detector. Brownlee columns were used for all reverse-phase HPLC (RP-HPLC). Elution was achieved using linear gradients of acetonitrile (Fisher HPLC grade) in Milli-Q grade water with 0.1% v/v HPLC grade TFA (Fisher) in each as described in the text.

**Mass spectrometry** - Matrix assisted laser desorption ionisation (MALDI) time of flight (TOF) mass spectra were recorded on a PerSeptive Biosystems Voyager Biospectrometry Workstation using either α-cyan-4-hydroxycinnamic acid or 3,5-dimethoxy-4-hydroxycinnamic acid as matrix.

Positive ionisation atmospheric pressure electrospray mass spectrometry (ESMS) was carried out using a Micromass Platform II single quadrupole mass spectrometer. Samples were prepared in acetonitrile/H₂O (1:1), 0.5% v/v formic acid. Cone voltages were typically in the range 45-55 V.

**Sequencing** - Protein N-terminal sequencing was performed by on an ABI 477A sequencer at the Welmet protein characterisation facility (University of Edinburgh) and on an ABI 476 sequencer at the University of Leicester.

DNA sequencing was performed on an ABI 377 XL sequencer following reaction with dRhodamine Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems).

**Differential scanning calorimetry** - Differential scanning calorimetry (DSC) studies were conducted on a MC-2 differential scanning calorimeter (Microcal, Northampton, MA). The cell volume was 1.5 ml, the rate of heating was 1 degree C/min and excess pressure was
Materials and Methods

kept equal to 80 bar. Protein concentration was approximately 1 mg/ml in the buffers described in the text. The partial molar heat capacity and melting curve were analysed using standard procedures \textsuperscript{199}. The data were processed using the software ORIGIN 2 (Microcal).

**Circular dichromism** - Circular dichromism (CD) studies were performed using a Jasco-600 spectropolarimeter (Japan Spectroscopic Company, Japan) with cylindrical quartz cell of path length 0.02 cm (at the EPSRC/BBSRC centre for CD at the University of Stirling). The protein concentration was approximately 0.2 mg/ml in the buffers described in the text.

**NMR** - NMR data were acquired on a Varian (Palo Alta, Ca, USA) Inova spectrometer operating at 600MHz frequency, in a 5mm triple-resonance pulsed field gradient probe. Sample conditions and concentrations are described in the text.

2.2 Chemical Synthesis

Further details concerning the synthesis of MCP-1 and fH-15 are described in Chapter Three.

2.2.1 Materials

Fmoc-L-amino acids were purchased from Novabiochem and \textit{p}-alkoxybenzylalcohol
Materials and Methods

(Wang) resin from Bachem. Peptide synthesis grade dimethylformamide (DMF), 1,4-dioxane and piperidine were supplied by Rathburn Chemicals. Peptide synthesis grade N,N-diisopropylethylamine (DIEA) and trifluoroacetic acid (TFA) were purchased from Applied Biosystems. 4-dimethylaminopyridine (DMAP), N-hydroxybenzotriazole (HOBt) and diisopropylcarbodiimide (DIC) were obtained from Aldrich and ethyl 1-hydroxy-1H-1,2,3-triazole-4-carboxylate (HOCt) synthesised in-house 148. Trifluoroacetic acid (TFA) was purchased from Applied Biosystems (ABI). The scavengers thioanisole, triisopropylsilane, phenol and ethanedithiol were purchased from Sigma. Fmoc-glycine (14C-UL)-OH was purchased from Sigma, with specific activity 4.2 mCi/mmol. N-hydroxysuccinimide was purchased from Aldrich and d-biotin and dicyclohexylcarbodiimide (DCC) from Sigma. (17-tetrabenz[a,c,g,i]fluorenyl-methylchloroformate (TbFmocCl) was synthesised in-house 160.

2.2.2 Solid phase synthesis

(a) Resin loading - p-alkoxybenzyl alcohol polystyrene (Wang) resin (1 mmol/g) and DMAP (10 mg) were swollen in dimethylformamide (DMF) (5 ml) (1 h, 20 °C). Excess solvent was removed prior to resin loading.

(1) Fmoc-Glu -resin for MCP-1 synthesis

Fmoc-Glu(OtBu)-OH (1.40 g, 3.15 mmol) and DIC (159 μl, 1.58 mmol) were suspended in DMF (5 ml) and mixed by ultrasonication for 20 min at 20 °C. The resulting solution was added to pre-swollen Wang resin and sonication continued for a further 30 min. The resin was isolated by filtration, washed sequentially with dichloromethane (DCM) (2 x 50 ml) and diethyl ether (2 x 50 ml) and dried overnight in vacuo.
(2) **Fmoc-Gly-resin for fH-15 synthesis**

Fmoc-Gly-OH (1.8 mmol, 0.535 g), HOCl (1.8 mmol, 283 mg and DIC (1.8 mmol, 282 µl) were suspended in DMF (1 h, 20 °C). The solution was added to pre-swollen Wang resin and mixed by ultrasonication for 1 h at 20 °C. The resin was isolated and washed as described above.

(b) **Determination of loading efficiency** - All loading efficiencies were determined by sonicating a portion of the peptidyl-resin (≈5 mg) in 20% v/v piperidine/DMF (10 ml) for 10 min at 20 °C. Loading efficiencies were determined by monitoring the absorbance of the resulting solution at 302 nm (ε = 15 400 M⁻¹cm⁻¹ for Fmoc-piperidine adduct).

(c) **Automated solid phase synthesis** - Peptides were synthesised using the Fmoc-Nα-strategy on an ABI 430A automated peptide synthesiser with on-line monitoring using an ABI 758A detector.

Fmoc-L-amino acids were side-chain protected as follows: Ser, Thr, Tyr: t-butyl ethers; Asp, Glu: t-butyl esters; Lys, Trp: t-butoxycarbonyl (Boc); Asn, Gln, His: τ-triphenylmethyl (Trt) and Arg - 2,2,5,7,8-pentamethylchroman-6-sulphonyl (Pmc).

The following procedure was used for each amino acid coupling:

(i) Capping (irreversible blocking of any unreacted amino groups)

(ii) Deprotection (removal of Fmoc-Nα-protection to release terminal amino group)

(iii) Coupling (activation of required amino acid, reaction with peptide on solid support)

The resin was washed with solvent between each cycle to remove excess reagents.
(i) Capping

The resin was vortexed with acetic anhydride (0.25 M), DIEA (0.125 M) and HOBt (0.2% w/v) in DMF/1,4-dioxane (1:1, 10 ml) for 10 min. The solution was drained from the reaction vessel and the resin washed with DMF/1,4-dioxane (1:1, 6 x 10 ml).

(ii) Deprotection

The resin was vortexed with a solution of 20% v/v piperidine/DMF (10 ml) for 10 min. The reaction vessel was drained and absorbance of the resulting solution monitored at 302 nm. The resin was washed with DMF/1,4-dioxane (1:1, 4 x 10 ml) before treating with 20% v/v piperidine/DMF (10 ml) for 1.5 min. The resin was then washed with DMF/1,4-dioxane (1:1, 7 x 10 ml).

(iii) Coupling

Each amino acid was converted to the corresponding activated ester using HOCt or HOBt (1 mmol) (as described in the text), with DIC (1 mmol) in DMF/1,4-dioxane (1:1, 10 ml). On addition of the activated amino acid to the resin the mixture was vortexed (30 min) before the vessel was drained and the resin washed with DMF/1,4-dioxane (1:1, 6 x 10 ml). For double/triple-coupling an additional portion(s) of Fmoc-amino acid (1 mmol) was activated and the coupling cycle repeated.

On completion of the synthesis the N-terminal Fmoc group was cleaved under the conditions described for the Fmoc loading test. The N-terminal residue was Fmoc-deprotected by sonicating the resin (≈1 g) in 20% (v/v) piperidine/DMF (10 ml) for 10 min at 20 °C. The resin was washed sequentially with DMF (3 x 50 ml), 1,4-dioxane (3 x 50 ml), and DCM (3 x 50 ml) and dried overnight under vacuum. The peptide was then either cleaved from the resin or subjected to N-terminal derivatisation.
(a) MCP-1 synthesis

Synthesis of MCP-1 (corresponding to residues 32-97 of the native human MCP sequence) was performed on a 0.1 mmol scale as described above. All amino acids were coupled using the HoCt active-ester method (1 mmol of amino acid) excluding the histidine residues (His 78 and His 85) which were coupled using the HOBr active-ester method. Residues Asp 61 to Lys 66 inclusive were double coupled.

(b) fH-15 synthesis

The synthesis of fH-15 (corresponding to human factor H residues 866-928) was carried out on a 0.1 mmol scale as described for MCP-1. All residues were single coupled as the HOBr active-ester excluding the regions Glu 866-Gln 875 and Phe 905-Thr 913 inclusive which were double coupled. In the modified synthesis, residues Glu 866-Ser 890 and Ser 908-Glu 912 were double-coupled and Phe 905-Ile 907 were triple-coupled. All remaining residues were single coupled.

(d) N-terminal derivatisation - Following Fmoc-deprotection, N-terminal derivatisation was performed as described in the text according to the following protocols:

(1) N-terminal derivatisation with Fmoc-glycine (14C-UL)-OH

HOBr (135 mg, 1 mmol) was dissolved in DMF (2 ml) and Fmoc-Gly-OH (297 mg, 1 mmol) added. Fmoc-Gly(14C-UL)-OH (typically ~0.4 MBq) was added followed by 1,4-dioxane (2 ml) and DIC (156 μl, 1 mmol). The resulting solution was mixed by sonication for 30 min at 20 °C. Pre-swollen resin (~1 g in minimum volume DMF) was added and the mixture sonicated for 3 h at 20 °C. The resin was isolated by filtration and washed sequentially with DMF (3 x 30 ml), 1,4-dioxane (3 x 30 ml), and DCM (3 x 30 ml) and
dried under vacuum overnight. The Fmoc-glycine($^{14}$C-UL) loading efficiency was
determined by Fmoc-deprotection.

(2) *Derivatisation with tetrabenzo[a,c,g,i]fluorenyl-17-methoxycarbonyl (TbFmoc)*

TbFmocCl (78 mg, 0.17 mmol) was sonicated, in a flask shielded from light, with DIEA
(9.8 μl, 0.06 mmol) in 1,4-dioxane (10 ml) for 20 min at 20 °C. Pre-swollen peptidyl-resin
was added and sonication was continued for a further 4 h. The resin was isolated and
washed as described above. The TbFmoc loading efficiency was determined as described
for Fmoc-deprotection.

(3) *Biotinylation*

N-hydroxysuccinimide-biotin (NHS-biotin) was prepared by dissolving d-biotin (500 mg,
2 mmol) in hot DMF (6 ml) followed by stirring with N-hydroxysuccinimide (235 mg, 2
mmol) and DCC (422 mg, 2 mmol) for 2 h (20 °C). The resulting precipitate was removed
by filtration and the filtrate concentrated to dryness in vacuo. Following resuspension in
diethyl ether (10 ml), NHS-biotin was recrystallised from propan-2-ol. Fmoc-deprotected
peptidyl-resin and NHS-biotin (2 mg/ mg resin) were suspended in DMF (10 ml) and shaken
for 22 h at 45°C and 240 rpm. The resin was isolated by filtration, washed sequentially with
DMF (2 x 50 ml), DCM (2 x 50 ml) and DCM/methanol (60:40, 2 x 50 ml) and dried in
vacuo.

(e) *Resin cleavage* - Peptidyl-resin (~1 g) was stirred under nitrogen with ethanedithiol
(2 ml), thioanisole (0.5 ml), triisopropylsilane (0.5 ml), water (0.5 ml) and phenol (0.75 g)
for 20 min at 20 °C. TFA (16 ml) was added and the resulting mixture was stirred for 4 h.
The resin was removed by filtration and washed with TFA (10 ml). The filtrate and
wettings were combined and triturated with ice-cold 2% v/v β-mercaptoethanol/diethyl ether (200 ml). The resulting white solid was isolated by centrifugation at 4000 g for 5 min and washed with ice-cold ether (50 ml). Following centrifugation the white solid was resuspended in the buffer described in the text.

2.2.3 Purification of MCP-1

(a) Size exclusion chromatography - Size exclusion chromatography (SEC) was performed on a Pharmacia HiLoad G30 Superdex 16/60 column with elution at 0.5 ml/min. Column calibration was performed with lysozyme, ubiquitin and insulin (all 1 mg/ml) in 8 M urea, 0.1 M TrisHCl (pH 8.8), 50 mM β-mercaptoethanol (βME).

After isolation from the cleavage reaction, the protein mixture was resuspended in 8 M urea, 0.1 M TrisHCl (pH 8.8), 0.2 M βME and incubated for 16 h at 37 °C. The solution containing the protein mixture was loaded onto the column in 1 ml aliquots and elution (in 8 M urea, 0.1 M TrisHCl (pH 8.8), 50 mM βME) monitored at 280 nm. Two major peaks were isolated and the fractions corresponding to each peak were concentrated by ultrafiltration (1 kDa MWCO membrane) to 10 ml. Following dilution (1:4) with 0.1 M TrisHCl (pH 8.2), 0.3 mM oxidised glutathione (GSSG), 3 mM reduced glutathione (GSH) and 1 mM di-sodium ethylenediaminetetraacetic acid (EDTA), the protein solution was dialysed against the same buffer (1 kDa MWCO membrane, 5 x 1 l, 48 h, 4 °C). The dialysed solutions were then concentrated by ultrafiltration (1 kDa MWCO) to 25 ml. The fractions corresponding to each of the major peaks were assayed by scintillation counting, RP-HPLC and ESMS.
(b) Anion exchange chromatography - Anion exchange chromatography (AIEX) was performed by applying the dialysed protein corresponding to the earliest eluting peak from SEC to a Pharmacia Mono Q HR5/5 column pre-equilibrated in 0.1 M TrisHCl (pH 8.2). The protein was eluted from the column with a linear gradient of 0-0.5 M NaCl (0.1 M TrisHCl, pH 8.2), over 40 column volumes. The major fractions were collected and analysed by RP-HPLC, ESMS and scintillation counting.

(c) Hydrophobic interaction chromatography - Hydrophobic interaction chromatography (HIC) was performed on a Phenyl Superose HR 5/5 column pre-equilibrated in 2M ammonium sulphate, 0.1 M TrisHCl (pH 8.2). Fractions collected from anion exchange chromatography were diluted 1:1 with 4 M ammonium sulphate and applied to the column. A gradient of 2 M to 0 M ammonium sulphate, 0.1 M TrisHCl (pH 8.2) was applied at 0.5 ml/min over 30 column volumes.

(d) Reverse-phase HPLC - All RP-HPLC chromatography was performed on a Aquapore Butyl (100 x 4.6 mm, 300 Å, 7 μm) column pre-equilibrated with 90% H2O (0.1% v/v TFA), 10% acetonitrile (0.1% v/v TFA). Elution was performed with a linear gradient to 50% acetonitrile (0.1% v/v TFA) over 30 min.

(e) Reduction of methionine sulphate - The protein was dissolved in acetic acid/H2O (10% v/v) to ~1 mg/ml. N-methyl-mercaptoacetamide (5 mole equivalents) was added and the solution incubated under nitrogen (37 °C, 16 h). The resulting solution was lyophilised.
Materials and Methods

2.2.4 Purification of fH-15

(a) Size exclusion chromatography - Column calibration was performed as described above with 6 M GuHCl, 0.2 M β-ME as the eluent. Following overnight incubation with 6 M GuHCl, 0.2 M βME, the protein mixture was partially purified by SEC with 6 M GuHCl, 50 mM βME as the eluent, as described for MCP-1. Fractions containing protein with the expected elution volume for fH-15 were dialysed against 0.1 M TrisHCl (pH 8.4), 0.3 mM GSSG, 3 mM GSH, 1 mM EDTA as described for MCP-1.

(b) RP-HPLC - Chromatography was carried out on either Aquapore Octyl analytical (250 x 4.6 mm, 300Å, 7 μm), semi-preparative (250 x 10 mm, 300 Å, 20 μm) or Aquapore ODS analytical (250 x 4.6 mm, 300 Å, 7 μm) or semi-preparative (250 x 10 mm, 300 Å, 20 μm) columns as described in the text. Elution was achieved with a linear gradient of 10-90% acetonitrile (0.1% v/v TFA) over 30 min.

(c) Anion exchange chromatography - AEX was performed as described for MCP-1. The sample was eluted with a gradient of 0-1 M NaCl (0.1 M TrisHCl, pH 7.4) over 30 column volumes.

(d) Amino acid analysis - Protein (~0.2 mg) was dissolved in 6 M HCl (1 ml) in a Carius tube. Air was evacuated from the tube under vacuum and the sealed tube incubated at 110 °C for 42 h. The sample was evaporated to dryness and resuspended in sodium citrate buffer (pH 2.2, 600 μl). Amino acid analysis was performed on a Pharmacia Biochrom 20 LKB 4150 alpha acid analyser.
2.3 Recombinant Expression of MCP-1

All standard molecular biology procedures were performed according to Molecular Cloning (Sambrook et al. 163).

2.3.1 Materials

MCP cDNA (BC2 isoform in pSG5 vector) was a gift from John Atkinson (Washington University School of Medicine, St. Louis). Synthetic oligonucleotides were purchased from Gibco or Genosys. Restriction enzymes and DNA modifying enzymes were purchased from New England Biolabs and Promega. Expression media components were obtained from Difco and [\(^{14}\)N]-ammonium sulphate from Goss Scientific. Isopropyl-\(\beta\)-D-thiogalactopyranoside (IPTG) was purchased from Europa Bioproducts and ampicillin from Boehringer Mannheim. Agarose was obtained from BioRad and DNA was extracted from agarose gels using the Qiagen Qiaquick Gel Extraction Kit. Plasmid DNA preparations were performed using the Qiagen Qiaprep Spin Miniprep and the Promega Wizard Plus Maxiprep kits. The expression vector pGEX-6P-1, glutathione Sepharose 4B and PreScission Protease were obtained from Pharmacia. \(P.\ pastoris\) expression vectors and strains were purchased from Invitrogen and the pGEM-T Easy Vector System from Promega. Endoglycosidase \(H_f\) was purchased from Boehringer Mannheim. Phenylmethylsulphonyl fluoride (PMSF) was purchased from Sigma. ConcavalinA Sepharose was obtained from Pharmacia.
2.3.2 Prokaryotic expression of MCP-1

(a) Construction of the plasmid pGEX-6P-1 encoding for MCP-1 - The DNA encoding for MCP-1 (comprising the native human sequence residues 31-98) was amplified by PCR using pSG5-MCP as a template. The 5' primer GCG GAA TTC GAT GCC TGTCGAG GAG CC contained an *EcoRI* site and the 3' primer TTT TCC TTT TGC GGC CGC CTA TTA TGT TTC TCT TCT ATA ACAGGC GTC a *NcoI* site for cloning into the expression vector pGEX-6P-1.

The PCR reaction mix consisted of 10x reaction buffer (10 μl), a mixture of dNTPs (200 μM each, final concentration), template DNA (~100 ng), primers (100 pmol each), Deep Vent DNA polymerase (1 μl, 5 units), dimethylsulphoxide (DMSO) (5 μl) and sterile deionised water (to 100 μl). Amplification was performed by PCR with an annealing temperature of 50 °C for 30 seconds. The amplified DNA fragment and the vector pGEX-6P-1 were each double digested with the restriction enzymes *EcoRI* and *NcoI* (37 °C, 16 h) and, following purification by agarose gel electrophoresis, ligated by incubation with T4 DNA ligase (16°C, 16 h).

The ligation product was transformed into calcium chloride competent *E. coli* cells (strains as described in the text) and resulting colonies screened for the correct plasmid by PCR followed by gel electrophoresis. The previously described PCR reaction was performed with the template DNA being replaced by a small amount of the bacterial colony. The primers used were either the cloning primers or pGEX-6P-1 sequencing primers to confirm the presence of the cloned insert and correct vector respectively. Positive transformants were sequenced to confirm correct construction of the plasmid.
(b) Expression of MCP-1 - Expression of MCP-1 in *E. coli* was performed with reference to the Pharmacia *GST Gene Fusion Manual*. 

Media were prepared as follows:

- **LB**: 10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl in deionised water, pH adjusted to 7.0
- For LB-Amp, ampicillin was added to a final concentration of 50 µg/ml.
- **YTA media**: Tryptone (16 g/l), yeast extract (10 g/l) NaCl (5 g/l) in deionised water and pH adjusted to 7. Ampicillin was added to a final concentration of 50 µg/ml. For YTGA media a sterile 20% w/v glucose solution was added to the YTA media to a final concentration of 2% w/v glucose.

1. **Expression in *E. coli JM109***

A single colony was used to inoculate LB-Amp (10 ml) and incubated with shaking at 37°C and 240rpm for 16 h. LB-Amp (100 ml) was inoculated with 500 µl of this culture and incubated with shaking for 24 h at 37°C and 240rpm. This culture was then diluted 1:100 in LB-Amp and the culture incubated (37°C, 240 rpm) to \( A_{600} \approx 0.7 \) AU. Protein expression was induced by the addition of IPTG to 0.5 mM and incubation continued under the same conditions for 4 h.

The cell pellet was harvested (7700 g, 10 min, 4°C) and the supernatant discarded.

Following resuspension in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na\(_2\)HPO\(_4\), 1.8 mM KH\(_2\)PO\(_4\), pH 7.3) (50 µl/ml culture) cells were lysed by sonication (8 bursts of 30 seconds duration). 20% v/v Triton-X100 was added to a final concentration of 1% v/v and the mixture incubated with gentle mixing at room temperature for 30 min. Following
centrifugation (1200 g, 10 min, 4°C), phenylmethylsulphonyl fluoride (PMSF) was added to the cell-free extract (to 50 μM) and the pellet stored at -20°C.

The above expression protocol was modified for the expression of GST-MCP~1 at reduced temperatures and with reduced concentrations of IPTG by changing the cultivation temperature to 28 or 25°C and inducing expression with 50, 150 and 250 μM IPTG as described in the text. Purification was performed as described above.

(2) Expression in E. coli JA221

Expression of GSTMCP~1 in E. coli JA221 was performed as described above. Protein expression was induced by the addition of IPTG to 0.1 mM for 3 h at 37 °C and 240 rpm.

(3) Expression of GSTMCP~1 in E. coli BL21

The plasmid pGEX-6P-1 encoding MCP~1 was transformed into calcium chloride-competent E. coli BL21 cells. A single colony was used to inoculate YTGA media and the culture incubated at 37°C for 16 hours with shaking at 240 rpm. The culture was diluted 1:100 into YTA medium and incubated at 28°C with shaking at 240 rpm to A600 ~ 0.5. Expression of GSTMCP~1 was induced by the addition of IPTG to a final concentration of 0.1 mM and incubation continued for 2 h. The cell-free extract was prepared as described above.

(c) Purification of soluble GSTMCP~1 - The cell-free extract was applied to a Glutathione Sepharose 4B column (1 ml column volume/100 ml cell-free extract) pre-equilibrated in PBS. The flow through was collected and the column washed with 3 x 10 column volumes of PBS. Adsorbed protein was eluted with 10 mM reduced glutathione,
Materials and Methods

50 mM TrisCl (pH 8.0) (1 ml/ml resin) following incubation with the same buffer for 15 min. The elution procedure was repeated twice.

(d) Cleavage of GST from MCP~1 - GST-fusion protein was dialysed against 50 mM TrisCl (pH 7.0), 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT) (5 x 1 l, 48 h, 4 °C). The GST fusion partner was cleaved from MCP~1 by incubation with PreScission Protease (20 μl, 40 U) at 4 °C for 16 h. The cleavage mixture was applied to a Glutathione Sepharose 4B column and the solution containing unbound proteins collected. MCP~1 was isolated from the column flow-through by RP-HPLC. Protein identity and homogeneity was checked by SDS-PAGE, ESMS and N-terminal sequencing. Typically, 0.25-0.5 mg of purified protein was isolated from 1 litre of cell culture.

(e) Isolation of solubilised inclusion bodies - The post-sonication pellet was resuspended in 0.5% v/v TritonX-100/1×PBS (9 μg pellet) and incubated without shaking at room temperature for 5 min. Following centrifugation (12000 g, 4 °C, 10 min) the remaining pellet was resuspended in 8 M urea, PBS, 2 mM GSH, 0.2 mM GSSG (9 ml/g pellet) and incubated without mixing at room temperature for 1 h. The resulting viscous solution was centrifuged (12000 g, 30 min, 4 °C) and the supernatant isolated. The supernatant was dialysed against 0.1 M TrisCl (pH 8.2), 2 mM GSH, 0.2 mM GSSG, 1 mM EDTA (5 x 1 L over 48 h, 4 °C).

GST was cleaved by PreScission Protease as described above. The cleavage mixture was purified by SEC on a Pharmacia G30 HiLoad G30 Superdex 16/60 column with 0.1 M TrisCl (pH 8.2) as eluent. Protein of the appropriate size for MCP~1 was further purified
by RP-HPLC as described above. Typically, 0.75-1 mg of purified protein was obtained from 1 litre of cell culture.

2.3.3 Eukaryotic expression of MCP-1

Expression in *P. pastoris* was performed with reference to the Invitrogen *Pichia pastoris* expression manual. The procedures used are outlined in brief below.

Media were prepared as described in the *P. pastoris* expression manual:

Buffered minimal glycerol (BMG) media - 100 mM potassium phosphate (pH 6.0), 1.34% w/v yeast nitrogen base (YNB) (with ammonium sulphate, without amino acids), 4 x 10⁻³% w/v biotin, 1% v/v glycerol

Buffered minimal methanol (BMM) media - as BMG, 0.5% v/v methanol substituted for 1% v/v glycerol.

For [¹⁵N]-media - YNB replaced with 1.34% w/v yeast nitrogen base (YNB) (without ammonium sulphate, without amino acids), supplemented with 0.2% w/v [¹⁵N]-ammonium sulphate.

(a) Construction of plasmid pPIC9 encoding MCP-1 - The amplified DNA fragment encoding for MCP-1 (as described previously) and *P. pastoris* expression vector pPIC9 (each digested *EcoRI-NotI*) were ligated (16 °C, 16 h) and the ligation product used to directly transform calcium chloride competent *E. coli Top10F*’ cells. Transformants were assayed by PCR screening (as described previously) and DNA sequencing. 20 μg of the DNA construct was linearised with *SalI* and transformed into the *P. pastoris* strains KM71.
and GS115 by electroporation using a BioRad Gene Pulser at 1500 V, 25 μF and 200 Ω.

(b) Test inductions of MCP-1 in *P. pastoris* KM71 and GS115 - Following screening for the His' phenotype, small-scale test inductions were performed by growing clones in BMG media (10ml) at 30 °C to $A_{600} > 6$. Following harvesting by centrifugation (500 g, 5 min, 4 °C) cells were resuspended in BMM media (2.5 ml) to induce protein expression. Growth was continued at 30 °C for 6 days and the media was supplemented daily with methanol to 0.5% v/v. The cells were harvested by centrifugation (500 g, 5 min, 4 °C) and the supernatant screened for MCP-1. Aliquots (1ml) of the supernatant were precipitated by the addition of a 30% w/v trichloroacetic acid (TCA) solution (to 10% v/v final concentration) and analysed by SDS-PAGE.

Expression of MCP-1 in *P. pastoris* KM71 was scaled up to permit analysis of the sample. pPIC9-MCP-1 was grown in BMG (100 ml) to $A_{600} \approx 20$ (48 h) at 30 °C, 280 rpm. Cells were harvested and resuspended in BMM (100 ml) and grown at 30 °C, 280 rpm for 7 days with the daily addition of methanol to 0.5% v/v. The cells were harvested by centrifugation (1500 g, 10 min, 4 °C), the supernatant isolated and concentrated by ultrafiltration (3 kDa MWCO membrane) to 5 ml. After 1:1 dilution with 0.1 M TrisHCl (pH 9), the protein was deglycosylated by the addition of Endoglycosidase H$_r$ (EndoH$_r$) (5 mU/mg protein) and incubated at 37 °C for 16 h. EDTA (to a final concentration of 10 mM) was added to limit proteolysis. The protein was isolated by TCA-precipitation and analysed by SDS-PAGE.

(c) Large scale expression of MCP-1 in *Pichia pastoris* - A single colony of pPICMCP-1 was used to inoculate BMG (5 ml) and grown at 30 °C for 2 days. This seed culture (1 ml) was used to inoculate BMG (100 ml) which was grown at 30 °C for 24 h.
The starter culture (20 ml) was then used to inoculate 2 l of BMG and growth continued at 30°C to A₅₀₀ ≈ 20 (~48 h). The cells were harvested (1500 g, 10 min, 4 °C) and resuspended in 1 l of BMG and incubation continued for a further 24 h to A₅₀₀ ≈ 60-65. The cells were harvested (1500 g, 10 min, 4 °C), resuspended in 2 l of BMM, and incubated at 30 °C for 6 days with daily methanol supplementation to 0.5% v/v. The protein was harvested by centrifugation (1500 g, 10 min, 4 °C), the supernatant isolated and PMSF added (to 80 μM) to inhibit proteolysis.

**Materials and Methods**

(d) **Purification of MCP-1** - Following concentration by ultrafiltration (3 kDa MWCO) the supernatant was applied to a PD10 desalting column (Pharmacia) and eluted in 0.1 M TrisHCl (pH 9.0). The protein was partially purified by AIEX on a Pharmacia Q-Sepharose Fast Flow (FFQ) column (50 x 0.5 mm) equilibrated in the same buffer and eluted with a gradient of 0-0.5 M NaCl over 40 column volumes. Fractions collected were analysed by SDS-PAGE and pooled as appropriate prior to concentration in an Amicon pressure cell (3kDa MWCO membrane) to 10 ml.

The protein was deglycosylated by the addition of EndoH₄ (5 mUnits/mg protein) and incubated at 37°C overnight in the presence of EDTA (10 mM). MCP-1 was isolated by RP-HPLC on a C4 analytical column and lyophilised. Following resuspension in 20 mM TrisHCl (pH 7.4), 0.5 M NaCl, 1 mM CaCl₂, 1mM MnCl₂ and 1 mM MgCl₂ the protein was applied to a Concavalin A (ConA) sepharose column (Pharmacia, 20 x 2.5 mm) pre-equilibrated in the same buffer. The column was washed with five column volumes of this buffer and fractions containing protein (as determined by UV absorbance at 280 nm) collected. MCP-1 was further purified by RP-HPLC and the fractions containing protein lyophilised. Typically 6-8 mg of purified protein was obtained per litre of culture.
(e) **Preparation of \[^{15}\text{N}]\text{-MCP-1**} - A \[^{15}\text{N}\]-labelled sample of MCP-1 was produced in the *P. pastoris* strain KM71. Cells were grown to $A_600 \approx 20$ as described previously. The harvested cells were resuspended in BMG containing 0.2% w/v \[^{15}\text{N}\]-ammonium sulphate as the sole nitrogen source and growth continued for 24 h to $A_600 \approx 60-65$. The cells were harvested and expression of \[^{15}\text{N}\]-MCP-1 induced by resuspending in BMM containing 0.2% w/v \[^{15}\text{N}\]-ammonium sulphate as the sole nitrogen source. The protein was expressed and purified as previously described.

2.3.4 **Eukaryotic expression of MCP-1(T85A)**

(a) **Construction of pPICZ\(\alpha\)MCP-1(T85A** - The N-linked glycosylation site (\[^{31}\text{NHT}\}) in MCP-1 was removed by site directed mutagenesis to replace threonine 85 with alanine. The MCP-1 gene cloned for expression of MCP-1 in *P. pastoris* was utilised as the template in the PCR reactions to yield the MCP-1(T85A) gene. Four primers were utilised in the reactions (all 5'-3'):

Primer 1  | GCG GAA **TTC** GAT GCC TGTCGAG GAG CC  
Primer 2  | ATT TGT GAT CGG AAT CAT **GCA** TGG CTA CCT GTC TCA GAT  
Primer 3  | ATC TGA GAC AGG TAG CCA **TGC ATG** ATT CCG ATC ACA AAT  
Primer 4  | TTT TCC TTT **TGC** GGC **CGC** CTA TTA TGT TTC TCT TCT ATA ACAGGC GTC

Primers 1 and 2 were as used for cloning MCP-1 and contain *EcoRI* and *NotI* restriction enzyme sites respectively. The complementary primers 2 and 3 contain the mutation ACA→GCA (TGC in the complementary strand) to afford the Thr to Ala change and introduce a unique *NsiI* restriction site (*ATG CAT*).
A two step PCR process introduced the mutation; the first step comprising of two separate reactions with the MCP-1 template and the second using the products of the first PCR step with primers 1 and 2.

**PCR Step I:**

Two reactions were performed with annealing temperature 50 °C for 30 seconds.

(a) 10x reaction buffer (10 µl), dNTPmix (200µM each, final concentration), template DNA (100 ng), primers 1 and 3 (100 pmol each), DMSO (5 µl), Taq DNA polymerase (1 µl, 5 units) and sterile deionised water (to 100 µl).

(b) As above, substituting primers 2 and 4 for primers 1 and 3.

**PCR step 2**

10x reaction buffer (10 µl), dNTPmix (200µM each final concentration), PCR step1 (a) product (0.5 µl), PCR step1 (b) product (0.5 µl), primers 1 and 4 (100 pmol each), DMSO (5 µl), Taq DNA polymerase (1 µl, 5 units) and sterile deionised water (to 100 µl).

The amplified product was ligated into the pGEM-T vector and transformed into competent *E. coli* BLX-1 cells (as described in the pGEM-T Easy Vector System manual).

Transformants were screened for the correct insert by amplification with primers 1 and 4 (as described for MCP-1). Screening for potential mutants was carried out by digestion with *NsiI*. Transformants containing an *NsiI* site were digested with the restriction enzymes *EcoRI* and *NotI* and the cloned MCP-1(T85A) gene ligated to the *P. pastoris* expression vector pPICza and transformed into chemically competent *E. coli* Top10F' cells. The correct construction of the plasmid was determined by DNA sequencing of the transformants.
Materials and Methods

*P. pastoris* strains KM71, GS115 and SMD1168 were transformed by electroporation with the pPICzα plasmid following linearisation with *BstXI*, as described in the Invitrogen manual. After electroporation the cells were incubated at 30°C for 2 hours without shaking. Zeocin resistance was used to identify positive transformants using plates containing 50 μg/ml zeocin. Transformed cells were screened for increased zeocin resistance using plates containing 100, 500, 1000, 2000 μg zeocin/ml to assay for potential jackpot clones.

Test inductions of MCP-1(T85A) were carried out as described for MCP-1. BMG and BMM were supplemented with histidine (0.004% w/v). Large scale (2 l) expression of MCP-1(T85A) in *P. pastoris* KM71 in both unlabelled and [15N]-labelled media was carried out as described for MCP-1.

**(b) Purification of MCP-1(T85A)** - Purification of the concentrated supernatant containing MCP-1(T85A) was performed by AIXEX chromatography as described for MCP-1. Fractions containing MCP-1 (T85A) were concentrated by ultrafiltration (1kDa MWCO membrane) to 10 ml and the protein isolated by RP-HPLC as described previously. The protein was further purified on a ConA column and by RP-HPLC as described for MCP-1. Typically 5-7 mg of purified protein was obtained per litre of culture.
2.4 Structural Characterisation of MCP-1

This section describes the techniques utilised for structural characterisation of MCP-1. An extensive description of NMR spectroscopy is given as this was the principal characterisation technique used throughout the project.

2.4.1 Differential scanning calorimetry

Scanning calorimetry measures the heat capacity of a sample by heating or cooling it at a constant rate. The change in heat capacity on thermal denaturation of a protein provides information about the conformational stability of the molecule.

All temperature induced changes are accompanied by a change in enthalpy; for example, the thermal denaturation of proteins by increasing the temperature is accompanied by heat absorption. The relationship between temperature \( T \) and enthalpy \( H \) can be determined experimentally by measuring the heat capacity of a substance at constant pressure \( (C_p) \) over a given temperature range, according to Equation 2.1. The enthalpy is then determined by integration of the heat capacity curve.

\[
C_p = \frac{\partial H}{\partial T} \quad [2.1]
\]

Since calorimeters measure the differential, rather than the absolute, heat capacity; the measured heat capacity of a sample is determined relative to the solvent. The protein
Materials and Methods

A solution is prepared, generally to a concentration of 1 mg/ml, and dialysed against a solvent until equilibrium between the low molecular weight compounds in the sample and the pure solvent is achieved. The dialysate is then used as the reference sample and its heat capacity is determined. The heat capacity of the protein solution is subsequently measured and the difference between these two values is the calorimetric partial heat capacity of the protein.

A peak in the heat capacity curve, corresponding to a large adsorption in heat, is considered to represent denaturation of the proteins tertiary structure. The exposure of hydrophobic groups to water dominates the partial heat capacity of proteins. Therefore, the heat capacity of the unfolded state is found to be greater than that of the folded state as the non-polar groups within the core of the protein are exposed upon thermal denaturation.

The enthalpy change on denaturation of the proteins tertiary structure is determined by integration of the heat capacity curve. The relationship between the calorimetrically determined enthalpy (\(\Delta H_{\text{cal}}\)) and the van’t Hoff enthalpy change (\(\Delta H_{\text{vH}}\)), the temperature dependence of the equilibrium constant, see Equation 2.2; provides information about the nature of the protein.

\[
d \left(\ln K\right) / d \left(1/T\right) = -\Delta H / R \tag{2.2}
\]

If \(\Delta H_{\text{cal}}\) and \(\Delta H_{\text{vH}}\) are approximately equal, the denaturation of the protein is considered to be a one-step co-operative process. Therefore, the protein denatures as a whole, such that at equilibrium only the fully folded and full unfolded states are present. If \(\Delta H_{\text{cal}}\) is found to be
considerably greater than $\Delta H_{\text{m}}$, the protein is considered to consist of two domains and in
the reverse situation ($\Delta H_{\text{m}} > \Delta H_{\text{w}}$) the protein is considered to be oligomeric.

DSC therefore provides a method for determining if the protein has any tertiary structure
and can be used to compare the conformational stabilities of, for example, wild-type and
mutant proteins. In addition, the nature of the protein in solution can be assessed.

2.4.2 Circular dichromism

Circular dichromism (CD) provides information about the differential absorption of left and
right circularly polarised components of plane-polarised light by optically active molecules.
CD bands occur in the near-and far-UV regions which both provide structural information.
In the far UV-region (190-250 nm) the spectrum is dominated by peptide bonds which can
yield secondary structure information about the protein. Although $\alpha$-helices have
characteristic spectra in this region, other secondary structure elements are much less well
defined.

The near-UV region bands (250-290 nm) are provided by aromatic side-chains and
disulphide-bonds. Aromatic residues are often in asymmetric environments within the
protein structure and therefore absorb left- and right-handed circularly polarised light to
different extents giving rise to the spectrum. In the absence of an ordered structure the CD
in this region is low so information about protein conformation is readily obtained. The
local environment of the aromatic rings determines both the sign and magnitude of the CD
bands, therefore individual absorbances cannot be attributed to transitions in the vicinity of
Materials and Methods

particular side-chains. However, the near-UV spectrum provides a sensitive means of assessing protein structure and can be used as a fingerprint of the correctly folded conformation.

The intensities of the CD bands cannot be predicted but they are known to be determined by three factors. Highly mobile side chains tend to result in CD-bands of lower intensities than those fixed in a specific orientation. The intensities are also affected both by the number and closeness of the aromatic side chains. Far-UV CD is particularly useful for comparing the tertiary structures of wild-type and mutant forms of a protein as small changes in the environment of the aromatic side-chains can be detected. The technique has also been used to study molten-globule states of proteins and can be used to assess renaturation of proteins.

2.4.3 NMR spectroscopy

The theory of NMR spectroscopy has been extensively detailed. This section outlines the principles behind the technique and describes the methods used in this project.

NMR active nuclei are defined as those with non-zero spin quantum numbers. For example, the nuclei of interest for biomolecular NMR (\(^1\)H, \(^{13}\)C, \(^{15}\)N and \(^{31}\)P) all have spin quantum number \(I=\frac{1}{2}\). These nuclei have corresponding magnetic moments, and when placed in a magnetic field (\(B_0\)) exhibit two orientations in space, aligned with or against the magnetic field. The population of the energy levels is given by the Boltzmann distribution with the lower energy level being slightly more populated in the equilibrium state. This results in net magnetisation (denoted \(M_z\)), in the direction of the applied field. In addition to being
aligned with the external magnetic field, the spinning charged nuclei precess around the field at a rate known as the Larmor frequency. The Larmor frequency is dependent on the chemical environment of the nucleus and the strength of the applied magnetic field.

The application of a second magnetic field ($B_1$), in the form of a radiofrequency (RF) pulse of the same frequency as the Larmor frequency of the nucleus, causes transitions between the two energy states. This results in rotation of the magnetisation from its equilibrium position towards the xy plane. The length and strength of the applied pulse $B_1$ determines the tilt of the magnetisation away from the $z$ direction; for example a "90° pulse" produces a zero $M_z$ and a non-zero $M_x$. The magnetisation rotating in phase in the xy plane at the Larmor frequency induces a current in the detection coil which varies sinusoidally. This current decays exponentially as the magnetisation returns to its equilibrium position. The signal is known as the free induction decay (FID) and is the primary NMR signal. Fourier transformation of the FID results in the NMR spectrum.

The relaxation of the spin system to its original equilibrium state occurs exponentially over a period of time. There are two relaxation mechanisms: longitudinal relaxation is the return of $M_z$ to its original value and is characterised by the longitudinal relaxation time, $T_1$, whilst the decay of $M_x$ to zero is known as transverse relaxation and is described by the spin-spin relaxation time, $T_2$. It is the spin-spin relaxation time that determines for how long the NMR signal can be collected in the form of an FID. The decay is faster for large molecules that tumble more slowly in solution.
Materials and Methods

The presence of electron clouds around the NMR-active nuclei shields them. Consequently the applied magnetic field, and the local magnetic fields actually experienced by the nuclei, are not necessarily the same. Therefore nuclei in different chemical environments within the molecule may resonate with different Larmor frequencies. This effect, known as chemical shift, enables nuclei in different chemical environments to be distinguished from each other. However, it is not possible to unambiguously identify resonances simply on the basis of their chemical shift as many effects contribute to the exact resonance frequency. The ‘ring current effect’ is a result of the π-electron system of aromatic rings introducing a secondary magnetic field and is frequently observed in biochemical NMR. Nuclei above and below the aromatic ring are shielded and their Larmor frequency is reduced, whilst other nuclei close to the ring are deshielded and resonate to a higher frequency. For example, shielding of the methyl groups of leucine and valine by aromatic residues within the hydrophobic core of a protein and the corresponding change in chemical shift of these groups can be indicative of the formation of a hydrophobic core.

2.4.3.1 Water suppression

Techniques for effective water suppression are essential for the success of biomolecular NMR in aqueous solution. Whilst the protein sample is at millimolar concentration the water concentration is 55 M and consequently overwhelms the protein resonances. Although all spectra could be recorded in D₂O, exchangeable amide protons would no longer be visible in ¹H spectra. Consequently, a number of alternative methods have been developed to suppress the water signal of the molecule. However, D₂O can be used as a solvent to simplify spectra and to identify those amide protons that are unavailable for
exchange due to hydrogen-bonding or burial in the core

One method of water suppression is pre-saturation, in which the sample is irradiated at the Larmor frequency of water using weak RF pulses. Consequently, there is almost no net magnetisation of water in the z direction and a significant reduction in the signal due to water is observed after application of the 90° pulse. Although this technique suppresses the water resonance it can also obliterate the signals due to α-protons which resonate below or close to water and it also affects other protons due to solvent-exchange.

The development of pulsed field gradients has resulted in significant improvements in water suppression techniques. The two most widely used gradient-based techniques are WATERGATE 169 and double pulsed field gradient spin echo (DPFGSE) 170. Both use a combination of water-selective pulses and pulsed field gradients to eliminate the water signal. At the same time, the protein signal undergoes a dephasing-rephasing cycle and is consequently restored prior to acquisition. The DPFGSE sequence in particular results in very good water suppression and is a robust method less susceptible to operator error than WATERGATE.

2.4.3.2 Multi-dimensional NMR

The introduction of two-dimensional (2D) NMR spectroscopy has proved crucial for structure determination of proteins and has enabled NMR to become a valid alternative to X-ray crystallography for the study of small proteins 171. The transfer of magnetisation from one spin to another is essential for 2D NMR techniques. Two spins are correlated either by through-bond interactions (scalar coupling) or through-space interactions (dipole-dipole
Materials and Methods

coupling). The transfer of magnetisation results in resonances being distributed in two dimensions. In addition to reducing the amount of overlapping resonances, the extra dimension enables the identification of chemical shift of pairs of nuclei participating in through-bond or through-space interactions within the molecule. These interactions are inaccessible in 1D-spectra of macromolecules which consist of a series of overlapping resonances from which little structural information can be obtained.

The basic 2D NMR experiment, see *Figure 2.1*, can be divided into 4 blocks:

![Generalised 2D NMR pulse sequence](image)

**Figure 2.1** *Generalised 2D NMR pulse sequence.* Details are described below.

[A] *Preparation* - the spins are allowed to relax towards equilibrium, and at the end of the relaxation period a 90° pulse is applied which flips magnetisation from the z axis into the xy plane and introduces phase coherence.

[B] *Evolution* - typically in homonuclear experiments this period (t₁) serves to label the magnetisation of spins by their chemical shift. For heteronuclear experiments a state suitable for the transfer of magnetisation is prepared during this period.
Materials and Methods

[C] Mixing - the transfer of magnetisation occurs during this period. In heteronuclear experiments this is followed by chemical shift labelling of the hetero-nuclei and transfer of the magnetisation back to protons.

[D] Acquisition - the FID is collected during the acquisition time $t_2$.

The cycle is repeated with systematic variation of $t_1$ to result in a matrix of data with each row representing one FID. The amplitude of the signals detected in $t_2$ is a function of the effects that occurred during $t_1$. Fourier transformation results in a 2D matrix with frequency axes $F_1$ and $F_2$ reflecting the Larmor frequencies of spins detected during $t_1$ and $t_2$. The 1D-spectrum forms a diagonal across the spectrum with off-diagonal crosspeaks representing the interactions between nuclei.

Although spreading resonances into a second dimension eases the problem of overlap in NMR spectra the problems re-emerge as the size of the studied molecule increases. A potential solution is to further increase the dimensionality, by adding another set of evolution and mixing times for each additional dimension. For 3D NMR, the data are dependent on three time periods; both $t_1$ and $t_2$ are systematically varied and the data is acquired during $t_3$. Three Fourier transformations are required to transform the 3D matrix. The data is therefore presented in a cube, planes of which can be taken to yield a stack of 2D spectra with fewer connectivities than are observed in conventional 2D spectra, see Figure 2.2. Pulse sequences for 3D spectroscopy are generally constructed by joining two 2D pulse sequences together.
Materials and Methods

Figure 2.2 Schematic of 2D plane extracted from a 3D heteronuclear NMR spectrum. $F_1 = ^{15}\text{N}$ or $^{13}\text{C}$, $F_2, F_3 = ^1\text{H}$.

Increasing the dimensionality of homonuclear experiments does not, however, result in reduced numbers of crosspeaks; in fact the number is significantly increased. In conjunction with limited digital resolution and inefficient polarisation transfer for larger proteins, 3D homonuclear experiments often fail to reduce the problem of crowded 2D spectra. The power of 3D spectroscopy is fully realised by running heteronuclear experiments to produce simplified spectra, after labelling the sample with the NMR-active nuclei $^{15}\text{N}$ and/or $^{13}\text{C}$. Unlike their more abundant $^{12}\text{C}$ and $^{14}\text{N}$ counterparts, the NMR active isotopes $^{13}\text{C}$ and $^{15}\text{N}$ are spin-$\frac{1}{2}$ nuclei and pulse sequences can therefore be designed to transfer phase coherence efficiently between the different types of NMR-active nuclei present in the molecule. Additionally, the larger heteronuclear coupling constants ($J_{\text{NH}} = 94$ Hz, $J_{\text{CH}} = 125 - 155$ Hz) and carbon-carbon coupling constants ($J_{\text{CC}} = 35 - 55$ Hz) allow more efficient transfer of magnetisation than is observed in homonuclear experiments.

The current size-limit of NMR for double-labelled proteins is approximately 30 kDa. As the size of the molecule increases in addition to resonances being overlapped, the efficiency of magnetisation transfer decreases. This is a consequence of faster relaxation caused by larger molecules tumbling more slowly in solution. In an attempt to circumvent these
difficulties, a ‘divide and conquer’ strategy has been developed for the study of large modular proteins\(^3\). Whilst the intact proteins containing many modules are too large for NMR studies, dissection into single-, double- or triple-modules facilitates structural elucidation by NMR.

2.4.3.3 Homonuclear NMR experiments

This section describes the homonuclear 2D experiments utilised for structural characterisation of MCP-1 expressed in \(P.\ pastoris\). Experimental details can be found in Chapter Five.

(a) Correlated spectroscopy (COSY)

The COSY experiment yields crosspeaks for pairs of directly coupled spins. For proteins, spectra are recorded in phase-sensitive mode and double-quantum filtration (DQF) is applied to remove the dispersive diagonal peaks observed in regular COSY experiments. Since the size of \(J_{1H1H}\) is small, magnetisation is only efficiently transferred through three bonds, therefore only vicinal correlations are observed. The COSY experiment is of particular use, in conjunction with total correlation spectroscopy (TOCSY), for determining the spin-systems of individual amino acids.
(b) **Total correlation spectroscopy (TOCSY)**

In the TOCSY experiment magnetisation is transferred through bonds as in the COSY sequence. However, transfer is not limited to directly coupled protons. The experiment is based on the homonuclear Harmann-Hahn (HOHAHA) effect. In the TOCSY pulse sequence, after application of the second $90^\circ$ pulse, a series of $180^\circ$ RF pulses are repeated rapidly during the mixing time. The application of this irradiation acts as a 'spin-lock', overriding the effects of chemical shift and causes all the rotating magnetisation to precess at the frequency of this applied radiation. Consequently all spins precess as if they have the same Larmor frequency and magnetisation is transferred between coupled nuclei.

Each residue has an isolated spin-system since magnetisation cannot be transferred through the peptide bond. Therefore, only intra-residue correlations are observed and there are characteristic patterns for each amino acid type. The NH region of the TOCSY spectrum is of particular importance. It is possible to transfer magnetisation from protons occurring progressively further along the amino acid side-chain to the amide proton resulting in a complete proton spin-system for each amino acid, excluding proline. The extent of magnetisation transfer through the spin-system is dependent both on the length of the side-chain and the mixing time. If a protein has favourable relaxation properties, a mixing time of $\sim 70$ ms generally results in the complete proton spin-system of a particular residue, at the frequency of the amide proton. The TOCSY experiment is therefore used to assign types of amino acids within a protein prior to sequential assignments.
Nuclear Overhauser effect spectroscopy (NOESY) provides information about through-space couplings and therefore provides essential information for determining the secondary and tertiary structures of proteins. The NOE results in the change in intensity of one resonance when the population of states of another resonance is perturbed. In steady-state NOE experiments, which are acquired as 1D experiments, this is usually achieved by saturation, when the populations of the upper and lower energy states are equalised by the application of a continuous radiofrequency pulse.

For protein NMR a 2D transient NOESY experiment is more appropriate. Following the $t_1$ period, part of the magnetisation is flipped to the $-z$ axis. This is a non-equilibrium state and during the subsequent mixing time the populations of the dipole-dipole coupled spins are perturbed relative to their equilibrium states.

The transfer of magnetisation occurs if the two nuclei are in chemical exchange or if they are close in space since there will be a dipole-dipole interaction between them. The intensity of NOEs has a $r^6$ dependence, where $r$ is the distance between the nuclei. NOEs can be observed between protons up to ~5 Å apart which provides essential information for determining the 3D structure of a protein.
2.4.3.4 Heteronuclear experiments

Isotopic labelling with $^{15}\text{N}$ and/or $^{13}\text{C}$ is frequently used to simplify NMR spectra of macromolecules. The experiments described were used to aid the structure determination of MCP~1.

(a) 2D Heteronuclear single quantum correlation (HSQC) spectroscopy

The basis of 2D and 3D heterocorrelated experiments is the transfer of magnetisation from protons to carbon or nitrogen for chemical shift-labelling followed by transfer back to protons. Since excitation and detection are carried out on protons, full proton sensitivity is maintained despite the temporary transfer of magnetisation to the heteroatoms. Once the magnetisation is transferred back to protons, the signal is acquired in the presence of carbon or nitrogen decoupling. Since the heteronuclear coupling is typically decoupled during the $t_1$ interval, only one crosspeak is observed for each pair of coupled heteronuclei.

(b) 3D Heteronuclear NMR spectroscopy

The HSQC experiment is frequently used in conjunction with other 2D experiments to create 3D experiments. For example, $^{15}\text{N}$-edited experiments such as $^1\text{H}-^{15}\text{N}$ 3D NOESY-HSQC and $^1\text{H}-^{15}\text{N}$ 3D TOCSY-HSQC can be readily constructed. After the initial chemical shift-labelling of proton resonances during $t_1$, a corresponding (NOESY or TOCSY) mixing scheme is applied to transfer the magnetisation to amide protons. The magnetisation is then transferred to $^{15}\text{N}$ nuclei for chemical shift labelling during $t_2$. Finally, the magnetisation is
transferred back to the amide protons and the signal acquired during the acquisition time \( t_1 \).

The cross-peaks spread within a 3D cube are depending on the \( ^{15}N \) chemical shift of the nitrogen attached to the amide proton. This significantly reduces the crosspeak overlap of 2D homonuclear spectra.

The 3D HNHA experiment was also used in this study. The experiment is designed so that the intensities of NH and H_\text{a} crosspeaks reflect the \( ^3J_{\text{HN,Ha}} \) coupling constants. The \( ^3J_{\text{HN,Ha}} \) coupling constants are related, via the Karplus equation, to the torsion angle \( \phi \) in polypeptides. These values can therefore be used as restraints in structure calculations.

2.4.3.5 Band selective 2D experiments for resonance assignment of proline residues

Band selective 2D TOCSY-TOCSY and 2D TOCSY-NOESY experiments have been used in various forms in high resolution NMR spectroscopy, particularly for the study of carbohydrates\(^\text{173-175}\). These techniques have been optimised in our laboratory\(^\text{176}\) to assist in the identification of proline spin systems of proteins. The chemical shift distribution histogram of proline protons in proteins\(^\text{177}\) indicates that most of the \( \delta \)-protons resonate in the region \( 3.6 \pm 0.4 \) ppm, whilst most \( \beta \)- and \( \gamma \)-protons are found in the region \( 2.0 \pm 0.4 \) ppm, see Figure 2.3. Although many other protons resonate in these regions, the probability of non-proline CH\(_2\) protons from the same residue resonating in both of these regions is low. Based on this analysis, it is possible to simplify the regular 2D spectra and aid assignment of proline resonances using band selective 2D TOCSY-TOCSY and 2D TOCSY-NOESY experiments.
Materials and Methods

Figure 2.3 *Chemical shift distribution histogram for proline residue.* This is a compilation of between 800 to 1100 chemical shift entries for each proline proton. Excitation profiles of 10.1 ms q-SNEEZE pulses centred around 3.6 and 2.0 ppm, respectively, are shown. Adapted from BioMagResBank (http://www.bmrb.wisc.edu/data_access/seg_histogram_plots/).

Band selective TOCSY transfer from proline δ-protons to β- and γ-protons (or in the opposite direction) was achieved by a q-SNEEZE pulse which has a 'top-hat' profile and provides clean excitation with pure phase across the selected region. At the end of this preparation block a regular 2D TOCSY or NOESY pulse sequence was appended. A comparison of the regular 2D spectra with their band selective TOCSY counterparts is shown in *Figure 5.9.*
CHAPTER 3

CHEMICAL SYNTHESIS OF CP-MODULES

3.1 Introduction

The work in this section assesses the feasibility of solid phase peptide synthesis (SPPS) for the production of CP-modules. Despite the reported successes of the solid phase methodology, the synthesis of proteins consisting of more than 60 residues continues to prove problematic, particularly when sufficient yield and purity for structure determination are required. The renaturation of synthesised proteins, particularly those containing disulphide bonds, may cause further complications. The synthesis of CP-modules, in sufficient yield for structural elucidation by NMR, therefore presents several challenges to peptide chemistry. There are currently no reported syntheses of CP-modules although other disulphide bond-containing proteins of a similar size have been synthesised. The synthesis of the 54-residue epidermal growth factor-like module of the human complement protease C1r and subsequent structural elucidation by NMR is one successful example of the methodology.

In the current project, the synthesis of two CP-modules, MCP~1 and fH~15, has been performed. The synthesis of MCP~1 was carried out in conjunction with the recombinant expression of the protein in the heterologous host E. coli. This dual approach was taken to enable studies on the renaturation of the protein. fH~15 was selected as a model for
developing the methodology as it is known from a previous structural study to be stable, soluble and to exist in a single conformation as an isolated single module.

3.2 Chemical Synthesis of MCP-1

The chemical synthesis of MCP-1 provides a comparative study to the recombinant expression of the protein as discussed in Chapter Four. The purification and renaturation of the protein is described.

3.2.1 Solid phase synthesis

MCP-1, corresponding to residues 32-97 of the native human MCP sequence, was synthesised on a 0.1mmol scale on Wang resin using step-wise N-α-Fmoc chemistry. The introduction of the C-terminal glutamine residue onto the solid support was effected by diisopropylcarbodiimide (DIC) carboxyl activation and the resin loading determined to be 0.37 mmol glutamine/g resin. All subsequent amino acids were coupled as the N-hydroxytriazole (DIC/HOCt) active ester with the exceptions of the histidine residues (His 78 and His 85). To minimise enantiomerisation, the trityl-protected histidine residues were coupled as the hydroxybenzotriazole (DIC/HOBt) active ester. Residues Asp 61 to Lys 66 were double coupled as a previous synthesis revealed a decrease in coupling efficiencies in this region. On completion of the synthesis, the loading of the N-terminal serine residue was determined by Fmoc deprotection to be 0.098 mmol/g resin, equivalent to
a 26% loading efficiency (relative to the C-terminal residue loading). The synthesised sequence is shown in Figure 3.1.

During the course of the synthesis of MCP-1 it was necessary to remove ~30% of the resin due to excessive swelling. This may have prevented sufficient vortexing of the resin with the reagents and possibly resulted in a decrease in coupling efficiencies. The portion of resin was removed and set aside before commencing the double coupling cycles. An increase in resin volume has frequently been observed during synthesis on Wang resin. Although this effect has not been quantified, it has been shown that for copoly(styrene-1\%divinylbenzene) resin the volume of a swollen, polypeptide-bearing resin bead is larger and can contain more solvent than a swollen unsubstituted bead. Resin-bound, protected polypeptides are often found to be highly solvated, but they are frequently insoluble in the same solvent after cleavage from the solid support. This high degree of solvation results in increased swelling of the resin, an effect which becomes more pronounced as the polypeptide is elongated and its properties dominate those of the resin.

The use of the transient Fmoc protecting group enables the progress of the synthesis to be monitored in real-time. Piperidine mediated cleavage of the N-terminal Fmoc at the end of each cycle releases the chromophoric Fmoc-piperidine adduct. The relative efficiency of each coupling cycle can be determined by monitoring the absorbance of this deprotection solution at 302 nm. Although no quantitative information is obtained at the level of each
individual cycle, a qualitative assessment of regions of unsatisfactory coupling can be made. The relative coupling efficiencies for the synthesis of MCP-1 are plotted in Figure 3.2.

![Figure 3.2 MCP-1 coupling efficiencies. Couplings were determined from monitoring Fmoc-piperidine adduct formation and are calculated as a percentage of the C-terminal residue loading.](image)

The limitations of this method are illustrated by the data obtained for MCP-1. A number of the couplings appear to be >100% efficient. This is possibly a consequence of excessive resin-swelling since retention of solvent by the peptidyl-resin can result in an increase in the apparent concentration of the Fmoc-piperidine adduct. Apparent decreases in coupling efficiencies were observed in three areas: at Asn 83; in the region Asp 61 to Lys 66 despite the introduction of double coupling cycles; and over the final ten residues of the synthesis (Ser 32 to Phe 41). Despite the approximate nature of the monitoring technique these regions of decreased coupling efficiency suggest that the synthesis of MCP-1 did not proceed with high efficiency.

Peptide synthesis frequently results in the accumulation of resin-bound deletion and
Results and Discussion

truncated peptides due to incomplete coupling reactions. Deletion peptides, missing one or more internal residues, can be formed during the chain assembly. The 'capping cycle' results in truncated peptides. After each coupling cycle, any unreacted terminal amino groups on the resin-bound polypeptide are acylated ('capped') to prevent participation in further reactions. In this manner the formation of deletion peptides is minimised and on completion of the synthesis only the target sequence should bear a free terminal amino group. Consequently, N-terminal derivatisation is often used to differentiate the target sequence from the truncated sequences. The introduction of chromatographic tags for affinity purification of the target sequence is a commonly used strategy.\textsuperscript{160, 182 - 186}

In this instance, derivatisation with \(^{14}\text{C}\)-glycine was selected to permit identification and quantification of the target sequence throughout purification. Labelling was carried out by isotopically enriching Fmoc-Gly-OH with Fmoc-Gly (\(^{14}\text{C-UL}\))-OH, see Figure 3.3, followed by coupling to the free N-terminus of the resin-bound protein.

\[
\text{Fmoc--HN}^*\text{O}^\ast \text{OH}
\]

Figure 3.3 \textit{Fmoc-Gly (}\(^{14}\text{C-UL})\text{-OH}. \(^{14}\text{C}\) atoms are denoted \(*\).

Subsequent analysis indicated that only 2% of the \(^{14}\text{C}\) label was incorporated into the protein. Since the \(^{14}\text{C}\)-enriched Fmoc-Gly-OH was used in ten-times excess over the resin-bound protein to drive the reaction to completion, this indicates that only 20% of the total resin-bound species possess a free terminal amino group. This provides further evidence
that the synthesis of the full-length sequence had not progressed satisfactorily and indicated
that the majority of the resin-bound species were truncated peptides.

3.2.2 Purification of MCP-1

Simultaneous cleavage of the protein from the resin and removal of the side-chain
protecting groups was achieved by treatment with trifluoroacetic acid (TFA) for 4 h. In
order to ensure clean and efficient cleavage a mixture of 'scavengers', thioanisole,
ethanedithiol (EDT), triisopropylsilane (TIS) and phenol, was added. Scavengers serve to
suppress alkylation of susceptible residues by t-butyl-cations released on cleavage of the
side-chain protecting groups. EDT is an effective scavenger of t-butyl cations, minimises
acid catalysed oxidation of tryptohan residues and aids cleavage of the trityl protecting
group from cysteine \(^{187}\). The addition of thioanisole minimises the oxidation of methionine,
a frequently observed side-reaction of acidolytic cleavage. Its presence also enhances the
cleavage of the protecting group 2,2,5,7,8-pentamethylchroman-6-sulfonyl from arginine \(^{187}\).
Phenol protects tyrosine and tryptophan residues \(^{188}\) whilst TIS has similar protective
properties to EDT. Additionally, scavengers minimise the reaction with cations produced
on cleavage of the peptide-resin linker which can result in irreversible re-attachment of the
peptide to the resin \(^{189}\).

Following cleavage from the resin, the crude protein mixture was dissolved in 8M urea,
0.1M TrisHCl, 0.2M β-mercaptoethanol (βME) and incubated at 37 °C to denature and
reduce the proteins. Analysis by scintillation counting confirmed incorporation of the
Results and Discussion

[14C]-label. From the specific activity of Fmoc-glycine (14C-UL)-OH (23.38 x 10^6 dpm/mmol) the yield of [14C]-labelled protein was calculated to be 14 µmoles. The total yield of crude cleaved peptide was calculated from its absorbance at 280 nm to be approximately 52 µmoles (calculated ε=13370 M^-1 cm^-1), indicating that the [14C]-labelled sequence represented 27% of the total protein cleaved from the resin.

Analysis of the crude protein mixture by RP-HPLC indicated the presence of a number of closely eluting protein species over the range 43-50% acetonitrile (0.1% v/v TFA) (see Figure 3.4). The [14C] label was found to be distributed throughout the major protein-containing fractions. The identity of these different [14C]-labelled species was unclear. It was initially postulated that the different retention times could be due to different states of oxidation of the protein or the presence of protecting groups. However, incomplete capping cycles during the synthesis could result in deletion peptides possessing a free N-terminus. N-terminal derivatisation would consequently be non-specific for the target sequence, thereby reducing the efficacy of the technique for isolating only the protein of interest.

Figure 3.4 RP-HPLC of crude peptide mixture. The protein was eluted from a C4 reverse phase column with a gradient of 10-50% acetonitrile (0.1% v/v TFA) in water over 30 mins.
3.2.2.1 Size exclusion chromatography

Initial purification of MCP-1 was carried out by size exclusion chromatography (SEC) to remove small truncated or deletion peptides and small molecules from the cleavage reaction. The column was calibrated prior to use with lysozyme (14.4 kDa), ubiquitin (8.5 kDa) and insulin (5.7 kDa) to enable a calculated elution volume for MCP-1 to be determined, see Figure 3.5.

![Figure 3.5](image)

Figure 3.5 Calibration of Pharmacia G30 HiLoad Superdex 16/60 column for SEC. The standards described in the text were eluted in 8 M urea, 0.1 M TrisHCl (pH 8.8) and 50 mM βME at 0.5 ml/min. The calculated elution volume for MCP-1 is indicated.

The crude peptide mixture was eluted from the column in two broad peaks with elution volumes 46-54 ml [A] and 55-62 ml [B] respectively, see Figure 3.6. Peak [A], which had the expected elution volume for MCP-1, comprised approximately 40% of the total protein eluted from the column. Analysis of the fractions collected indicated that the [14C] label was divided approximately equally between peaks [A] and [B]. Since peak [B] corresponds to proteins of lower molecular weight than calculated for MCP-1, the N-terminal
derivatisation of deletion peptides is the probable cause of the different protein species purified by RP-HPLC. The presence of [\(^{14}\)C]-labelled non-target sequences prevented the use of the label to quantify the yield of MCP-1. However, the label was used qualitatively in subsequent purification procedures.

Figure 3.6 SEC of MCP-1. The crude protein mixture was eluted from a Pharmacia G30 HiLoad Superdex 16/60 column in 8 M urea, 0.1 M TrisHCl (pH 8.8) and 50 mM βME at 0.5 ml/min. Peak [A] has the expected elution volume for MCP-1.

Fractions collected from SEC were collected and analysed by RP-HPLC. Proteins from peak [A] were found to be more strongly retained than those from peak [B] (44-50% and 41-48% acetonitrile, 0.1% v/v TFA respectively), see Figure 3.7. In each case the chromatograms were broad, suggesting the presence of a variety of protein species requiring further purification.

As the by-products of chemical synthesis are often physically and chemically similar to the target sequence, a refolding protocol was implemented. It was anticipated that the target
Results and Discussion

and truncated sequences would adopt different conformations and therefore have different physical properties which would facilitate their separation.

![Graph](image)

Figure 3.7 RP-HPLC analysis of fractions from SEC peak [A] and [B]. Elution from a C4 analytical column was performed with a gradient of 10-50% acetonitrile (0.1% v/v TFA) in water over 30 mins.

3.2.2.2 Renaturation of MCP-1

The fractions corresponding to peak [A] isolated by SEC were concentrated to approximately 10 ml by ultrafiltration. A five-fold dilution of the concentrated protein samples was performed, using a refolding buffer consisting of 0.1 M TrisHCl (pH 8.2), 3 mM reduced glutathione (GSH), 0.3 mM oxidised glutathione (GSSG) and 1 mM EDTA, in order to dilute the protein to ~100 μg/ml. It was anticipated that the presence of the glutathione redox pair would provide an appropriate oxidising environment for the formation of the disulphide bonds of MCP-1. Removal of the denaturing and reducing agents was achieved by dialysis against the buffer described above. The application of this protocol to a version of MCP-1 expressed in *E. coli* is described in Chapter Four.
Renaturation of disulphide bond-containing proteins frequently proves problematic and the difficulties can be exacerbated by the presence of contaminating proteins. No precipitation was observed when the proteins within sample [A] were subjected to the refolding protocol described. In contrast, treatment of the fractions corresponding to peak [B] of the SEC gave a precipitate that could only be re-solubilised by incubation at 37 °C with 8 M urea and 0.2 M βME. Since sample [B] was believed to contain deletion and truncated proteins, mis-folding and subsequent aggregation and precipitation was not unexpected.

Following the attempted renaturation of sample [A], the sample was analysed by RP-HPLC. Elution occurred in overlapping broad peaks over the range 45-53% acetonitrile (0.1% v/v TFA), see Figure 3.8. Fractions were collected at 1 minute intervals for analysis by ESMS. A mass spectrum for an aliquot collected from the main peak is shown in Figure 3.9. Although a range of molecular weights was observed, some material of the mass expected for synthetic MCP−1 (7755 Da) was identified. The masses observed for other aliquots are recorded in Table 3.1. The other proteins co-eluting with MCP−1 are close in mass illustrating how similar the synthetic by-products are to the target sequence.

![Figure 3.8 RP-HPLC of SEC peak [A] following renaturation. The protein was eluted from a C4 reverse phase column with a gradient of 10-50% acetonitrile (0.1% TFA) in water over 30 mins.](image-url)
Results and Discussion

Attempts were made to identify the truncated sequences by considering the omission of various N-terminal residues. Since synthesis progresses from the C-terminus, this approach allows for a likely cause of the low molecular weight species although it cannot take into account deleted internal residues. Putative deleted residues and the calculated masses of these truncated sequences are detailed in Table 3.1.

RP-HPLC and ESMS analysis was also performed on sample [B]. The protein mixture was eluted over the range 45-49% acetonitrile (0.1% v/v TFA) and subsequent analysis by ESMS carried out as described above. Only peptides of masses in the range 3.5-4.8 kDa were identified confirming that this sample contained deletion and truncated peptides. The major observed products had masses of 4421 Da and 4348 Da. These species correlate to termination of the synthesis at D61 (calculated mass 4421.0 Da) and Y62 (calculated mass 4305 Da) where a decrease in coupling efficiencies was observed.
<table>
<thead>
<tr>
<th>Observed Mass / Da</th>
<th>Calculated difference in mass from MCP-1 / Da</th>
<th>Putative deleted/modified residues</th>
<th>Calculated Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>6844.4</td>
<td>-915</td>
<td>[G(^{25})SDACEEP + E]</td>
<td>6841</td>
</tr>
<tr>
<td>6860.3, 6862.0,</td>
<td>-899</td>
<td>[G(^{25})SDACEEP + C/I/L]</td>
<td>6857 - 6867</td>
</tr>
<tr>
<td>6864.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7041.9</td>
<td>-718</td>
<td>- [G(^{25})SDACE + R]</td>
<td>7041</td>
</tr>
<tr>
<td>7058.7, 7059.4, 7064.5</td>
<td></td>
<td>- [G(^{25})SDACEE]</td>
<td>7068</td>
</tr>
<tr>
<td>7085.2, 7086.1</td>
<td>-674</td>
<td>- [G(^{25})SDACE +I/L]</td>
<td>7084</td>
</tr>
<tr>
<td>7118.0,7118.3, 7126.3</td>
<td></td>
<td>- [G(^{25})SDACE + A]</td>
<td>7126</td>
</tr>
<tr>
<td>7316.0, 7316.2</td>
<td>-443</td>
<td>- [G(^{25})SDAC]</td>
<td>7316</td>
</tr>
<tr>
<td>7699.0</td>
<td>-56 #</td>
<td>MCP-1, all cysteines oxidised</td>
<td>7699</td>
</tr>
<tr>
<td>7755.5</td>
<td>-</td>
<td></td>
<td>7755</td>
</tr>
<tr>
<td>7766.7</td>
<td>+9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7771.4</td>
<td>+16 #</td>
<td>MCP-1 + oxidised Met</td>
<td>7771</td>
</tr>
<tr>
<td>7797.4, 7797.6</td>
<td>+42 #</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7814.7</td>
<td>+59 #</td>
<td>MCP-1 + 'Bu</td>
<td>7812</td>
</tr>
<tr>
<td>7821.3</td>
<td>+66 #</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1 ESMS data for SEC peak [A]. Calculated mass for MCP-1 is 7759 Da (all cysteines reduced), 7755 Da (all cysteines oxidised). Mass differences based on calculated reduced mass unless denoted #.

Despite the presence of contaminating proteins, as indicated by ESMS, the mixture containing protein of the mass expected for synthetic MCP-1 was analysed by 1D \(^1\)H NMR, see Figure 3.10. The spectrum shows evidence for some folded material particularly in the methyl region where resonances have shifted to ~0ppm. However, the majority of the protein is random coil. The unfolded protein could be a consequence of MCP-1 mis-folding under the conditions used for renaturation, or due to deletion or truncated peptides that cannot fold correctly due to missing residues.
Results and Discussion

Table 3.10 1H-NMR spectrum of synthetic MCP-1. The spectrum was measured at 37 °C and 600 MHz. The sample was ~0.2 mM in 0.1 M TrisHCl (pH 8.2), 0.2 mM GSSG/2 mM GSH, 10% D₂O. The sharp resonance at 5.8 ppm is residual urea following renaturation of the sample.

3.2.2.3 Anion exchange chromatography

Anion exchange chromatography (AIEX) was selected as a means of isolating MCP-1 from the peptide mixture by exploiting any differences in net charge between the full length and truncated peptides. The major MCP-1-containing fraction eluted at 0.19-0.25 M NaCl (0.1 M TrisHCl, pH 8.2) and good separation from a number of other protein-containing fractions was achieved, see Figure 3.11. The [¹³C]-label was found to be distributed across the chromatogram although the label was most concentrated in the major protein peak.
Results and Discussion

Figure 3.11 *AEX chromatography on MCP-1*. The protein was eluted from a Mono Q HR5/5 column with a linear gradient of 0-0.5 M NaCl (0.1 M TrisHCl, pH 8.2). The labelled (*) peak was found to contain material of the correct mass for MCP-1 by ESMS.

Isolation of the protein contained within the major peak by RP-HPLC and subsequent analysis by ESMS confirmed the removal of the majority of contaminating proteins fractions, see *Figure 3.12*. The major protein identified had a mass of 7757.3 Da, in good agreement with the calculated mass for oxidised synthetic MCP-1 (7755.7 Da). Of the remaining contaminants one could be identified as MCP-1 with oxidised methionine (7773.5 Da).

Methionine oxidation frequently occurs during acidolytic cleavage of resin from synthetic proteins. Reduction of the methionine sulfoxide was attempted using N-methyl-mercaptoacetamide. Following the reaction, isolation of the product by RP-HPLC proved unsuccessful. A series of peaks were eluted from the column and analysed by ESMS but no material of the expected mass for MCP-1 was identified.
Results and Discussion

3.2.2.4 Further purification of MCP-1

Hydrophobic interaction chromatography (HIC) was selected to further purify those fractions collected from the AEX column that contained MCP-1, as determined by ESMS. HIC was selected to exploit the different hydrophobicity of any unfolded peptide contained within the sample compared to the folded peptide. Application of the major protein fraction obtained from AEX to the hydrophobic matrix resulted in a broad peak (eluting at 0.78-0.8 M ammonium sulphate, 0.1 M TrisHCl, pH 8.2) with a significant shoulder (eluting at 0.8-0.82 M ammonium sulphate, 0.1M TrisHCl, pH 8.2) indicating the presence of several different protein species, see Figure 3.13.

Figure 3.12 ESMS analysis of synthetic MCP-1 purified by AEX chromatography.
Figure 3.13  

HIC purification of MCP-1. A shallow gradient of 0.058-0.062 M TrisHCl (pH 8.2) over 10 column volumes was used to elute the protein from a Phenyl Superose HR5/5 column.

Analysis of the major protein-containing fractions A and B from HIC by ESMS indicated that purification of MCP-1 to homogeneity had not been achieved. Peak A was found to contain material of the calculated mass for MCP-1 (observed mass 7754.3 Da, calculated mass 7755 Da) and another major species of 7853.5 Da, see Figure 3.14. This species had not been observed previously. The mass difference from that calculated for MCP-1 is consistent with that for O-trifluoroacetylation of serine, a common by product of TFA cleavage reactions 191. Another species was potentially identified as MCP-1 missing the N-terminal residues GSDA (observed mass 7423.5 Da, calculated mass 7429.5 Da). Peak B was also found to contain material of the expected mass for MCP-1 and the O-trifluoroacetylated product. Structural analysis of the protein at this stage was not possible due to insufficient material.
Results and Discussion

Figure 3.14 (a) ESMS analysis of HIC peak [A].

Figure 3.14 (b) ESMS analysis of HIC peak [B].
3.3 Chemical Synthesis of fH-15

The 3D solution structure of fH-15 expressed in *Saccharomyces cerevisiae* showed the molecule to be stable, soluble and to exist in a single conformation. The module was therefore considered a valid target for developing CP-module synthesis methodology.

3.3.1 Solid phase synthesis

fH-15 (corresponding to human factor H residues 866-928) was synthesised on a 0.1 mmol scale on Wang resin as described for MCP-1. The resin was functionalised with the C-terminal glycine residue to 0.13 mmol/g resin. All residues were single-coupled with the exception of residues Glu 866-Gln 875 and Phe 905-Thr 913 which were double coupled, underlined in Figure 3.15. Double coupling of these residues was performed as a previous synthesis indicated that the coupling efficiency in these regions was unsatisfactory. On completion of the synthesis the loading efficiency of the N-terminal glutamate residue was determined to be 0.047 mmol/g resin by Fmoc-deprotection, equivalent to a 36% loading efficiency (relative to the C-terminal residue).

![Synthesised sequence of fH-15](image)

Figure 3.15 Synthesised sequence of fH-15. Regions of double coupling are underlined.
The relative coupling efficiencies for the synthesis of fH−15 were determined by monitoring the formation of the Fmoc-piperidine adduct, as described for the synthesis of MCP−1, and are recorded in Figure 3.16.

The synthesis was observed to proceed satisfactorily until Ile 907 when a significant decrease in coupling efficiency was observed, despite the introduction of double coupling cycles in this region. A further decrease in coupling efficiency was observed in the region Ile 881-Gln 888. From these observations it was anticipated that a number of deletion and truncated peptides would be present on the resin. The large decrease in coupling efficiency observed at the end of the synthesis is possibly a consequence of the sampling technique used. The loading of the N-terminal residue was manually determined, whilst all other couplings were automatically determined during the synthesis. As discussed above, (see Section 3.2.1), resin swelling can affect the apparent concentration of the Fmoc-piperidine adduct, resulting in inaccurate coupling efficiency data. The terminal, manually determined coupling efficiency of 36% is likely to represent more accurately the final resin loading.

Figure 3.16 Coupling efficiencies for H15. Couplings were determined from monitoring Fmoc-piperidine adduct formation and are calculated as a percentage of the C-terminal residue loading.
3.3.2 Purification of fH-15

Acidolytic cleavage of the protein from the resin was achieved by a 4 h incubation with the TFA/scavenger mixture described earlier for MCP-1. Following cleavage, the crude protein mixture was isolated by ether precipitation and resuspended in 6 M guanidinium hydrochloride (GuHCl) and 20 mM dithiothreitol (DTT). Analysis by RP-HPLC indicated the presence of a number of closely related protein species eluting over the range 31-44% acetonitrile (0.1% v/v TFA), see Figure 3.17.

As described for MCP-1, SEC was used to remove low molecular weight contaminants from the crude protein mixture. Calibration of the column (Pharmacia HiLoad Superdex G30 16/60) with lysozyme, ubiquitin and insulin in 6 M GuHCl, 20 mM DTT resulted in a calculated elution volume of 50 ml for fH-15. The protein mixture eluted in two major peaks with elution volumes of 46-60 ml (peak [A]) and 60-72 ml (see Figure 3.18).

RP-HPLC analysis of each SEC peak indicated the separation of a number of contaminating peptides. Fractions corresponding to peak [A] were found to elute at 41-44% acetonitrile (0.1% TFA) whilst those from peak [B] had shorter retention times and eluted at 32-42% acetonitrile (0.1% TFA), see Figure 3.19. Analysis of the proteins within peak [A] by MALDI-TOF identified a mixture of proteins of masses in the range 6.8-7 kDa but no protein of the exact mass for synthetic fH-15 was found (data not shown).

Although RP-HPLC analysis of fractions from peak [A] indicated the removal of a number of contaminating peptides from the sample, further purification was required. An attempt was made to refold fH-15 to enable exploitation of the different chromatographic properties
Results and Discussion

**Figure 3.17** Analysis of the crude protein mixture by RP-HPLC. The protein was eluted from a C8 analytical column with a linear gradient of 10-90% acetonitrile (0.1% v/v TFA) over 30 min.

**Figure 3.18** SEC purification of crude protein mixture. Peak [A] has the appropriate elution volume for fH−15.

**Figure 3.19** RP-HPLC analysis of fractions from SEC peak [A] and [B] respectively. The samples were eluted from a C8 column with a linear gradient of 10-90% acetonitrile (0.1% v/v TFA) over 30 min.
of folded and unfolded protein. The protein mixture was dialysed against 0.1 M TrisHCl (pH 8.4), 1 mM EDTA, 3 mM GSH and 0.3 mM GSSG. The proteins were isolated by RP-HPLC and eluted as a broad peak consisting of several overlapping peaks (20-35% acetonitrile, 0.1% v/v TFA), see Figure 3.20. To determine if fH-15 had been renatured under the conditions described, despite the presence of contaminating proteins, various protein containing fractions were collected and lyophilised prior to analysis by 1D 1H-NMR.

The lyophilised protein samples isolated by RP-HPLC were resuspended in 10 mM sodium phosphate (pH 6.0) in D2O. An example of a 1D 1H-NMR spectrum acquired is shown in Figure 3.21. It bears little resemblance to those acquired for recombinant fH15 or synthetic MCP-1 and is not characteristic of a CP-module in a single, folded conformation. Although there are some broad resonances in the region 0-0.8 ppm, which may indicate the partial formation of a hydrophobic core, on the basis of line width and chemical shift dispersion the protein is in a random coil state. This could be a consequence of mis-folding of the fH-15 sequence or indicative of contaminating peptides within the sample that were unable to renature correctly.

ESMS analysis of the samples confirmed the presence of protein of the expected size for fH-15 with both disulphides formed (6965.9 Da). The major contaminants in the sample were fH-15 containing oxidised methionine (6983.0 Da) and a truncated peptide corresponding to termination of the synthesis at Ile 876 (5878 Da), see Figure 3.22.
Results and Discussion

Figure 3.20  *RP-HPLC purification following renaturation.* Fractions were collected from this peak and analysed by NMR. Figure 3.21 shows the NMR spectrum acquired on the peak labelled *.

Figure 3.21  *1D 1H-NMR spectrum of FH~15.* The spectrum was acquired in 10 mM sodium phosphate (pH 6.0) in D$_2$O at 25 °C and 600MHz.
Extensive attempts were made to isolate fH~15 from the protein mixture by anion and cation exchange chromatography and further RP-HPLC analysis, however, no separation was achieved. Although SEC succeeded in removing a number of smaller peptides from the protein mixture, the separation of impurities more similar to fH~15 proved unsuccessful under the conditions explored. Consequently, an attempt was made to derivatise the N-terminus of fH~15 to facilitate its isolation.

### 3.3.3 N-terminal TbFmoc derivatisation of fH~15

In an attempt to facilitate the purification of fH~15 from the numerous closely related protein species, tetrabenzo[a,c,g,i]fluorenyl-17-methoxycarbonyl (TbFmoc) was introduced to the free N-terminus of the resin-bound protected protein. TbFmoc is a large, planar, highly aromatic molecule that acts as a hydrophobic chromatographic probe. The hydrophobicity of the molecule results in the labelled protein species being more strongly retained by RP-HPLC than non-labelled proteins. The labelled species can also be affinity-
purified on porous graphitised carbon (PGC) which adsorbs large, flat aromatic ring systems with high affinity. Additionally the specific UV/vis absorbance of the TbFmoc group at 364 nm allows identification of labelled sequences.

![Figure 3.23](image.png)

Figure 3.23  *(17-tetrabenzo[a,c,g,i]fluorenyl)-methylchloroformate (TbFmocCl)*

Before introduction of TbFmoc onto the resin-bound peptide, the synthesis ‘capping cycle’ was repeated by sonicating the peptidyl resin in acetic anhydride, to ensure all free amino groups were acylated and therefore unable to participate in the TbFmoc coupling reaction. After cleavage of the N-terminal Fmoc-group, the TbFmoc group was introduced onto the free N-terminus by sonication of the peptidyl resin with *(17-tetrabenzo[a,c,g,i]fluorenyl)-methylchloroformate (TbFmocCl)*. This reaction, and where possible all other procedures involving TbFmoc-labelled proteins, was carried out in vessels shielded from light as the TbFmoc group is light sensitive.

After cleavage from the resin and resuspension in 6 M GuHCl and 20 mM DTT, RP-HPLC analysis identified a number of closely related peptide species eluting over the range 60-71% acetonitrile, 0.1% v/v TFA. Due to the nature of the solid phase methodology only one TbFmoc-labelled species was anticipated. However, a number of peptides were found to have been derivatised (eluting at 62, 64, 66, 70% acetonitrile, 0.1% v/v TFA), see Figure 3.24. The possible identity of these species was unclear. It was postulated that
they could be full-length sequences with protecting groups still attached or deletions and truncated sequences that, despite the acylation capping reaction, were derivatised with TbFmoc.

![Graph](image)

**Figure 3.24 Analysis of TbFmoc-fH-15 at 280 and 364 nm.** A number of species were found to absorb at 364 nm suggesting that TbFmoc-derivatisation was non-specific for the target sequence.

Partial purification by SEC resulted in elution of the protein mixture in two peaks (45-60 and 62-70 ml) as described previously, data not shown. With regard to the partial purification of the non-derivatised fH-15 it was anticipated that protein of the size expected for TbFmoc-fH-15 would elute in the first peak. RP-HPLC analysis of the fractions corresponding to this peak confirmed the presence of a number of TbFmoc-labelled proteins. These were isolated by semi-preparative RP-HPLC and lyophilised. Semi-preparative RP-HPLC purification of the major TbFmoc-labelled protein species is shown in **Figure 3.25.**
Results and Discussion

Figure 3.25 *RP-HPLC purification of TbFmoc labelled fH–15*. This corresponds to one fraction collected from SEC purification of the crude protein mixture. Other fractions were found to contain TbFmoc-labelled species with different retention times. The protein was eluted from a C8 semi-preparative column with a linear gradient of 10-90 % acetonitrile over 30 min.

The TbFmoc group was cleaved from the various labelled protein species by incubation with 6 M GuHCl, 0.1 M TrisHCl (pH 8.6), 100 mM DTT at 37 °C for 4 h. Base-catalysed cleavage of TbFmoc occurs by β-elimination in an analogous manner to the cleavage of the Fmoc group by piperidine. After removal of free TbFmoc by gel filtration the denaturing and reducing agents were removed by dialysis of the protein solution against 0.1 M TrisHCl (pH 8.6). The isolated protein species were further purified by semi-preparative RP-HPLC. The retention times of the proteins were found to decrease after removal of the hydrophobic TbFmoc group (from 50-54 to 46-50 % acetonitrile, 0.1% v/v TFA), see Figure 3.26.

Analysis by MALDI-TOF MS and ESMS failed to identify any protein of the expected molecular mass for fH–15. However, N-terminal sequencing of the protein (EKIPCSQPQPQ)
was consistent with that for fH~15. Analysis of the resulting protein by 1D $^1$H-NMR provided no evidence of the protein being folded, see Figure 3.27. Further attempts to renature the module by dialysis against buffers containing oxidised and reduced glutathione proved unsuccessful, as monitored by NMR.

Figure 3.26 Analysis of the isolated protein after cleavage of the TbFmoc group. Although semi-prep RP-HPLC resulted in the isolation of one major TbFmoc-labelled species, cleavage of the TbFmoc group resulted in a mixture of proteins that were not resolved by RP-HPLC.

The attempted purification of fH~15 indicated that the crude protein mixture, when cleaved from the resin, did contain protein of the expected size and with the appropriate N-terminal sequence for fH~15. However, mass spectrometry indicated the presence of numerous truncated and deletion sequences. This was illustrated by SEC of the mixture of proteins cleaved from the resin which showed a broad molecular weight range. The isolation of fH~15 proved difficult due to the similar properties of the target and truncated sequences. In an attempt to increase the yield of fH~15 the synthesis was modified and repeated.
Results and Discussion

3.3.4 Modified synthesis of fH-15

Despite the introduction of double coupling cycles, the initial synthesis of fH-15 showed a significant decrease in coupling efficiency at Ile 907 and in the region Ile 881-Gln 888. Mass spectrometry identified a number of peptides with masses in the region 4.5-5.5 kDa that correspond to the synthesis having terminated in the region Gly 879-Gln 888.

To increase the proportion of full-length fH-15 the synthesis was modified to double couple the regions Glu 912-Ser 908 and Ser 890-Glu 866 and to triple couple residues Ile 907-Phe905. Glycine 879 was not double coupled as this can result in the formation of glycylglycine. The synthesis was carried out on a 0.1 mmol scale as described previously.
The resin loading efficiency of the C-terminal residue was 0.255 mmol/g resin and for the N-terminal residue it was 0.07 mmol/g, equivalent to a loading efficiency of 27% relative to the C-terminal residue loading. Coupling efficiency data are recorded in Figure 3.28.

Figure 3.28 Coupling efficiency data for fH−15. The sharp decrease in coupling efficiency at Gly 918 was due to an error in the automated sampling procedure and is not believed to accurately represent the coupling efficiency of that step.

To determine the success of the modified synthesis a small portion of the resin was cleaved. Since RP-HPLC analysis indicated that, as with the previous synthesis, the crude mixture consisted of a number of closely related protein species (see Figure 3.29), N-terminal derivatisation was attempted to isolate fH−15.

3.3.4.1 N-terminal derivatisation of fH−15

To assist in the purification of fH−15 three methods of N-terminal derivatisation were used. Two methods, the introduction of biotin and TbFmoc, altered the chromatographic properties of the labelled sequence whilst the third, radiolabelling, was used to facilitate
monitoring of the target sequence throughout the purification.

Figure 3.29 *RP-HPLC analysis of crude protein mixture.* The protein eluted from a C8 reverse phase column at 46-54% acetonitrile (0.1% v/v TFA).

(a) **Biotinylation of fH-15**

The specific interaction between biotin and avidin has been utilised for the purification of synthetic peptides \(^{184,195}\). Avidin has a very high (\(K_d \approx 10^{-15} \text{ M}\)) affinity for biotin and the introduction of a small molecule such as biotin to a polypeptide should not have a significant effect on conformation or activity. The biotin labelled polypeptide can be isolated from the crude peptide mixture by avidin affinity chromatography.

There are several reported methods for the introduction of biotin to the free N-terminus of the resin-bound polypeptide. An N-hydroxysuccinimide derivative of biotin was prepared \(^{161}\) and introduced onto the terminal amino group after Fmoc deprotection. The
protein was cleaved from the resin and isolated following the standard protocol. Following resuspension in 0.15 M NaCl and 0.1 M TrisHCl (pH 7.8), RP-HPLC analysis of the crude protein mixture showed the presence of a number of species eluting at 42-54% acetonitrile (0.1% v/v TFA), see Figure 3.30. By comparison to Figure 3.29, it can been seen that N-terminal biotinylation has little chromatographic effect.

The protein mixture was applied to an avidin-agarose column and non-specifically adsorbed material removed by washing the column with 0.15 M NaCl and 0.1 M TrisHCl (pH 7.8). Biotinylated-proteins were eluted with 8 M GuHCl and 0.1 M sodium citrate (pH 1.6). Subsequent analysis by RP-HPLC showed a single but broad peak (49-53% acetonitrile, 0.1% v/v TFA), see Figure 3.31.

Figure 3.30 RP-HPLC analysis of crude biotinylated H15. The protein was eluted from a C8 reverse with a linear gradient of 10-90% acetonitrile (0.1% v/v TFA) over 30 min.

Figure 3.31 RP-HPLC analysis of biotinylated-fH-15 purified on avidin-agarose column. The protein was eluted from a C8 reverse with a linear gradient of 10-90% acetonitrile (0.1% v/v TFA) over 30 min.
Although RP-HPLC analysis indicated that some purification was achieved by the method, no data was obtained when the material was analysed by MALDI-TOF mass spectrometry and there was insufficient material for other methods of characterisation.

The method was not found to offer any advantages over other methods of N-terminal derivatisation. In addition, avidin/biotin affinity chromatography is prohibitively expensive and low-yielding for the production of proteins in sufficient quantities for structural analysis by NMR. Typical binding capacities of avidin columns are 25-50μg biotin/ml avidin-agarose resin making large scale use of the technique unfeasible. Consequently, an attempt was made to purify fH~15 using the TbFmoc strategy described earlier.

(b) N-terminal TbFmoc derivatisation

TbFmoc-labelling of the protein produced in the modified synthesis was carried out as described previously. Analysis of the cleaved protein mixture by RP-HPLC showed that the TbFmoc group afforded separation of a group of derivatised peptides from non-labelled material. However, as was observed with the first synthesis of fH~15 a number of peaks were found to have been labelled. This indicates that there continued to be difficulties with the synthesis resulting in incomplete cycles and the failure to cap terminated sequences.

Partial purification of the crude protein mixture by SEC resulted in elution of the peptide mixture in two peaks, with elution volumes 42-60 and 62-70 ml, as previously described. The proportion of the two peaks relative to each other remained unchanged from the first attempt at synthesising fH~15, with material having the expected elution volume for fH~15 comprising ~40% of the total crude protein mixture (data not shown). This suggested that
modifications to the synthesis had not resulted in improving the ratio of full-length material to deletion/truncated peptides. RP-HPLC analysis of the material collected after SEC showed that the peak with the expected elution volume for TbFmoc-fH~15 contained one major TbFmoc-labelled component (which eluted at 65-70% acetonitrile, 0.1% v/v TFA), see Figure 3.32.

Following base-mediated removal of the TbFmoc group, analysis by RP-HPLC resulted in a broad peak eluting at 48-52% acetonitrile (0.1% v/v TFA). After lyophilisation, the sample was fully denatured and reduced by incubation with 6 M GuHCl, 0.1 M TrisHCl (pH 8.6) and 0.3 M βME (37 °C, 4 h). The reducing and denaturing agents were removed by dialysis against 0.1 M TrisHCl, (pH 8.2) 3 mM GSH, 0.3 mM GSSG and 1 mM EDTA. RP-HPLC analysis of the protein after this treatment resulted in a slight shift in retention time from 48-52 to 46-51% acetonitrile (0.1% v/v TFA), see Figure 3.33.

![Figure 3.32 (a)](image1)

![Figure 3.32 (b)](image2)

Figure 3.32 *RP-HPLC analysis of TbFmoc-fH~15 partially purified by SEC monitored at (a) 280 nm and (b) 364 nm. The protein was eluted from a C8 analytical column. TbFmoc-labelled species eluted in one peak as identified by absorbance at 364 nm.*
Results and Discussion

Figure 3.33 RP-HPLC analysis of fH-15 (a) following TbFmoc cleavage and (b) after renaturation.

It was anticipated that, if the protein had refolded, a change in retention time would be observed due to the change in protein conformation. However, no change in the peak shape was observed and the slight change in retention time suggested that the protein had not been renatured under the conditions used.

In order to characterise the partially purified protein, the sample was hydrolysed in 6 M HCl (110 °C, 42 h) and amino acid analysis performed. The composition of the sample was in reasonable agreement with that expected for synthesised fH-15 (see Table 3.2). The major discrepancies observed are for cysteine, proline, serine and tyrosine. However, this technique does not provide any information on the composition of the impurities present in the sample and consequently amino acid analysis was not used when mass spectrometry data was available.

Analysis by 1D 1H-NMR suggested that the protein was not in a single, stable folded conformation, see Figure 3.34. Efforts to characterise the sample by MALDI-TOF MS,
ESMS and liquid chromatography ESMS failed to identify any material of the expected molecular weight for fH~15.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Calculated composition (%)</th>
<th>Actual composition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>1.6</td>
<td>1.88</td>
</tr>
<tr>
<td>Arg</td>
<td>3.2</td>
<td>2.72</td>
</tr>
<tr>
<td>Asx</td>
<td>3.2</td>
<td>2.63</td>
</tr>
<tr>
<td>Cys</td>
<td>6.3</td>
<td>2.01</td>
</tr>
<tr>
<td>Gix</td>
<td>19.0</td>
<td>16.00</td>
</tr>
<tr>
<td>Gly</td>
<td>9.5</td>
<td>8.48</td>
</tr>
<tr>
<td>His</td>
<td>3.2</td>
<td>3.70</td>
</tr>
<tr>
<td>Ile</td>
<td>6.3</td>
<td>4.79</td>
</tr>
<tr>
<td>Leu</td>
<td>1.6</td>
<td>1.95</td>
</tr>
<tr>
<td>Lys</td>
<td>4.8</td>
<td>4.22</td>
</tr>
<tr>
<td>Met</td>
<td>1.6</td>
<td>1.39</td>
</tr>
<tr>
<td>Phe</td>
<td>1.6</td>
<td>2.06</td>
</tr>
<tr>
<td>Pro</td>
<td>7.9</td>
<td>15.62</td>
</tr>
<tr>
<td>Ser</td>
<td>15.9</td>
<td>10.90</td>
</tr>
<tr>
<td>Thr</td>
<td>7.9</td>
<td>6.18</td>
</tr>
<tr>
<td>Trp</td>
<td>1.6</td>
<td>-</td>
</tr>
<tr>
<td>Tyr</td>
<td>4.8</td>
<td>1.19</td>
</tr>
<tr>
<td>Val</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 3.2  Amino acid analysis of fH~15 purified using TbFmoc as a chromatographic probe

Figure 3.34  1D 1H-NMR analysis of protein purified using TbFmoc as a chromatographic probe.
(c) **N-terminal derivatisation with $[^{14}\text{C}]-\text{glycine}$**

As described for MCP-1, radiolabelling of fH-15 was performed to facilitate isolation of the target sequence. The $[^{14}\text{C}]-\text{glycine}$ labelled protein was subsequently derivatised with TbFmoc prior to cleavage from the resin. Following resuspension in 6 M GuHCl, 40 mM βME it was found that 17% of the total crude protein was derivatised with $[^{14}\text{C}]-\text{glycine}$. Partial purification by semi-preparative RP-HPLC resulted in a single peak for TbFmoc -labelled protein (as determined by absorbance at 364nm) eluting at 60-65% acetonitrile, 0.1% v/v TFA, see *Figure 3.35*.

![Figure 3.35 (a)](image1)

![Figure 3.35 (b)](image2)

*Figure 3.35* **RP-HPLC of TbFmoc-$[^{14}\text{C}]\text{Gly}-f\text{H-}15$**.

(a) The crude protein mixture was eluted from a C8 analytical column and the TbFmoc-labelled protein (denoted *) was found to be separated from the non-labelled proteins.

(b) The TbFmoc-labelled protein was isolated by semi-preparative RP-HPLC on a C8 column.

Cleavage of the N-terminal TbFmoc group was performed as previously described. Subsequent RP-HPLC analysis of the peptide resulted in a broad chromatogram with elution over the range 53-59% acetonitrile (0.1% v/v TFA), see *Figure 3.36*.
Results and Discussion

Figure 3.36 RP-HPLC analysis of partially purified protein. The protein was eluted from a C18 column with a linear gradient of 10-90% acetonitrile.

Anion exchange chromatography was used in an attempt to purify fH-15 from any remaining deletion peptides. Application of the protein mixture to a MonoQ HR5/5 anion exchange column resulted in elution of the major protein-containing fraction at 0.12-0.19 M NaCl (50 mM TrisHCl, pH 7.1), see Figure 3.34. Separation from a number of other protein-containing fractions was achieved. The [14C]-label was found to be distributed over the range of proteins separated. No material of the expected mass for fH-15 was identified by ESMS.

Figure 3.37 Attempted purification of fH-15 by AEX.
Despite the use of various purification techniques it was not possible to purify synthetic fH~15 to homogeneity. Renaturation of the protein proved problematic, and difficulties in obtaining mass spectrometry data on the sample limited monitoring of the purification. Extensive efforts were made to characterise the sample by ESMS. It was anticipated that derivatisation of the cysteine residues with iodoacetamide would prevent any cross-linking of the peptide species through disulphide bonds, enabling mass spectrometry data to be more readily acquired. However, this failed to resolve the difficulties encountered.

3.4 Discussion

Despite the reported successes of solid phase synthesis for the production of proteins in the literature, in this study the synthesis of the CP-modules MCP~1 and fH~15 proved highly problematic. The multi-step purification strategy resulted in isolation of MCP~1 that was approximately 95% pure as determined by ESMS. The low yield of purified protein prevented any further structural characterisation, although initial analysis by 1D ¹H-NMR spectroscopy indicated that the renaturation protocol was partially successful. The application of the refolding protocol to recombinant MCP~1 recovered from inclusion bodies in a denatured state is discussed in Chapter Four.

It was not possible to isolate fH~15 in a pure form despite the utilisation of a number of chromatographic techniques. Renaturation of the protein proved problematic, and difficulties in obtaining mass spectrometry data on the sample limited monitoring of the purification. Extensive efforts were made to analyse the sample by ESMS using a variety of
conditions, however, the difficulties were not resolved. It appears that similar difficulties were encountered for both proteins. In each case the synthesis was low in yield, as indicated by the monitoring of coupling efficiency data, radiolabelling and the use of mass spectrometry.

It is recognised that the formation of the target sequence in high yield is necessary for successful purification. This is frequently the case for small proteins. For example, in the synthesis of the 31-residue neuropeptide hormone β-endorphin the target sequence comprised 84% of the crude product after cleavage from the resin. In the syntheses of the CP-modules described here the target sequences comprised a minor proportion of the total crude product, as illustrated by SEC. The nature of solid phase synthesis results in the by-products being chemically and physically similar to the target sequence, since it is possible that they differ by only one internal residue. Therefore, despite exploiting various chromatographic properties of the protein, purification to homogeneity, of a minor component of the crude protein mixture, in good yield may not be a realistic proposition.

Incomplete peptide bond formation during the synthesis results in the accumulation of deletion and truncated peptides on the resin and a highly heterogeneous crude product. Various studies of ‘difficult sequences’ have been implemented to determine the cause of incomplete peptide bond formation. Difficult couplings have been divided into ‘random’, due to the slow reactivity of certain residues, particularly β-branched amino acids, and ‘non-random’ which are caused by aggregation of the nascent peptide chains due to the formation of β-sheet secondary structure.
Results and Discussion

One method for predicting difficult sequences was described by considering the propensity of each amino acid to contribute to aggregation. Aggregation potentials, \( <P_a> \), were determined by synthesising 87 peptides unrelated in sequence and monitoring the decrease in the volume of the swollen resin. The proposed analysis of these aggregation profiles suggests that a sequence becomes potentially difficult to synthesise after the values of \( <P_a> \) exceed 1.1 for two or three consecutive residues. The calculated aggregation profiles for MCP-1 and fH-15 are illustrated in Figure 3.38.

Within the sequences synthesised here there are a number of pairs of consecutive residues with \( <P_a> \) values greater than 1.1 suggesting that couplings may be difficult. The 10 residues at the carboxyl terminus have been shown to dominate the character of the synthesis. For MCP-1 and fH-15 although consecutive residues do not have \( <P_a> \) values greater than 1.1, three residues within the C-terminal decapeptides show a propensity to aggregate. According to this analysis, both sequences would be expected to have a tendency to aggregate. However, the physical characteristics of on-resin aggregation were not observed. Aggregation results in a decrease in the swelling of the resin, however in both these cases the resin was found to swell excessively. In this case, it is therefore possible that aggregation is not the cause of the low yielding syntheses.
Figure 3.38 (a) Aggregation potentials for synthetic MCP-1. Two or three consecutive residues with $\langle P_a \rangle$ values $>1.1$ suggest that the sequence has a propensity to aggregate.

Figure 3.38 (b) Aggregation potentials for synthetic fH-15
Results and Discussion

An alternative method for predicting difficult sequences considers the tendency of each residue to adopt a random coil conformation when coupled to the nascent peptide chain rather than forming β-sheet or α-helix. The formation of secondary structure elements is a major cause of aggregation and incomplete coupling reactions. The calculated coil formation parameters, $<P_c>$, for MCP-1 and fH-15 are plotted in Figure 3.39. Difficult coupling reactions are associated with $<P_c>$ values less than 0.9. It is reported that near-quantitative incorporation of the subsequent residue occurs if the $<P_c>$ value for that residue is greater than 1.00.

In the case of MCP-1, 24 residues have $<P_c>$ values less than 0.9, equivalent to 40% of the total sequence. This suggests that the synthesis may prove difficult due to the formation of secondary structure elements reducing the accessibility of the N-terminus to the incoming residue. Although the number of residues within fH-15 that have $<P_c>$ values less than 0.9 is slightly lower (18 residues, 29% of the total sequence) the overall profile suggests that this sequence may also prove difficult.

From a consideration of both studies, the residues most likely to result in poor couplings are Ala, Glu, Phe, Ile, Lys, Leu, Met, Gln, Val and Trp. In the case of the MCP-1 synthesis, according to coupling efficiency data, the region Asp-Tyr-Lys-Cys-Lys-Lys appeared problematic despite the introduction of double coupling cycles. Similarly, in the synthesis of fH-15 the region Ile-Asn-Ser-Ser-Arg-Ser-Ser-Gln continued to prove problematic despite the introduction of triple coupling cycles. Each of these difficult regions contains some of the residues implicated in difficult couplings. However, since ten of the twenty naturally occurring amino acids are described as ‘difficult’ by these studies, it is highly probable that most sequences synthesised contain some, or all, of these residues. Therefore,
Figure 3.39 (a) *Coil formation parameters for synthetic MCP-1*. Difficult couplings are associated with $<P_a>$ values < 0.9.

Figure 3.39 (b) *Coil formation parameters for synthetic fH-15*
the difficulties experienced with the syntheses of MCP~1 and fH~15 cannot be attributed solely to the presence of these so-called difficult residues.

The proteins synthesised in this study are on the boundaries of the size considered 'routine' for SPPS. In theory the yield of each coupling reaction is 99.9%, therefore the stepwise addition of each residue results in a theoretical yield of 94% for the synthesis MCP~1 and fH~15. In practice, this was clearly not achieved, since the yield of free N-terminus in each case was ~20% (as determined by coupling of [14C]-glycine) and even this material did not comprise solely of the target sequence. Although N-terminal derivatisation of the target sequence should facilitate its isolation from a complex protein mixture, the method relies on efficient capping of truncated sequences. The formation of a number of derivatives observed for both MCP~1 and fH~15 reduced the efficacy of this strategy. With reference to the calculated <Pc> values, this could be a consequence of secondary structure formation inhibiting coupling and subsequent capping reactions.

The syntheses described here indicate that chemical synthesis does not provide a suitable method for the routine production of CP-modules in sufficient yield for structural elucidation by NMR. Analysis of the two sequences has shown them both to conform to some of the criteria that describe 'difficult sequences'. However, any proteins of this length would be expected to contain these difficult residues. Although the synthesis of MCP~1 was not optimised, the overall yield of the target sequence would have to be greatly improved before sufficient protein could be synthesised for structural analysis by NMR or X-ray diffraction. Even if this was possible, the synthesis of each CP-module would require extensive optimisation which would not enable the method to be used routinely. A possible
solution would be to chemically ligate shorter peptides, which are more reliably synthesised in good yield, for the production of proteins of this size \(^{149-151}\).
4.1 Expression of MCP-1 in *E. coli*

The work in this section describes the expression of a single CP-module, the N-terminal module of membrane co-factor protein (MCP-1), in the heterologous host *Escherichia coli*. Although a vast number of genes from different sources have been efficiently expressed in *E. coli*, the expression of CP-modules in this organism has proved difficult. One successful attempt involved the expression of the three N-terminal CP-modules from CR1 (CR1-1-3). The protein was deposited in inclusion bodies, which consist of dense particles of aggregated protein complexed with RNA. After solubilisation and renaturation, active CR1-1-3 was isolated and shown to be capable of inhibiting complement mediated lysis of red blood cells.

The expression of disulphide bond-containing proteins, such as CP-modules, in *E. coli* is notoriously problematic. As a prokaryote, *E. coli* is not equipped to introduce post-translational modifications typical of eukaryotic proteins such as glycosylation. Expression of disulphide bond-containing proteins in the bacterial cytosol often results in mis-folding and aggregation as the local environment does not provide the appropriate redox conditions to correctly form the disulphides.
The deposition of MCP-1 in inclusion bodies observed in the current study, despite attempts to express the protein in a soluble form, necessitated the development of an *in vitro* refolding protocol. This study of recombinant protein therefore provides an instructive comparison with work on chemically synthesised CP-modules which are presumably produced in a fully denatured state.

### 4.1.1 Expression in *E.coli* JM109

The use of gene fusion vectors is often reported to enhance the solubility of the heterologous protein as the fusion partners are generally highly soluble and hydrophilic. Examples of fusion partners include glutathione-S-transferase\(^{200}\) maltose-binding protein\(^{201}\) and Protein A\(^{202}\). Moreover, the fusion partner assists in the identification and purification of soluble fusion protein by providing a chromatographic ‘handle’.

The DNA encoding MCP-1 (comprising residues 31-98 of the native human MCP sequence) was constructed in the gene fusion vector pGEX-6P-1, wherein the protein is expressed as a fusion protein with the amino-terminal partner glutathione-S-transferase (GST) from *Schistosoma japonicum*. *E. coli* JM109 harbouring the expression plasmid were grown at 37 °C and expression of GST-MCP-1 induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to 0.5 mM for 4 hours. Analysis of cell lysates by sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE) showed over-expression of a protein of the expected size for GST-MCP-1 (~34 kDa) which was absent from non-induced cells, see *Figure 4.1*. The cells were lysed by sonication and Triton X-100 (1% v/v) added to aid solubilisation of the fusion protein. After removal of insoluble proteins and
Results and Discussion

Figure 4.1 Expression of GST-MCP-1 in E.coli JM109. [A] uninduced cells; [B] over-expression of GST-MCP-1 after induction with IPTG (0.5 mM) for 4 h.

Figure 4.2 Purification of MCP-1. [A] Cell-free extract; [B] GST-MCP-1 eluted from glutathione sepharose column; [C], [D] MCP-1 following enzymatic cleavage of GST fusion partner; [E] GST
other debris by centrifugation, soluble GST-MCP−1 was partially purified from the cell-free extract by affinity chromatography on a glutathione sepharose column at room temperature. GST-MCP−1 was isolated by competitive elution with reduced glutathione. Analysis of the purified fusion protein by SDS-PAGE indicated that the protein was of the expected molecular weight of approximately 34 kDa and that purity was ~90%, see Figure 4.2.

Cleavage of the GST tag from the purified fusion protein was achieved by incubation at 4 °C with the commercially available ‘PreScission Protease’ enzyme (Pharmacia), a fusion of GST and human rhinovirus 3C protease. The cleavage reaction was monitored by SDS-PAGE and it was found that an incubation of 16 hours with the enzyme resulted in two products of the expected sizes for GST and MCP−1 (~26 kDa and 8.5 kDa respectively). A band of lower molecular weight (<6 kDa) was also observed. The reaction mixture did not appear to contain uncleaved GST-MCP−1 fusion protein. GST and Precission Protease were removed by applying the cleavage mixture to a glutathione sepharose column. Further purification of MCP−1 by RP-HPLC resulted in minor and major protein species eluting at 41-43 and 43-49 % acetonitrile (0.1% v/v TFA) respectively, see Figure 4.3. N-terminal sequencing (GPLGSP) of each protein species confirmed cleavage of GST at the expected position. ESMS yielded a mass of 8584.6Da for the major component which corresponds to the calculated mass of the expressed protein with both disulphides formed (8584.7Da), see Figure 4.4. The minor component had a mass of 4722Da and was found to have the expected N-terminal sequence for MCP−1. It was postulated that this species corresponds to cleavage of the C-terminal residues Lys 66 - Thr 98 from MCP−1, see Figure 4.5. A 1D 1H-NMR spectrum of the intact protein was recorded which confirmed, on the basis of chemical shift dispersion and narrow line widths, that the purified protein was in a folded conformation, see Figure 4.6.
Results and Discussion

Figure 4.3  *RP-HPLC purification of soluble MCP-1 expressed in E. coli JM109*. The protein was eluted from a C4 reverse phase column with a gradient of 10-50% acetonitrile (0.1% v/v TFA) over 30 mins.

Figure 4.4  *ESMS analysis of the protein expressed in E. coli JM109 following enzymatic cleavage of GST*. The major product has the calculated mass for MCP-1 expressed in *E. coli*. 
Results and Discussion

Figure 4.5 Sequence for MCP–1 expressed in E. coli. The N-terminal residues GPLGSPE are residual from cloning and cleavage of the GST tag. The proposed site of cleavage, resulting in protein of mass 4722 Da, is indicated [1].

Figure 4.6 1D $^1$H-NMR spectrum of MCP–1 expressed in E. coli JM109. The sample was 0.06 mM in 20 mM potassium phosphate pH 6.0 and the spectrum recorded at 37 °C and 600MHz.

Although the introduction of GST had facilitated the purification of GST-MCP–1 isolated from the sonicated E. coli cells, the majority of the fusion protein was found to remain insoluble. In addition, a proportion of the soluble protein was degraded. Since solubilisation and renaturation of protein deposited in inclusion bodies often results in a loss of yield due to aggregation and/or mis-folding, production of the fusion protein in a soluble
form may be advantageous. To increase the yield of soluble GST-MCP-1 several strategies were utilised that have been reported to minimise the formation of inclusion bodies.

4.1.2 Minimising formation of inclusion bodies

In addition to the expression of gene fusions, alternative methods for minimising the formation of inclusion bodies include: decreasing the rate of protein synthesis, the use of alternative strains of *E. coli* and the co-expression of molecular chaperones. In this study, various *E. coli* hosts were sampled, and efforts were made to slow the rate of heterologous gene expression in order to increase the proportion of GST-MCP-1 synthesised that was soluble.

4.1.2.1 Expression under sub-optimal conditions

It has been reported that as the rate of protein synthesis decreases, the relative yield of native protein increases. Additionally, the yield of native protein is dependent on the competition between folding and aggregation. Reduction of the cultivation temperature has also been reported to increase the proportion of expressed protein that has a native conformation although a decrease in total yield of protein is observed. Therefore, the cultivation of *E. coli* JM109 was undertaken at the sub-optimal temperature of 25°C. Expression of GST-MCP-1 was induced with various concentrations (50, 150 and 250 µM) of IPTG.
Results and Discussion

Analysis of the culture by SDS-PAGE over a period of 24 hours indicated a reduction in the total protein expressed under these conditions compared to the growth at 37 °C using higher levels of IPTG. The yield of soluble fusion protein was estimated to be <100 μg/l compared to approximately 0.25-0.5 mg/l when expressed under the conditions described above (data not shown). Therefore, although useful quantities of soluble protein could be produced in this strain, yields are low from the point of view of structural studies and this was not pursued further.

4.1.2.2 Employment of alternative strains of E. coli

An attempt to increase the yield of soluble fusion protein was made by expressing GST-MCP-1 in different strains of E. coli under the sub-optimal growth conditions purported to increase the relative yield of expressed protein that has a native conformation. Despite the close relationship between different strains of E. coli the presence of different gene products can result in increased solubility of recombinant proteins. To explore this further, the expression plasmid was transformed into E. coli JA221 and E. coli BL21. E. coli JA221 has previously been used successfully to express soluble proteins when grown under sub-optimal conditions. The E. coli strain BL21 lacks cytoplasmic protease gene products and its use should, therefore, minimise the degradation of heterologous proteins expressed in the bacterial cytoplasm.

Expression of GST-MCP-1 in JA221 by induction with 0.1 mM IPTG for 3 hours at 37 °C was performed. The GST-MCP-1 fusion was purified as described previously. The proportion of fusion protein that was soluble did not increase relative to expression carried out in E. coli JM109. Following isolation of the fusion protein and enzymatic cleavage of
the GST fusion partner, the protein was purified by RP-HPLC, see Figure 4.7. A number of protein species eluting over the range 33-48% acetonitrile (0.1% v/v TFA) were observed which were subsequently analysed by ESMS, Figure 4.8. Protein of the calculated mass for MCP-1 was not detected. It appeared that the soluble protein produced was proteolysed. The major purified proteins had masses of 5074.0 Da [1] and 6089.6 Da [2] corresponding to C-terminal proteolysis of MCP-1 at Tyr 69 (calculated mass 5070.74 Da) and His 77 (calculated mass 6110.97 Da), see Figure 4.9.

![Figure 4.7 RP-HPLC analysis of proteins expressed in E. coli JA221 following enzymatic cleavage of GST fusion partner.](image)

![Figure 4.8 ESMS analysis of proteins expressed in E. coli JA221.](image)
Expression in *E. coli* BL21 was carried out under sub-optimal conditions with the expectation of enhancing the solubility of the fusion protein and producing full-length MCP-1. The culture was incubated at the reduced temperature of 28 °C and expression of GST-MCP-1 was induced by the addition of IPTG to 0.1 mM. Furthermore, in an attempt to maximise the yield of soluble fusion protein, the culture was induced at a lower cell density (A_{600} ≈ 0.5) and for a shorter period of time (2 hours) relative to expression in *E. coli* JM109.

As anticipated, expression under these conditions resulted in a reduction of the total yield of target protein synthesised. SDS-PAGE analysis of the soluble and insoluble fractions after sonication indicated that, under these conditions, the fusion protein was expressed in a soluble form and was not deposited in inclusion bodies, see *Figure 4.10*. Purification of the fusion protein and subsequent enzymatic removal of the GST fusion partner were performed as described previously. Further purification by RP-HPLC resulted in a number of poorly resolved species eluting over the range 38-44% acetonitrile (0.1% v/v TFA) (data not shown). The retention times of these species did not correspond to those previously observed for intact MCP-1. ESMS confirmed that no intact MCP-1 was present in the mixture. The major species had masses of 5071.11 [1], 5681.9 [2] and 5869.76Da [3]. Species [1] and [3] correspond to those observed when the fusion protein was expressed in

\[\text{Figure 4.9 Putative sequences of proteins expressed in *E. coli* JA221. The proposed sites of cleavage of MCP-1, resulting in proteins of masses 5070.74 Da and 6110.97 Da, are indicated [1], [2] respectively.}\]
E. coli JA221. Species [2] corresponds, using the same analysis as previously described, to cleavage at Pro 73 (calculated mass 5688.48 Da).

Since the fusion protein expressed in E. coli BL21 appeared by SDS-PAGE to be of the same size as the fusion protein expressed in E. coli JM109, (see Figure 4.11), attempts were made to inhibit adventitious proteolysis during cell disruption, and the subsequent protein purification procedure. Purification of the protein was carried out in the presence and absence of the protease inhibitor phenylmethylsulphonyl fluoride (PMSF) and the products analysed by ESMS. Expression of the fusion protein was also carried at 25 °C, (c.f. 28 °C) and the purification performed in the presence and absence of PMSF to determine if a reduction in the temperature of cultivation improved the stability of the protein. ESMS analysis of the purified protein indicated that under these conditions no material of the expected mass was present and that the addition of this protease inhibitor did not prevent significant degradation of the protein.

![Figure 4.10](image)

Results and Discussion

A B C U
- - - 1 •

Gst-fusion protein

Figure 4.11 Comparison of expression in E. coli JM109 and E. coli BL21. [A] Uninduced E. coli JM109; [B] Expression of GST-fusion protein induced in E. coli JM109; [C] Uninduced E. coli BL21; [D] Expression of GST-fusion protein induced in E. coli BL21.

Hence, under the conditions described, expression of MCP-1 as a soluble protein in high yield has proved unsuccessful. Although it was possible to minimise the formation of inclusion bodies, the soluble protein proved susceptible to degradation by host proteases, despite the use of a protease deficient E. coli strain. It was therefore concluded that a strategy for the solubilisation and renaturation of the protein deposited in inclusion bodies should be developed.

4.1.3 Solubilisation and renaturation of inclusion bodies

The formation of inclusion bodies is not an occurrence exclusive to expression of recombinant proteins in E. coli. Inclusion body formation has been observed when endogenous proteins are expressed at high levels. Providing renaturation is not problematic the formation of inclusion bodies can prove advantageous. Inclusion bodies
Results and Discussion

primarily consist of the over-expressed protein often resulting in facile purification. Additionally the aggregated protein is less susceptible to proteolysis than soluble proteins. The deposition of GST-MCP-1 in inclusion bodies has a further advantage, from the viewpoint of this project, in terms of optimising strategies for the production of CP-modules. The ability to refold MCP-1 expressed in inclusion bodies has ramifications for the challenge of refolding proteins produced by chemical synthesis.

The deposition of GST-MCP-1 in inclusion bodies was encouraged by expressing the protein in *E. coli* JM109, as described previously, with the concentration of inducing agent (IPTG) increased to 1 mM. After lysis and removal of the cell-free extract the insoluble proteins were incubated with Triton-X100 (0.5 %) to solubilise the cell membrane and lipids. Following removal of this cell debris, the inclusion bodies were solubilised by the addition of 8 M urea, 1xPBS, 2 mM reduced glutathione (GSH) and 0.2 mM oxidised glutathione (GSSG) to denature the protein and disrupt disulphide bonds. Centrifugation of the resulting solution yielded soluble protein of the approximate mass of GST-MCP-1, as judged by SDS-PAGE, in reasonable yield (2 mg/l).

Refolding of cysteine-containing proteins can be highly problematic and frequently results in loss of protein. As with *in vivo* folding, the formation of the correct fold is in competition with protein aggregation. The presence of contaminating proteins which aggregate under the refolding conditions can cause additional aggregation of the protein of interest. Therefore it is often necessary to perform a purification step prior to renaturation. In the case of GST-MCP-1, SDS-PAGE analysis indicated the presence of few contaminating proteins permitting the instigation of a refolding protocol without prior purification.
A glutathione redox system was utilised to create an oxidising environment for the formation of disulphide bonds \(^{212}\). These systems facilitate disulphide bond shuffling until the correct tertiary structure is obtained \(^{213}\). Renaturation was achieved by dialysis of the denatured and reduced protein solution against 100 mM TrisHCl (pH 8.2), 3 mM GSH, 0.3 mM GSSG, 1 mM EDTA. The addition of EDTA to the renaturation buffer prevents metal-catalysed air oxidation of the protein \(^{214}\). It was observed that the protein concentration needed to be <100 μg/ml to avoid aggregation.

Following renaturation, an attempt was made to purify the fusion protein from minor contaminating proteins by exploiting the affinity of GST for glutathione. However, GST did not bind to the column indicating that the refolding conditions did not enable GST to refold correctly. Despite this, proteolytic cleavage of MCP-1 from the fusion partner was possible. Therefore no attempt was made to optimise conditions for the renaturation of GST. Instead, cleavage of GST was carried out and MCP-1 purified. In this way, if MCP-1 was also incorrectly folded, a refolding protocol specific to MCP-1 could be developed.

After cleavage of the GST fusion partner, isolation of MCP-1 was carried out by size exclusion chromatography and RP-HPLC. Size exclusion chromatography indicated that the isolated protein was of the expected mass for monomeric MCP-1. The renatured protein showed a similar retention time to that observed for MCP-1 expressed in a soluble form, see Figure 4.12. After RP-HPLC the protein was lyophilised prior to analysis. The molecular weight of the protein was determined to be ~8.5 kDa by SDS-PAGE and the protein was found to be >95 % pure. N-terminal sequencing confirmed cleavage of GST,
and ESMS yielded a mass of 8584.6 Da indicating that the protein was of the correct mass for MCP−1 with both disulfides formed, see Figure 4.13.

Figure 4.12 RP-HPLC of MCP−1 isolated from inclusion bodies. The protein eluted at 45-49% acetonitrile (0.1% TFA).

Figure 4.13 ESMS analysis of MCP−1 isolated from inclusion bodies following SEC.

1D ¹H NMR was used to investigate the conformation of the renatured MCP−1, see Figure 4.14. The spectrum is characteristic of CP-modules and compares well to the spectrum.
recorded on the protein expressed in a soluble form. The resonances are sharp and well-dispersed and there is little evidence of random coil. The resonances at 0-0.5 ppm provide evidence for the formation of a hydrophobic core indicating that the protein has refolded into a stable conformation. This preliminary investigation demonstrates that MCP-1 can be refolded from a fully denatured state and that it is a stable and soluble protein fragment. Further structural characterisation performed by calorimetry and circular dichromism is described in Chapter 5.

Figure 4.14 1D $^1H$-NMR spectrum of MCP-1 isolated and renatured from inclusion bodies. The sample was 0.15 mM in 20 mM potassium phosphate (pH 6.0) and the spectrum recorded at 37 °C and 600 MHz.
4.2 Expression of MCP-1 in *Pichia pastoris*

This section describes the expression of MCP-1 in the methylotrophic yeast *Pichia pastoris*. Since eukaryotes provide an appropriate environment for correct disulphide formation in heterologous proteins, this study was implemented to provide a comparison with MCP-1 that had been refolded *in vitro*. The secretory pathway of *P. pastoris* results in glycosylation of MCP-1 at its naturally occurring N-linked glycosylation site, Asn 83. For the purposes of structural elucidation by NMR, the sugars were removed enzymatically. In an alternative approach, the N-linked glycosylation site was removed by site directed mutagenesis (T85A) and the mutant form of MCP-1 was also expressed in *P. pastoris* to study the effect of the carbohydrate moiety on the conformation of the protein.

Since expression of CP-modules in *E. coli* in high yield has generally proved difficult eukaryotic expression systems have been considered. For example, expression of individual and pairs of CP-modules from human factor H in the yeast *S. cerevisiae* produced folded protein for structural elucidation by NMR. However, the level of expression tended to be low, for example in the case of the 16th CP-module of Factor H, the yield of purified protein was typically 0.05 - 0.1 mg/l. Expression in *P. pastoris* can result in increased yields relative to *S. cerevisiae* as production of the heterologous protein is controlled by a strong, highly-regulated methanol-inducible promoter. Additionally, multiple insertion of the gene of interest into the Pichia genome can result in greatly increased levels of expression. These rarely (~1% of transformants) occurring 'jackpot clones' can result in expression levels of up to 1 g/l. *P. pastoris* is frequently used for the expression of recombinant proteins in high yields and has been successfully used for the expression of CP-modules.
The heterologous protein expressed in *P. pastoris* can be directed either to the cytosol or into the secretory pathway. During secretion, folding, disulphide bond formation and glycosylation occurs. Since MCP-1 possesses an N-linked glycosylation site it was directed into the secretory pathway by fusion with the α-factor secretion signal from *S. cerevisiae*.

In *P. pastoris*, although glycosylation occurs at the natural glycosylation site, the carbohydrate is often different to the native type being of the high mannose-type and often excessive in length. The reported average length of the carbohydrate chain in *P. pastoris* is 8-14 residues\(^{156}\). Although the glycosylation of proteins could prove problematic for structural elucidation by NMR, the carbohydrate can be cleaved enzymatically or the potential glycosylation site can be removed by site-directed mutagenesis. Consequently, the recombinant protein produced is similar to the native protein and additionally is suitable for structural elucidation by NMR.

To aid structure determination by NMR, labelling proteins with NMR-active isotopes is highly advantageous for simplifying spectra. Since *P. pastoris* grows on minimal medium the cost of labelling with \(^{15}N\) and/or \(^{13}C\) is relatively low. Media containing \(^{15}N\)-ammonium sulphate as the sole nitrogen source and \(^{13}C\)-glucose or \(^{13}C\)-glycerol as the sole carbon source, in addition to salts and a low concentration of biotin, provide adequate conditions for growth. For \(^{13}C\)-labelling, induction with \(^{13}C\)-methanol is required. The *P. pastoris* expression system therefore potentially offers a means of producing isotopically labelled protein in good yield.
The expression of MCP-1 in *P. pastoris* was expected to produce properly folded protein in sufficient quantities for structural studies. This would then enable comparisons to be made to the versions of the protein refolded *in vitro*.

### 4.2.1 Expression and purification of MCP-1

The *P. pastoris* expression vector pPIC9 encoding for MCP-1 was transformed into *P. pastoris* KM71 and GS115. Test inductions were performed on a number of transformants in each cell line. After two days growth in buffered minimal glycerol (BMG) the cells were harvested and expression of MCP-1 induced by resuspending in buffered minimal methanol (BMM). Growth was continued for six days with the daily addition of methanol (0.5% v/v) to the medium. The cells were harvested by centrifugation and the supernatant collected. The presence of MCP-1 was tested for by SDS-PAGE after TCA precipitation of the supernatant. MCP-1 was difficult to detect; in the KM71 cell line a number of species were observed which formed a smear across the molecular weight range of the gel, whilst in the GS115 cell line the level of expression of MCP-1 was too low to detect by Coomassie staining.

Although the level of expression of MCP-1 in *P. pastoris* KM71 was low and resulted in a number of protein products, a dominant band of ~22k Da was detected by SDS-PAGE. The presence of numerous protein species of higher than expected mass for MCP-1 suggested the presence of various glycoforms. Since glycosylation in *P. pastoris* is of the high mannose-type, the supernatant was incubated with the deglycosylating enzyme
Endoglycosidase H (EndoH) which specifically cleaves high mannose carbohydrates leaving a single N-acetylglucosamine (GlcNAc) on the asparagine of the protein.

SDS-PAGE analysis of the protein following treatment with EndoH resulted in a band of the expected size for MCP-1 (~8.5 kDa) confirming that the protein produced was heavily glycosylated during secretion. Additionally, the removal of the carbohydrate resulted in the glycoforms resolving into a single band on a gel which indicated that the expression of MCP-1 was in fact high (yielding ~7 mg/l) despite the initial analysis. Analysis of deglycosylated MCP-1 by SDS-PAGE indicated that the protein produced was >85 % pure in the growth medium, see Figure 4.15.

Expression of MCP-1 in KM71 was scaled up to provide sufficient protein for structural characterisation. The initial growth in BMG (2 l) was carried out at 30 °C to A₆₀₀ ≈ 20. The

![Figure 4.15 SDS-PAGE analysis of MCP-1 expressed in P. pastoris KM71 before (A) and after (B) treatment with EndoH.](image)
cells were harvested and resuspended in fresh BMG (1 l) and incubated for a further 24 h. This enabled the cell density to increase to $A_{650} \approx 65$. The cells were then harvested and expression of MCP-1 induced by resuspension in BMM (2 l). As with the test inductions, the media was supplemented daily with methanol to 0.5% v/v. After six days the cells were harvested and the supernatant collected. SDS-PAGE analysis of the supernatant showed the dominant protein species to be ~22 kDa, as observed previously.

In addition to post-translational modifications, the high purity of the protein of interest in the growth medium is a further advantage of utilising the eukaryotic secretory pathway. Few native *P. pastoris* proteins are secreted and the minimal medium used for expression contains only low levels of protein. Consequently, concentration of the growth medium by ultrafiltration serves as the first stage in isolating the secreted protein, which is generally >80% pure. A potential disadvantage of this approach is that the recombinant protein is more susceptible to proteolysis than one which is expressed intracellularly. However, addition of protease inhibitors to the samples throughout purification minimised protein degradation. The various glycoforms of MCP-1 produced by *P. pastoris* were all found to behave in a similar way when subjected to anion exchange (AEX) chromatography on "Fast Flow" Q-Sepharose (FFQ) media facilitating their separation from contaminants. The multiple glycoforms did however result in a rather broad chromatogram eluting over a range of 0.27-0.37 M NaCl (0.1 M TrisHCl, pH 9), see Figure 4.16. Fractions containing MCP-1 glycoforms were concentrated by ultrafiltration prior to enzymatic deglycosylation with Endo H (20 units for 16 hours at 37°C). Deglycosylated MCP-1 was isolated by RP-HPLC and lyophilised. The protein was found to have a similar retention time (40-46 % acetonitrile, 0.1 % v/v TFA) to MCP-1 expressed in *E. coli* (45-49% acetonitrile, 0.1 % v/v TFA) and synthetic MCP-1 (46-47% acetonitrile, 0.1 % v/v TFA), see Figure 4.17.
Figure 4.16 *Purification of MCP-1 glycoforms by ALEX chromatography.*

Figure 4.17 *RP-HPLC of deglycosylated MCP-1*

Following deglycosylation, and desalting by RP-HPLC, analysis by SDS-PAGE showed the protein to be of the expected size for MCP-1 (~8.5 kDa) and to be >95% pure. The N-terminal sequence was consistent with that expected for MCP-1 expressed in this system.
(YVEFSDACEE). Furthermore, since the N-terminus was homogeneous, cleavage of the secretion signal sequence appeared complete. ESMS confirmed the presence of a single species of 8541.4Da, corresponding precisely to the calculated mass of the expressed protein with both disulphides formed and a single GlcNAc moiety, see Figure 4.18.

![ESMS analysis of deglycosylated MCP-1](image)

Figure 4.18 ESMS analysis of deglycosylated MCP-1.

The purified protein was analysed by 1D ¹H-NMR to confirm that it was correctly folded, see Figure 4.19. A spectrum characteristic of CP-modules was recorded showing good dispersion of resonances. However, although the protein appeared to be homogeneous by SDS-PAGE and ESMS, NMR spectroscopy detected the presence of a number of carbohydrates. The presence of non-covalently attached carbohydrates has previously been observed when proteins are expressed in *Pichia pastoris* \(^{158}\). The protein was applied to a Concavalin A column which, in the presence of divalent metal ions, reversibly binds molecules containing α-D-mannopyranosyl and α-D-glucopyranosyl. MCP~1 was isolated
Results and Discussion

Figure 4.19 (a) $^{1}H$-NMR spectrum of MCP-1 expressed in *P. pastoris* KM71. The sample was 1.0 mM in 50 mM potassium phosphate and the spectrum recorded at 600 MHz and 37 °C.

Figure 4.19 (b) $^{1}H$-NMR spectrum of MCP-1 expressed in *P. pastoris* KM71 after purification on Concavalin A.
Results and Discussion

Analysis by 1D $^1$H-NMR confirmed the removal of the carbohydrate moieties as indicated by the decrease in resonances at 3.6-4.4 and 4.8-5.4 ppm.

Preliminary characterisation of MCP-1 by NMR showed the module to be folded and there was little evidence of random coil. The resonances are sharp and well dispersed and characteristic of CP-modules. Further structural characterisation by calorimetry, circular dichromism and multi-dimensional NMR is discussed in Chapter Five.

As the protein was expressed in high yield, isotopic labelling with [15N] ammonium sulphate as the sole nitrogen source was feasible to assist in structural elucidation by NMR. The protein was expressed in medium containing 0.2 % w/v [15N]-ammonium sulphate and purified as described above. The level of expression of MCP-1 (~7 mg/ml) was not affected by the reduced nitrogen content of the medium.

4.2.2 Expression of MCP-1 (T85A)

The N-linked glycosylation site of MCP-1 was removed by site directed mutagenesis to determine if glycosylation affects the conformation of MCP-1. The mutation of threonine 85 to alanine was selected to remove the N-linked glycosylation site (NH$_{85}$T). Additionally, this mutation introduces a unique restriction enzyme site (NsI) to enable the rapid screening of potential mutants. The mutation was introduced by a two stage PCR reaction as outlined in Figure 4.20. In the first round, two separate reactions were carried out to introduce the mutation to both strands of DNA. These two fragments then serve as the template in the
second PCR stage when the primers used are those which define the termini of the amplified fragment.

![Diagram of PCR strategy](image)

1. Primers 1 + 3
2. Primers 2 + 4
3. Products anneal

Template DNA
Amplified fragment
Site of mutation
Primers

1. Two separate reactions are carried out with primers 1 + 3 and 2 +
2. The products of the first round of PCR each contain the mutation.
3. In the second PCR step, the products anneal and are amplified using the cloning primers (1 + 4).
The amplified fragment contains the mutation in each strand.

Figure 4.20 *PCR strategy used to introduce mutation to MCP-1(T85A)*

The amplified fragment was initially cloned into the pGEM-T vector and positive transformants selected by blue/white colour screening. Colonies were further screened by restriction digestion with *NsiI* before cloning into the *P. pastoris* expression vector pPICZα. The correct construction of the plasmid was confirmed by DNA sequencing prior to transformation into the *P. pastoris* cell lines KM71, GS115 and SMD1168.
Two different vectors, pPIC9 and pPICZα, were used for the construction of the wild-type and mutant plasmids respectively. Although there are some differences between the vectors both permit the induction of high-level protein expression by methanol and possess the α-factor secretion signal from *S. cerevisiae* resulting in secretion of the protein of interest into the medium. The presence of a zeocin resistance gene in pPICZα enables positive transformants to be screened rapidly. Additionally, this can be used to screen for ‘jackpot’ clones as they have an increased resistance to zeocin. However, despite screening a number of transformants for each plasmid encoding MCP~1 (T85A) no jackpot clones were identified. This was also the case for the wild-type where all transformants tested expressed MCP~1 at the same level.

The use of the pPICZα vector required the addition of histidine (0.004 % w/v) to the growth media as, unlike pPIC9, pPICZα does not possess the HIS4 gene. With the exception of this modification, expression of MCP~1 (T85A) was carried out as described for MCP~1. A series of test inductions were carried out and expression of MCP~1 (T85A) analysed by SDS-PAGE. Although the level of expression in GS115 was not detectable by Coomassie staining, the expression levels in KM71 and SMD1168 were comparable (~5 mg/l). Each transformant tested expressed at the same level. The protein was found to be of the expected size for MCP~1 (T85A) (~8.5 kDa) and to be >85 % pure in the growth medium as judged by SDS-PAGE.

Expression of MCP~1 (T85A) in KM71 was scaled up to 2 litres and the growth carried out under the conditions described for MCP~1. After harvesting and concentrating the supernatant by ultrafiltration MCP~1 (T85A) was purified by anion exchange chromatography. MCP~1 (T85A) eluted from a FFQ column in one major peak
(0.41-0.51 M NaCl, 0.1M TrisHCl, pH 8.2) while the contaminating proteins were not retained on the column, see Figure 4.21. The protein isolated from anion exchange chromatography was desalted and concentrated by RP-HPLC, (37-47% acetonitrile, 0.1%v/v TFA), see Figure 4.22, and lyophilised. The mutant protein was more strongly retained than MCP-1 in anion exchange chromatography although this is probably a consequence of the lack of carbohydrates rather than the single amino acid change.

RP-HPLC retention times were similar for both MCP-1 and MCP-1(T85A) suggesting that the properties of the protein were not significantly altered. As described for MCP-1, the protein was applied to a Concavalin A column to remove any non-covalently bound carbohydrates.

![Graph](image1)

Figure 4.21 AIEX chromatography of MCP-1(T85A)

![Graph](image2)

Figure 4.22 RP-HPLC purification of MCP-1(T85A)
SDS-PAGE analysis showed the protein to be of the expected size for MCP-1 (T85A) (8.5 kDa) and to be >95 % pure. However, analysis of the protein by ESMS indicated the presence of three protein species of molecular weight 8410.25 Da, 8248.11 Da, 8048.12 Da that could not be resolved by SDS-PAGE. The predicted mass for the expressed protein is 8046.2 Da. N-terminal sequencing indicated incomplete processing of the signal sequence which can result in additional pairs of glutamate-alanine residues being left N-terminal to the recombinant protein. The two major species could be identified as EAEFSDACEE and EFSDACEEPP, the masses of which are calculated as 8246.2 Da and 8046.2 Da respectively. This correlates well with the ESMS data. These two protein species could not be resolved chromatographically using the techniques described above. The identity of the third species, of mass 8410.25 Da, was unclear and it was not identified by N-terminal sequencing.

Figure 4.23 ESMS analysis of MCP-1(T85A)
The formation of the major and minor forms of MCP-1 (T85A) differing by EA at the N-terminus is a consequence of incomplete processing of the signal sequence. The α-factor sequence that signals for the protein to be secreted is cleaved in a two step process. The pro-sequence (underlined) is cleaved by the Kex2 enzyme between arginine and glutamine in the sequence E K R * E A E A E F S D A. A second cleavage then occurs after the EA repeats mediated by the STE13 gene product. The cleavage efficiency appears to depend on the N-terminal sequence of the recombinant protein. For MCP-1 in the vector pPIC9 the signal sequence was efficiently cleaved resulting in a homogeneous N-terminus with no EA repeats. However, expression of VCP-3,4 resulted in two proteins differing by three N-terminal residues due to incorrect cleavage of the signal sequence.

1D 1H NMR analysis of MCP-1 (T85A) resulted in a spectrum characteristic of CP-modules, see Figure 4.24.

![Figure 4.24 1D 1H NMR spectrum of MCP-1(T85A). The sample was 0.9 mM in 50 mM potassium phosphate (pH 6.0). The spectrum was acquired at 600 MHz and 37 °C.](image)
The resonances are sharp, well-dispersed and the resonances at 0.5-0ppm are indicative of a folded CP-module. No discernible differences with respect to the 1D spectrum recorded of enzymatically deglycosylated MCP–1 were noted. To enable a more rigorous comparison with MCP–1, a [\(^{15}\)N] labelled sample was prepared to probe any structural differences by 2D NMR. Additionally, MCP–1(T85A) has been further characterised by calorimetry and circular dichromism, as described in Chapter Five.

4.3 Discussion

Despite the known difficulties of expressing disulphide bond containing proteins in *E. coli*, MCP–1 has been successfully isolated and renatured from inclusion bodies. Deposition of the protein in inclusion bodies facilitated its purification as few contaminating proteins were observed. Therefore, although under the conditions described the renaturation of the GST fusion partner was not achieved, its use as an affinity purification tag was not required. The presence of the fusion partner may in fact prove to be disadvantageous during renaturation of the target protein as its aggregation may facilitate aggregation of the protein of interest. The refolding protocol developed has not been applied to other CP-modules expressed in *E. coli* and therefore it is not possible to cite it as a generally applicable protocol. However, application of a similar protocol to chemically synthesised MCP–1 also yielded some folded protein, despite the presence of several contaminating species, as described in Chapter Three.

The renaturation of proteins, particularly those that contain disulphide-bonds, can result in
loss of yields due to aggregation and mis-folding. Consequently, the expression of proteins in a soluble form is often favoured. The employment of some strategies reported to increase the proportion of soluble protein synthesised proved unsuccessful in the case of MCP~1. Any soluble protein synthesised was observed to be proteolytically degraded. Proteases target abnormal or mis-folded proteins for degradation. It is possible therefore that the soluble fusion protein expressed in *E. coli* was incorrectly folded and susceptible to degradation. Strategies for minimising proteolysis of recombinant proteins include growth at reduced temperatures, construction of fusion proteins and the use of protease-deficient strains (reviewed by Makrides\(^{154}\)). In this study these methods failed to produce soluble full-length MCP~1. A more extensive study may have resulted in the expression of the protein in a soluble form. However, in the context of this project the deposition of MCP~1 in inclusion bodies proved advantageous as comparative refolding studies of recombinant and synthetic versions of the protein could be performed.

Targeting expression of MCP~1 to the periplasm was not attempted, despite reports of the periplasmic expression of several correctly folded disulphide-bond containing proteins including human epidermal growth factor\(^{215}\) and bovine pancreatic trypsin inhibitor\(^{202}\). Expression in the periplasm facilitates purification of recombinant proteins as only a small percentage of host proteins are expressed in this cellular compartment\(^{216}\). In addition, the protein is exposed to far fewer proteases compared to the cystolic environment. Despite these theoretical advantages, efforts to target the expression of a number of CP-modules to the periplasm proved unsuccessful. The expressed proteins were not detectable by Coomassie staining or Western blotting\(^{219}\).

Although it was possible to produce folded MCP~1 from *E. coli*, expression in *P. pastoris*
produced folded protein in significantly higher yields. The high level of expression and ease of purification makes *P. pastoris* an ideal system for producing CP-modules in sufficient quantities for structural elucidation by NMR. The most significant drawback to this system is the glycosylation of those proteins naturally possessing glycosylation sites. However, as demonstrated, the use of deglycosylating enzymes or site-directed mutagenesis can eliminate this problem.

The expression of MCP-1 both in *E. coli* and *P. pastoris* has confirmed the module to be stable and soluble as a single entity. The comparison of 1D NMR spectra of the protein isolated and renatured from inclusion bodies, and of the proteins expressed in *P. pastoris* indicated that the protein can be successfully refolded. Since it is possible to successfully renature MCP-1, refolding can be disregarded as a potential problem with the chemically synthesised protein.
5.1 Introduction

The work in this chapter describes the structural characterisation of MCP-1 by differential scanning calorimetry, circular dichromism and NMR spectroscopy. Versions of the protein expressed in prokaryotic and eukaryotic systems are compared and progress towards the NMR solution structure of MCP-1 is described.

5.2 Differential Scanning Calorimetry Studies

Differential scanning calorimetry (DSC) provides information about the conformational stability of a protein. The technique is useful for comparing the conformational stabilities of wild-type and mutant forms of a protein. DSC can also be useful for the study of fragments from multi-domain proteins, for example, in the case of CP-modules the stability of the inter-modular interface can be estimated.

DSC was performed on recombinant MCP-1 expressed in both prokaryotic and eukaryotic systems. As described in Chapter Four, expression in E. coli resulted in deposition of the
protein in inclusion bodies requiring the development of a renaturation protocol. It was anticipated that DSC, in conjunction with other structural studies, would enable the success of the renaturation protocol to be assessed. A further study was made of MCP-1(T85A) expressed in *P. pastoris* to compare the conformational stability of this protein to the wild-type protein expressed in the same organism.

5.2.1 Study of MCP-1 expressed in *E. coli*

Following isolation from inclusion bodies and *in vitro* renaturation, as described in Chapter Four, a sample of MCP-1 expressed in *E. coli* was analysed by DSC, see Figure 5.1.

![Differential scanning calorimetry of recombinant (E. coli) MCP-1.](image)

Figure 5.1 *Differential scanning calorimetry of recombinant (E. coli) MCP-1.* The sample was 0.14 mM in 20 mM potassium phosphate (pH 6.0). Experimental data after base-line subtraction are shown by the solid line, the dotted line indicates the computer-fitted data.
The peak observed in the heat capacity curve is indicative of thermal unfolding of the proteins tertiary structure and would not be measurable for protein in a random coil conformation. The calorimetric enthalpy change on unfolding was determined to be 54 kcal/mol. This value is comparable to those observed for other single CP-modules. This study therefore corroborates the 1D 'H-NMR data acquired on this sample (see Figure 4.14) and suggests that the module was refolded correctly from a fully denatured state. A relatively high melting temperature (61 °C) was observed which is again comparable to the values previously observed for single modules. This provides evidence that this version of MCP~1 is relatively stable to denaturation by heat and indicates that the protein has a well-defined tertiary structure.

5.2.2 Studies of MCP~1 expressed in P. pastoris

MCP~1 expressed in P. pastoris was analysed by DSC as described above, see Figure 5.2 (a). The calorimetric enthalpy of this sample was determined to be 58.4 kcal/mol and the melting temperature to be 62 °C. These results are consistent with a protein that has a stable, well-defined tertiary structure. Furthermore, the values observed for this version of MCP~1 are in good agreement with those observed for MCP~1 isolated from inclusion bodies. This suggests that the two versions of the protein are conformationally similar and provides further evidence that the latter was successfully renatured in vitro.

The observed melting temperature of 59 °C for MCP~1(T85A), see Figure 5.2 (b), indicates that this version of the protein is significantly less stable to heat denaturation than...
the wild-type. This suggests that a destabilising conformational change has been caused by the single amino acid mutation. The calorimetric enthalpy of MCP-1(T85A) was determined to be 82 kcal/mol, approximately 1.4 times greater than that observed for MCP-1. The reason for this is unclear although it is postulated that the mutation has altered the hydrophobic nature of the protein resulting in aggregation through exposed hydrophobic patches.

Figure 5.2(a) DSC studies of MCP-1 expressed in P. pastoris. The sample was 1.3 mM in 50 mM potassium phosphate pH 6.0.

Figure 5.2(b) DSC studies of MCP-1(T85A) expressed in P. pastoris. The sample was 1.2 mM in 50 mM potassium phosphate pH 6.0.
Analysis of the X-ray crystal structure of MCP-1,2 provides a possible insight into the effect of the mutation. In the crystal structure, Thr 85 is close in space to Asp 81 and it is possible that there may be a hydrogen bond between these residues. The mutation of Thr 85 to Ala would result in disruption of this bond which may have a destabilising effect on the protein, as illustrated in the reduced melting temperature of MCP-1(T85A). The mutation may also have an effect on the hydrophobic core of the protein since it is adjacent to the consensus tryptophan residue (Trp 86). Removal of the hydrogen bond could result in a change in the orientation of Trp 86 and the potential exposure of this hydrophobic residue may result in aggregation of the protein. A possible alternative explanation is that removal of the glycosylation site affected the processing and folding of the protein prior to secretion. The effect of the mutation on the conformation of MCP-1 was further explored by circular dichromism and NMR studies.

5.3 Circular Dichromism Studies

Circular dichromism (CD) provides a rapid means for assessing protein conformation in solution. Although only low-resolution structural information is obtained it allows the secondary structure content of a protein to be estimated. The spectral properties of a protein are highly dependent on the environment and mobility of its chromophores. Therefore the technique can be used to study conformational changes of the protein under different conditions or to compare the properties of wild-type and mutant forms of a protein. As described for DSC studies, versions of MCP-1 expressed in E. coli and P. pastoris (both
wild-type and mutant) were analysed by CD to probe any conformational differences between them.

### 5.3.1 Study of MCP-1 expressed in *E. coli*

The secondary structure of the refolded protein isolated from inclusion bodies was investigated by CD, see Figure 5.3. The far-UV spectrum shows a minimum at ~195nm and positive ellipticity in the region 220-235nm, with a maximum at 230nm. The near UV CD shows two minima at ~285nm and 295nm and a local maximum at 290nm.

![Figure 5.3 (a)](image)

![Figure 5.3 (b)](image)

**Figure 5.3** Circular dichroism spectra in of recombinant (*E. coli*) MCP-1 (a) the far-UV and (b) the near-UV regions. The sample was 0.18 mM in 20 mM potassium phosphate (pH 6.0).

The spectra are in good agreement with those recorded for other CP-modules and show positive ellipticity in the far UV-region which has been identified as a signature for CP-modules \(^{33, 157, 222}\). It has been reported that tryptophan residues in β-strands contribute to positive ellipticity in the far-UV region \(^{223}\). Since each CP-module contains a tryptophan as
part of its consensus sequence it is postulated that this results in this characteristic CD-profile of CP-modules in this region.

5.3.2 Studies of MCP-1 expressed in *P. pastoris*

CD analysis of wild-type MCP-1 resulted in a minimum in the far-UV spectrum at ~195nm and a strong positive ellipticity in the region 220-240 nm with a maximum at 230 nm. In the near-UV spectrum two minima were observed at 280 and 285 nm and a local maximum at 295 nm, see *Figure 5.4*. These results are characteristic of CP-modules and additionally are in good agreement with those observed for MCP-1 isolated from inclusion bodies. This therefore provides further evidence that the secondary structures of the two versions of the protein are similar.

Although analysis of MCP-1(T85A) showed a similar overall CD-profile to the wild-type version of the protein, the mutant form showed a stronger positive molar ellipticity, see *Figure 5.4*.

As previously indicated by DSC studies, this suggests that the single amino acid change resulted in a change in the conformation of the protein. It was previously postulated that the mutation had resulted in a change to the hydrophobic core of the protein. As discussed earlier, this could result in the aromatic residues being in different environments in the wild-type and mutant versions of MCP-1, possibly resulting in different ellipticity values.
Figure 5.4 CD spectra of MCP~1 and MCP~1(T85A). The samples were 0.09 and 0.11 mM in 50 mM potassium phosphate (pH 6.0) respectively.

5.4 Analysis of MCP~1 by NMR Spectroscopy

NMR spectroscopy was the principal technique used to assess protein structure throughout this project. This section compares data acquired on the different recombinant versions of MCP~1. Progress towards the NMR structure determination of the enzymatically deglycosylated version of MCP~1 expressed in *P. pastoris* is also described.
5.4.1 1D 'H-NMR studies

One dimensional 'H-NMR studies were initially performed to assess the different versions of MCP-1. These results have been described in Chapters Three (for synthetic MCP-1) and Four (for recombinant versions of MCP-1).

No discernible differences were observed between the spectra recorded on the recombinant versions of MCP-1 expressed in *E. coli* and *P. pastoris*. However, due to the increased yield of protein and other advantages of eukaryotic expression over expression in *E. coli*, the version of MCP-1 expressed in *P. pastoris* was selected for further structural analysis. The wild-type and mutant forms of the protein were studied by 2D NMR to further probe the differences observed between them in both DSC and CD studies.

5.4.2 Comparison of MCP-1 and MCP-1(T85A) by 2D 'H-'15N heteronuclear spectroscopy

Following isotopic enrichment of MCP-1 and MCP-1(T85A) with ['5N]-ammonium sulphate, 2D 'H-'15N HSQC spectra were recorded to enable a more thorough comparison of the two versions of the protein, see Figures 5.5 and 5.6 (a).

Analysis of the 2D 'H-'15N HSQC spectrum of MCP-1 indicate that the protein is in a stable conformation with a well-defined structure. The resonances are dispersed and there is little evidence of random coil. In contrast, the HSQC spectrum of MCP-1(T85A), acquired under the same conditions, indicated that the structure of the mutant version of the protein is less-well defined. Many of the resonances are overlapped and clustered in the centre of the
Results and Discussion

Figure 5.5 $^1H$-$^{15}N$ HSQC spectrum of MCP-1. The sample was 1.0 mM in 50 mM potassium phosphate (pH 6.0) and the spectrum recorded at 600 MHz and 37 °C.

Figure 5.6 $^1H$-$^{15}N$ HSQC spectra of MCP-1(T85A). The sample concentration was 0.9 mM and spectra were acquired at 600 MHz and 37 °C.

(a) Sample in 50 mM potassium phosphate (pH 6.0)
(b) Sample in 50 mM potassium phosphate (pH 6.0), 100 mM NaCl

(b) Sample in 50 mM potassium phosphate (pH 6.0), 200 mM NaCl
Results and Discussion

In an attempt to minimise any potential aggregation, spectra were recorded with the addition of 100 and 200mM NaCl. Although the spectra recorded showed some improvement, resonances remained overlapped and were not dispersed to the same extent as for MCP–1, see Figures 5.6 (b) and (c).

The physical data acquired on MCP–1 and MCP–1(T85A) indicate that the mutation to remove the glycosylation site had a severe effect on the protein conformation. DSC and CD studies provided evidence that the single amino acid mutation resulted in a change in the conformational stability of the protein. Potential exposure of the hydrophobic core and subsequent aggregation is suggested by the NMR data acquired on the sample.

Consequently, further structural characterisation was performed on the enzymatically deglycosylated version of MCP–1 expressed in P. pastoris. The following sections describe the data acquired on the sample and progress towards the NMR solution structure of MCP–1.

The solution structure of MCP–1 forms part of an ongoing project to study the structure of CP-module containing proteins. In spite of the X-ray crystal structure of MCP–1,2 the NMR solution structure of this protein remains a valid target due to the flexibility of the protein in solution. It is anticipated that this preliminary NMR structure characterisation of MCP–1 may aid the solution structure determination of MCP–1,2 and provide further structural insights into its function as the measles virus receptor.
5.4.3 Analysis of MCP-1 by 2D and 3D NMR spectroscopy

A series of 2D and 3D NMR experiments were acquired on the version of MCP-1 expressed in *P. pastoris* and enzymatically deglycosylated as described in Chapter Four. The experimental details are outlined in Table 5.1. All experiments were recorded at 600 MHz and 37 °C in 50 mM potassium phosphate at pH 6.0. The sample concentration was 1.0 mM. Multidimensional spectra were processed using the Azara package 224.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Solvent</th>
<th>Solvent suppression</th>
<th>Mixing Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D TOCSY</td>
<td>H₂O, D₂O</td>
<td>DPFGSE</td>
<td>40 ms</td>
</tr>
<tr>
<td>2D COSY</td>
<td>H₂O,</td>
<td>RAW-SCUBA</td>
<td>-</td>
</tr>
<tr>
<td>2D NOESY</td>
<td>H₂O, D₂O</td>
<td>DPFGSE</td>
<td>150</td>
</tr>
<tr>
<td>2D H-15N HSQC</td>
<td>H₂O</td>
<td>PFG</td>
<td></td>
</tr>
<tr>
<td>1H-15N 3D -NOESY-HSQC</td>
<td>H₂O</td>
<td>PFG</td>
<td>100</td>
</tr>
<tr>
<td>1H-15N 3D -TOCSY-HSQC</td>
<td>H₂O</td>
<td>PFG</td>
<td>70, 35</td>
</tr>
<tr>
<td>HNHA</td>
<td>H₂O</td>
<td>Jump and return</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1 NMR experiments acquired on MCP-1.

5.5 Resonance Assignment of MCP-1

Prior to the development of 2D NMR techniques, resonance assignments of protein spectra relied largely on the availability of an X-ray crystal structure of the protein. The introduction of multi-dimensional NMR methods increased the amount of information
Results and Discussion

obtainable from NMR spectra of macromolecules and enabled the development of a systematic method for resonance assignment. This sequential assignment method begins with the identification of spin systems using TOCSY and COSY spectra. Subsequently, the method uses the sequential through-space connectivities observed in the NOESY experiment to link spin systems to their neighbours.

5.5.1 Sequential assignment of MCP-1

Spectra were displayed and cross-peak lists maintained in ANSIG, in which spectra are viewed as 2D planes or projections of 2D planes. Initial assignments of spin systems were made using the 2D 'H-15N HSQC and 'H-15N 3D -NOESY-HSQC spectra. Peaks were picked in the 15N-HSQC spectrum and the associated TOCSY resonances identified using 'peak_find' to direct peak picking. Since there is no spin-spin coupling across the peptide bond, the spin systems could be divided into various categories according to the number of resonances and chemical shifts observed. These categories and random coil chemical shifts are well documented. The 'H-15N 3D -NOESY-HSQC spectrum was then used to connect spin systems via sequential through-space NOEs. If the resulting chains of sequential spin-systems were unique within the primary structure of MCP-1 they were specifically assigned.

This approach resulted in assignment of the backbone and the majority of side-chain resonances for MCP-1, excluding the proline residues, the N-terminal residues Tyr 28 and Val 29, and the asparagine residue bearing the GlcNAc moiety following deglycosylation (Asn 83). Further sidechain assignments were made using the 2D-NOESY, TOCSY and
COSY spectra. The assigned $^1$H-$^{15}$N HSQC spectrum is shown in Figure 5.7 and Table 5.2 documents the resonance assignments made. Figure 5.8 illustrates some of the sequential assignments made from the $^1$H-$^{15}$N 3D -NOESY-HSQC spectrum.
Figure 5.7 The 2D $^{15}$N-$^1$H HSQC with resonance assignments. In order to maximise the digital resolution in 3D experiments the spectrum was folded to reduce the sweep width in the $^{15}$N dimension.
<table>
<thead>
<tr>
<th>Residue</th>
<th>$^{15}$N</th>
<th>NH</th>
<th>$\alpha$H</th>
<th>$\beta$H</th>
<th>$\gamma$H</th>
<th>$\delta$H</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E30</td>
<td>125.3</td>
<td>8.35</td>
<td>4.07</td>
<td>1.82</td>
<td>1.80</td>
<td>2.09</td>
<td></td>
</tr>
<tr>
<td>F31</td>
<td>120.6</td>
<td>7.89</td>
<td>4.52</td>
<td>2.99</td>
<td>2.84</td>
<td></td>
<td>HD* 7.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HE* 6.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HZ* 7.02</td>
</tr>
<tr>
<td>S32</td>
<td>116.9</td>
<td>8.10</td>
<td>4.30</td>
<td>3.78</td>
<td>3.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D33</td>
<td>122.6</td>
<td>8.15</td>
<td>4.54</td>
<td>2.65</td>
<td>2.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A34</td>
<td>122.6</td>
<td>7.57</td>
<td>4.74</td>
<td>1.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C35</td>
<td>117.0</td>
<td>8.99</td>
<td>4.53</td>
<td>2.80</td>
<td>2.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E36</td>
<td>120.8</td>
<td>8.41</td>
<td>4.90</td>
<td>2.14</td>
<td>2.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E37</td>
<td>118.0</td>
<td>8.42</td>
<td>4.40</td>
<td>2.44</td>
<td>2.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P38</td>
<td>3.26</td>
<td>1.50</td>
<td>1.55</td>
<td>3.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P39</td>
<td>4.14</td>
<td>1.95</td>
<td>1.33</td>
<td>2.81</td>
<td>2.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T40</td>
<td>127.6</td>
<td>7.69</td>
<td>4.52</td>
<td>4.03</td>
<td></td>
<td>$\gamma$CH3 1.16</td>
<td></td>
</tr>
<tr>
<td>F41</td>
<td>120.0</td>
<td>8.77</td>
<td>4.73</td>
<td>3.51</td>
<td>2.50</td>
<td></td>
<td>HD* 7.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HE* 6.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HZ* 6.72</td>
</tr>
<tr>
<td>E42</td>
<td>120.6</td>
<td>8.73</td>
<td>4.11</td>
<td>2.08</td>
<td>2.01</td>
<td>2.29</td>
<td></td>
</tr>
<tr>
<td>A43</td>
<td>116.9</td>
<td>8.41</td>
<td>3.52</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M44</td>
<td>125.6</td>
<td>6.87</td>
<td>5.13</td>
<td>1.86</td>
<td>1.65</td>
<td>1.98</td>
<td></td>
</tr>
<tr>
<td>E45</td>
<td>118.9</td>
<td>9.32</td>
<td>4.99</td>
<td>2.12</td>
<td>1.87</td>
<td>2.13</td>
<td></td>
</tr>
<tr>
<td>L46</td>
<td>125.7</td>
<td>8.54</td>
<td>3.69</td>
<td>1.45</td>
<td>1.35</td>
<td>$\delta$CH3 0.37</td>
<td></td>
</tr>
<tr>
<td>I47</td>
<td>124.0</td>
<td>8.21</td>
<td>3.93</td>
<td>1.63</td>
<td>0.72</td>
<td>$\gamma$CH3 0.82</td>
<td></td>
</tr>
<tr>
<td>G48</td>
<td>128.2</td>
<td>8.07</td>
<td>4.15</td>
<td>3.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K49</td>
<td>120.9</td>
<td>8.11</td>
<td>4.62</td>
<td>1.77</td>
<td>1.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P50</td>
<td>4.37</td>
<td>2.21</td>
<td>1.93</td>
<td>2.06</td>
<td>3.86</td>
<td>3.66</td>
<td></td>
</tr>
<tr>
<td>K51</td>
<td>122.2</td>
<td>7.81</td>
<td>4.56</td>
<td>1.45</td>
<td>1.15</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>P52</td>
<td>4.28</td>
<td>2.21</td>
<td>1.72</td>
<td>1.97</td>
<td>3.61</td>
<td>3.60</td>
<td></td>
</tr>
<tr>
<td>Y53</td>
<td>123.7</td>
<td>6.52</td>
<td>5.19</td>
<td>2.92</td>
<td>2.60</td>
<td></td>
<td>HD* 7.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HE* 6.80</td>
</tr>
<tr>
<td>Y54</td>
<td>121.2</td>
<td>8.99</td>
<td>4.52</td>
<td>2.85</td>
<td>2.63</td>
<td></td>
<td>HD* 7.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HE* 6.80</td>
</tr>
</tbody>
</table>

Table 5.2: Resonance assignments for MCP-I (in ppm).
<table>
<thead>
<tr>
<th>Residue</th>
<th>15N</th>
<th>NH</th>
<th>αH</th>
<th>βH</th>
<th>γH</th>
<th>δH</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>E5</td>
<td>121.7</td>
<td>9.13</td>
<td>4.45</td>
<td>2.13</td>
<td>2.37</td>
<td>1.91</td>
<td></td>
</tr>
<tr>
<td>I56</td>
<td>121.1</td>
<td>8.34</td>
<td>3.41</td>
<td>1.70</td>
<td>1.51</td>
<td>0.52</td>
<td>0.33</td>
</tr>
<tr>
<td>G57</td>
<td>117.3</td>
<td>9.37</td>
<td>4.37</td>
<td>3.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E58</td>
<td>122.8</td>
<td>8.31</td>
<td>4.27</td>
<td>2.36</td>
<td>2.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R59</td>
<td>124.8</td>
<td>8.38</td>
<td>5.42</td>
<td>2.08</td>
<td>1.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V60</td>
<td>122.7</td>
<td>8.81</td>
<td>4.47</td>
<td>1.62</td>
<td>0.52</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>D61</td>
<td>122.6</td>
<td>8.26</td>
<td>5.29</td>
<td>2.54</td>
<td>2.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y62</td>
<td>117.5</td>
<td>8.60</td>
<td>5.09</td>
<td>2.68</td>
<td>2.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K63</td>
<td>116.6</td>
<td>9.12</td>
<td>4.80</td>
<td>1.85</td>
<td>1.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C64</td>
<td>120.7</td>
<td>8.78</td>
<td>4.85</td>
<td>2.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K65</td>
<td>124.2</td>
<td>8.02</td>
<td>4.10</td>
<td>1.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K66</td>
<td>121.1</td>
<td>8.36</td>
<td>4.18</td>
<td>1.84</td>
<td>1.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G67</td>
<td>128.2</td>
<td>8.84</td>
<td>4.21</td>
<td>1.78</td>
<td>1.07</td>
<td>0.58</td>
<td>0.76</td>
</tr>
<tr>
<td>Y68</td>
<td>8.31</td>
<td>5.16</td>
<td>3.29</td>
<td>2.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F69</td>
<td>120.3</td>
<td>9.48</td>
<td>5.06</td>
<td>3.10</td>
<td>2.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y70</td>
<td>122.2</td>
<td>8.27</td>
<td>5.26</td>
<td>3.26</td>
<td>2.81</td>
<td>6.99</td>
<td></td>
</tr>
<tr>
<td>I71</td>
<td>117.4</td>
<td>8.39</td>
<td>4.29</td>
<td>1.78</td>
<td>1.07</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>P72</td>
<td></td>
<td>2.47</td>
<td>2.15</td>
<td>1.82</td>
<td>3.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P73</td>
<td></td>
<td>4.55</td>
<td>1.99</td>
<td>1.67</td>
<td>3.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L74</td>
<td>125.1</td>
<td>8.12</td>
<td>4.15</td>
<td>1.87</td>
<td>1.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A75</td>
<td>124.7</td>
<td>8.54</td>
<td>4.49</td>
<td>1.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T76</td>
<td>125.0</td>
<td>8.87</td>
<td>4.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H77</td>
<td>117.3</td>
<td>7.06</td>
<td>5.80</td>
<td>3.36</td>
<td>3.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T78</td>
<td>127.1</td>
<td>8.78</td>
<td>4.65</td>
<td>4.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I79</td>
<td>124.0</td>
<td>8.10</td>
<td>5.42</td>
<td>1.69</td>
<td>1.20</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>C80</td>
<td>127.2</td>
<td>8.14</td>
<td>3.76</td>
<td>2.81</td>
<td>1.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D81</td>
<td>121.7</td>
<td>8.86</td>
<td>4.73</td>
<td>2.76</td>
<td>2.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R82</td>
<td>116.4</td>
<td>8.26</td>
<td>3.97</td>
<td>1.80</td>
<td>1.72</td>
<td>1.63</td>
<td>3.19</td>
</tr>
<tr>
<td>N83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H84</td>
<td>128.4</td>
<td>8.71</td>
<td>4.14</td>
<td>3.48</td>
<td>3.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T85</td>
<td>7.23</td>
<td>4.49</td>
<td>4.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Results and Discussion*
<table>
<thead>
<tr>
<th>Residue</th>
<th>15N</th>
<th>NH</th>
<th>αH</th>
<th>βH</th>
<th>γH</th>
<th>δH</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>W86</td>
<td>125.8</td>
<td>8.37</td>
<td>4.95</td>
<td>3.53</td>
<td>2.86</td>
<td></td>
<td>HD1 7.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HE1 9.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HE3 7.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HZ2 6.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HZ3 6.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HH2 6.09</td>
</tr>
<tr>
<td>L87</td>
<td>124.9</td>
<td>8.40</td>
<td>4.61</td>
<td>1.98</td>
<td>1.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P88</td>
<td></td>
<td>4.44</td>
<td></td>
<td>2.24</td>
<td>1.90</td>
<td></td>
<td>2.16 4.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.86</td>
<td></td>
<td>1.90</td>
<td>2.09</td>
<td></td>
<td>3.68</td>
</tr>
<tr>
<td>V89</td>
<td>122.1</td>
<td>7.90</td>
<td>4.29</td>
<td>1.90</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S90</td>
<td>119.7</td>
<td>7.85</td>
<td>4.69</td>
<td>3.92</td>
<td>3.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D91</td>
<td>121.6</td>
<td>8.05</td>
<td>4.53</td>
<td>2.65</td>
<td>1.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D92</td>
<td>118.1</td>
<td>7.85</td>
<td>4.29</td>
<td>2.61</td>
<td>2.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A93</td>
<td>118.7</td>
<td>7.32</td>
<td>4.16</td>
<td>1.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C94</td>
<td>128.6</td>
<td>7.42</td>
<td>5.43</td>
<td>2.60</td>
<td>2.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y95</td>
<td>125.8</td>
<td>9.31</td>
<td>4.77</td>
<td>2.61</td>
<td>2.56</td>
<td>7.23</td>
<td>HD* 7.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HE* 7.05</td>
</tr>
<tr>
<td>R96</td>
<td>125.4</td>
<td>8.15</td>
<td>3.88</td>
<td>1.52</td>
<td>1.40</td>
<td>1.31</td>
<td>2.79 2.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.37</td>
</tr>
<tr>
<td>E97</td>
<td>125.0</td>
<td>8.26</td>
<td>4.31</td>
<td>1.90</td>
<td>2.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T98</td>
<td>121.3</td>
<td>7.47</td>
<td>4.15</td>
<td></td>
<td></td>
<td></td>
<td>1.16</td>
</tr>
</tbody>
</table>
Figure 5.8  *Sequential assignment of 1H-15N 3D-NOESY-HSQC spectrum*. Residues Thr 40 - Glu 45 of MCP-1 are shown.
5.5.2 Assignment of prolines

The ring structure of proline imposes conformational restraints on the secondary structure of proteins. Since its presence causes a turn and breaks helical and sheet structure the resonance assignment of proline residues provides important structural information.

As proline does not have an amide proton, the $^1$H-$^{15}$N 3D TOCSY-HSQC spectrum cannot be used to identify its spin system. However, the $^1$H-$^{15}$N 3D NOESY-HSQC spectrum can be used since, for trans prolines, strong NOEs are observed from the δ-proline protons to the α-protons of the preceding residue. The 2D-NOESY, TOCSY and COSY spectra can then be used to assign the remainder of the proline spin-system.

In the case of MCP-1, which contains seven proline residues, assignment of the side chains proved problematic. To reduce overlap in the spectra and assist assignment, band selective 2D-TOCSY-TOCSY and 2D-TOCSY-NOESY spectra were acquired by Dr. Dusan Uhrin. Selective excitation of the proline δ-protons and transfer to the β- and γ-protons resulted in greatly simplified 2D-spectra. These are illustrated, with their non-selective counterparts, in Figure 5.9.

Using a combination of the regular 2D experiments and the band selective excitation experiments resonance assignments for the majority of the proline protons were made as detailed in the resonance assignment table (see Table 5.2). Despite the simplification of the spectra using the band selective techniques, resonances remained overlapped and complete resonance assignment proved problematic. Selected assignments are shown in the 2D-TOCSY-NOESY spectrum, see Figure 5.9 (d).
Figure 5.9 (a) 2D TOCSY spectrum
Results and Discussion

Figure 5.9 (b) Band selective 2D TOCSY-TOCSY spectrum.
Results and Discussion

Figure 5.9 (c) 2D NOESY spectrum.
Figure 5.9 (d) 2D TOCSY-NOESY spectrum. Selected proline assignments are shown.
5.5.3 Chemical Shift Index for MCP-1

It has been observed that 'H chemical shifts are dependent on the secondary structure of a protein. For all naturally occurring amino acids the Hα proton is found to shift upfield relative to its random coil shift when in a helical configuration, denoted -1 in the chemical shift index (CSI)\(^2\), and downfield when in a β-strand (denoted +1). The chemical shift indices are then used to identify secondary structure elements in the protein. Four or more consecutive residues denoted ‘-1’ are considered to represent an α-helix, whilst four or more consecutive residues designated ‘+1’ represent a β-strand.

The CSI for MCP-1 is illustrated in Figure 5.10. From this, four β-strand regions (residues Glu59-Lys 64, Gly 68-Tyr 71, Ala 76- Thr 79 and Thr 85-Leu 87) and one helical region (residues 30-33) were identified. From consideration of NMR-solution structures of other CP-modules, MCP-1 is expected to consist primarily of β-sheets. This is confirmed by the X-ray crystal structure of MCP-1,\(^2\) which shows that each module is comprised of five β-strands. The β-strands of MCP-1 identified in the crystal structure are shown in Figure 5.10 (a). Whilst reasonable agreement between the CSI and crystal structure of MCP-1 is observed, the CSI provides only an estimate of protein secondary structure elements and cannot be used absolutely.
5.5.4 Regular secondary structures from spin-spin couplings $^{3}J_{HN,Ha}$

The relationship between vicinal spin-spin couplings, $^{3}J_{HN,Ha}$, and the torsion angle $\phi$, corresponding to the rotation about the N-C$_{a}$ bond, correlate to protein secondary structure elements. These relationships are shown in Table 5.3.

<table>
<thead>
<tr>
<th>Secondary Structure Element</th>
<th>$\phi$ /degrees</th>
<th>$^{3}J_{HN,Ha}$ /Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$-helix</td>
<td>-57</td>
<td>3.9</td>
</tr>
<tr>
<td>$\beta_{10}$ helix</td>
<td>-60</td>
<td>4.2</td>
</tr>
<tr>
<td>antiparallel $\beta$-sheet</td>
<td>-139</td>
<td>8.9</td>
</tr>
<tr>
<td>parallel $\beta$-sheet</td>
<td>-119</td>
<td>9.7</td>
</tr>
</tbody>
</table>

Table 5.3 Relationship between secondary structure elements and $^{3}J_{HN,Ha}$ values. Adapted from Wüthrich.

The $^{3}J_{HN,Ha}$ coupling constants for MCP-1 are plotted in Figure 5.10 (b). The $\beta$-sheets identified from the crystal structure are also shown. In these regions the $^{3}J_{HN,Ha}$ coupling constants are generally in the range 8-10 Hz providing further evidence that the protein contains mostly $\beta$-sheets.
Figure 5.10 (a) CSI for MCP-1. β-sheets identified from the crystal structure of MCP-1,2 are denoted with an arrow.

Figure 5.10 (b) $^{1}J_{H,H_1}$ couplings for MCP-1. β-sheets identified from the crystal structure of MCP-1,2 are denoted with an arrow.
Hydrogen bonds from D$_2$O exchange experiments

Amide protons that are involved in hydrogen bonds or are inaccessible to the solvent can be identified by acquiring spectra on a sample freshly dissolved in D$_2$O since these protons are slow to exchange with deuterium. $^{15}$N-labelled MCP-1 was lyophilised and dissolved in D$_2$O and a series of $^{15}$N-HSQC experiments recorded. The amide protons were found to rapidly exchange as illustrated in Figure 5.11 which shows two HSQC spectra recorded after 16 and 28 minutes in D$_2$O. Assignments were made from resonance positions in the $^{15}$N-HSQC acquired in H$_2$O. In the first spectrum recorded, 23 amide resonances are observed, decreasing to 12 in the second. After 39 minutes only one resonance, corresponding to the overlapped R59/L87 resonance, was observed. These data can be incorporated into structure calculations as distance restraints.

It was observed that the slowly exchanging amide protons identified by this analysis are located within the β-sheets determined from the crystal structure of MCP-1,2. This data is illustrated in Figure 5.12. Hα-Hα resonances identified within the 2D-spectra acquired on MCP-1 show good correlation to the secondary structure elements of the crystal structure.
Results and Discussion

Figure 5.11 (a) Non-exchanged amide protons after 16 min in D₂O.

Figure 5.11 (a) Non-exchanged amide protons after 28 min in D₂O.
Figure 5.12. Secondary structure diagram for MCP-1. β-Sheets were identified from the crystal structure of MCP-1.2. Hydrogen bonds and Hα-Hα NOEs were identified from D2O exchange experiments and the 2D-NOESY data acquired on MCP-1 respectively.
5.5.6 NMR Structure Determination of MCP-1

The analysis of the secondary structure of MCP-1 based on the CSI, calculation of the $^{3}J_{\text{HN,Ha}}$ coupling constants and determination of those residues involved in hydrogen bonds, indicates that MCP-1, as a single module in solution has a β-sheet structure. The determination of the tertiary structure of the protein depends largely on a consideration of inter-proton distances derived from NOESY experiments. Since nuclear Overhauser effects (NOEs) have an $r^6$ distance dependency, these experiments provide information about pairs of protons that are close together in space. Generally an inter-proton distance of up to 5Å gives rise to an NOE. By determining large numbers of distance restraints it is possible to calculate the overall tertiary structure of a protein.

Structure calculations for MCP-1 were performed within the program CNS [229] using a simulated annealing protocol. The protocol incorporated floating stereochemistry and included active swapping of prochiral centres in random order using a metropolis-style acceptance criterion [230]. Following backbone and resonance assignments in ANSIG the NOESY spectra were integrated to determine the peak intensities and consequently the distance restraints which were set as strong, 2.7Å; medium 3.3Å; weak, 5.0Å; very weak 6.0Å.

A total of 453 NOEs were utilised in the calculation along with distance restraints to represent thirteen hydrogen bonds (identified on the basis of slowly exchanging amides and supporting NOEs) and the two disulphides. The distribution of experimentally determined derived distance restraints as a function of residue number is shown in Figure 5.13. Since only a portion of the total NOEs were assigned in the time available the distribution is
Figure 5.13 Number of distinct NOEs used in structure calculation as a function of residue number. Black bars correspond to intra-residue NOEs; dark grey, sequential NOEs; light grey, short range NOEs \([i \leftrightarrow i+(2-4)]\); white, medium and long range NOEs \([i \leftrightarrow i+(>4)]\).
uneven. Some residues, for example Tyr 62 and Trp 86, have many NOEs reflecting their consensus nature and positions within the hydrophobic core. No NOEs however were assigned for some residues such as Gly 48, Lys 49, Pro52 and the unassigned residues Tyr 28, Val 29 and Asn 83.

A total of 30 structures were generated. Each structure has an associated energy according to the extent to which it satisfies the constraints of both good covalent geometry and the experimental data. Of the thirty structures calculated the thirteen lowest energy structures were selected. The total energies of these structures ranges from 666-1012 kcalmol\(^{-1}\). The average structure was calculated and the thirteen structures superposed on this, as shown in Figure 5.14. Only the residues between the N- and C-terminal cysteine residues (residues Cys35-Cys94) are illustrated. The superposition of the structures was performed on residues 35-70 and 84-94. The average backbone r.m.s.d (based on superposition of residues 35-94 only) of these structures is 2.20. The r.m.s.d. per residue (C\(\alpha\)) is plotted as a function of residue number in Figure 5.15. Only residues 35-94 are illustrated. The best defined parts of the structure correspond to residues for which the most NOEs were assigned. These tend to lie within the \(\beta\)-strands. In contrast, the loops and are least well-defined and exhibit the lowest number of NOEs. Somewhat surprisingly, the "hypervariable loop" is not as relatively poorly defined as it is in several other CP-modules studied previously.

Although the structure determination is at a preliminary stage the overall fold of the protein shows reasonable correlation to that observed for other CP-modules. The superposition of the average structure of MCP~1 (shown in blue) and VCP~4 (shown in yellow) is illustrated
Figure 5.14 Superposition of the thirteen lowest energy structures calculated for MCP–1 on the average structure. The structures were overlaid using Insight II on residues Cys 35-Tyr 70 and His 84-Cys 94.
Figure 5.15 Graphical representation of the r.m.s deviation of the Ca atoms of residues 35-94 as a function of residue number, based on superposition of these Ca atoms from thirteen calculated structures on the average structure.
Figure 5.16 Superposition of the average structure of MCP-1 and VCP-4. The structures were overlaid using InsightII on the disulphides.
in Figure 5.16. Short β-strands run parallel and anti-parallel with the long axis of the molecule and contribute side-chains to the hydrophobic core. The “hypervariable loop” projects to the side of the molecule. N- and C-termini lie at opposite poles of the ellipsoid shape.

5.6 Discussion

Structural characterisation of MCP~1 by DSC, CD and NMR indicates that the protein has a well-defined tertiary structure. MCP~1 shows similar structural characteristics to other CP-modules studied. The in vitro refolded MCP~1 and the wild-type version of the protein expressed in *P. pastoris* show similar structural characteristics. However, site-directed mutagenesis to remove the N-linked glycosylation had a destabilising effect on the protein structure. Possible interpretations of this data were discussed previously. Further structural analysis may identify the nature of this conformational change.

Whilst the structure determination of MCP~1 is not yet complete the overall fold of the protein is typical of other CP-modules studied to date. In conjunction with other structural analysis performed the protein has been shown to be stable, soluble and to have a defined tertiary structure as an isolated CP-module. Studies of the dynamics of the protein are underway.

The structure determination of MCP~1 is part of a study to determine the structure of the extracellular portion of MCP. The crystal structure of MCP~1,2 indicated that this portion
of the protein is flexible. NMR studies are ideally suited to assess this. Understanding the structural characteristics of MCP~1 is important for interpretation of the structural data acquired on MCP~1,2. An HSQC spectrum of MCP~1,2 is shown in Figure 5.17. The spectrum shows some overlap and clustering of resonances. The high resolution solution structure determination of MCP~1 may aid analysis of this data and facilitate the structure determination of MCP~1,2.

Figure 5.13 HSQC spectrum of MCP~1,2. The protein was enzymatically deglycosylated. The sample was analysed under the same conditions used for MCP~1.
6.1 Chemical Synthesis and Recombinant Expression of CP-modules

The production of CP-modules in sufficient yield and purity for structure determination by NMR has previously caused some difficulties. For example, the expression of CP-modules from factor H in *Saccharomyces cerevisiae* proved to be low yielding and consequently did not permit isotopic labelling to aid structure determination. Expression of CP-modules in *P. pastoris* has resulted in higher yields and isotopically (^15N and ^13C) labelled samples have been produced.

This project has explored both synthetic and recombinant methods of protein production. Chemical synthesis offers the only method of introducing NMR-active isotopes site specifically, which could prove advantageous for analysing protein:protein or protein:virus interactions by NMR. One objective of this study was to develop a general protocol for the synthesis, purification and renaturation of CP-modules. However, the studies of MCP-1 and fH-15 indicate that it would be difficult to produce these modules routinely in high yield and purity. In both cases, the target sequence was found to comprise a minor component of the total protein synthesised. Additionally, analysis of the synthetic by-products by mass spectrometry indicated that they were often similar to the target
Conclusions

sequence. Therefore, purification to homogeneity was complex, and in the case of fH-15 was not achieved.

Although the isolation of some pure MCP-1 was achieved, the yield was too low for structural analysis. The synthesis of MCP-1 was not optimised, however, using the methods described here the production of this protein in sufficient yield for high resolution structural studies is unfeasible. Developments in synthetic protein chemistry, such as native chemical ligation and expressed protein ligation, may enable the potential advantages of the methodology to be explored more thoroughly.

The renaturation of disulphide bond containing proteins can prove problematic and result in aggregation and reduced yields. Synthetic proteins are presumably produced in a denatured state requiring the development of a refolding protocol. Therefore, a comparative study was made to assess the renaturation of MCP-1 by expressing a version of the protein in E. coli. The deposition of the protein in inclusion bodies proved advantageous, in the context of this project, the renaturation of the synthetic and recombinant versions of the protein could be compared. Whilst preliminary structural analysis of the protein isolated and renatured from inclusion bodies indicated that it had refolded correctly, a major proportion of the synthetic material was found to be unfolded. This could be a consequence of the heterogeneity of the synthetic sample and the inability of truncated peptides to renature correctly. The successful refolding of the recombinant protein eliminates renaturation as a potential problem with the production of synthetic MCP-1. The attempted renaturation of synthetic fH-15 proved less successful, as monitored by NMR. Again, this could be due to the highly heterogeneous nature of the sample.
The isolation and renaturation of MCP-1 deposited in inclusion bodies produced protein in reasonable yield. However, expression of MCP-1 in *P. pastoris* resulted in significantly higher yields and consequently this version of the protein was selected for structural analysis by NMR. Expression in *P. pastoris* had the additional advantages of producing folded protein with a high degree of purity in the growth medium. The protein was found to be soluble and stable as a single module thus permitting NMR analysis.

A potential disadvantage of expression in *P. pastoris* for NMR studies is glycosylation of the protein. In this study, enzymatic deglycosylation and site-directed mutagenesis to remove the N-linked glycosylation site were compared. In this case it was observed that the mutation selected had a detrimental effect on the conformational stability of the protein. However, this could be a consequence of the choice of mutation and does not necessarily indicate that the carbohydrate is structurally important. Structural characterisation of the wild-type protein may enable the nature of the conformational change to be determined.

### 6.2 Structural Characterisation of MCP-1

Low resolution structure analysis of the recombinant versions of MCP-1 indicated that, in all cases, the module had a stable, well-defined tertiary structure. DSC and CD analysis showed that the protein is similar to other CP-modules studied. 1D $^1$H-NMR analysis indicated, by comparison to MCP-1 expressed in *P. pastoris*, that the version of MCP-1 expressed in *E. coli* was successfully renatured *in vitro*. Furthermore, preliminary NMR analysis of synthetic MCP-1 indicated that it was possible to renature a proportion of this sample.
Analysis of MCP-1(T85A) indicated that mutagenesis to remove the N-linked glycosylation site affected the conformation of the protein. Since this version of the protein appeared to be aggregated, enzymatically deglycosylated wild-type MCP-1 was selected for high resolution structural analysis by NMR. The 2D and 3D NMR spectra acquired on MCP-1 confirm that the protein is stable, soluble and has a well-defined tertiary structure. Structure calculations show that MCP-1 has a tertiary structure similar to those of other CP-modules studied. Further work is ongoing to complete the tertiary structure determination of MCP-1 and to study the dynamics of the protein.

The studies performed thus far form part of an on-going study to determine the solution structure of CP-module-containing proteins. Although the crystal structure of MCP-1,2 has been solved, analysis of the molecule in solution is important due to the flexible nature of the protein. It is anticipated that the structural characterisation of this single module will aid the structure determination of MCP-1,2 in solution. This solution structure, in conjunction with other structural and functional data on the protein may assist in determining the nature of the interaction with measles virus.

6.3 Summary

Comparison of different methods of protein production has indicated that recombinant expression of CP-modules in *P. pastoris* provides a reliable means for producing these proteins in sufficient yield and purity for structure determination. The production of MCP-1 in this organism produced protein that is stable, soluble and has a well-defined
tertiary structure. Structural characterisation of MCP–1 indicates that the module is characteristic of other CP-modules studied.
REFERENCES

17. Barlow, P. N., *Personal communication*
76. Seya, T., Hara, T., Matsumoto, M. G. and Akedo, A., *J. Immunol.*, (1990), 145, 238-245

208


164. GST Gene Fusion System Manual, Pharmacia

165. Pichia Expression Kit Manual, Invitrogen


173. Homans, S. W., *J. Magn Reson.*, (1990), 90, 557


176. Uhrin, D. *Personal communication,*
179. Rothwell, N., *BSc (Chemistry) Honours Project,* (1995), University of Edinburgh
180. Shaw, K., *Personal communication*
187. *Synthesis Notes,* Novabiochem

211
217. Mullin, N. P., *Personal communication*
221. Kirkitadze, M. D., *Personal communication*
225. Wüthrich, K. *NMR of Proteins and Nucleic Acids*, (1986), John Wiley and Sons
230. Raine, A. R. C., Smith, B. O. and Damaillie, P. J., *Unpublished results*
231. Herbert, A., *Personal communication*
LECTURES AND COURSES ATTENDED

Organic Section Research Seminar Series, Department of Chemistry, University of Edinburgh, 1995 - 1999

Departmental Colloquia, Department of Chemistry, University of Edinburgh, 1995 - 1999

Edinburgh Centre for Protein Technology Seminar Series, University of Edinburgh 1997-1999

RSC Perkin Division Scottish Regional Meeting, University of Glasgow, December 1995
RSC Perkin Division Scottish Regional Meeting, University of Edinburgh, December 1996

SCI Graduate Symposium (Scotland) on Novel Organic Chemistry, University of St. Andrews, March 1996

Structures: Form and Function Symposium, University of Edinburgh, March 1996

British Biophysical Society and RSC NMR Discussion Group - NMR of proteins and their Interactions, University of Leicester, April 1996

Ames Symposium - Combinatorial Libraries, University of Edinburgh, May 1996
- Biomolecular Structure and Mechanism, University of Edinburgh, June 1998

Merck, Sharp and Dohme Postgraduate Lecture Series, University of Edinburgh, April 1997


Mass Spectrometry - the Emerging Tool for Proteomics, University of Edinburgh, September 1998

XVIIth International Complement Workshop, Rhodes October 1998
Poster: Synthesis and expression of Complement Control Protein Modules

Edinburgh and Dundee Scientific Exchange and Lectures, University of Edinburgh, February 1999

Structural Biochemistry Symposium, University of Edinburgh, July 1999