Crystallographic studies on Macrophage Inflammatory Protein-1α

John K F Maclean

Submitted in part fulfilment for the degree of Doctor of Philosophy

Department of Chemistry
University of Glasgow

September 1997

DECLARATION

This thesis has been written in accordance with University regulations and is less than 50,000 words in length. The work contained herein was performed by the author unless otherwise stated.

John K F Maclean

September 1997

ABSTRACT

Macrophage Inflammatory Protein-1α is a member of the Chemokine family of inflammatory mediators. In addition to its chemotactic properties, MIP-1α is an inhibitor of the proliferation of primitive stem cells, and is therefore potentially useful in cancer therapy. It has also recently been shown to protect certain subsets of human cells from infiltration by HIV-1 virus.

The crystal structures of three structural mutants of the Chemokine Macrophage Inflammatory Protein-1α have been determined. The structure of a monomeric mutant has been determined to 2.9Å, that of a dimeric mutant to 2.3Å, and the structure of a tetrameric mutant has been determined in two crystal forms at 2.05Å and 1.65Å.

This thesis outlines the process by which these mutant structures were solved and compares them with other Chemokine structures. The potential implications of the structures of the MIP-1α mutants are investigated. The complex network of Chemokines and Chemokine receptors is described, and attempts are made to rationalise receptor-binding specificities of MIP-1α and related proteins on the basis of the MIP-1α structures.

Acknowledgements

Firstly I would like to thank Neil Isaacs for his help and encouragement throughout the project. In particular I am grateful to him for letting me work on such an interesting project, and for allowing me so much freedom to choose the direction the project was going in.

I am indebted to Paul Emsley, Steve Prince and Dina Fotinou for their patience and for teaching me so much about proteins and crystallography. I am also indebted to a large number of people at Glasgow who have helped me throughout the project, and for generally making the lab a nIcE place to work. In no particular order, they are (nicknames where possible, of course): Old Andy, LB, Gerry, Zorb, Pix, Steve, Purp, Spoons, Jeremy, Tino, Isobel, Aleks, Dangerous Dave, David Adam, Asghar Torabi, Ken Muir, Garry... and Betty. My apologies to anyone whom I have omitted to mention. I also have to thank John Coggins for paying me money to do something else and then turning a blind eye when I had to concentrate on the thesis-thing...

I am also very grateful to Neil, Purp, Aleks and Sharon for their patience in working their way through my initial attempts at this "thing", and for their comments and advice.

I would also like to thank Gerry Graham and his group at the Beatson for providing me with such pure protein, and thereby ensuring that I didn't really have to do any of that biochemistry nonsense...

Contents

1. Introduction. 9

1.1 The Immune System. **9**

1.2 Cytokines. **9**

1.2.1 Definition and Historical Significance. 10

1.2.2 Biological Activities. 11

1.2.3 Location and Targets. 12

1.3 Chemokines **13**

1.3.1 Definition and Physical Properties of the Family. 13

1.3.2 Sequence Comparison. 16

1.3.3 Biological Activities of the Chemokines. 22

1.3.4 Known Chemokine Structures. 32

1.4 Biology and Biochemistry of MIP-1α **33**

1.4.1 Physical Properties. 33

1.4.2 Known Biological Activity. 34

1.4.3 Receptor Interactions. 34

1.4.4 SCI Activity of MIP-1α. 35

1.4.5 HIV-suppressant Activity of MIP-1α. 36

1.5 Mutants of MIP-1α **36**

1.5.1 Rationale behind the Mutants Produced. 37

1.5.2 Biological Activities of Mutants. 37

1.5.3 Other Mutants. 37

2. Crystallisation 39

2.1 Crystallisation: Materials and Methods **39**

2.1.1 Analysis of Mutant Molecular Weight and Purity by SDS-PAGE 43

2.1.2 Analysis of Mutant Sample Monodispersity by Dynamic Light Scattering 44

2.2 Crystallisation of Monomeric Mutant **44**

2.3 Crystallisation of Dimeric Mutant **45**

2.4 Crystallisation of Tetrameric Mutant **47**

2.5 Final Crystallisation Conditions **50**

2.6 Discussion **50**

3. Structure Solution 53

3.1 Data Collection and Processing **53**

3.2 Merging and Scaling of Data **55**

3.3 Initial Molecular Replacement Attempts **55**

3.4 Derivatisation **56**

3.5 Solution of Monomeric Mutant **57**

3.6 Solution of Dimeric Mutant **58**

3.7 Solution of Tetrameric Mutant **59**

3.8 Molecular Replacement: Why did it not work? **62**

3.8.1 Molecular Replacement using NMR Search Models 63

3.8.2 Molecular Replacement using MCP-1 63

3.8.3 Fundamental Differences between NMR and X-ray Structures 64

3.9 Refinement **65**

4. Chemokine Structures 68

4.1 Descriptions of other Chemokine Structures. **68**

4.1.1 Introduction 68

4.1.2 α-Chemokine Structures. 68

4.1.3 β-Chemokine Structures. 70

4.1.4 Structures of other Chemokine Types. 73

4.2 Descriptions of MIP-1α Structures. **73**

4.2.1 Structures of Monomeric and Dimeric Mutants. 73

4.2.2 Structure of Orthorhombic Form of Tetrameric Mutant. 80

4.2.3 Structure of Monoclinic Form of Tetrameric Mutant. 85

4.3 Comparison of Chemokine Structures. **86**

4.3.1 Comparison of the MIP-1α Monomers. 86

4.3.2 Comparison between MIP-1α and β-Chemokine Structures. 87

4.3.3 Comparison between β- and α-Chemokine Structures. 90

4.4 Conclusion **91**

5. Implications of MIP-1α Structures 92

5.1 Aggregation Mechanism **92**

5.1.1 Analysis of Surface Accessibility Data 94

5.1.2 Comparison with other Chemokines 96

5.1.3 Shapes of Potential Tetramers 97

5.1.4 Specific Interactions within Potential Tetramers 98

5.1.5 Do the Mutations make sense? 99

5.1.6 Higher Aggregation States 102

5.1.7 Conclusion 107

5.2 The Potential Role of Ca^{2+} Ions. **110**

5.2.1 Are Ca^{2+} Ions involved in Aggregation? 111

5.2.2 Are Ca^{2+} Ions necessary for Crystallisation? 111

5.2.3 Are Ca^{2+} Ions involved in MIP-1α Interactions with Receptor? 112

5.3 Areas of the Structures involved in Receptor-Binding **112**

5.3.1 Chemokine Receptors 112

5.3.2 β-Chemokine Receptors 114

5.3.3 CCR5 115

5.3.4 CCR1 120

5.3.5 CCR2 121

5.3.6 CCR3 121

5.3.7 CCR4 122

5.3.8 CCR6, CCR7 and CCR8 . . . 122
5.3.9 CCR9, CCR10 and other receptors . . . 122

5.4 Conclusion **123**
5.4.1 Stem Cell Inhibition by MIP-1α . . . 123
5.4.2 HIV-1/CD4 Interactions of MIP-1α . . . 123
5.4.3 GAG-Binding by MIP-1α and β-Chemokines . . . 124
5.4.4 Redefinition of Chemokine Family Boundaries . . . 127

6. Summary . . . 133

Appendix **134**

A. Alternate Names of Chemokines . . . 135

B. Crystallisation Theory . . . 136

B.1 Theory of Crystallisation **136**

B.2 Factors affecting Crystallisation **138**
B.2.1 Temperature . . . 138
B.2.2 pH . . . 139
B.2.3 Precipitant . . . 139
B.2.4 Protein Concentration . . . 140
B.2.5 Detergent . . . 140
B.2.6 Additives . . . 140

C. Molecular Replacement . . . 141

C.1 Overview **141**

C.2 Rotation Function **141**

C.3 Translation Function **141**

C.4 Rigid-Body Refinement **142**

C.5 Modifications of the basic MR technique **142**

D. Glossary 144

E. Structure Validation 145

List of Figures

2.1 Dimeric mutant: Absorption Spectrum 1 40
2.2 Dimeric mutant: Absorption Spectrum 2 41
2.3 SDS-PAGE Gel 43
2.4 Trigonal Bipyramidal crystal of dimeric mutant 46
2.5 Orthorhombic crystal form of tetrameric mutant 48
2.6 Monoclinic crystal form of tetrameric mutant 49

3.1 Difference Patterson: Harker Section z=$\frac{1}{3}$ 57
3.2 Rotation function solutions for dimeric mutant 59
3.3 Translation function solutions for dimeric mutant 60
3.4 Translation function solutions for dimeric mutant 60
3.5 Rigid-body fitted solutions for dimeric mutant 61
3.6 Rigid-body fitted solutions for dimeric mutant 61

4.1 Monomer structures of IL-8 and bPF4. 69
4.2 Dimer structures of IL-8 and MIP-1β. 71
4.3 MCP-1 tetramer. The two distinct MIP-1β-type dimers are coloured differently. . . 72
4.4 Secondary structure elements of MIP-1α. 74
4.5 Hydrogen bonding pattern at the dimerisation interface. 76
4.6 Arg 21 H-bonding to Asn 64 and to β-strand I. 77
4.7 Hydrogen-bonding pattern in the antiparallel β-sheet. 78
4.8 Aromatic residues in the hydrophobic core of MIP-1α. 79
4.9 Structure of the MIP-1α tetramer. 81
4.10 Electron density surrounding the Calcium ion. 83
4.11 Calcium coordination: View down crystallographic two-fold axis. 84

5.1 Tet2: View showing the single dimer-dimer contact. 94
5.2 Influence of mutated residues on aggregation 101
5.3 Structure of Aggregation Model Agg1 106

5.4 Interactions involved in formation of Aggregation Model Agg2 108
5.5 Structure of Aggregation Model Agg2 109
5.6 Structure of Aggregation Model Agg2 109
5.7 Schematic diagram of CC receptor 113
5.8 Models of MCP and MIP interactions with receptors 119
5.9 Residues involved in the proposed cationic cradle in MIP-1α 125
5.10 Summary of available data on β-Chemokines and their receptors 131

B.1 Phase diagram for protein in solution. 137
B.2 Phase diagram for possible crystallisation experiments.. 138

E.1 Ramachandran plot: Dimeric mutant 146
E.2 Ramachandran plot: Tetrameric mutant, monoclinic form 147
E.3 Ramachandran plot: Tetrameric mutant, orthorhombic form 148

List of Tables

1.1 Sequence alignment of representative members of each of the Chemokine families 15
1.2 Sequence identities between proteins of the β-Chemokine families 17
1.3 Sequence Alignment of all α-Chemokines 18
1.4 Sequence Alignment of all β-Chemokines 19
1.5 Cluster plot based on sequence identities of β-Chemokines 20
1.6 Defining characteristics of the Chemokine subfamilies 22
1.7 Chromosomal location of β-Chemokine Genes. 23
1.8 Known Chemokine Receptors and their Ligands. 30
1.9 Murine MIP-1α mutants 37

2.1 Initial and final protein concentrations 39
2.2 Results of initial crystallisation trials for monomeric mutant 45
2.3 Results of further crystallisation trials for monomeric mutant 45
2.4 Results of crystallisation trials for dimeric mutant 47
2.5 Results of crystallisation trials for tetrameric mutant 47
2.6 Final Crystallisation Conditions for each mutant 50

3.1 Statistics for datasets collected from crystals of monomeric mutant 54

3.2 Statistics for datasets collected from crystals of dimeric mutant 54
3.3 Statistics for datasets collected from crystals of tetrameric mutant 54
3.4 Unit Cell parameters and contents for all crystal forms of MIP-1α mutants 55
3.5 Refinement statistics for MIP-1α mutants 66

4.1 RMS deviations between MIP-1α monomers 87
4.2 RMS deviations between MIP-1α (dimer mutant) and other β-Chemokines 88
4.3 RMS deviations between other β-Chemokine monomers 88
4.4 RMS deviations of Cysteine atoms in the various MIP-1α monomers 89
4.5 RMS deviations of Cysteine atoms in the β-Chemokine family. 90
4.6 RMS deviations between α- and β-Chemokine monomers 91

5.1 Solvent-Accessible Areas for Tetrameric Subunits of MIP-1α 95
5.2 Surface Areas buried by potential MIP-1α multimerisation interactions 95
5.3 Solvent-Accessible Areas for subunits of other Chemokines 97
5.4 Areas buried on formation of multimers for other Chemokines 97
5.5 Sequence alignment: Residues that may be involved in CCR5/CCR2 binding . . . 117

E.1 Geometric statistics for the refined dimer and tetramer structures. 145

Abbreviations

Ac	Acetate
AMS	Ammonium Sulphate
CAC	Cacodylate (Sodium Dimethylarsinate) (Buffer)
CIT	Sodium Citrate (Buffer)
GPCR	G-protein coupled (7-transmembrane-helical) receptor
HEPES	Piperazineethanesulphonic Acid (Buffer)
IL	Interleukin
IPR	Isopropanol
*k*D	kiloDaltons
MES	4-Morpholinoethanesulphonic Acid (Buffer)
MIR	Multiple Isomorphous Replacement
MIRAS	Multiple Isomorphous Replacement with Anomalous Scattering
MPD	2-Methyl-2,4-Pentanediol
MR	Molecular Replacement
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline (150mM NaCl, 15mM Phosphate buffer)
PEG	Polyethylene Glycol
PIPES	Piperazinebis(Ethanesulphonic Acid) (Buffer)
SDS	Sodium Dodecyl Sulphate
SIR	Single Isomorphous Replacement
SIRAS	Single Isomorphous Replacement with Anomalous Scattering

1. INTRODUCTION.

1.1 The Immune System.

Normal function of the immune system is essential to life. It encompasses the movement, targeting, adhesion, activation and eventual removal of a vast array of cell types involved in defending the host organism from hostile invasion. At the heart of this process is the inflammatory response, where leukocytes or white blood cells are transported to a site of injury and activated in order to alleviate the problem. However, the balance between a beneficial inflammatory response and a harmful one is often very fine. If the magnitude of the immune response is out of proportion to the injury it is trying to combat, the outcome can be more harmful to the host organism than the original injury would have been. Examples of detrimental immune responses include arthritis, asthma and the ability of retroviruses such as HIV to infiltrate and assume control of immune cells.

The recruitment and targeting of the specific subset of leukocytes required to challenge a particular injury is therefore of paramount importance. Of equal, and perhaps even greater, import is the subsequent removal of those cells from the site of injury and the downregulation of leukocyte production to prevent harmful reaction to the immune response.

It is therefore clear that the mediators of inflammation are of extreme importance in the survival of many organisms. Knowledge of the structures of these molecules and their receptors will allow greater understanding of the interactions they make with each other and with leukocytes, and may shed light on their roles in immune and autoimmune disorders.

1.2 Cytokines.

The Cytokines are a broad family of proteins, ranging in size from 8 to over 50*k*D. They are all Inflammatory mediators, and often have other properties which influence the Immune response, such as the stimulation or inhibition of the growth and proliferation of progenitor cells. In general, they are structurally unrelated, although they can be divided into several distinct superfamilies based on the three-dimensional structures of their monomers. Membership of the Cytokine family is therefore based on biological activity rather than structural definitions.

1.2.1 Definition and Historical Significance.

The term "Cytokine" was originally intended to refer to molecules with some form of modulatory activity over immune cells[1,2,3]. Derived from terms for "cell" and "motion", it was appropriate for proteins which mediated the proliferation and distribution of a great variety of immune-system cell types.

More recently it became clear that any attempt to clarify the definition of the family on the basis of the source, target or activity of its members would be impossible[3,4]. Although some Cytokines are produced by only one cell type, most have multiple cellular sources. Similarly, the majority of Cytokines are active on more than one cell type, and many have a variety of different activities on different types of cell. While most of the known Cytokines are fairly small soluble proteins, membrane-bound forms of several Cytokines, have also been identified[5,6].

Today the term refers to a large and diverse group of over 120 proteins[7] involved in regulating the normal activity of the immune system, controlling the activity and growth of individual cells, and in mediating interaction and communication between cells.

Within the large family of proteins to which the term "Cytokine" can be applied, three broad structural groups can be defined:

- α-Cytokines — Consist of various types of four helix bundles.
- β-Cytokines — A more diverse collection of primarily β-sheet proteins.
- α + β — A more clearly defined group of proteins containing significant amounts of both helix and sheet.

Each of these consists of many distinct proteins and can be further subdivided[8].

Within the α-Cytokines, more than ten proteins have been structurally characterised, and have been shown to belong to three subfamilies. The Long-chain four-helix bundle structures are characterised by human Growth Hormone (*h*GH) and Granulocyte-Colony Stimulating Factor (G-CSF), and the Short-chain four-helix bundles by Interleukins-2 and -4 (IL-2 and IL-4). Both groups contain a novel "up-up-down-down" four-helix motif, with very little additional secondary structure. The third family contains Interferon-γ and IL-10, and is distinct from the other two as its members dimerise to form two adjacent four-helix bundles, with at least one helix in each coming from the other monomer.

The β-Cytokines also consist of three subfamilies, although there is little similarity between them. The Cysteine knot growth factors, such as Platelet-derived Growth Factor (PDGF)[9] and

human Chorionic Gonadotropin (*h*CG)[10, 11], are heterodimers containing a knotted central cluster of three disulphides. They have little secondary structure, containing just a few strands of β-sheet. The β-trefoil structures, such as IL-1β[12] and human Fibroblast Growth Factor (*h*FGF)[13] are trimeric β-barrels. The third group are β-sandwich structures, the only example of this type to date being the Tumour Necrosis Factor (TNF) trimer[14].

In some classifications, the α+β-Cytokines comprise only the Chemokines, while in others proteins such as Angiogenin[15] are also included. The Chemokines all have a monomer which consists of three strands of β-sheet and a single α-helix[16], whereas Angiogenin has a fold similar to that of Ribonuclease A[17], containing a long β-sheet and three α-helices.

1.2.2 Biological Activities.

As their names often reflect, the majority of the Cytokines were originally identified by their possession of a distinct biological activity. However it is also a general property of the Cytokines that there is a very high degree of cross-reactivity and hence redundancy in their activities. In fact there are many examples where two or more Cytokines can replace one another with little or no reduction in biological activity or specificity[18].

Consequently it has been difficult in many cases to precisely determine the natural function of many of the Cytokines. There are several whose primary biological activities remain a mystery.

It now seems clear that the overlap in function observed between so many Chemokines can not be accidental. However, it is not clear why such a complex and frequently overlapping network of chemotactic proteins is necessary. It has been argued that these functional overlaps are extremely important in providing as flexible an immune response as possible[2]. While many of the Cytokines have apparently identical major functions, some of their minor ones differ. As Cytokines seldom occur individually, the overall activity of a Cytokine mixture could be subtly altered by the production of different combinations of proteins.

While all of the proteins defined as Cytokines are involved in some way in modulating the normal function of the immune system, their specific activities are often very different. The family contains a huge variety of proteins involved in chemotaxis and cell signalling, as well as Growth Factors and glycoprotein hormones[19].

Amongst the chemotactic proteins there are examples which preferentially attract neutrophils, monocytes, lymphocytes, almost any other immune system cell type, or any combination thereof[3]. There are also vast differences in the relative potencies and specificities of many of those whose chemotactic targets are similar. Most of them will only show chemotactic activity within a fairly

narrow range of concentrations[2, 20, 3], and if the concentration is too low or too high they will be unable to attract their target cells, further enhancing the flexibility of the Cytokine network.

For example, TGF-β, the most potent monocyte chemoattractant known, is most effective at a concentration less that one millionth of the optimal concentration for MCP-1, which is also a highly potent monocyte chemoattractant[3, 4]. Both proteins can be produced together in an immune response, and would have relatively high concentrations close to the site of production. It is possible to envisage a scenario where the local concentration of TGF-β is so high that it is relatively inactive as a chemoattractant, whereas MCP-1 is still extremely effective. As both proteins diffuse away from the site of production, TGF-β concentration is reduced to the level at which it becomes an exceedingly effective monocyte chemoattractant, and therefore is able to recruit monocytes from distances far beyond the range of MCP-1. Thus TGF-β and MCP-1 can act in a complementary fashion, although at first glance it might appear that they would compete with one another.

Among the Cytokines which are also growth factors are proteins of fundamental importance in development, including Nerve Growth Factor (NGF), Transforming Growth Factor-β (TGF-β) and the glycoprotein hormones[19] - Chorionic Gonadotropin (CG), Leutinising Hormone (LH), Thyroid Stimulating Hormone (TSH) and Follicle Stimulating Hormone (FSH). The proteins of the TGF-β family affect the normal cell cycle, as well as stimulating or inhibiting cell growth because of their ability to modulate the effects of other proteins which act as growth factors. The specific role of NGF is less well understood, although it is known to control the development and survival of certain subsets of neurons.

The glycoprotein hormones are a group of four related heterodimeric proteins. All of them have important roles in endocrinilogy. CG, FSH and LH are regulators of the reproductive cycle, whereas TSH is involved in metabolic control[21].

1.2.3 Location and Targets.

The Cytokines should be considered primarily to be mediators of immune system cells, controlling chemotaxis, cell interaction, cell growth and in some cases cell motility. The plethora of other properties belonging to individual Cytokines are generally secondary to this. However there is a potentially important subset for which these "secondary" effects seem to be the major, if not sole biological activities that can be ascribed[22].

The sites of production of the Cytokines reflect their primary activities. Those with chemotactic properties are rapidly produced at sites of inflammation and wound-healing, and then can

control the migration of cells of the immune system towards those sites. Almost all cell types are able to produce some population of Cytokines, and most of these proteins can be generated by many different cells. In addition, many Cytokines have a vast range of functions, as described above, making the definition of specific cellular targets and sites of production relatively meaningless.

For this type of rationale to be of use, the Cytokines must be considered as groups of related molecules, expressed close to one another, and having similar but non-identical profiles of biological activity. Only in this way can the intricate Cytokine network be simplified to any extent.

1.3 Chemokines

As mentioned above, the Chemokines are a distinct Cytokine subset consisting mainly of small proteins with molecular weights from 8-15*k*D. They can be subdivided into four distinct families, based on their relative distributions of Cysteine residues, and on some of their physical properties. The first and second families, which are called the α- and β-Chemokines, have been known for around twenty years and were thought to be the only two branches of the Chemokine subfamily tree. However, there have recently been reports of two other proteins which do not conform to the Cysteine distribution pattern of any other known Chemokine. Lymphotactin[23] was proposed as the first and so far, only, member of the γ-Chemokine family, and Neurotactin[5] as the single member of the δ-Chemokines.

1.3.1 Definition and Physical Properties of the Family.

The majority of the Chemokines are small proteins, typically containing 70 amino acids and having molecular weights of approximately 8*k*D. However there are some slightly larger Chemokines, which are discussed in Section 1.3.2.

Members of the four Chemokine subfamilies share a high degree of sequence homology. Indeed it is the patterns of certain conserved residues which have been used to define the boundaries between them.

The α-Chemokines were formerly known as the "CXC Chemokines", since the two cysteine residues closest to the N-termini of these proteins were always separated by a single residue (See Table 1.1). Conversely, in the β-Chemokines, these two cysteines are adjacent in the sequence (See Table 1.1). In both cases there are two other completely conserved cysteine residues and two conserved disulphides, which are formed between Cys^{I}-Cys^{III} and Cys^{II}-Cys^{IV}, where Cys^{I}

is the Cysteine closest to the N-terminus. In this and subsequent sequence alignments, human Chemokine sequences are used.

In Lymphotactin, Cys^{I} and Cys^{III} are absent, which results in just a single disulphide in this molecule (See Table 1.1). The alternate notation for the γ-Chemokines is therefore the "C Chemokines".

In Neurotactin, Cys^{I} and Cys^{II} are separated by three residues (See Table 1.1), and hence have been referred to as "CX_3C Chemokines". It has been proposed that Neurotactin and any related proteins be classified as δ-Chemokines[5]. The disulphide bonding pattern observed for the first four Cysteines in this protein is identical to those of the α- and β-Chemokines. However Neurotactin is distinct from these families in having another three Cysteine residues close to its C-terminus (not shown in Table 1.1, two of which form a third disulphide bond. Neurotactin is also referred to as Fractalkine in some publications[24].

Aside from the pattern of conserved disulphides, there are several other characteristics that define the boundaries between the families. Although there is some overlap in biological activities, particularly between the α- and β-Chemokines, the sites of action and specific activities of the proteins of the four subfamilies are fairly distinct. The physical properties of the proteins are also a defining feature. The proteins of the α- and β- families are small, soluble and heparin-binding, and generally have a molecular weight in the range from 8-10*k*D[25]. Lymphotactin is slightly larger than this, with a molecular weight of 12.5*k*D[23]. Neurotactin, on the other hand, is considerably larger than any other Chemokine, and unlike the others is proposed to be a membrane protein with a C-terminal cytoplasmic domain. From its sequence[5], the expected molecular weight will be in excess of 40*k*D. However, it has also been reported that the Chemokine-like fragment of Neurotactin can be detached from its membrane anchor under certain conditions to produce a soluble Chemokine similar in size to the α- and β-Chemokines[24].

Despite the differences outlined above, the sequence similarity between most members of the Chemokine family is significant. As would be expected, the greatest degree of sequence identity is found between members of the same Chemokine subfamily, and can be as high as 88% in the case of the α-Chemokines *gro*-α, -β and -γ[3]. Sequence identity within the β-Chemokine family is as high as 71%, between Monocyte Chemotactic Protein-1 (MCP-1) and Monocyte Chemotactic Protein-3 (MCP-3)[3]. However there are also sequence identities of 20-40% between α- and β-Chemokines[16], suggesting some structural similarities between these two subfamilies. Sequence comparisons between Lymphotactin and various α- and β-Chemokines showed a maximum sequence identity of 26% with MCP-2. The Chemokine domain of Neurotactin shows a maximum

```
                1       10        20        30        40        50        60        70        80
                |-------+---------+---------+---------+---------+---------+---------+---------+-
   MIP-1alpha   ASLAADTPTAC---CFS-YTSRQIPQNFIADY-FETSSQCSKPGVIFLTKRSRQVCADPSEEWVQKYVSDLELSA
        Il-8    ...VLPRSAKELRC--QCIKTYSKPFHPKFIKELRVIESGPHCANTEIIVKLSDGRELCLDPKENWVQRVVEKFLKRAENS
Lymphotactin    ...IVEGVGSEVSDKRTCVSLTTQR-LPVSRIKTYTITEGS-L--RAVIFITKRGLKVCADPQATWVRDVVRSMDRKSN--TRN...
 Fractalkine    ...LAGQHHGVTKCNITC-SKMTSK-IPVALLIHYQQNQAS-CGKRAIILETRQHRLFCADPKEQWVKDAMQHLDRQAAALTRN...
     Consensus            C  C s yt r iP  i  y  e  s C k  !If tkr r vCaDP e WVq  v  l  a
```

Table 1.1: Sequence alignment of representative members of each of the Chemokine families

sequence identity of 40% with MCP-1[5], 30% with Macrophage Inflammatory Protein-1α (MIP-1α) and only 23% with RANTES (Regulated on Activation, Normal T-cell Expressed and Secreted). Comparison between Lymphotactin and Neurotactin shows a sequence identity of 26%. A summary of the sequence identities between the various members of the β-Chemokine family is given in Table 1.2.

1.3.2 Sequence Comparison.

Table 1.3 shows a sequence alignment for all known members of the α-Chemokine family. One member of each of the β, γ and δ is also shown for comparison. This and subsequent sequence alignments were performed using MultAlin[26] and the original Blosum62[27] symbol comparison table.

Table 1.4 shows a similar alignment for all known members of the β-Chemokine family. Table 1.5 shows a cluster plot based on the sequence alignment shown in Table 1.4. Sequence identity increases towards the left of the diagram, with MIP-1β being most similar to MIP-1α, and TECK being least so. It is clear from this diagram that the MIP-like proteins cluster together, as do the MCP-like proteins. The Exodus proteins also seem to form a distinct cluster. This diagram suggests that sequence comparison may indicate subtle differences between groups of β-Chemokines, which could lead to further subdivision of the family. These concepts are discussed in greater detail in Section 5.4.4.

Another striking fact regarding the sequence homology within the Chemokine family is apparent when proteins from different species are compared. The majority of the Chemokines identified to date have been detected in both mice and humans, and some for quite a variety of species. However there are some exceptions, most notably the α-Chemokine IL-8, for which only a human form is known[4], although the murine IL-8 receptor has been identified[28]. The sequence identity between homologous proteins from different species is generally very high; for example, the sequence identity between the human and murine MIP-1α proteins is 75%, and between the corresponding MIP-1β proteins this figure is 78%.

For Lymphotactin, the identity between the murine and human versions of the proteins is slightly lower than this at around 60%. The figure is around 76% for the Chemokine portion of Neurotactin, very similar to the figures for MIP-1α and MIP-1β. Lymphotactin and Neurotactin are somewhat larger than the other Chemokines, and therefore only a part of their sequences actually comprise the "Chemokine domain". There appears to be no significant similarity between the residues in the extra portions of these molecules. In the case of Lymphotactin, there is no

Chemokine	MIP-1β	RANTES	DC-CK1	HCC-1	MCP-1	MCP-2	MCP-3	MCP-4	MCP-5	Eot	Eot-2
MIP-1α	**57%**	**48%**	**60%**	**51%**	38%	39%	36%	**40%**	32%	**40%**	**44%**
MIP-1β	-	**45%**	**43%**	**50%**	36%	37%	38%	**40%**	34%	**42%**	38%
RANTES	-	-	37%	39%	34%	37%	35%	32%	34%	**43%**	30%
DC-CK1	-	-	-	**52%**	36%	35%	**42%**	38%	38%	37%	37%
HCC-1	-	-	-	-	30%	**43%**	38%	**44%**	30%	39%	38%
MCP-1	-	-	-	-	-	**62%**	**72%**	**60%**	**68%**	**68%**	**41%**
MCP-2	-	-	-	-	-	-	**58%**	**56%**	**57%**	**66%**	**44%**
MCP-3	-	-	-	-	-	-	-	**60%**	**59%**	**70%**	**42%**
MCP-4	-	-	-	-	-	-	-	-	**56%**	**62%**	**48%**
MCP-5	-	-	-	-	-	-	-	-	-	**56%**	33%
Eotaxin	-	-	-	-	-	-	-	-	-	-	38%

Chemokine	Exo	Exo-2	Exo-3	MPIF-1	MIP-1γ	I-309	MDC	TARC	HCC-2	ILINCK	TECK
MIP-1α	36%	35%	30%	**54%**	**50%**	38%	32%	31%	**59%**	36%	*41%*
MIP-1β	32%	**45%**	35%	**44%**	**40%**	32%	34%	28%	39%	**45%**	30%
RANTES	34%	27%	31%	33%	36%	23%	33%	38%	**43%**	37%	31%
DC-CK1	35%	30%	39%	**49%**	**41%**	36%	36%	35%	**50%**	36%	32%
HCC-1	36%	32%	30%	**49%**	**44%**	37%	28%	34%	**44%**	**48%**	33%
MCP-1	30%	**41%**	30%	**40%**	30%	**41%**	32%	29%	38%	31%	28%
MCP-2	32%	**40%**	34%	**43%**	35%	39%	33%	28%	**43%**	35%	*40%*
MCP-3	31%	**40%**	27%	**41%**	31%	**45%**	35%	34%	**41%**	37%	30%
MCP-4	24%	38%	29%	**41%**	32%	**40%**	33%	37%	37%	36%	*35%*
MCP-5	27%	**41%**	29%	33%	29%	39%	**42%**	38%	35%	33%	27%
Eotaxin	29%	**42%**	31%	**40%**	34%	**42%**	34%	35%	**40%**	31%	38%
Eotaxin-2	31%	37%	36%	**41%**	34%	37%	31%	24%	**45%**	35%	*37%*
Exodus	-	36%	36%	33%	37%	-%	26%	28%	30%	**42%**	31%
Exodus-2	-	-	37%	37%	28%	32%	29%	27%	39%	39%	**40%**
Exodus-3	-	-	-	30%	36%	27%	35%	33%	35%	32%	31%
MPIF-1	-	-	-	-	**46%**	36%	28%	31%	**70%**	37%	35%
MIP-1γ	-	-	-	-	-	37%	28%	26%	**41%**	38%	37%
I-309	-	-	-	-	-	-	34%	30%	**40%**	33%	-%
MDC	-	-	-	-	-	-	-	*40%*	34%	35%	-%
TARC	-	-	-	-	-	-	-	-	32%	31%	33%
HCC-2	-	-	-	-	-	-	-	-	-	28%	35%
ILINCK	-	-	-	-	-	-	-	-	-	-	39%

Table 1.2: Sequence Alignment of all β-Chemokines

Table 1.3: Sequence Alignment of all α-Chemokines

Table 1.4: Sequence Alignment of all β-Chemokines

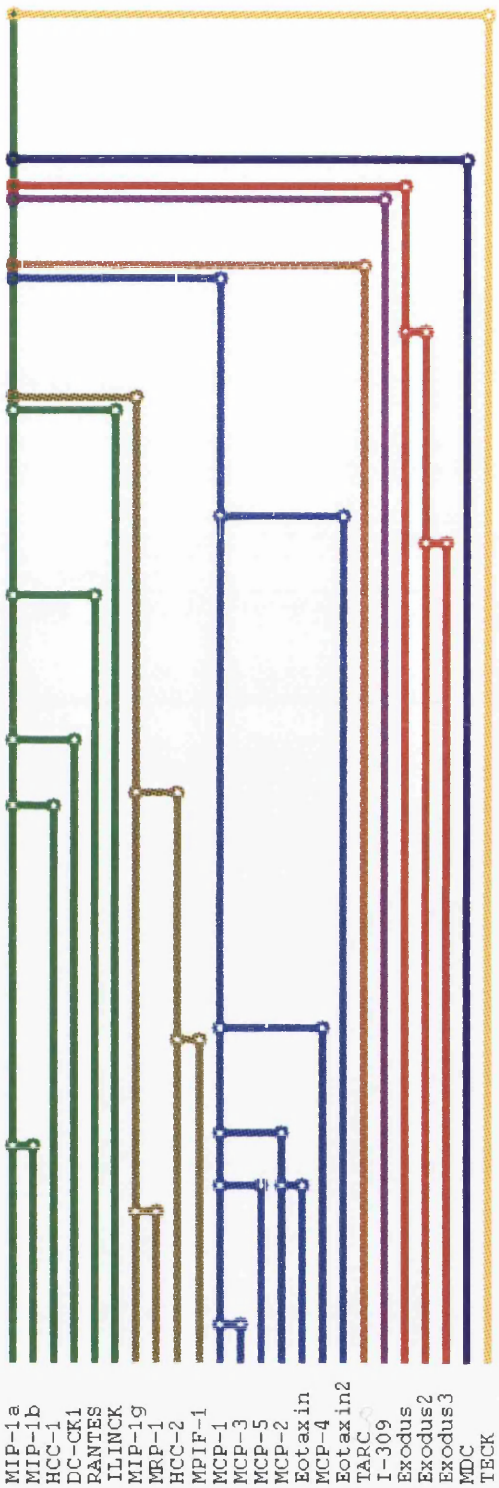


Table 1.5: Cluster plot based on sequence identities of β-Chemokines

indication from the sequence as to the probable conformation of the additional peptide. Conversely, for Neurotactin, one stretch of nineteen conserved hydrophobic residues seems to indicate a membrane-spanning domain, and therefore the residues to the C-terminal side of this region correspond to a cytoplasmic domain[5]. There are three "extra" Cysteine residues in Neurotactin, one of which is encompassed by the potential membrane- spanning region, while the other pair are expected to form a disulphide bond within the cytoplasmic domain. The remaining residues, connecting the Chemokine domain to the membrane-spanning domain, have been suggested simply to be a flexible linker region, even though this region is larger than the others combined.

In addition to these two proteins, there are several other recently identified members of the β-Chemokine family that are somewhat larger than those already known. Murine MIP-1γ, which was first identified[29] in 1996, consists of 100 amino acids, including an extra twenty-five amino-terminal residues not present in MIP-1α and MIP-1β. Human Macrophage Inflammatory Protein Related Protein-1 (MRP-1[30]) consists of 96 residues, and has insertions at both the N and C termini which do not occur in other β-Chemokines. MPIF-1[31] and HCC-2[32] also appear to be related to MIP-1γ. These proteins are distinct from the other members of the family in having six cysteine residues within the portions of their sequences which correspond to the full lengths of the other β-Chemokines. The two extra conserved Cysteine residues occur in positions that would be proximal were MIP-1γ assumed to have the same secondary structure elements as the other Chemokines, suggesting that a third conserved disulphide bond would be formed. These proteins may represent the first members of a new subdivision of the β-Chemokine family. Some of the MCP-1 homologues isolated from species other than human have also been shown to have a substantial number of extra residues at their C-termini, giving molecular weights up to around 14*k*D.

The final piece of evidence which suggests partitioning of the Chemokines into distinct subfamilies is perhaps the most compelling of all. When the chromosomal location of the various Chemokine genes is compared, it is apparent that the Chemokines are subdivided in a way which corresponds exactly with partitioning on the basis of cysteine distribution (See Table 1.6). Thus the "CXC", "CC", "C" and "CX_3C" Chemokines are found at distinct and different chromosomal locations. In addition, there is similar evidence to show that β-Chemokine genes occur in several discrete clusters which are close together, but nevertheless distinct, on human chromosome 17. This suggests that the β-Chemokine family may consist of several subfamilies with identical cysteine distribution but which can be distinguished according to their exact chromosomal location.

All the α-Chemokine genes map to human chromosome 4 and to mouse chromosome 5. The majority of the β-Chemokine genes map to human chromosome 17 and to mouse chromosome 11,

although there are several recently discovered exceptions (See Table 1.7). Lymphotactin differs from both families as its gene maps to mouse chromosome 1[23] and also to human chromosome 1[33]. Finally Neurotactin is different again[5], its gene being located on human chromosome 16q and on mouse chromosome 8. Two β-Chemokine genes have also been located on human chromosome 16q[34,35], and it is possible that there may be some relationship between these proteins and Neurotactin.

Chemokine Family	Representative Protein	Cys Distribution	Chromosomal Location	
			Human	Murine
α	IL-8	CXC	4q12-q21	5
β	RANTES	CC	17q11-q21	11
γ	Lymphotactin	C	1	1
δ	Neurotactin	CXXXC	16q	8

Table 1.6: Defining characteristics of the Chemokine subfamilies

1.3.3 Biological Activities of the Chemokines.

1.3.3.1 History.

The first Chemokines were identified as long ago as 1977, when the isolation and crystallisation of Platelet Factor 4 (PF4) was reported[36]. PF4 had been known for substantially longer than this however, and had first been described as a heparin-neutralising activity in extracts from blood platelets in 1951[37]. Until very recently it was common for Chemokines to first be identified in this way, due to a new or interesting biological activity rather than similarity to related proteins. However, enormous advances in genetic screening capabilities coupled with an increased understanding of the important physical properties of the various Chemokine subfamilies have led to the discovery, production and sequencing of several new Chemokines in a relatively short space of time. In addition, DNA sequences of even more candidate Chemokines have been identified[4].

While PF4 and another product of blood platelet α-granules, β-Thromboglobulin (β-TG), were being studied in the late 1970s[38], it was not until the late 1980s that it became clear that the Chemokines were a family of related proteins. The identification, in rapid succession, of IL-8[39] (then called Macrophage Derived Neutrophil Chemotactic Factor - MDNCF), Interferon-γ inducible Protein-10 (γ-IP10)[40] and 9E3[41] brought about an understanding that there was a structural relationship between all of these molecules. The characterisation of the two disulphide bonds in β-TG[38] and the recognition that the four cysteine residues were conserved in each molecule[39]

Chemokine	Chromosomal Location	
	Human	Murine
MIP-1α	17q11-17q21	11
MIP-1β	17q11-17q21	11
HCC-1	17q11.2	*not known*
DC-CK1	17q11.2	?
RANTES	17q11.2-q12	11
ILINCK	17q11.2	*not known*
MIP-1γ	*not known*	11
MRP-1	*not known*	11
HCC-2	17q11.2	*not known*
MPIF-1	?	*not known*
MCP-1	17q11.2-q12	11
MCP-2	17q11.2	?
MCP-3	17q11.2-q12	11
MCP-4	17q11.2	*not known*
MCP-5	*not known*	11
Eotaxin	17q11.2-17q21	11
Eotaxin-2	?	*not known*
TARC	16q13	*not known*
I-309	17q11.2	11
Exodus	2q33-q37	*not known*
Exodus-2	9p13	*not known*
Exodus-3	9p13	*not known*
MDC	16	*not known*
TECK	?	*not known*

Table 1.7: Chromosomal location of β-Chemokine Genes.

In Table 1.7, *not known* implies that the protein has not been isolated for that species, whereas '?' implies that the protein is known but the chromosomal location has not been reported.

led to the notion of the "CXC Chemokine" family[42].

Following these discoveries regarding the nature of the protein family itself, several other members were identified. First the proteins Connective Tissue Activating Peptide (CTAP-III)[43], Platelet Basic Protein (PBP)[44] and Neutrophil Activating Peptide-2 (NAP-2)[45], which were also known to be platelet α-granule products, were shown to be cleavage products belonging to the same root. PBP is the intact molecule, with CTAP-III, β-TG and finally NAP-2 being formed by successive N-terminal processing.

The next of the α-Chemokines to be discovered was *gro*-α[46] (then called Melanoma Growth Stimulatory Activity - MGSA). Two closely related proteins with similar biochemical activities, *gro*-β and *gro*-γ, were reported some time after this. Several other α-Chemokines have been identified since, including Epithelial cell-derived Neutrophil-Activating protein (ENA-78)[47] for which only human and bovine forms are known. Other α-Chemokines for which reliable protein sequences are available include Granulocyte Chemotactic Protein-2 (GCP-2)[48], LIX[49], stromal cell-derived factor-1 (SDF-1)[50] and Monokine induced by Interferon-γ (Mig)[51].

While the first α-Chemokines have been known for a considerable time due to their obvious biological activities, it was not until 1986 that the first member of the β-Chemokine family was isolated. LD78 was the first to be discovered[52], followed in 1988 by Act-2[53]. The subsequent identification of two closely related murine proteins, MIP-1α[54] and MIP-1β[55] led to the proposal that LD78 and Act-2 were the human homologues, and today LD78 is known as human MIP-1α and Act-2 as human MIP-1β.

MCP-1[56] and RANTES[57] were the next β-Chemokines to be isolated. These two proteins have been studied extensively since their discovery, and are by far the best-characterised of this family of proteins. Recently a substantial number of other β-Chemokine proteins and sequences have been identified, and these have often been named according to their similarity to one of the known proteins. Hence MCP-2[58], MCP-3[58], MCP-4[59] and MCP-5[60] are now characterised, in addition to I-309[61], Eosinophil Chemotactic Protein (Eotaxin)[62], Eotaxin-2[63, 31], HCC-1[64], HCC-2[32], Exodus[25], Exodus-2[65], (also called SLC), Exodus-3[66], (also called ELC, or EBI1-ligand Chemokine), Myeloid Progenitor Inhibitory Factor-1 (MPIF-1)[31], Macrophage-derived Chemokine (MDC)[34], Thymus and activation-regulated Chemokine (TARC)[35], Dendritic-cell-derived C-C Chemokine (DC-CK1)[67], ILINCK[68] and TECK[69]. Based on analysis of their sequences, they are all expected to be structurally similar to MCP-1, RANTES and the MIP-1 proteins.

Very recently though several other sequences have been identified that while possessing the definitive CC motif, are somewhat larger than the other β-Chemokines and do not follow patterns

of other conserved residues that occur throughout the group. This suggests that there may be at least one more group of "CC-Chemokines" which are structurally distinct from the others. Murine MIP-1γ[29], and the human proteins MRP-1[30, 70], MPIF-1[31] and HCC-2[32] have all been shown to possess an additional pair of conserved cysteine residues that are expected to form a third disulphide, and I-309 also possesses a third pair of cysteine residues which occur at positions different to those proteins. The identification, in addition, of the first members of the γ- and δ-Chemokines, suggests that there is still a great deal to be discovered regarding variations in structure and sequence within the Chemokine family.

1.3.3.2 Alternate Names.

As will already have become clear, the field of Chemokine study is now littered with defunct and confusing alternate terms. This is a problem that is likely to diminish as understanding of the underlying organisation of the various subfamilies improves, particularly as many of the newly discovered Chemokines are being categorised and named due to their similarity to existing proteins.

The problem has been engendered as a result of the way in which the first known Chemokines were isolated and identified. Proteins such as MIP-1α, RANTES, PF4 and β-TG were named according to their first known activities, and in some cases they were known by several names until it became clear that the same factor was responsible for each activity. However, as it becomes apparent that most of the Chemokines have a variety of biological activities, it is clear that a few of the existing names refer to their minor properties and hence are not particularly appropriate.

Nevertheless, continuing rationalisation of the naming conventions of the Cytokines in general should ensure that things will continue to become easier for those with an interest in the subject. As the majority of Chemokines are now being identified from searches of DNA databases, the problem is unlikely to arise again. The current convention is for each uncharacterised Chemokine to be given a temporary label, with the real name of the protein following once it has been biochemically characterised. There are two new α-Chemokines, CKα1 and CKα2, and at least thirteen new β-Chemokines, CKβ1-CKβ13, that fall into this category[4], although some of these have now been characterised and renamed. Appendix A contains an extensive list of the alternative names by which the various Chemokines are and have been known.

1.3.3.3 Classification based on Activity.

As has already been explained, the Chemokines were initially identified by biochemical purification of particular biological activities from cell supernatants. For this reason they tended to be associated purely with a single specific activity, until the breadth of function of IL-8 was realised[3], and led others to explore the full range of activities mediated by the Chemokines.

IL-8 was originally identified as a selective chemoattractant for neutrophils that had no activity towards monocytes. Subsequently, a range of experiments showed that IL-8 was able to activate as well as mobilise neutrophils, inducing shape change, a rise in cytosolic free Ca^{2+} concentration and an increase in the respiratory burst, amongst other activities[3]. It was also shown to be chemotactic and mildly activating for basophils, chemotactic for a subset of T-lymphocytes, and to have minor but significant non-chemotactic effects on monocytes, keratinocytes and melanoma cells.

The other α-Chemokines which were known relatively early, such as PF4, Gro-α and β-TG, were initially grouped together with IL-8 as they were also shown to be chemotactic for neutrophils. Indeed, one early definition of the α- and β-Chemokines was that α-proteins were chemotactic for neutrophils but not monocytes, β-proteins were chemotactic for monocytes but not neutrophils. More careful study showed this partitioning to be slightly flawed, since the α-Chemokine PF4, for example, is chemotactic for both neutrophils and monocytes. However, PF4 is chemotactic for neutrophils only at concentrations much higher than required for IL-8, and lacks the neutrophil-stimulating activity of IL-8 as well. PF4 has been shown to be chemotactic for fibroblasts and to have effects on megakaryocytopoiesis and angiogenesis, but its prime biological activity is likely to be as a Glycosaminoglycan (GAG)-binding factor and consequently as a controlling factor in the coagulation process.

Of the other α-Chemokines, only NAP-2 shows neutrophil chemotactic and activating activity to rival IL-8. PBP, CTAP-III and β-TG, the precursors of NAP-2, are relatively inactive, although β-TG does function as a fibroblast chemoattractant. CTAP-III was reported to function as a growth factor[43], but subsequent studies have cast doubt on these findings[44]. The *gro*-proteins are also mildly chemotactic and activating for neutrophils, and *gro*-α has been shown to have growth stimulatory activity for melanoma cells and fibroblasts. SDF-1 has recently been shown to be a lymphocyte chemoattractant, although it was also known as PBSF (pre-B cell growth stimulating factor) as it stimulates B-cell progenitor proliferation. It also stimulates monocytes, neutrophils and peripheral blood lymphocytes, and appears to be a very specific ligand for the important α-Chemokine receptor CXCR4. γIP-10 and Mig have been shown to be specific ligands for the α-Chemokine receptor CXCR3, which is expressed only on activated T-lymphocytes,

suggesting they are involved in the regulation of lymphocyte recruitment. The other members of the α-Chemokine family have been studied to a lesser extent and these proteins are not well characterised biochemically.

Of the β-Chemokines, only MIP-1α has been shown to have any stimulatory effect on neutrophils, and even then it is not chemotactic. However, the majority of the β-Chemokines have significant effects on monocytes. Monocyte chemotactic properties have been ascribed to RANTES, I-309, MCP-1, MCP-2 and MCP-3, and monocyte stimulation has been observed both by these proteins and by MIP-1α and MIP-1β.

In addition, MCP-1, RANTES and MIP-1α, but not MIP-1β, are highly effective stimulators of basophils. Of these proteins, RANTES has been shown to be the most effective basophil chemoattractant, whereas MCP-1 is the most powerful stimulant.

Eosinophils are also much more susceptible to β- rather than α-Chemokines. RANTES is a powerful chemoattractant and activator, with MIP-1α being slightly less effective, and MCP-1 and MIP-1β inactive. The β-Chemokines Eotaxin and Eotaxin-2 are both extremely potent and selective eosiniphil chemoattractants.

Some other studies have suggested that several β-Chemokines have chemotactic effects on certain types of T-lymphocytes, and that several also inhibit the proliferation of certain progenitor cells, but their results have at times been contradictory. RANTES, MIP-1α and MIP-1β certainly appear to have some effects, although the specific properties ascribed to each protein differ, whereas MCP-1 is inactive. Exodus-3 is a highly specific ligand for the receptor CCR7, and has been shown to be a chemoattractant and activator for lymphocytes. Both MPIF-1 and Eotaxin-2 have been shown to be chemotactic for resting T-lymphocytes and neutrophils, while MPIF-1 also attracts monocytes. Both of these proteins affect hematopoietic progenitor cells, MPIF-1 suppresses proliferation of progenitors of the granulocyte and monocyte lineages, whereas Eotaxin-2 suppresses the proliferation of multipotential hematopoietic progenitor cells. MDC is produced specifically by cells of macrophage lineage, and is a chemoattractant for monocytes, macrophage-derived dendritic cells and for natural killer cells. Exodus is chemotactic for peripheral blood mononuclear cells, and is expressed by both lymphocytes and monocytes. It inhibits the proliferation of myeloid progenitor cells. TARC is chemotactic for some human T cell lines, and has been shown to be a highly specific ligand for the receptor CCR4. HCC-1 has weak activity on human monocytes and also inhibits the proliferation of some myeloid progenitor cells, but is inactive on some lymphocyte and leukocyte populations. DC-CK1 is chemotactic for T cells, but unlike RANTES and MIP-1α, it specifically attracts naive or inactivated T cells.

Like the α-Chemokines, many of the β-Chemokines are only recently identified and have therefore not been studied extensively. In a few cases all that is available is a protein sequence in a database, predicted from a cDNA sequence. However, even from the limited experimental results available to date, it is clear that the Chemokines as a family have an extensive variety of overlapping functions. It will therefore be extremely difficult to identify the specific properties of each protein, but it will be vitally important to do so in order to fully understand the synergy and the complexity inherent in their interactions.

1.3.3.4 Cellular Sources.

While the specific cellular targets of the various Chemokines are many and varied, the same is also true of the cellular sources of most of the proteins. This complicates the picture still further, particularly as there is no obvious partitioning between α- and β-Chemokines with respect to their cellular sources, in contrast to the obvious differences in their target specificities.

Monocytes and macrophages are sources of many α- and β-Chemokines, with IL-8, *gro*-α,-β,-γ, MCP-1 and MCP-2 among the major products. Lymphocytes are important sources of many of the β-Chemokines, including MIP-1α, MIP-1β and RANTES, but are less significant with respect to the α-Chemokines. Neutrophils have also been shown to produce some α-Chemokines, including IL-8, *gro*-α and *gro*-β.

Chemokine expression has also been observed in a great variety of other cell-types, most notably endothelial cells, epithelial cells and fibroblasts. The production of IL-8, *gro*-α, *gro*-β, *gro*-γ, ENA-78, γ-IP10, MCP-1, MIP-1α and MIP-1β has been reported in one or more of them.

1.3.3.5 Receptors.

Only very recently have the Chemokine receptors begun to be understood to any extent. For several years it was assumed that there would probably be a specific receptor related to each Chemokine protein. It was not until their incredible cross-reactivity became apparent that the reality of receptors specific for certain subsets of Chemokines was recognised. Knowledge of the complete Chemokine receptor system is still sketchy at best, although significant recent advances have been made in receptor identification, cloning and sequencing that will enhance understanding. In addition, the recent implication of some Chemokine receptors in the study of the HIV virus[71] will ensure that the field remains in the forefront of research for the foreseeable future.

All of the known Chemokine receptors are seven-transmembrane-helical G-protein coupled receptors (GPCRs). Although quite a number of them are now known, it is very likely that a

significant number remain to be discovered. They can be partitioned into two distinct subsets, the α-Chemokine receptors which have sequence identities between 36-77%, and the β-Chemokine receptors, with sequence identities between 46-89%. Between the two families, average sequence identities are around 25%. In addition to the obvious relationship between many of these receptors hinted at by the above statistics, there are several important motifs that are conserved throughout either family.

A series of experiments has been able to demonstrate that, in general, it is the amino-terminal extracellular domain of a Chemokine receptor that determines specificity. The α-Chemokine receptors CXCR1 and CXCR2 and the β- Chemokine receptor CCR2B have been shown to obey this rule , although the β-Chemokine receptor CCR5 does not. Four α-Chemokine receptors are currently known, as well as eight β-Chemokine receptors. The majority of them bind two or more known Chemokines, although a few appear to be specific for a single molecule. Viral receptors, which bind both α- and β-Chemokines have been discovered, in addition to the Duffy Blood Group antigen, a "promiscuous" receptor that binds both α- and β-Chemokines, and is implicated in malarial infection. The known Chemokine receptors and their ligand-binding profiles are summarised in Table 1.8.

Identification of the β-Chemokine receptor CCR5 as a coreceptor for macrophage-tropic HIV strains, and the α-Chemokine receptor CXCR4 (formerly *fusin*) as a coreceptor for T-cell-tropic variations of the virus gave research into Chemokine receptors a massive boost. Both of these receptors and their various ligands therefore became the subject of substantial speculation regarding their potential utility in anti-HIV therapies. More recent research has shown that other Chemokine receptors are also implicated in binding different strains of the virus[72, 73], with the result that Chemokine research has assumed a high profile within the research community. Primary HIV-1 infection affects only cells of macrophage lineage, and hence the primary isolates of HIV are known as macrophage-tropic. Disease progression is associated with a switch from macrophage-tropic to T-cell-tropic virus, and associated sequence differences in the gp120 subunit of the viral envelope glycoprotein.

The macrophage-tropic stage is often considered a latent phase, as monocytes and macrophages support only low levels of virus replication. Infected individuals can remain in the macrophage-tropic state for many years, but progression to T-cell-tropic virus brings a dramatic increase in virus levels, as $CD4^+$ T-lymphocytes support much higher levels of virus replication. Typically 99% of viral particles present in infected individuals were produced by T-cells and just 1% by monocytes and macrophages[74].

α-Chemokine receptors	
Receptor	Ligand
CXCR1	IL8
CXCR2	All α-Chemokines containing ELR motif
CXCR3	γIP-10, mig
CXCR4	SDF-1

β-Chemokine receptors	
Receptor	Ligand
CCR1	MIP-1α, RANTES, MCP-3, HCC-2
CCR2	MCP-1, MCP-2, MCP-3, MCP-4, MCP-5
CCR3	Eotaxin, Eotaxin-2, RANTES, MCP-3, MCP-4, HCC-2
CCR4	*MIP-1α, RANTES*, TARC
CCR5	MIP-1α, MIP-1β, RANTES
CCR6	Exodus
CCR7	Exodus-3
CCR8	I-309
CCR9	?
CCR10	MCP-1, MCP-3 *MCP-4, RANTES*

Others		
Receptor	Description	Ligand
CMV US28	Receptor encoded by Cytomegalovirus	MCP-1, RANTES, MIP-1α, MIP-1β
HSV ECRF3	Receptor encoded by Herpes Saimiri virus	IL-8, gro-α, NAP-2
DBGA	Duffy Blood Group Antigen "promiscuous" receptor	Many α- and β-Chemokines

Table 1.8: Known Chemokine Receptors and their Ligands.

Minor ligands are in italics

The recent identification of CXCR4 as the coreceptor for HIV on T-tropic cell lines[75], and of SDF-1 as the specific ligand for this receptor[50], brings an opportunity to use Chemokines or Chemokine analogues to combat both macrophage-tropic and T-cell-tropic HIV strains. One of the most exciting recent studies in the field of HIV treatment has been the continuing progress in the use of triple combination therapy. This is a cocktail of three of the most effective known anti-HIV drugs, including a reverse transcriptase inhibitor and a protease inhibitor. The results suggest that this cocktail suppresses viremia to less than 1% of the level before treatment[74], but even continued administration does not eradicate the virus. One possibility is that the cocktail is efficiently controlling viral replication in T-cells, but is unable to do so in monocytes and macrophages. However, the β-Chemokine cocktail of MIP-1α, MIP-1β and RANTES has already been shown to be able to control viral replication in these cell types[71]. It is possible that a blend of triple combination therapy and β-Chemokine therapy would be able to clear the virus from T-cells, monocytes and macrophages.

1.3.3.6 Importance of Fusin and CCR5.

A landmark paper[71] in 1995 revealed that a cocktail of three β-Chemokines, MIP-1α, MIP-1β and RANTES, was able to inhibit binding of the macrophage-tropic strains of the HIV virus to human cells. Subsequently, it was shown simultaneously by several research groups that a receptor specific for those three Chemokines was responsible for the observations, and was the elusive coreceptor for macrophage-tropic HIV strains. This had a number of vital implications for those studying β-Chemokines and their receptors. Firstly, since the proteins themselves had been shown to be HIV-suppressants, a great deal of interest was generated in the potential use of native or modified proteins in inhibiting the HIV virus. Since the macrophage-tropic strain of HIV is responsible for the initial stages of infection, there was the potential to arrest the spread of the virus through the body before it managed to infiltrate any other cell-types.

Secondly, the receptor also became a viable target for possible mutagenesis and for the generation of inhibitors that would block the binding of the HIV virus without side effects. A CCR5 mutation that was quite prevalent in Northern Europe was discovered to severely inhibit the ability of HIV-1 strains to infiltrate macrophages[76]. It was subsequently shown that the mutation in question was a 32-base-pair deletion which resulted in a severely truncated CCR5 receptor lacking the final three transmembrane segments[76]. The truncated CCR5 receptor did not appear to be able to reach the cell surface[77]. Even more importantly, there appeared to be absolutely no detrimental effects on host fitness associated with this mutation. The CCR5 receptor was therefore proposed

to be redundant as it had already been shown that each of its ligands bound to more than one other Chemokine receptor[76].

Individuals heterozygous for this mutation (denoted CCR5/Δ*ccr*5) were found to have partial resistance to HIV infection[77, 78, 79]. Several studies reported examples of individuals homozygous for this mutation (denoted Δ*ccr*5/Δ*ccr*5) who had been exposed to the virus on numerous occasions but had remained uninfected. It was therefore proposed that normal CCR5 was a regulator of both HIV-1 infection and of disease progression, and that homozygous individuals were therefore highly protected from HIV-1 infection, and were perhaps even immune. Hopes of immunity were short-lived though, as it was subsequently shown that a handful of homozygous individuals have been infected by HIV-1[80]. Careful investigation showed that macrophage-tropic strains are capable of using CCR2b and CCR3 as fusion cofactors, in addition to CCR5. While CCR5 is undoubtedly responsible for macrophage infection by the majority of HIV-1 strains, these other receptors might allow a small but significant degree of infection by minor HIV-1 strains, giving the virus an initial foothold in the body. However, it is also possible that the infected homozygous individuals were directly exposed to T-cell-tropic virus, from which they had no protection. While this phenomenon is rare, it is thought to be a plausible explanation.

1.3.4 Known Chemokine Structures.

Given the recent surge of interest in the biochemistry of the Chemokines, it is no surprise that there have been significant advances in the elucidation of Chemokine structure within the same period. While the majority of the structures solved to date have been of α-Chemokines, this is due mainly to the fact that they were discovered earlier and have been studied more extensively. The β-Chemokine structures that are known have provided many surprises, and the continuing identification of many new and potentially different β-Chemokines promise more.

All of the α-Chemokines have proved to have highly similar structures so far, the only real difference being the aggregation state of the protein involved, with both dimers and tetramers known. In each case the dimers have the same topology, and the tetramers are built up in the same way, as described in greater detail in Section 4.1.1.

This is very much in contrast to the situation observed for the β-Chemokines, where the presence of two types of dimer and one type of tetramer is already known from just four structures, and where a second type of tetramer is reported in Section 4.2.2. These variations lend further weight to the suggestion that the β-Chemokine family as currently defined could be divided into at least two and possibly several more subfamilies. The α-Chemokines seem at present to comprise

a single well-defined group of proteins.

1.4 Biology and Biochemistry of MIP-1α

As was explained in Section 1.3.3.1, the human form of MIP-1α was the first β-Chemokine to be identified. The structural studies presented here relate to the murine protein, which was identified slightly later. As it was one of the earliest known β-Chemokines, there is a substantial amount of biochemical data available for the protein, but recent focus has tended to be placed on RANTES and MCP-1 rather than MIP-1α, as the effects of those proteins seem to be more ubiquitous.

1.4.1 Physical Properties.

Murine MIP-1α is a fairly typical member of the β-Chemokine family, consisting of 69 amino acids and having a native molecular weight of 7883Da by Electrospray Mass Spectrometry[81]. Despite being a heparin-binding protein, it is overall slightly acidic, with a pI of 5.14[I]. The aggregation state of the protein under physiological conditions is difficult to quantify, although several reports have suggested that, in common with the β-Chemokines MIP-1β[82], γ-IP10[40] and RANTES[83], MIP-1α forms large aggregates with mass of at least 100*k*D, which correspond to a potential dodecamer. This property, allied to the propensity that many Chemokines demonstrate to bind tightly and non-specifically to plastics and glycans, has made biochemical characterisation difficult. Nevertheless, MIP-1α was shown to exist as an equilibrium mixture of several aggregation states, including homogeneous monomer and tetramers[84], in addition to the much larger aggregates. It was also shown that the active form of the protein was the monomer[85], and that the aggregation phenomenon was completely reversible[86]. All this biochemical evidence pointed to a particularly interesting equilibrium between the various aggregation states, but also predicted that crystallographic analysis of the protein would be extremely difficult, due to inhomogeneity. In addition, any potential clinical use for the protein would be tempered by the probability that the aggregated states would have varying efficacy and tissue penetration, thereby making specific dosing difficult. Previous investigation of the associative properties of MIP-1α had shown that the higher aggregation states could be disrupted to tetramers in high salt concentrations, suggesting that the interactions responsible for their formation were electrostatic[86], whereas those responsible for formation of the tetramer were considerably more hydrophobic[84].

[I] Theoretical pI from the SWISS-PROT database

1.4.2 Known Biological Activity.

As the outline of the biological activities of the Chemokines in Section 1.2.2 has already demonstrated, each of these proteins has several types of activity on a variety of cells, and MIP-1α has probably the broadest known activity spectrum.

It has recently become apparent that MIP-1α has one property that is shared by very few of the closely related Chemokines. It was shown that the protein Stem Cell Inhibitory factor (SCI), a potent inhibitor of stem cell proliferation and differentiation, was identical to MIP-1α[22]. It therefore occupies a pivotal role in the control of the stem cell compartment and in the regeneration of cells of hematopoietic lineage. Not only is MIP-1α an extremely potent inhibitor, its effects are also reversible and specific, suggesting a potential clinical role as described in more detail below. While several other β-Chemokines are inhibitors of various progenitor cell types, only MIP-1α inhibits the most primitive hematopoietic stem cells. The potential effects of MIP-1α on leukocyte population are therefore substantially more important than the possible effects of proteins such as MPIF-1, Exodus and Eotaxin-2.

In addition to this novel activity, MIP-1α displays a broad range of chemotactic properties. It has been shown to be chemotactic for monocytes, T-cells, basophils and eosinophils, in common with many other β-Chemokines. However, in general these effects are only apparent at concentrations higher than required to obtain similar results from a variety of other Chemokines, suggesting that these chemotactic properties may represent minor biological activities. MIP-1α also appears to have effects on mast cells, natural killer cells, dendrocytes, B-cells and neutrophils[3, 20, 18], which are affected by few, if any, of the other β-Chemokines. However, the major activity of MIP-1α actually appears to be stem cell inhibition[22].

1.4.3 Receptor Interactions.

MIP-1α has been shown to bind to and induce a signal from several of the known β-Chemokine receptors. The binding profile is similar to but not identical to those of MIP-1β and RANTES. These three proteins have all been shown to be ligands for the receptor CCR5[87, 88, 89, 90], while MIP-1α and RANTES, but not MIP-1β are also ligands for CCR1[28] (originally called the "MIP-1α/RANTES receptor") and CCR4[4]. This suggests a greater similarity between MIP-1α and RANTES, despite the greater sequence identity between MIP-1α and MIP-1β, as was shown in Table 1.2. However, RANTES is also a ligand for the receptor CCR3[4] ("the Eotaxin receptor") while MIP-1α is not, so the profile of receptor interaction for these two proteins is also different.

1.4.4 SCI Activity of MIP-1α.

Bone marrow contains a population of immortal self-replicating "stem cells", which are hematopoietic progenitors. This means that these cells are the progenitors of all the body's leukocytes. Each stem cell has the potential to divide when it is active, or in cycle, producing another immortal stem cell and a leukocyte progenitor. The new stem cell will remain in the bone marrow and the leukocyte progenitor will develop into a particular type of leukocyte as it is transported out of the bone marrow and becomes a part of the body's immune system. This is an ongoing process, continually refreshing and replenishing the cells of the immune system.

In normal bone marrow, around 10% of stem cells are in cycle, while the remainder are said to be quiescent, and are unable to proliferate. The administration of MIP-1α inhibits the proliferation of the stem cells by forcing some of the active cells into the quiescent state. This provides a potential clinical use for MIP-1α in chemotherapeutic treatments[91], as outlined below.

The majority of chemotherapeutic agents are cytotoxic for normal cells as well as tumour cells, and therefore the administration of the agent kills a percentage of the patient's immune cells, including those stem cells which are in cycle. These are replenished by an increase in the percentage of active stem cells in bone marrow, leading to increased immune cell production.

Unfortunately, in these treatments it is generally necessary to administer serial doses of the chemotherapeutic agent over a period of weeks or months. This can lead to a catastrophic reduction in the numbers of available bone marrow stem cells in a short period of time, as the number of active stem cells will continue to rise, placing more and more of them at risk. Eventually, they will be unable to properly replenish the immune cells and the patient will become vulnerable to secondary infection.

MIP-1α could be administered shortly before each dose of the chemotherapeutic agent, reducing the number of stem cells in cycle at that particular time, and effectively protecting many of those cells which would otherwise have been killed. The deterioration of stem cell numbers would be less dramatic, and additionally there would be the potential to use higher doses of the chemotherapeutic agent with a resultant reduction in the number of doses required. MIP-1α has been shown to effectively inhibit stem cell proliferation both *in vitro* and *in vivo* in clinical tests on mice[91].

1.4.5 HIV-suppressant Activity of MIP-1α.

As was explained in Section 1.3.3.6, MIP-1α, MIP-1β and RANTES are the ligands for the β-Chemokine receptor CCR5, and a cocktail of all three proteins is able to inhibit the infiltration of human cells by the macrophage-tropic strain of the virus[71]. Each of these molecules is therefore potentially of great use in understanding the mechanism of HIV infection and in designing useful therapies. An understanding of the specific structural requirements for binding to the receptor CCR5 would allow the design of small molecules which could imitate the HIV-suppressant activities of the protein cocktail. Such an understanding could come only from the three-dimensional structure of the receptor itself, or from careful study of the three-dimensional structures of the three ligands, and comparison between them and related proteins which are unable to bind to CCR5.

The structure solution of CCR5, while possible, would be extremely difficult. Many previous attempts to discern the structures of seven-transmembrane helical GPCRs have met with failure. As well as the difficulties inherent in handling transmembrane proteins which have been removed from their membranes, most of these proteins can only be expressed at very low levels, and seldom can sufficient material be produced for structural investigation to be feasible. The structures of RANTES and MIP-1β, however, are known[82, 92], and several structures of MIP-1α mutants are presented within this thesis. Taken together, these structures should provide a great deal of information regarding the regions of these molecules critical for HIV inhibition. Specifically, comparison with the recently reported structures of the non-CCR5 binding β-Chemokines MCP-1 and MCP-3 should allow the definition of specific portions of the structures which interact with certain receptors. Previous biochemical studies have identified regions of the MCP-1 molecule that are implicated in receptor-binding, and by comparison of the structures of MIP-1α and MCP-1 it will be possible to determine which of these regions are similar in the two molecules and hence may be involved in receptor interactions. It may also be possible to discern a structural basis for the different receptor-binding profiles of the various β-Chemokines. These comparisons will be explored in greater detail in Section 5 after the MIP-1α structures have been described.

1.5 Mutants of MIP-1α

Since the aggregation of MIP-1α had been shown to be mediated by electrostatic interactions, several mutants were prepared in which one or more charged residues had been neutralised[86]. The intention was to discover whether specific regions of the molecule, or even specific residues, were responsible for aggregation. The results indicated that there were several residues which did seem

to be critical to the aggregation interactions.

1.5.1 Rationale behind the Mutants Produced.

It was noticed that the proteins MIP-1α, MIP-1β and RANTES, in contrast to non-aggregating β-Chemokines, had several acidic residues close to their C-termini. It was therefore these residues, specifically Glu 60, Asp 64 and Glu 66, which were the initial targets for neutralisation[86]. The effects of these neutralisations is shown in Table 1.9.

Mutant	C-terminal sequence													Aggregation State	Mol. wt.
Wild type	W	V	Q	E	Y	I	T	D	L	E	L	N	A	Aggregate	100*k*D
1	W	V	Q	E	Y	I	T	D	L	**Q**	L	N	A	Tetramer	32*k*D
2	W	V	Q	E	Y	I	T	**N**	L	**Q**	L	N	A	Dimer	16*k*D
3	W	V	Q	**Q**	Y	I	T	**N**	L	**Q**	L	N	A	Monomer	8*k*D
	57	*58*	*59*	*60*	*61*	*62*	*63*	*64*	*65*	*66*	*67*	*68*	*69*		

Table 1.9: Murine MIP-1α mutants

The most dramatic effect was the neutralisation of Glu 66. This single mutation completely prevented formation of aggregated species larger than around 30*k*D.

1.5.2 Biological Activities of Mutants.

While the biological activities of the mutants have not been studied to the same degree as those of the native protein, it has been shown that each of the mutants has an almost identical capacity to inhibit the proliferation of haemopoietic stem cells[86]. This is probably not unexpected given previous evidence that MIP-1α is functional as a monomer and that it spontaneously disaggregates in solution as required. These mutants therefore provide a potential means of administering a specific MIP-1α dose, which would be impossible using the native protein[85]. While the biological activity of the mutants is almost identical to that of the native protein *in vitro*, it is entirely possible that this might not be the case *in vivo*, due to differences in protein transport and storage[86]. The structures of the mutants may help to explain some of these observations and perhaps allow the design of other mutants with higher potency than native MIP-1α.

1.5.3 Other Mutants.

The production of two other MIP-1α mutants has been reported. The mutant BB10010[93], which involves neutralisation of Asp 26, resulted in a homogeneous dimer. This implied that residue Asp

26 was involved in dimer-dimer interactions to form tetramers, and was potentially involved in tetramer-tetramer interactions giving larger aggregates as well.

Another mutant, called Hepmut, involving neutralisation of Lys 44 and Arg 45, has been reported[94]. This particular variant also appeared to have a molecular weight corresponding to a dimer, although its elution peak in a gel filtration experiment was broad, suggesting the dimer is present in equilibrium with other higher aggregates. These two residues were also proposed to be implicated in the electrostatic aggregation process[86], and the fact that a dimer results on their neutralisation supports that hypothesis. A second result of this neutralisation is that Hepmut does not bind heparin or any other proteoglycan to a significant extent. The patch of positive potential that residues Lys 44 and Arg 45 represent may therefore constitute the major heparin binding determinant in MIP-1α. Since Hepmut was inactive on human monocytes, and was shown to be unable to bind to the human receptor CCR1[94], it is likely that this region is also involved in interaction between MIP-1α and CCR1, and hence is important to the monocyte chemoattractant property of MIP-1α.

It is possible to explain precisely why both of these mutants result in dimers by comparing them with the three mutant structures presented in Section 4.1.1.

2. CRYSTALLISATION

The first stage in the structural determination of the MIP-1α mutants was to obtain three-dimensional crystals suitable for diffraction from each of them. The growth of protein crystals is often difficult, and has proved to be a bottleneck in many previous protein structure determinations, but in the case of the MIP-1α mutants it proved to be relatively straightforward. The theory of crystal generation and growth is outlined in Appendix B.

2.1 Crystallisation: Materials and Methods

Each of the mutants was supplied in standard Phosphate-Buffered Saline (PBS) and 30% Acetonitrile. The Acetonitrile was removed by blowing nitrogen gas across each of the samples. The original and resulting protein concentrations were as shown below:

Mutant	Initial concentration (mg/ml)	Final concentration (mg/ml)
PM3	0.47	0.77
PM2	0.51	0.71
PM1	0.42	0.59

Table 2.1: Initial and final protein concentrations

To prepare the samples for crystallisation trials, 1ml of each was added to 9ml of HPLC grade water and centrifuged at 5000*g* using a 3kD-cutoff Centricon microconcentrator, to give a ten-fold dilution of the PBS buffer. The volume of each sample was reduced to around 100 μl, to give a theoretical protein concentration of 4-7 mg ml^{-1}. However, the final concentrations after this stage were somewhat lower than this, as the mutants bound readily to the Centricon membrane. It was estimated from Ultraviolet absorption measurements (See Figures 2.1 and 2.2) that around 25% of each of the samples was lost in this way, giving final protein concentrations for crystallisation trials of 3-5 mg ml^{-1}.

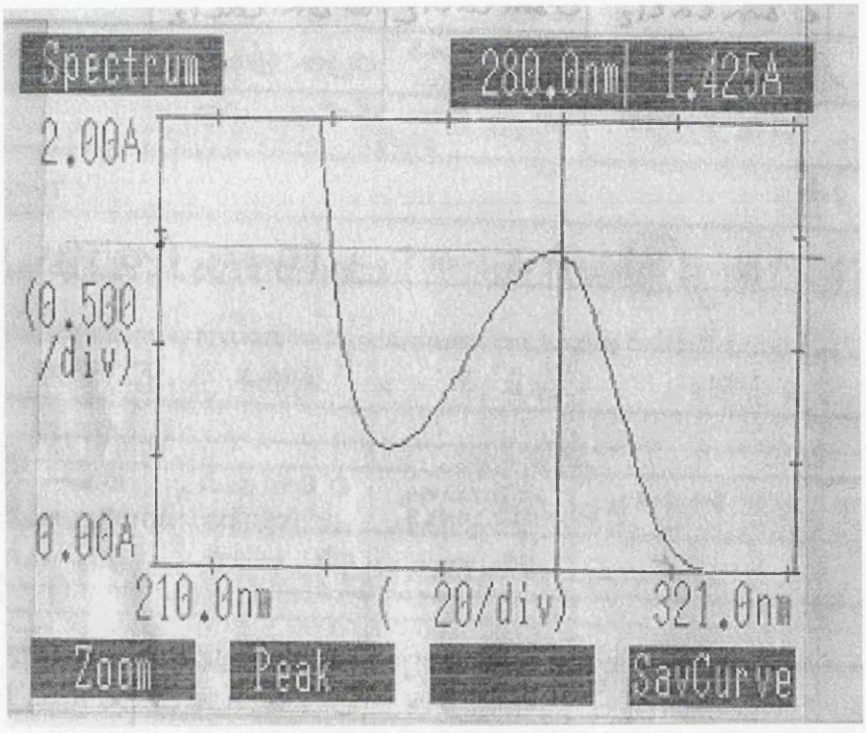


Figure 2.1: Dimeric mutant: Absorption Spectrum 1

UV Absorption Spectrum at 280nm of original dimeric mutant sample prior to concentration. Estimated protein concentration: 0.7 mg ml^{-1}.

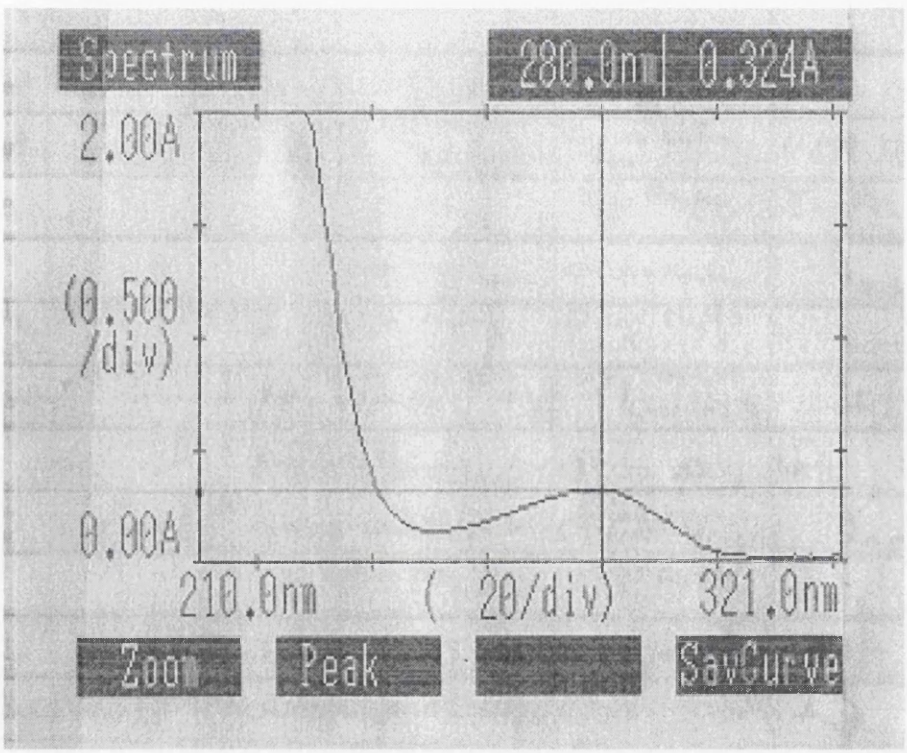


Figure 2.2: Dimeric mutant: Absorption Spectrum 2

UV Absorption Spectrum at 280nm of concentrated sample of dimeric mutant. Sample diluted 50 times in order to remain within Absorbance limts of Spectrophotometer. Estimated protein concentration before 50-fold dilution: 5.1 mg ml^{-1}.

In order to determine the Molar Extinction Coefficient ε for the dimeric mutant, the Absorption Spectrum shown in Figure 2.1 was measured. Assuming the estimated protein concentration of 0.7 $mgml^{-1}$ to be approximately correct for this sample, ε was calculated using the Beer-Lambert law:

$$A = c\,\varepsilon\,l$$

where A : Absorbance

c : Protein concentration ($mol\ l^{-1}$)

ε : Molar Extinction Coefficient ($l\ mol^{-1}cm^{-1}$)

l : Path length of sample^{l} (cm)

This gave an estimated ε of 1.98 $l\ mol^{-1}cm^{-1}$. This value was then used in conjunction with the second Absorption Spectrum (Figure 2.2) in order to determine the actual concentration of protein present in the second sample. The resulting figure, 3.8 mg ml^{-1}, was then compared with the estimated protein concentration of 5.1 mg ml^{-1}, in order to determine the proportion of the sample which was lost on concentration. Similar experiments were used to determine ε values for monomeric and tetrameric samples.

2.1.1 *Analysis of Mutant Molecular Weight and Purity by SDS-PAGE*

Molecular weights and purities of the mutant samples were assessed by SDS-PAGE[95]. Reduced and non-reduced samples of each of the three mutants were analysed on a 12% SDS-PAGE gel (See Figure 2.3). The reduced samples of monomeric, dimeric and tetrameric mutants each had the same apparent molecular weight of around 8kD, suggesting that the dimerisation and tetramerisation interactions were disrupted by the high temperature and the presence of SDS. The majority of each of the the non-reduced samples also migrated as a band corresponding to 8kD, although there were faint bands at 15kD corresponding to the dimer and 30kD corresponding to the tetramer. The presence of SDS in the gels themselves is likely to have disrupted the equilibrium between the aggregation states, to the extent that only a small amount of the dimeric and tetrameric mutants remained in the aggregated state.

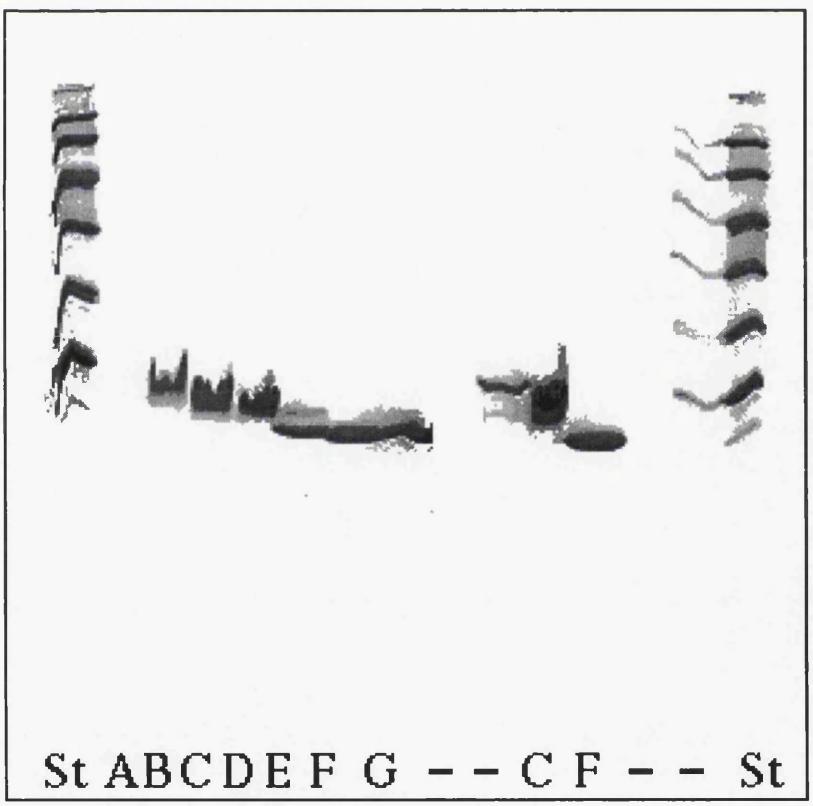


Figure 2.3: SDS-PAGE Gel

B:	PM1 without DTT	E:	PM1 with DTT
C:	PM2 without DTT	F:	PM2 with DTT
D:	PM3 without DTT	G:	PM3 with DTT

2.1.2 *Analysis of Mutant Sample Monodispersity by Dynamic Light Scattering*

It has been shown[96, 97] that the monodispersity of a crystallisation sample is often correlated to the likelihood that crystals will be obtained from it. Samples which are monodisperse, or contain a single species in solution, are more likely to provide crystals than samples which are polydisperse. Polydispersity can reflect the presence of multiple protein isoforms or of differing protein aggregation states, as well as impurity within the solution.

The monodispersity of concentrated samples of each of the three mutants was investigated by Dynamic Light Scattering (DLS) using the Dynapro-801 instrument. Unfortunately, in the DLS technique, the required sample concentration is inversely proportional to the sample molecular weight. The monomeric and dimeric mutants could not be concentrated to a sufficient extent to allow readings to be obtained, but a sample of the tetrameric mutant at 0.7 mg ml^{-1} gave satisfactory results. The solution showed two main components, one with an estimated molecular weight of around 30*k*D, corresponding closely to the predicted molecular weight of 31.2*k*D, the other with an estimated molecular weight of more than 1000*k*D, which suggested slight contamination of the sample. Previous spectroscopic evidence suggested that the molecular weight of the MIP-1α aggregate was between 100 and 250*k*D[86, 85, 18]. However, there was no evidence for any species in the range 36-200*k*D, suggesting that the tetrameric mutant did not exist in a state of equilibrium with higher aggregation states. The estimated molecular weights from five separate runs ranged from 24*k*D to 36*k*D, as the low protein concentration made calibration of the instrument difficult and gave an unstable sample baseline.

2.2 *Crystallisation of Monomeric Mutant*

The monomeric mutant was crystallised from the "Magic 50" sparse matrix screen[98, 99]. Of the 48 conditions tried, 6 gave crystals, and 3 of these gave large, single, trigonal bipyramidal crystals. The other conditions gave stacks of thin rhombus-shaped plates or fine needles. The conditions which gave single crystals all contained Mg^{2+}, were buffered at mildly acidic pH and had a fairly high concentration of a mild precipitant. The successful screening conditions are summarised in Table 2.2.[I]

A second crystallisation screen, consisting of 60 conditions, gave a further 4 conditions under

[I] It was later discovered that the HEPES buffer used in this initial crystallisation screen had been prepared incorrectly. The pH of the standard HEPES solution from which all conditions containing HEPES were prepared was 5.7, which is well outside the useful buffer range for this compound. When conditions 12 and 23 were repeated using Citrate (CIT) buffer pH 5.7 instead of HEPES, crystals of the same size and morphology were obtained.

Condition	Salt (0.2M)	Buffer (0.1M)	Precipitant	Result
9	NH_4Ac	CIT pH 5.6	30% PEG 4000	Irregular plates (0.3mm) and Long needles (0.7mm)
12	$MgCl_2$	HEPES pH 5.7	30% IPR	Large trigonal bipyramids
21	MgAc	CAC pH 6.5	30% MPD	Sharp trigonal bipyramids (0.2mm)
23	$MgCl_2$	HEPES pH 5.7	30% PEG 200	trigonal bipyramids (0.2mm)
26	NH_4Ac	CIT pH 5.6	30% MPD	Small sharp, irregular crystals
34	none	NaAc pH 4.6	2.0M NaForm	Stacks of round-edged plates

Table 2.2: Results of initial crystallisation trials for monomeric mutant

which the monomeric mutant could be crystallised (Shown in Table 2.3). Each of these gave small cuboidal crystals, with a morphology that appeared different to the crystals obtained previously.

Condition	Salt (2M)	Buffer (0.1M)	Precipitant	Result
N6	none	MES pH 6.1	1.1M AMS	Sharp cuboids (0.1mm)
N7	none	MES pH 6.1	1.4M AMS	Small cuboids (0.05mm)
N59	NaCl	PIPES pH 6.8	10% PEG 3400	Small cuboids (0.1mm)
N60	NaCl	PIPES pH 6.8	30% PEG 3400	Cuboids & small trigonal bipyramids

Table 2.3: Results of further crystallisation trials for monomeric mutant

Diffraction tests on the three distinct crystal morphologies obtained for this mutant showed the trigonal bipyramidal crystals to be the most promising. Small crystals taken directly from the initial screen showed diffraction to 3.5Å. Finer crystallisation screens around those initial conditions yielded large single crystals up to 0.5mm in each dimension. These showed diffraction to 2.5Å at Station 9.5 of the Synchrotron Radiation Source at Daresbury Laboratory (DL9.5). The diffraction limit was further improved by cryocooling, and a dataset extending to 2.0Å was collected at 100 K at Station 7.2 of the Synchrotron Radiation Source (DL7.2).

2.3 Crystallisation of Dimeric Mutant

The dimeric mutant was also crystallised from the "Magic 50" crystallisation screen[98,99]. Large single crystals were obtained from 7 separate conditions, and crystals or microcrystals were observed in 24 of the 48 wells. The largest of these were trigonal bipyramids (See Figure 2.4) similar to those observed for the monomeric mutant, although rhombus-shaped plates and some irregularly-shaped crystals were also present. Similar morphologies were observed for the smaller crystals, although small cuboids and trigonal prisms were present in some wells. The successful

screening conditions are summarised in Table 2.4.

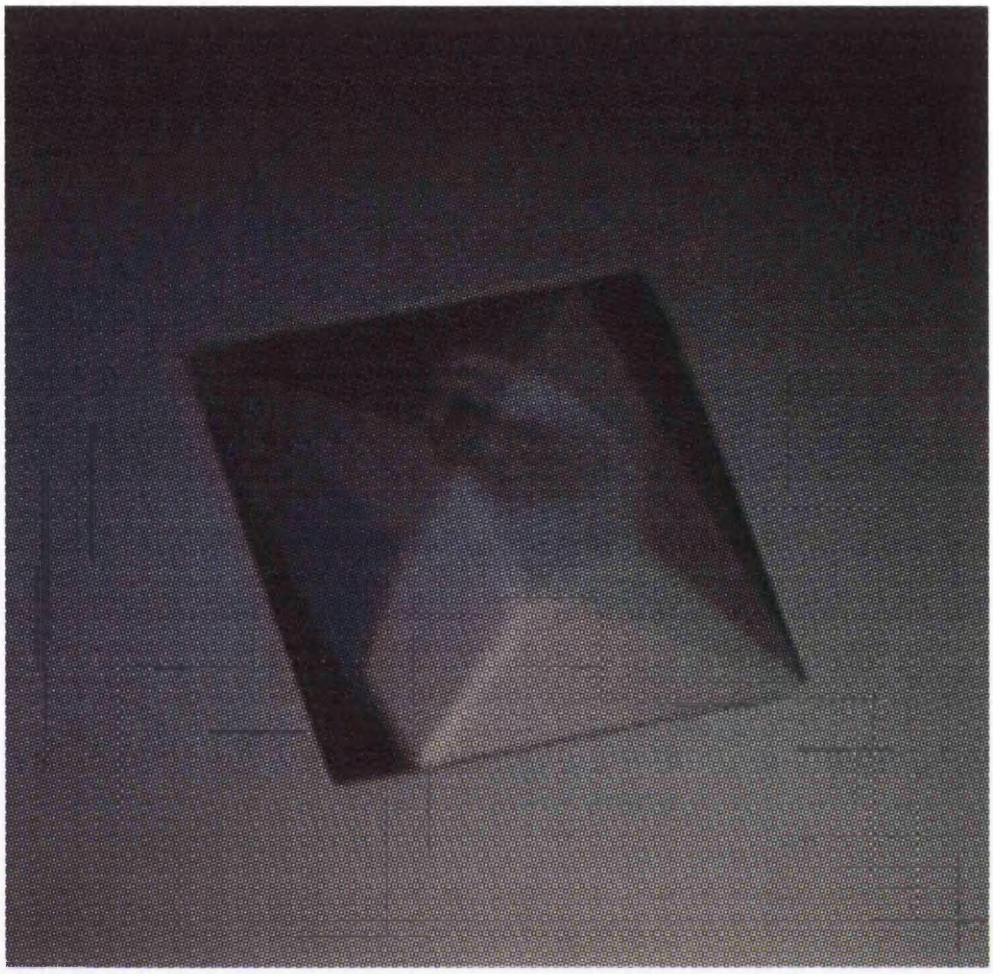


Figure 2.4: Trigonal Bipyramidal crystal of dimeric mutant

The two conditions which gave large single trigonal bipyramidal crystals were similar to each other and resembled those which gave similar crystals for the monomeric mutant. Once again the buffer was mildly acidic and the precipitant was around 30% of MPD, IPR or PEG 200. These three precipitants are mild and are often considered interchangeable. The only significant difference between the conditions which gave crystals of the monomeric mutant and the corresponding conditions for the dimeric mutant was the presence of $MgCl_2$ in the former and $CaCl_2$ in the latter. A series of screens around conditions 1 and 14 for the dimeric mutant showed that $CaCl_2$ was not a prerequisite for crystallisation, and that replacing $CaCl_2$ with $MgCl_2$ or one of several other divalent metal chlorides resulted in similar crystals. A similar screen for the monomeric mutant showed that $MgCl_2$ could be replaced by $CaCl_2$ to give larger crystals with sharper faces

Condition	Salt (0.2M)	Buffer (0.1M)	Precipitant	Result
1	$CaCl_2$	NaAc pH 4.6	30% MPD	Sharp trigonal bipyramids (0.2mm)
7	none	CAC pH 6.8	1.4M NaAc	Large plates (0.2mm x 0.2mm)
8	NaCIT	CAC pH 6.8	30% IPR	Rhombus-shaped plates (0.1mm)
14	$CaCl_2$	HEPES pH 7.5	28% PEG 200	Trigonal bipyramids (0.3mm)
18	MgAc	CAC pH 6.8	20% PEG 8000	Large plates (0.3mm x 0.1mm)
19	NH_4Ac	Tris pH 8.5	30% IPR	Irregular single crystals (0.1mm)
20	AMS	NaAc pH 4.6	25% PEG 4000	Small irregular crystals (0.05mm)

Table 2.4: Results of crystallisation trials for dimeric mutant

and edges.

Diffraction tests demonstrated diffraction to 3.7Å for the trigonal bipyramidal dimer crystals, and to 3.0Å for the largest of the plate crystals. Further improvement of the crystallisation parameters resulted in trigonal bipyramidal crystals which showed diffraction to 2.3Å. The diffraction limit was improved to 1.8Å at DL7.2 by cryocooling.

2.4 Crystallisation of Tetrameric Mutant

Obtaining initial crystallisation conditions for the tetrameric mutant proved considerably more difficult than for the other two. Several different sparse matrix crystallisation screens were attempted, but only 3 of the 228 conditions tried gave promising results. The successful screening conditions are summarised in Table 2.5.

Condition	Salt	Buffer (0.1M)	Precipitant	Result
19	0.2M NH_4Ac	Tris pH 8.5	30% IPR	Large hexagonal plates (0.5mm)
N54	none	none	2.0M NaCIT pH 6.4	Small rods (0.1mm)
M67	0.05M $CaCl_2$	Tris pH 8.75	5% PEG 8000	Orthorhombic disphenoid (0.4mm)

Table 2.5: Results of crystallisation trials for tetrameric mutant

The crystals from conditions 19 and N54 were tested with Coomassie Blue. Protein crystals, but not salt crystals, are generally stained blue by this compound, although the test is far from definitive. The crystal from well N54 gave a positive result, whereas the crystal from 19 appeared not to stain at all. Crystals from condition N54, which grew in around a week, were then tested and found to diffract to around 8Å. A crystal from condition 19, which took around two weeks to grow, was also tested and no diffraction was observed.

Crystals from condition M67 grew in 24 hours, but began to crack shortly afterwards and

shattered within 2-3 days of being set up. This is characteristic of very swift equilibration between well and drop, which can probably be attributed to the presence of PEG 8000, which is a powerful dehydrating agent. Therefore a finer crystallisation screen was tried around condition M67, in order to find conditions where the rate of equilibration was slower and the crystals would be likely to be more stable. A condition was found which gave large single crystals in around 24 hours. These were stable for several weeks, but did eventually crack in a similar way to those obtained from the initial screen. Crystals from this condition were used for subsequent diffraction tests, and the crystallisation parameters were improved over some time, eventually giving large single crystals in around a week which survived indefinitely. These crystals were orthorhombic in habit (See Figure 2.5), and showed diffraction to 2.4Å in house on a Siemens Area detector, and to 1.8Å at DL9.5. The resolution limit was improved to 1.5Å on finding suitable cryocooling conditions.

Figure 2.5: Orthorhombic crystal form of tetrameric mutant

In addition, crystals were occasionally observed to be hollow, which is again characteristic of crystal growth being too rapid. However, standard techniques for slowing down the rate at which new molecules can add to a growing crystal, such as the addition of glycerol to the drop, or dilution of the drop in order to slow down the rate of equilibration, did not appear to improve this particular problem. It was eventually solved by increasing the protein concentration from around 2 mg ml^{-1} to 5 mg ml^{-1}. Protein concentration, although not particularly high for crystallisation of any of the three mutants, did appear to be one of the most important parameters.

The addition of glycerol to certain crystallisation trials did have some unexpected effects. A

second crystal form of the tetrameric mutant (See Figure 2.6) was observed to grow alongside the previously observed orthorhombic crystals when glycerol was included in the crystallisation mixture. This form was monoclinic and showed diffraction to at least 2.0Å in house on a DIP 2020 image plate.

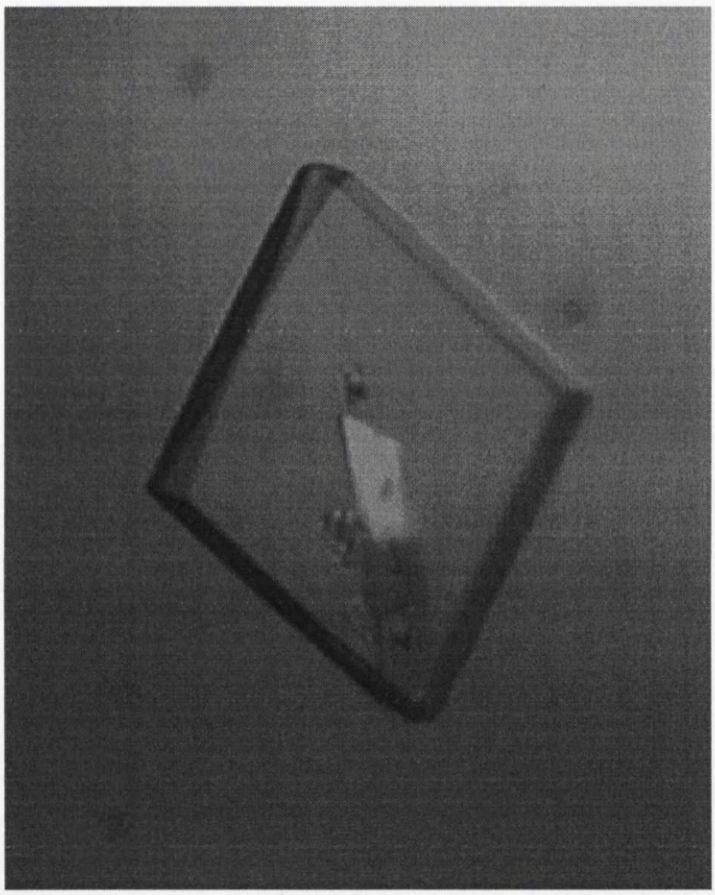


Figure 2.6: Monoclinic crystal form of tetrameric mutant

Unlike the other two mutants, the crystals of the tetramer could not be reproduced by swapping $MgCl_2$ or some other divalent metal chloride for the $CaCl_2$. In all crystallisation trials which produced orthorhombic or monoclinic crystals, $CaCl_2$ was present, suggesting that the Ca^{2+} ion might have some role in mediating crystal contacts between tetramers and might therefore be necessary for crystallisation.

2.5 Final Crystallisation Conditions

The final crystallisation conditions used for each of the mutants is given below in Table 2.6. All crystals were grown using the hanging-drop vapour diffusion method. Crystallisation drops were prepared by mixing 1μl of protein solution with 1μl of the appropriate well solution. Once final conditions were determined, larger crystals were grown from larger drops composed of the same proportions of protein and well solutions. Crystals were grown at 19°C.

Mutant	Salt	Buffer (0.1M)	Precipitant	Protein Conc.	Morphology
Monomer	none	none	2.0M AMS	1mg/ml	Cuboid
Monomer	0.2M $CaCl_2$	Acetate pH 4.35	30% PEG 200	1mg/ml	Hex. Bipy.
Dimer	0.2M $CaCl_2$	Acetate pH 4.35	30% PEG 200	2mg/ml	Hex. Bipy.
o-Tetramer	0.2M $CaCl_2$	Tris pH 7.4	8% PEG 6000	3mg/ml	Orthorhombic disphenoid
m-Tetramer	0.2M $CaCl_2$	Tris pH 7.4	8% PEG 6000 & 10% Glycerol	5mg/ml	Trapezoid

Table 2.6: Final Crystallisation Conditions for each mutant

2.6 Discussion

A crystal was obtained from the dimeric mutant which suprisingly had very similar physical appearance to the orthorhombic crystals grown from the tetrameric mutant. A diffraction test showed that this crystal had an identical unit cell and spacegroup to the orthorhombic crystals of the tetrameric mutant. This result implied that interaction between dimeric mutants could form tetramers under certain conditions, and seemed at odds with the gel filtration evidence which predicted this was highly unlikely[86]. The implications of this observation are discussed in more detail in Section 4.1.1, where the structures of the dimeric and tetrameric mutants are compared.

The condition from which this crystal was obtained was 0.2M $MgCl_2$, 0.1M HEPES pH 5.7, 1% PEG 200. As mentioned above, the HEPES buffer used in many of these conditions was at a pH outwith the useful range for this compound. The crystal took over a year to appear, and the drop from which it grew was almost dry by the time it eventually appeared. Therefore it is likely that the final conditions within the drop were very different to the original ones. In particular, the protein concentration at the time the crystal began to grow would have been very much higher than was used in any of the crystallisation trials.

This condition also contained $MgCl_2$, not $CaCl_2$. This would seem to contradict the other evidence suggesting a requirement for $CaCl_2$ in formation of crystals of the tetrameric mutant. However, Mg^{2+} is similar to Ca^{2+} in many of its physical properties, and should be able to replace

it in some cases, despite having a smaller ionic radius and preference for higher coordination states[100]. It is possible that tetramer-tetramer interaction via Ca^{2+} is considerably more favourable than similar interactions involving other ions, but that the unfavourability of those interactions can be overcome by the presence of a sufficiently high concentration of protein.

Given that the three mutants are generated simply by neutralisation of charged residues, the pH at which the various crystals are obtained is also of interest. Since the pK_a of lysine is 10.0 and that of arginine is 12.0, residues of these types are always protonated and positively charged in each of the mutant crystals. Glutamic acid and aspartic acid have pK_a values of around 4.4. Therefore the monomeric and dimeric crystals contain both protonated and unprotonated acid residues. While approximately half of these residues will be charged at this pH, it is impossible to calculate the precise precentage of Glu and Asp residues that will be charged as the pK_a value of a particular residue also depends on temperature, ionic strength and local chemical environment. Glutamic and aspartic acids in the two crystal forms of the tetrameric mutant will always be unprotonated, and therefore charged, whereas only some of those in the crystals of monomer or dimer will be charged. While at first glance this might hint that some of the interactions involved in formation of the various types of crystal are pH-dependant, this is not true for the monomeric and dimeric mutants, as crystals of identical symmetry and morphology have been obtained at much higher pH values. The tetramer may require a reasonably high pH value to ensure that the glutamic and aspartic acids are charged, which would imply electrostatic contacts involving residues of that type being involved in crystal contacts.

It is interesting that each of the mutants can be crystallised using $CaCl_2$, but that only the tetrameric mutant seems to absolutely require it for crystals to form. A series of trials on the monomeric, dimeric and orthorhombic tetrameric crystal forms tested the ability of various other substances, divalent metal chlorides for the most part, to replace $CaCl_2$. Crystals were obtained from monomeric and dimeric mutants using $MgCl_2$, $SrCl_2$, $MnCl_2$, $NiCl_2$, $CdCl_2$, $CoCl_2$ and $Lu_2(SO_4)_3$. However, none of these compounds resulted in crystals of the tetrameric mutant.

The original intention, when growing crystals in which Ca^{2+} ions had been substituted, was to discover whether Ca^{2+} had been incorporated into those crystals. This would allow the substitution of Ca^{2+} by heavier metal ions, such as Lu^{3+}, and allow phase information to be derived using either the anomalous signal from the Lu^{3+}, or from the isomorphous differences between Ca^{2+}-containing and Lu^{3+}-containing crystals. However, the range of ions which seemed to be effective in crystallisation of monomeric and dimeric mutants suggested that the Ca^{2+} was unlikely to be incorporated into those crystals in a well-ordered fashion. Comparison of data obtained from

crystals grown in $CaCl_2$, $SrCl_2$ and $Lu_2(SO_4)_3$ via difference Patterson maps proved that Sr^{2+} and Lu^{3+} were not present at specific points within the crystals, and that Ca^{2+} therefore was also unlikely to be bound at specific locations.

Conversely, the tetrameric mutant apparently depended upon the presence of Ca^{2+} for crystallisation, which suggested that Ca^{2+} might well have been incorporated into these crystals in a specific fashion. The fact that it proved impossible to replace those ions in cocrystallisation experiments, despite using other ions of very similar charge and ionic radius, suggested that a specific interaction involving Ca^{2+} might be involved in mediating tetramer-tetramer contacts within the crystal. While all of the ions which were tried were similar to Ca^{2+}, there were inevitable differences in both charge and ionic radii which could easily account for the observations. It was later discovered that Ca^{2+} ions were indeed present in the crystals, as discussed in more detail in Section 5.2

A later experiment which appeared to give a promising result was a conventional $Lu_2(SO_4)_3$ soak of the orthorhombic tetramer crystals. Data were collected from such a soaked crystal and after comparison with native data the soaked crystals appeared to represent a single-site Lu^{3+} derivative. This experiment is outlined and the results described in Section 5.2.2.

3. STRUCTURE SOLUTION

This chapter describes the structure solution process for each of the MIP-1α mutants. No attempt has been made to discuss most aspects of crystallographic theory, as the subject has been dealt with extremely well in a variety of excellent textbooks[100, 101, 102, 103]. However, the Molecular Replacement technique is outlined in Appendix C, as some problems encountered using this particular method are discussed in Section 3.8. Although certain programs are mentioned and referenced explicitly, the majority of the programs used during the structure determination were from the CCP4 suite[104].

The great similarity between the monomers of both α- and β- Chemokines is apparent in their crystal and NMR structures[16]. Therefore it was intended to solve the structures of the MIP-1α mutants by Molecular Replacement[105], using as a search model one or more of the existing Chemokine structures[106, 107, 108, 92, 82, 109, 110]. Unfortunately, for reasons discussed later in this chapter, this proved impossible. Initial phase information was therefore obtained by the Heavy Atom Isomorphous Replacement method; more specifically the Single Isomorphous Replacement with Anomalous Scattering (SIRAS) technique. Once one of the mutants had been solved by SIRAS, the resulting model was used to find Molecular Replacement solutions for the others. This chapter explains in greater detail the collection and processing of all diffraction data. It also investigates the initial failure of the Molecular Replacement technique and shows how the structures of the various mutants were eventually solved.

3.1 Data Collection and Processing

Native data were collected using a variety of sources and several types of detector. Relevant information for datasets used in the structure solution of each of the mutants is contained in the tables below. The following abbreviations are used:

- Temp. : Temperature at which data were collected. At 100 K, cryoprotectant was 15% Glycerol plus the equivalent drop solution.
- Osc. Rng. : Width of each diffraction image in degrees.

- Comp. : Overall completeness of data.
- Mult. : Overall multiplicity of data.
- N. Ref. : Total number of unique reflections measured.
- R_{merge} : Defined as

$$R_{merge} = \Sigma_h \Sigma_j \frac{|I(h) - I(h)_j|}{\Sigma_h \Sigma_j I(h)_j},$$

(where $I(h)$ is the mean intensity)

- DL7.2 : Daresbury Laboratory Synchrotron Radiation Source Station 7.2
- (o) : Orthorhombic crystal form.
- (m) : Monoclinic crystal form.

Sample	Detector	Location	Resolution	Temp.	Osc. Rng.	Comp.	Mult.	N. Ref.	R_{merge}
Native	DIP 2020	St Andrews	20.0 - 2.9Å	298K	2.0	93.1%	1.5	3195	6.4%
UO_2Ac soak	Siemens AD	Glasgow	28.0 - 4.0Å	298K	0.2	93.4%	1.5	1218	6.6%
Native	MAR 18cm	DL7.2	20.0 - 2.2Å	100K	1.5	80.4%	2.0	5728	5.4%

Table 3.1: Statistics for datasets collected from crystals of monomeric mutant
All data were collected at Cu K_α wavelength (1.5418Å), with the exception of DL7.2 data (1.448Å).

Sample	Detector	Location	Resolution	Temp.	Osc. Rng.	Comp.	Mult.	N. Ref.	R_{merge}
Native	DIP 2020	St Andrews	20.0 - 2.85Å	298K	2.0	99.6%	3.0	3425	5.7%
Native	Mar 18cm	DL7.2	16.6 - 2.24Å	100K	1.0	94.0%	4.9	6640	5.7%

Table 3.2: Statistics for datasets collected from crystals of dimeric mutant
All data were collected at Cu K_α wavelength (1.5418Å), with the exception of DL7.2 data (1.448Å).

Sample	Detector	Location	Resolution	Temp.	Osc. Rng.	Comp.	Mult.	N. Ref.	R_{merge}
Native (o)	DIP 2020	Glasgow	20.0 - 2.0Å	100K	2.0	94.9%	3.5	10168	6.6%
Native (o)	Mar 30cm	DL7.2	20.0 - 1.6Å	100K	1.5	85.7%	3.0	17182	4.8%
$Lu_2(SO_4)_3$ soak (o)	DIP 2020	Glasgow	30.0 - 3.0Å	100K	2.0	85.2%	6.0	3172	8.9%
Native (m)	DIP 2020	Glasgow	30.0 - 2.05Å	100K	1.0	87.8%	1.5	21339	6.7%

Table 3.3: Statistics for datasets collected from crystals of tetrameric mutant
All data were collected at Cu K_α wavelength (1.5418Å), with the exception of DL7.2 data (1.448Å).

Area detector data were processed and scaled using the programs XDS and XSCALE[111, 112]. Image plate data were processed using both DENZO[113, 114] and MOSFLM[115].

3.2 *Merging and Scaling of Data*

As a consequence of the strong diffraction observed from crystals of each of the three mutants, and also to the long exposure times necessary to collect high resolution data, low resolution spots were often saturated on synchrotron Image Plate data. Low resolution data were therefore collected separately in several instances, and later merged with the high resolution observations.

Where more than one dataset had to be combined, every measurement from each dataset was input to scaling. All data processed with DENZO and XDS were combined using SCALEPACK[113]. Where one of the datasets to be merged had been processed with MOSFLM, AGROVATA was used for scaling. Data were reduced using the program TRUNCATE[116]. Indexing and space-group determination for XDS data used the program IDXREF[111]. For other datasets, spacegroup was determined using DENZO. Data processed with MOSFLM were indexed with IDXREF. Data processed with DENZO were also indexed with DENZO.

The spacegroup and unit cell parameters obtained for each of the mutants is given in Table 3.4. The following abbreviations are used:

- Z : Number of molecules (MIP-1α monomers) per asymmetric unit.
- Solvent : Solvent contents of unit cell, as estimated by the method of Matthews[117].

Mutant	Spacegroup	Z	Solvent	*a*	*b*	*c*	α	β	γ
Monomer	$P3_121$	1	75.62%	59.780	59.780	66.580	90.0	90.0	120.0
Dimer	$P3_121$	1	74.05%	59.101	59.101	64.002	90.0	90.0	120.0
Tetramer	$C222_1$	2	47.81%	52.205	89.873	63.178	90.0	90.0	90.0
Tetramer	$P2_1$	4	54.76%	52.431	62.003	53.800	90.0	102.12	90.0

Table 3.4: Unit Cell parameters and contents for all crystal forms of MIP-1α mutants

3.3 *Initial Molecular Replacement Attempts*

It was expected that it would be possible to find Molecular Replacement (MR) solutions for each of the mutants using the coordinates of one of the Chemokines whose structure was known.

MIP-1β was initially used as a search model since there was a high degree of sequence homology between it and MIP-1α. Despite trying the whole protein, fragments of it and even a MIP-1α homology model based on it as search models, it proved impossible to solve the MR using MIP-1β. The other CC-Chemokine structure known at the time, that of RANTES, gave similar results.

This failure seems most likely to be due to the problems associated with the use of NMR structures as search models in solving crystal structures[118]. Two other β-Chemokine structures, those of MCP-1 and MCP-3, were published after the Molecular Replacement experiments were performed. As described in Sections 3.6 and 3.7, the dimeric and tetrameric MIP-1α mutants could both be phased by MR using the structure of either of the MCP-1 crystal forms as a search model.

3.4 Derivatisation

Once it became apparent that Molecular Replacement was unlikely to provide a way to solve the mutant structures, attempts were made to generate heavy atom derivatives in order to obtain initial phase information using the Isomorphous Replacement technique. Initially, this involved the use of selected heavy atom compounds during crystallisation, as previously described. Although this was generally unsuccessful, crystals of the monomeric mutant were obtained from crystallisation conditions containing two heavy atom compounds; Strontium Chloride and Lutetium Sulphate. Datasets were collected using crystals from both conditions, but comparison with existing native datasets via difference Patterson maps indicated that heavy atoms were not present in either. Attempts to grow crystals of the monomeric mutant from conditions containing Platinum, Mercury and Uranium compounds were unsuccessful.

As it appeared unlikely that heavy atoms would be able to be incorporated during crystallisation, several conventional heavy atom soaks were performed. Millimolar quantities of the heavy atom compound were added to the appropriate cryoprotectant solution, so that heavy atom and cryoprotectant entered the crystals concurrently. Platinum and Mercury compounds were again the first to be tried, but both proved ultimately unsuccessful. Crystals of the monomeric mutant generally survived for only a matter of minutes in soaks containing less than millimolar concentrations of Ammonium Tetrachloroplatinate. While this was encouraging in that it implied that the Pt compound was entering the crystals and binding to the protein, even crystals soaked for around ten seconds showed no observable diffraction. Reduction of the heavy atom concentration did not alleviate the problem.

Crystals soaked in Mercury Iodide, by contrast, did not appear to be adversely affected by the soaking condition, and showed diffraction comparable to that observed from native crystals. However, comparison with native diffraction data proved that these crystals had not been derivatised.

A successful heavy atom soak was eventually performed using a combination of Uranium-containing compounds, as described below in Section 3.5.

When it became apparent that crystals of the tetrameric mutant might contain Ca^{2+} ions bound

at specific positions, these crystals were soaked with $Lu_2(SO_4)_3$, in an attempt to specifically replace the Ca^{2+} ions with Lu^{3+} ions. The results of this experiment are described in Section 5.2.2.

3.5 Solution of Monomeric Mutant

The monomeric mutant was solved by soaking crystals in a solution containing 1mM each of $UO_2(Ac)_2$, $K_3(UO_2F_5)$ and $Mg(Ac)_2$. Patterson and difference Fourier maps showed that the resulting derivative had a single UO_2^{2+} ion bound close to the C-terminus of the protein. Uranyl compounds have successfully been used to derivatise proteins in several other cases[100], but the UO_2^{2+} ion has a reputation for binding to many sites per molecule, often in a non-specific fashion. However, the presence of a low concentration of Mg^{2+} ions in a solution is often sufficient to disrupt these weak, non-specific interactions, leaving bound only those Uranyl ions involved in strong specific interactions with the protein[119]. From a crystallographic viewpoint, the result is a much cleaner derivative, unlikely to have minor binding sites in addition to the major ones, and therefore considerably easier to interpret.

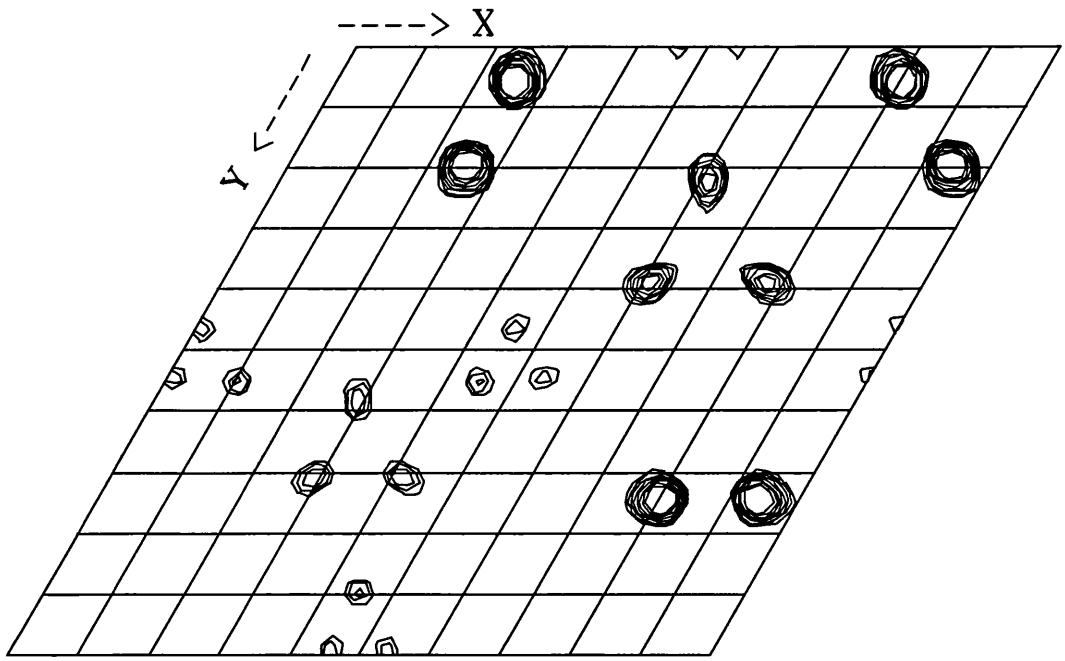


Figure 3.1: Difference Patterson: Harker Section z=$\frac{1}{3}$
The six largest peaks on this section are due to Uranium→Uranium vectors.

Data from a Uranyl-soaked crystal were collected at room temperature and extended to 4.0Å.

Comparison with a room temperature native dataset via a difference Patterson map (See Figure 3.1) clearly showed several promising peaks on each of the Harker sections. The Uranium atom was located using the program SHELX[120], and difference Fourier maps phased using the atomic position showed that there were no other sites. (Maximum Likelihood) Refinement of the phasing using the program MLPHARE[121] suggested that there was a detectable anomalous signal from the Uranium atom, allowing the anomalous occupancy of the derivative to also be refined. Upon convergence of the phase refinement, the techniques of solvent flattening and histogram matching were applied using the density modification program DM[122], thus improving the quality of the electron density maps. This made a dramatic improvement to the quality of the protein electron density maps, as would be expected for crystals of high solvent content. At this stage the hand of the solution had still to be determined, but it was obvious after solvent flattening and examining both possible maps that only the correct one contained a meaningful signal. Attempts to cross-phase the HgI_4 and $Lu_2(SO_4)_3$ datasets using the Uranyl phases showed that neither of the other two datasets was in fact a derivative.

Inspection of the solvent-flattened maps enabled visual identification of secondary structure elements in the density, and allowed the positioning of the main-chain atoms of the MIP-1β NMR structure within the density. This gave an initial model for the monomeric mutant, although the density was continuous and of sufficient quality that the main-chain atoms could easily have been traced without the model. Even at 4.0Å, many side-chains could be clearly distinguished. In addition, it was clear from the fit between the MIP-1β model and the electron density maps that while MIP-1α and MIP-1β overall were very similar, there were several regions in which the molecules were clearly different. These regions were presumably responsible for the failure to solve MIP-1α by MR using MIP-1β as a search model. Fitting of the initial MIP-1α model to the electron density map and subsequent refinement of the model are described in Section 3.9.

3.6 Solution of Dimeric Mutant

The dimeric mutant was solved by Molecular Replacement using the initial main-chain trace of the monomeric mutant as a search model. A rotation function, calculated using the program AMoRe[123], showed a single clear solution (See Figure 3.2). A translation search, and subsequent rigid-body refinement provided a single solution which was significantly better than any other both in terms of correlation coefficient and R-factor[124] (See Figures 3.3, 3.4, 3.5 and 3.6). The search model was transformed by the orientation matrix output from AMoRe using the program LSQKAB[125], and was used to calculate initial electron density maps. Fitting of this initial model

into these maps and the refinement of the structure of the dimeric mutant is described in Section 3.9. All MR calculations were performed using data of resolution 10 - 4.0Å, with the integration sphere radius automatically determined by the program.

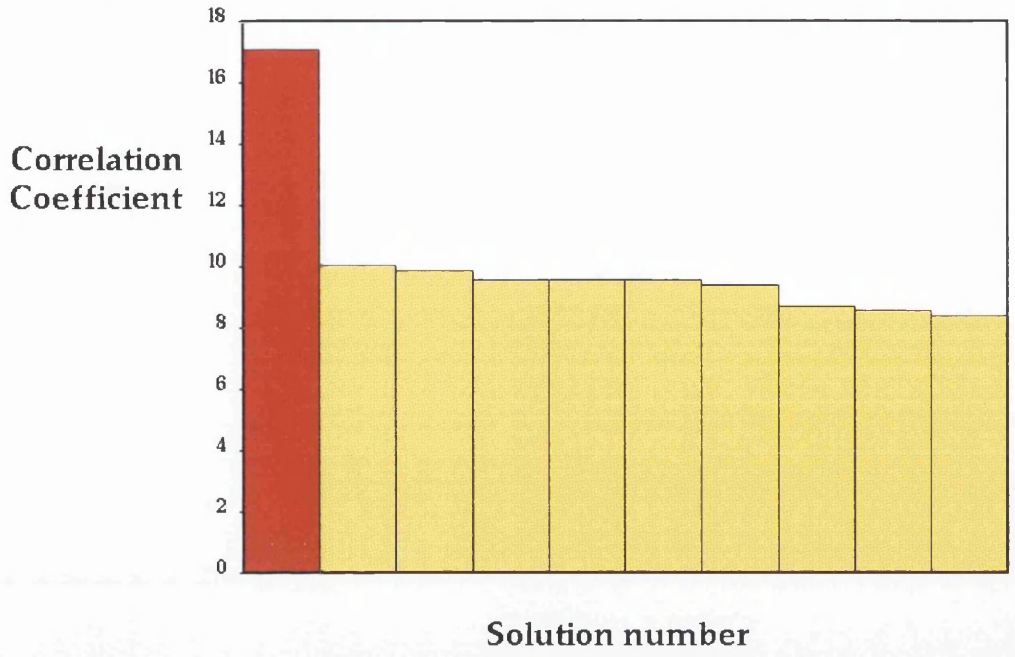


Figure 3.2: Rotation function solutions for dimeric mutant

3.7 Solution of Tetrameric Mutant

Unlike the monomeric and dimeric mutants, in the case of both tetrameric crystal forms there was non-crystallographic symmetry (NCS). In other words, the crystallographic asymmetric unit contained more than one MIP-1α monomer. In the case of the orthorhombic crystal form, there was two-fold non-crystallographic symmetry, and the tetramer was therefore formed by the intersection of the NCS axis with a crystallographic two-fold axis. In the monoclinic crystal form, there were four monomers within the asymmetric unit. Therefore, in order to form the same tetramer as seen in the orthorhombic crystal form, there must be two NCS two-fold axes, one which corresponds to the orthorhombic NCS two-fold and one which corresponds to the orthorhombic crystallographic two-fold. Although the presence of non-crystallographic symmetry can be extremely beneficial at certain stages of the structure solution process, it can be a hindrance at the Molecular Replacement stage. Effectively the presence of two-fold NCS represents a dilution of the useful signal relative

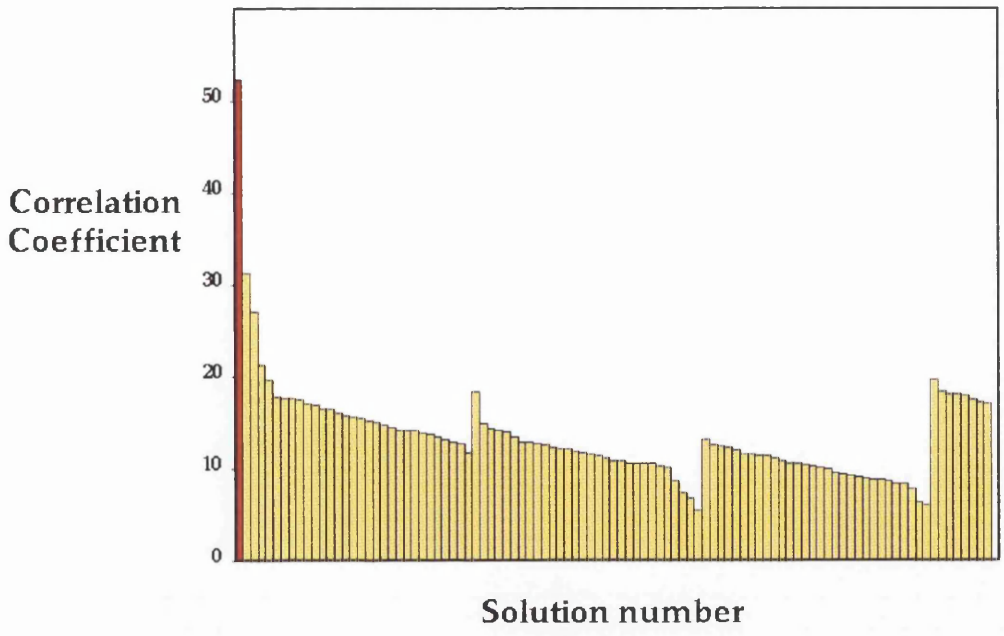


Figure 3.3: Translation function solutions for dimeric mutant

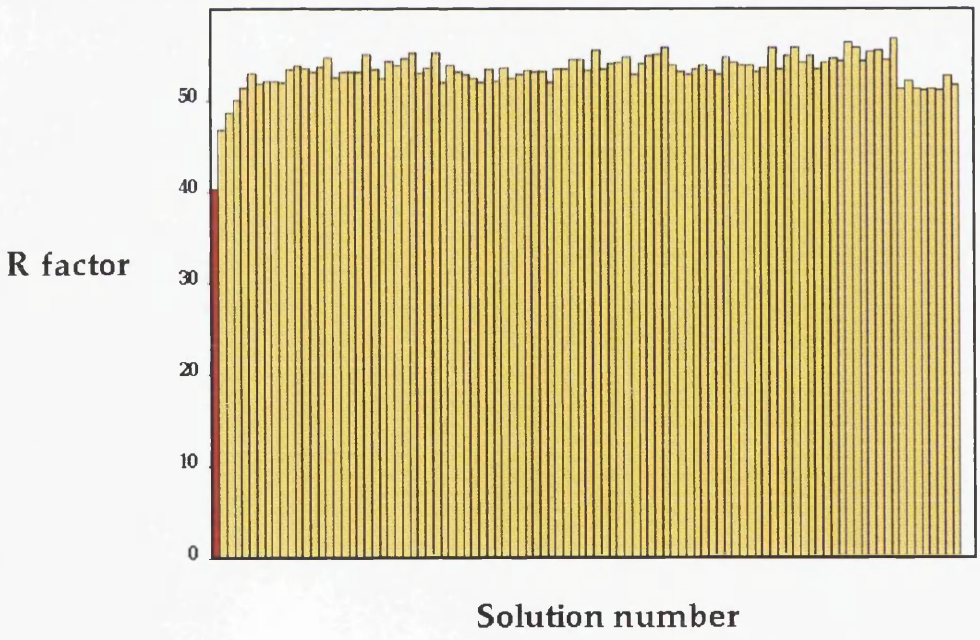


Figure 3.4: Translation function solutions for dimeric mutant

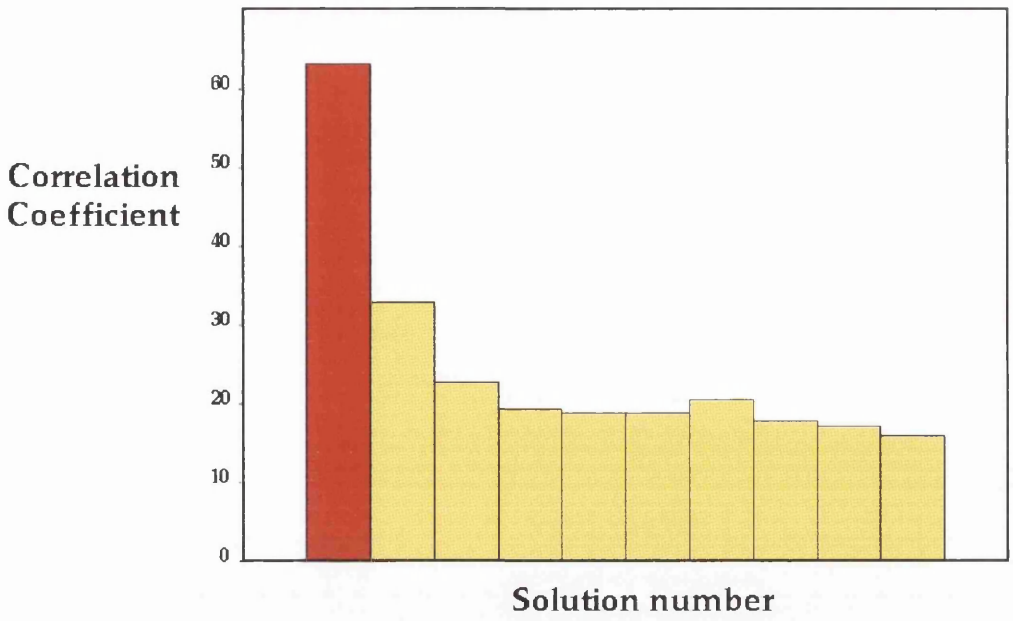


Figure 3.5: Rigid-body fitted solutions for dimeric mutant

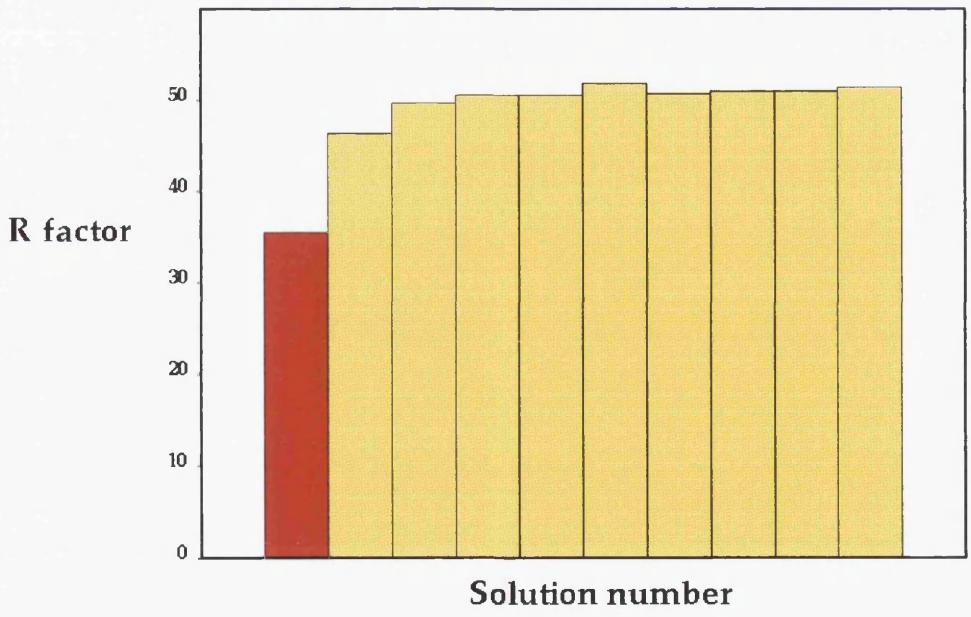


Figure 3.6: Rigid-body fitted solutions for dimeric mutant

to the noise level. In the case of four-fold NCS, the situation becomes even more difficult.

The orthorhombic crystal form of the tetrameric mutant was also solved by Molecular Replacement using the AMoRe program, with the initial main-chain trace of the monomeric mutant as a search model (The solutions are not shown graphically in this case). In the rotation function, two solutions clearly stood out from the rest. These were seen to be non-crystallographically related, as described above. A translation function search for two molecules, followed by rigid-body refinement, gave an outstandingly clear MR solution for two molecules. These two monomer molecules formed half of the MIP-1α tetramer, the other half being related by a crystallographic two-fold rotation. Calculation of a self-rotation function, also using the program AMoRe, in addition to inspection of the MR solution, showed the two monomer solutions to be related by a non-crystallographic rotation. Fitting of the initial model into electron density maps and refinement of the orthorhombic form of the tetrameric mutant is described in Section 3.9.

The monoclinic form of the tetrameric mutant was also solved using AMoRe and with the initial main-chain trace of the monomeric mutant as a search model. In this case there were four independent monomers in the asymmetric unit, so the AMoRe search was for four molecules. The four correct solutions appeared as second, third, ninth and sixty-fifth highest solutions, respectively, in the rotation function. None of the correct rotation function peaks were visible above the noise level, but in the case of the first, second and third MR solutions, there was a clear distinction between the correct solution and the incorrect solutions after a translation search. In the case of the fourth MR solution, even a translation search did not distinguish between the correct solution and the incorrect ones, and it was not until rigid-body fitting was performed that this solution also became apparent. Fitting of the initial model into electron density maps and refinement of this crystal form of the tetrameric mutant is described in Section 3.9.

3.8 Molecular Replacement: Why did it not work?

Initial MR attempts used the structures of MIP-1β and RANTES as search models, but were completely unsuccessful. Once it became clear that using the entire RANTES or MIP-1β molecule would not result in success, various truncated versions of both molecules were tried, again without success. While it is clear from the MIP-1β and RANTES structures that there are regions in which even very similar β-Chemokines differ, the core structure, and particularly the secondary structure elements, were expected to be very similar. It was presumed therefore that the major reason for the lack of success using MR was the difficulty in identifying which parts of these two molecules would be inappropriate for inclusion in the search model. Given the comparisons between the

MIP-1α structures and those of the other β-Chemokines presented in Section 4.3.2, it was possible to identify portions of the MIP-1β and RANTES molecules with the greatest similarity to MIP-1α, and to use only those parts in MR calculations. The intention was to solve the MIP-1α structure again, and to try to obtain some explanation for the apparent unsuitability of NMR structures as MR search models.

3.8.1 Molecular Replacement using NMR Search Models

The areas of the NMR search models identified as being highly similar to the correct, refined MIP-1α coordinates were effectively the areas of defined secondary structure, and as can be seen in Table 4.2, the RMS deviations between MIP-1α and the equivalent portions of the MIP-1β and RANTES molecules were just 1.175Å and 0.995Å respectively. These figures were sufficiently low as to imply that the truncated versions of both MIP-1β and RANTES could successfully be used to solve MIP-1α by MR.

However, when the experiment was performed with AMoRe, neither search model provided any solutions. Although both models contained considerably fewer atoms than the respective unabbreviated proteins, their contents were sufficient that the correct solution should have been visible above the noise level. Solutions were obtained from both crystal forms of MCP-1 using truncated search models containing identical numbers of residues. It is therefore unclear precisely why the MR could not be solved using the MIP-1β and RANTES models.

3.8.2 Molecular Replacement using MCP-1

As part of the investigation of the failure of MIP-1β and RANTES as MR search models, an attempt was made to solve the MIP-1α structure using both crystal forms of the Chemokine MCP-1 as search models. Although this protein has significantly lower sequence identity relative to MIP-1α than the two NMR models, it was possible to solve the MR using either crystal form.

Perhaps surprisingly, a clear solution was obtained using the entire *p*-form of MCP-1, while the entire *i*-form gave no result. However, when the two MCP models were truncated in exactly the same way as the NMR models had been, both gave very clear MR solutions. As shown in Table 4.2, the RMS deviations between the refined dimeric MIP-1α and *p*-MCP-1 structures was 1.615Å and that between MIP-1α and *i*-MCP-1 was 1.596Å. This in itself reflects the unpredictability of the technique, as based on these figures these two models would be expected to behave identically. What is also surprising is that the RMS deviation for the *i*-form is almost identical to that observed for the entire RANTES molecule. In addition, on truncation of both MCP-1 forms to leave just

secondary structure elements, the observed RMS deviations of 0.899Å (*p*-form) and 0.749Å (*i*-form) are again similar to those observed for RANTES and MIP-1β after similar treatment. The overall implication seems to be that there is an inherent problem with the use of NMR search models in MR, and that use of a crystal structure as a search model is to be encouraged, even when NMR structures of higher homology are available.

3.8.3 Fundamental Differences between NMR and X-ray Structures

Several previous studies[118, 126] have noted that NMR protein structures make notoriously poor search models in MR. Despite apparently high coordinate similarity between the solved MIP-1α structure and both search models, the NMR structures perhaps reflect a lower effective resolution and a greater "breathability", or internal flexibility.

Table 4.2 shows that both MCP-1 structures, which were successfully used in MR, had only slightly lower RMS deviations from MIP-1α for the same areas of defined secondary structure. This suggests that the significantly different MR results for NMR and X-ray search models were caused by extremely subtle differences between those models. While this is probably partly attributable to the greater internal flexibility of the NMR models, it may equally be a reflection of the innate and certainly unnatural inflexibility of proteins within crystals.

For this reason, it may be unreasonable to expect an NMR structure, which is generally an averaged structure based on many individual solutions, to be highly similar to a crystal structure. The averaged NMR structure inherently reflects the much greater flexibility of the protein in solution relative to the protein within a crystal. The use of each individual NMR solution as a separate search model may well be more likely to provide a solution, although it would obviously consume considerably more time.

The individual NMR solutions represent a population distributed around the average NMR structure. Clearly some of them will be considerably more similar to the crystal structure than the average NMR structure is. Therefore, while the average NMR structure represents a state which can be considered flexible, and the crystal structure represents a state which can be considered inflexible, individual solutions within the population of NMR structures are likely to resemble the inflexible crystal structure, purely as there is a finite probability that an individual NMR structure would have such a conformation in a randomly distributed population.

One aspect of the MR attempts using RANTES and MIP-1β that is likely to reduce the chances of obtaining a solution is the lack of temperature factors for NMR models. This necessitates the use of uniform temperature factors over the entire search model, which is clearly unrealistic. It has

been proposed that the use of pseudo-temperature factors[118] (pseudo B-factors) can overcome this problem in certain cases. Pseudo-temperature factors can be derived from an NMR structure if each of the original NMR structures used to generate the "averaged" NMR structure is available. Every atom in each of the original structures can be assigned a temperature factor based on its deviation from the corresponding average position. This results in a set of search models in which there is a fairly realistic distribution of temperature factors, and has been shown to be crucial in the solution of some proteins using NMR search models[118, 127]. It is possible that application of this particular technique to the case of MIP-1α would lead to solution of the MR using an NMR model.

It can also be argued that in certain cases the calculation of pseudo-temperature factors is analogous to truncation of the search model, and may therefore make little real difference the Patterson function calculated from the model. In NMR structures, as in crystal structures, the secondary structure elements and the core are the most well-defined regions, and would therefore have low pseudo-B factors. Poorly ordered loops are generally even more flexible in NMR structures than in crystal structures, and would therefore tend to have high pseudo-B factors. As this temperature factor effectively determines the weight a particular atom has in the MR calculation, loops with very high values will effectively make no contribution to the calculation. While admittedly the pseudo-B calculation provides a more realistic model of the less ordered regions of an NMR structure, if there is a sharp demarcation between ordered and unordered regions it is unlikely to be useful. In addition, it fails to address the more fundamental problem - that even the most well-ordered regions of an NMR solution seem to have considerably more conformational freedom than the same regions in a crystal structure. Clearly there is no easy way to deal with this problem. However, it is possible that the combination of the pseudo-B technique with other methods such as determination of sulphur distribution from native anomalous Pattersons may lead to more NMR models being successfully used in Molecular Replacement.

3.9 Refinement

Each of the three mutants was refined using the program REFMAC[128], where possible using the Free R-factor[129, 130] in order to validate the significance of individual refinement steps. As the monomeric mutant was refined using data to 2.9Å, there were not sufficient reflections to allow the use of Free R validation. However, for the dimeric mutant and for both crystal forms of the tetrameric mutant, around 5% of reflections were selected as a Free R set and were excluded from

refinement. Final refinement parameters are shown in Table 3.5.

Mutant	Final	Final	Resolution		R_{merge}		Completeness	
	R_f	R_{free}	Low	High	Overall	High res.	Overall	High res.
Monomer	19.41%	NA	20.4Å	2.90Å	6.4%	22.7%	98.6%	99.4%
Dimer	22.52%	27.75%	27.1Å	2.30Å	5.7%	42.2%	94.0%	99.5%
Tetramer *o*	23.15%	26.69%	26.1Å	1.65Å	6.4%	43.6%	91.6%	74.1%
Tetramer *m*	20.23%	26.11%	29.4Å	2.05Å	6.7%	25.5%	87.8%	86.8%

Table 3.5: Refinement statistics for MIP-1α mutants

Abbreviations used in Table 3.5 are as follows:

R_{merge} is defined as

$$R_{merge} = \sum_h \sum_j \frac{|\bar{I}(h) - I(h)_j|}{\Sigma_h \Sigma_j \, I(h)_j}$$

(where $\bar{I}(h)$ is the mean intensity)

R_f or reliability index is defined as

$$R_f = \frac{\Sigma_{h \notin X} ||F_o(h)| - |F_c(h)||}{\Sigma_{h \notin X} |F_o(h)|}$$

and the Free R-factor (R_{free}) is defined as

$$R_{free} = \frac{\Sigma_{h \in X} ||F_o(h)| - |F_c(h)||}{\Sigma_{h \in X} |F_o(h)|}$$

where $F_o(h)$ and $F_c(h)$ are observed and calculated structure factors, respectively. X represents the "Free R set" excluded from refinement calculations.

R_{merge} and Completeness values are given over the entire resolution range ("Overall") and for the highest resolution bin used in refinement ("High res.").

In each case, a bulk-solvent correction was performed using the program XPLOR[131]. Bulk-solvent terms were output from XPLOR as partial structure factors, which were then included in the REFMAC refinement calculations. Hydrogen atom contributions to the calculated structure factors were included in a similar way, partial structure factors being calculated using the CCP4 programs HGEN and SFALL.

Refinement used all available data, as REFMAC assigns weights to individual reflections based on their standard deviations, and is therefore considerably more robust than other refinement programs when including weak data. For instance, although the R_{merge} in the highest resolution shell

of the orthorhombic tetramer dataset was over 40%, it was not necessary to cut back the resolution to the point where R_{merge} was around 25%, as is conventionally done in other refinement programs. However, in the case of the dimeric mutant, where the highest resolution shells had R_{merge} values of over 70%, the resolution had to be cut back to 2.3Å, where the R_{merge} was 50%.

Data were also scaled anisotropically during refinement, and this correction in particular was observed to considerably improve R_{free} relative to R_f. Minimisation was by the conjugate direction method, using the "–log-likelihood" residual. Geometric restraints were calculated using PROTIN. At the end of each refinement cycle, maps were calculated from the coefficients output by REFMAC, using the CCP4 program FFT. The maps were analysed using the program ARP[132], in order to assign solvent molecule positions. Initially, a 4.5σ cutoff was used to determine whether electron density represented solvent molecules, but this was lowered to 3σ as the refinement proceeded and the model improved. Individual refinement stages were interspersed with manual rebuilding using the program O[133].

4. CHEMOKINE STRUCTURES

4.1 Descriptions of other Chemokine Structures.

4.1.1 Introduction

To date the structures of several Chemokines of both the α and β subfamilies are known. No structures have yet been obtained for proteins of the proposed γ or δ subfamilies. While there are similarities between the α- and β-Chemokines, analysis of their relative structures provides one way of differentiating between them.

The first structure of an α-Chemokine was obtained in 1989, and the first β-Chemokine structure in 1993. There are now several examples of each type, but there have been many surprises along the way. Given the vast recent expansion in the amount of relevant sequence information available to the scientific community, and the consequent increase in the rate at which new Chemokines and even new Chemokine families are being identified, it seems likely that the future will hold many more surprises.

4.1.2 α-Chemokine Structures.

Given the high degree of sequence similarity throughout the Chemokine family, it was expected that their monomers, dimers and higher aggregates would all have the same topology. Hence when the first Chemokine structure, of bovine Platelet Factor 4, was solved in 1989[107], it was expected that the remaining proteins could all be modelled on it. That structure revealed the interactions that gave rise to both dimer and tetramer, and showed how the simple monomer building-blocks could be arranged to form the rather beautiful tetramer. The monomer structure consisted of somewhat less α-helix and more β-sheet than had been predicted for the human form of the protein. Overall there was 18% helix and 29% sheet in the *b*PF4 structure.

This structure also hinted at some of the difficulties experienced by many investigators attempting structural analyses of Chemokines. From initial crystallisation[134] and diffraction to 2.8Å, it took nearly five years to solve the structure, and then only at 3.0Å. The first 23 residues were missing, and the final R-factor was 28.0%.

Nevertheless, the secondary structure of the protein was revealed. The basic topology was a three-stranded antiparallel β-sheet, arranged in a Greek key, surmounted by a C-terminal α-helix, as shown in Figure 4.1. From the N-terminus, the first 23 residues were unobserved; this was presumed to be due to proteolytic processing and disorder. The CXC motif was involved in twin disulphides which held together the secondary structure elements. Extending from the disulphides towards the C-terminus, there was a large open loop turning into strand I, a hairpin loop into strand II, a reverse turn into strand III and finally an 11-residue α-helix.

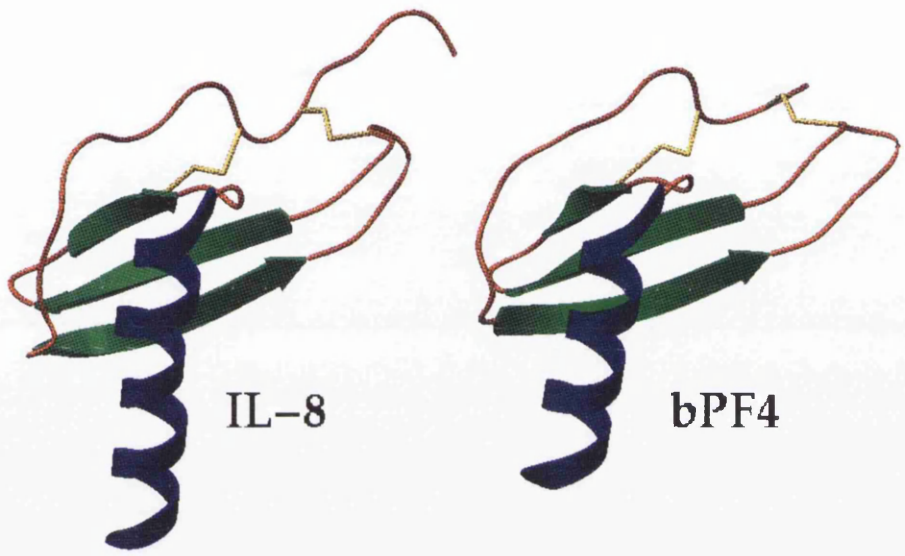


Figure 4.1: Monomer structures of IL-8 and bPF4.

The interactions which gave rise to the dimer primarily involved strand I and the final few residues of the helix. The dimer packing allowed the three-stranded β-sheets of two monomers to come together and form an extended six-stranded β-sheet. This resulted in the helix from one monomer overhanging the β-sheet of another, making stabilising interactions. The tetramer was then formed by packing two of these dimers back to back, stabilised by salt-bridges and electrostatic interactions.

The solution structure of a second α-Chemokine, Interleukin-8, was next to be described[135, 136]. IL-8 was known to be dimeric, and as expected the dimer was formed in exactly the same way as the PF4 dimer. The crystal structure of Interleukin-8 was solved shortly afterwards[137] by Molec-

ular Replacement using the IL8 solution structure as a search model. While the crystal structure was broadly similar to the solution structure, there were some specific differences, most notably in the monomer-monomer interaction and the distance between the two helices. It is notable that even several years later there are still very few proteins that have been solved by MR from an NMR model.

More recently, the structures of several other α-Chemokines have been described. The crystal structure of the human homologue of bovine PF4 was solved to a resolution of 2.4Å in 1994[108]. The structure was similar to that of *b*PF4, with four monomers forming a tetramer with pseudo-222 symmetry.

The solution structure of Melanoma Growth Stimulating Activity (*gro*-MGSA) was also published in 1994[106]. MGSA had been expected to form a dimer similar to that of IL-8, and that was exactly what was observed. MGSA was found to be more similar to the crystal structure of IL-8 than to any of the other α-Chemokine structures.

The most recently described α-Chemokine structure was that of NAP-2[138] which was solved to 1.9Å by X-ray crystallography. This particular molecule was actually a mutant in which Met 6 had been replaced by Leu in order to enable expression. The structure was a tetramer with identical topology to the PF4 structures, and had RMS deviations of 1-2Å when compared to the other known α-Chemokine structures. It is assumed that PBP, the precursor of NAP-2, and its other truncation products, CTAP-III and β-TG, will have essentially the same structure.

4.1.3 β-Chemokine Structures.

While the first β-Chemokine structure was not solved until 1995, a model of the MCP-1 (MCAF) structure, based on the IL-8 solution structure, was produced in 1991[139]. It predicted that MCP-1 in particular, and all β-Chemokines by extrapolation, would adopt the monomer, dimer and tetramer architecture seen for the α-Chemokines.

It was therefore something of a surprise when the first actual β-Chemokine structure, the solution structure of the MIP-1β, revealed a completely new method of dimerisation[82]. While the MIP-1β monomer, as expected, showed the same secondary structure elements that were present in the α-Chemokines, the dimer interface was the N-terminal region. The resulting dimer was elongated and rod-like, in contrast to the compact and approximately globular IL-8 dimer (See Figure 4.2). The MIP-1α structures were expected to be more similar to the MIP-1β structure, on the basis of sequence identities. The differences between the expected and observed structures were rationalised by comparing the relevant regions of the α- and β-Chemokine sequences. Strand

I of MIP-1β contains hydrophilic residues that would be buried in a hydrophobic pocket on formation of the IL-8 dimer, but remain exposed on formation of the MIP-1β dimer. The orientation of the N-terminal peptide is also very different in these two monomers, making it easier for MIP-1β but very difficult for IL-8 to dimerise in this way.

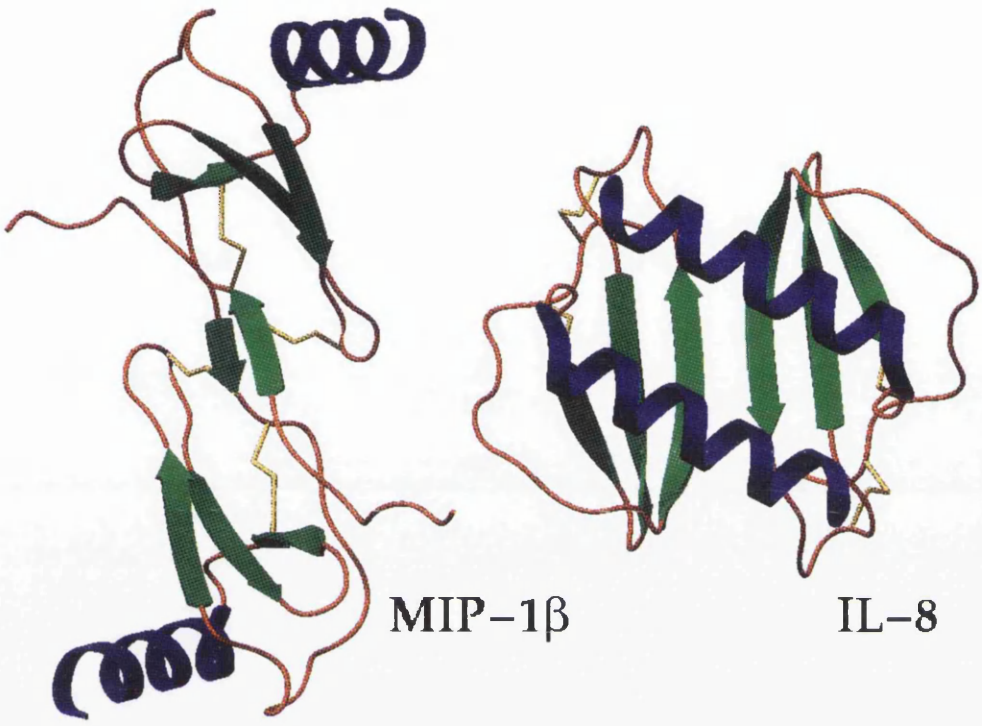


Figure 4.2: Dimer structures of IL-8 and MIP-1β.

Following this result, there was an expectation, based on sequence comparison, that the β-Chemokines would all form this type of dimer. To date most known structures have done so, but there have been further surprises as well. The solution structure of RANTES[92] in 1995 showed the MIP-1β-type dimer and supported this theory. RANTES has also been crystallised and preliminary diffraction data collected[140], but the crystal structure is yet to be reported. The first crystal structure of a β-Chemokine to be reported was that of MCP-1[109] which was solved in two crystal forms. In each of these an MCP-1 dimer similar to that of MIP-1β was observed, but in one form a dimer similar to that of IL-8 could also be seen. MCP-1 actually formed a tetramer, with one of the nominal dimers of the α type and the other of the β type (See Figure 4.3). The α-Chemokine type dimer did not have a continuous six-stranded β-sheet however, as the distance between the two

MCP-1 molecules was around 2Å greater than in IL-8.

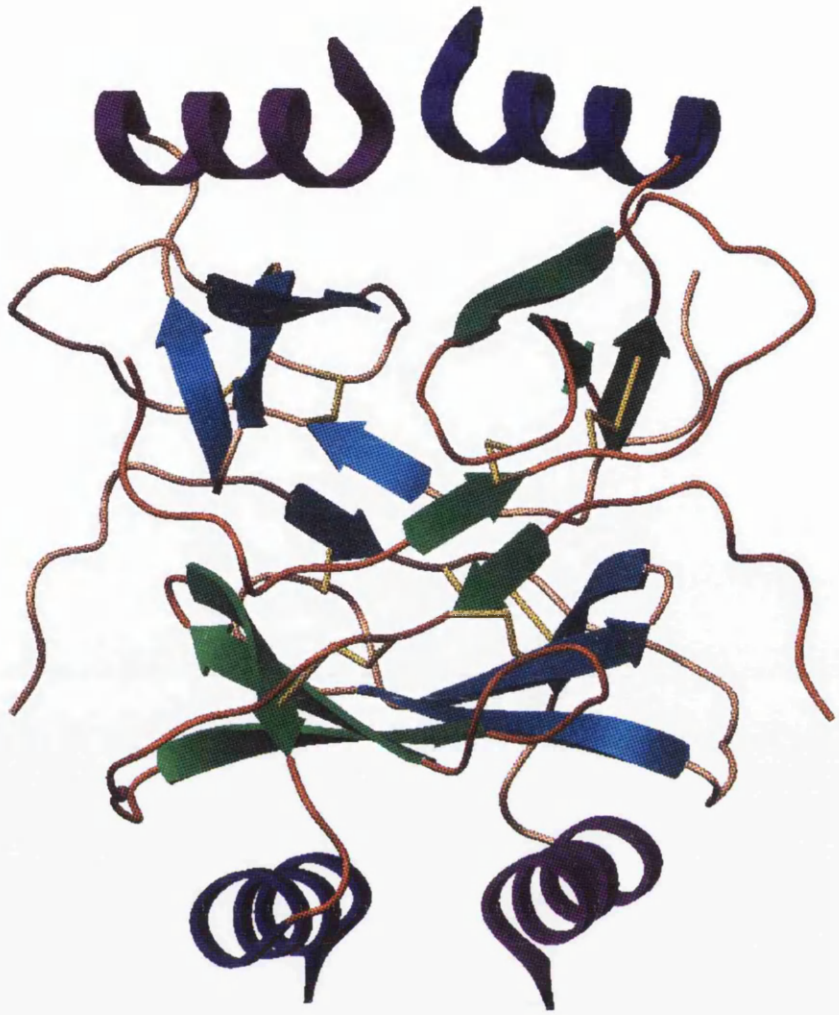


Figure 4.3: MCP-1 tetramer. The two distinct MIP-1β-type dimers are coloured differently.

Another β-Chemokine, exhibiting another distinct type of tertiary structure, has also been solved. Monocyte Chemoattractant Protein-3 (MCP-3) was predicted to be unable to form dimers or higher aggregates on the basis of ^{1}H NMR spectroscopy and analytical ultracentrifugation[141]. However, MCP-3 was subsequently shown to form the α-type dimer, similar to that observed for MCP-1 when the full solution structure was determined[110].

4.1.4 *Structures of other Chemokine Types.*

To date, no γ- or δ-Chemokine structures have been published. Sequence comparison with both α- and β-Chemokines suggests that these molecules are likely to have very similar monomer structures. However there will be interesting differences as a consequence of the different Cysteine distributions. In particular, the structure of a γ-Chemokine should indicate the importance of the Cys^{II}-Cys^{IV} disulphide in maintaining the integrity of the representative Chemokine monomer.

4.2 *Descriptions of MIP-1α Structures.*

The various mutants and crystal forms have provided a total of eight distinct MIP-1α monomer structures. Comparison within this group has shown that the differences between these structures are small and can generally be explained in terms of differences in crystal packing and crystal contacts. Comparison of each of the MIP-1α structures with the coordinates of other members of the β-Chemokine family highlights the regions of the proteins which vary most markedly. Therefore these regions are probably responsible for the differences in receptor affinity and biological activity.

4.2.1 *Structures of Monomeric and Dimeric Mutants.*

Given the identical Spacegroup and near identical Unit Cells observed for crystals of the monomeric and dimeric mutants, it was expected that there would be little or no difference between these two structures. This fact is borne out by the RMS deviation between the two structures of just 0.345Å for all residues. Since the dimer structure was refined at slightly higher resolution, only it will be described here, although the monomer structure can be considered to be essentially identical but less well defined.

The monomer structure is entirely as expected, showing the characteristic three-stranded Greek Key and single α-helix motif of all other Chemokines (See Figure 4.4). The short fourth strand of β-sheet near the N-terminus, expected to be involved in dimerisation and also observed in the structures of MIP-1β and RANTES, is also present.

As in both of those proteins, the first few residues at the N-terminus of MIP-1α are highly flexible and are not seen in the crystals due to disorder. This is unfortunate as these residues are almost certainly important if not crucial for many of the biological functions ascribed to MIP-1α. The N-terminus is present from residue Gly 4 on, stabilised at this point by a hydrogen bond from Asp 6 O to Cys 50 N. This interaction is conserved in MIP-1β, RANTES and MCP-1,

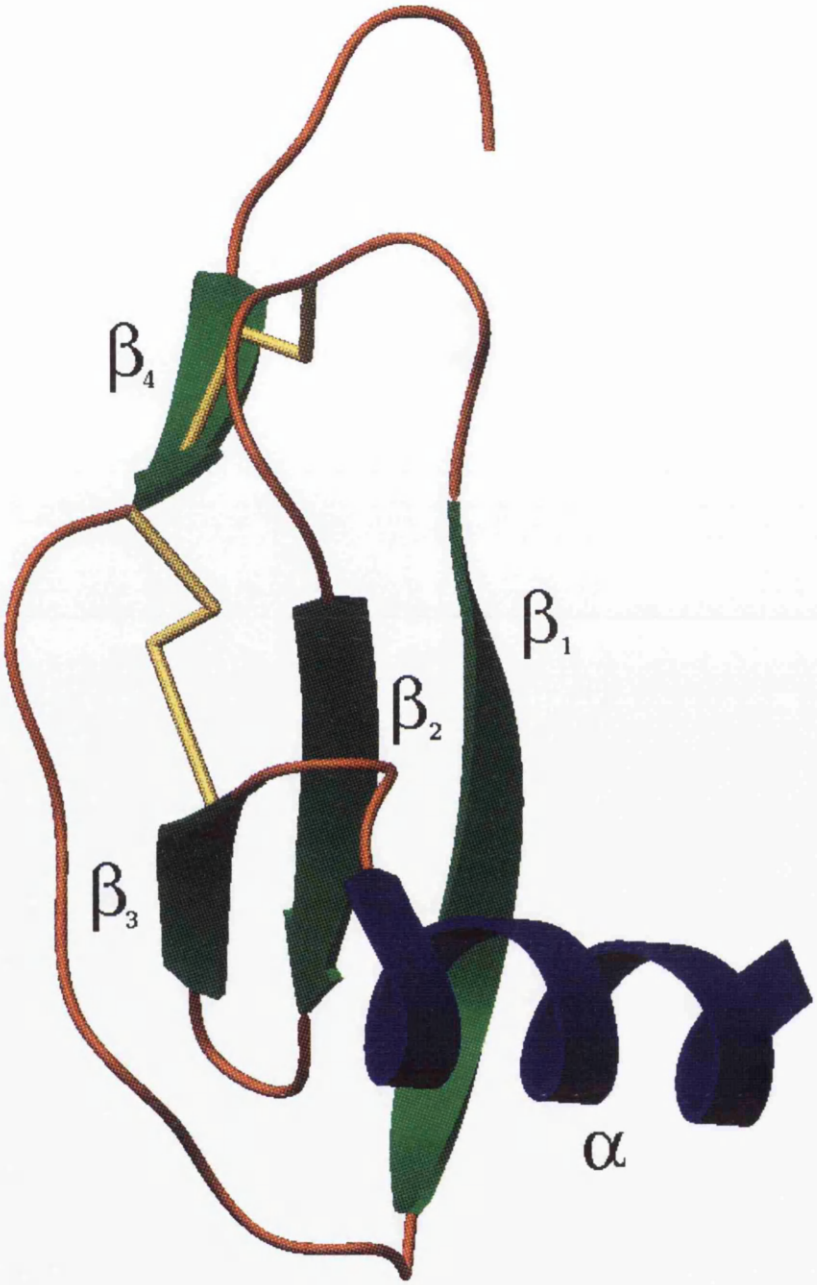


Figure 4.4: Secondary structure elements of MIP-1α.

and is probably very important in maintaining the integrity of this dimer. The next few residues form a short β-strand, and are extremely well defined in the density. Temperature factors for this part of the molecule are lower than average, indicating that this region is effectively buried in the structure. The β-strand interacts with an identical region of a symmetry-related molecule in antiparallel fashion, forming four hydrogen bonds. The same H-bonding pattern is also seen in all the β-Chemokine structures in which the dimer is formed by this method (See Figure 4.5). The four H-bonds are: Thr A9 N to Cys B111 O; Thr A9 O to Cys B111 N; Cys A11 N to Thr B109 O; Cys A11 O to Thr B109 N. Cys 11 makes a disulphide bond with Cys 34 and Cys 12 makes a disulphide with Cys 51. Both have the left-handed spiral conformation, characterised by dihedral angles[I] $\chi_1=\chi_1'=-60^o$, $\chi_2=\chi_2'=\chi_3=-90^o$. As previously reported[82], the conformation of the first disulphide differs between α and β Chemokines, and partially accounts for some of the differences in spatial distribution of secondary structure elements between the two families.

After the disulphide pair, a long unstructured loop extends to the first β-strand. Phe 13, at the start of the loop, inserts its side-chain into a hydrophobic pocket and therefore appears to be important in stabilising the dimer. Tyr 15 appears to be involved in stabilising the loop, both by hydrogen-bonding through OH to Pro 37 O, and by acting as a spacer between the loop and strands II and III. Both Arg 17 and Lys 18 are solvent-exposed, and Arg 17 in particular is poorly defined in the density. Ile 19 also helps to stabilise the loop, this time where it packs against the hydrophobic core of the monomer (See Figure 4.8). The side-chain of Arg 21 adopts an unusual conformation, curving back round on itself to facilitate hydrogen-bonds to Asn 64 OD2 and to Ile 24 O, and appears to be important in anchoring the helix onto the rest of the molecule (See Figure 4.6). The two residues at the apex of the loop, Asn 22 and Phe 23, are the least well-defined, mainly due to their high temperature factors. Residues Arg 21 to Phe 23 actually form a single turn of 3_{10} helix. The dihedral angles are distorted slightly from the ideal values, but the characteristic hydrogen bonds from $n \rightarrow n+3$ are present for Pro 20 O → Phe 23 N and Arg 21 O → Ile 24 N.

The first β-strand begins at Asp 26 and extends as far as Thr 30 (See Figure 4.7), making five hydrogen-bonds to strand II. In addition to these, strand II, which runs from Gly 38 to Thr

[I] These are defined as follows

- χ_3 : Cβ - Sγ - Sγ′ - Cβ′.
- χ_2 and χ_2' : Cα - Cβ - Sγ - Sγ′.
- χ_1 and χ_1' : N - Cα - Cβ - Sγ.

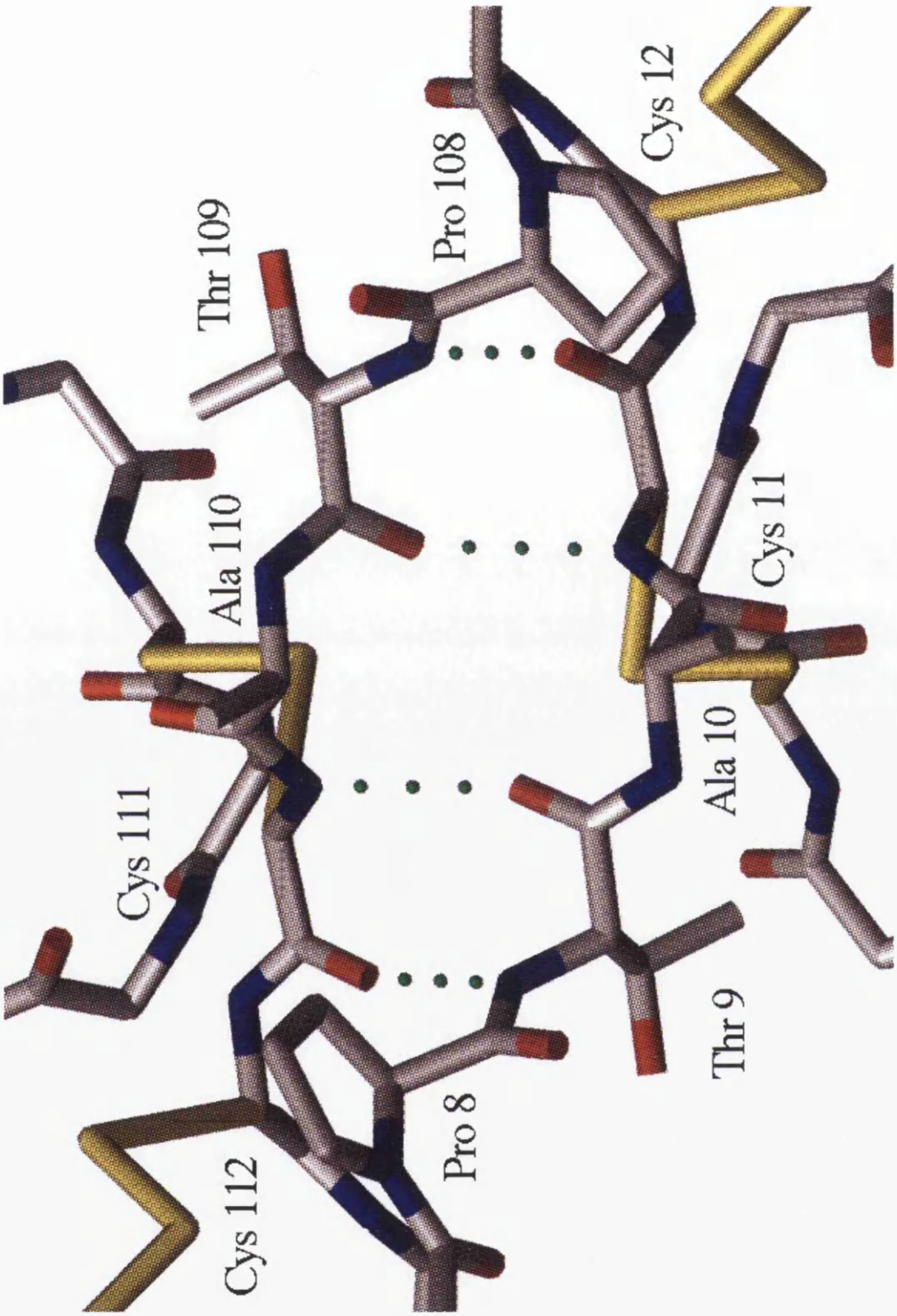


Figure 4.5: Hydrogen bonding pattern at the dimerisation interface.

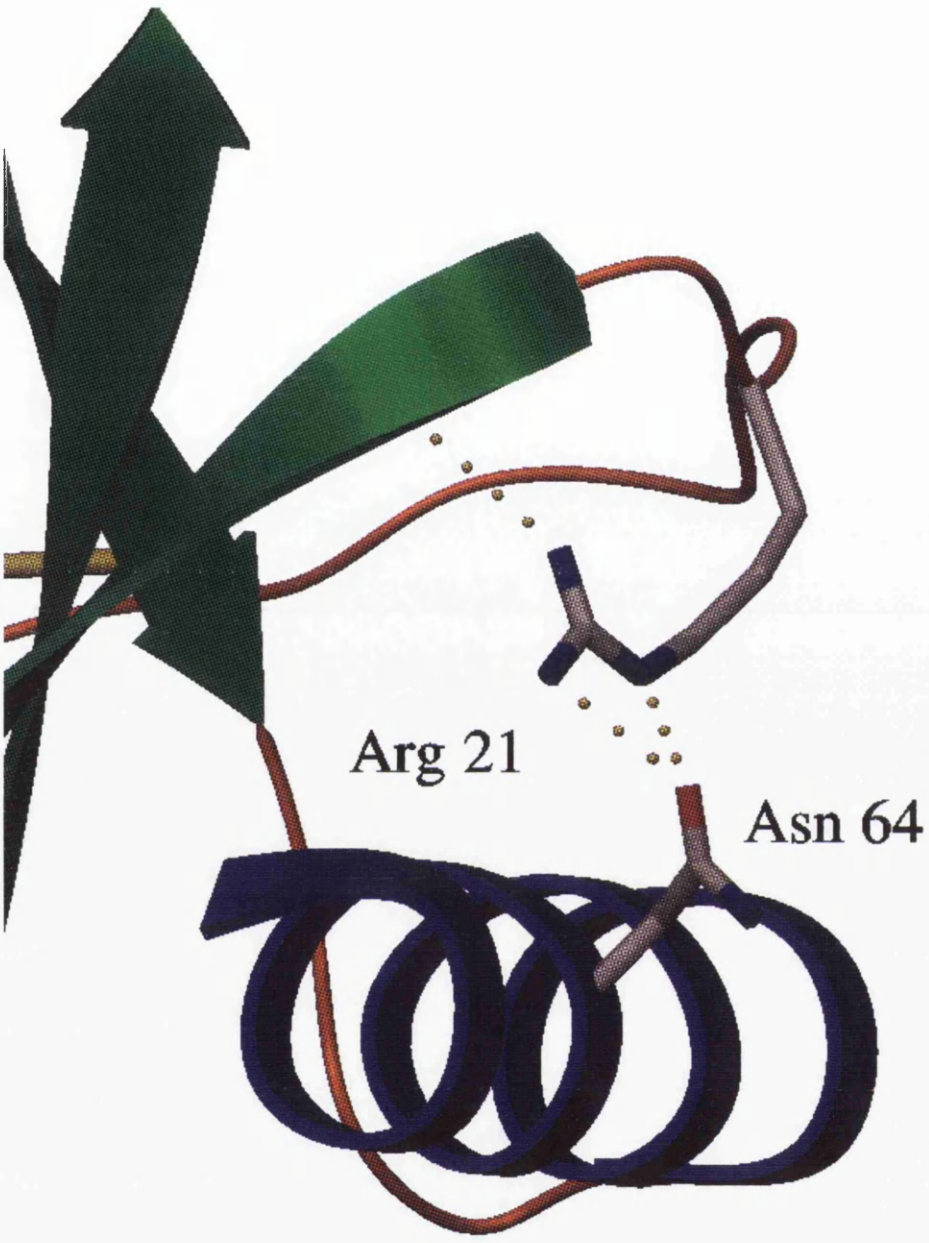


Figure 4.6: Arg 21 H-bonding to Asn 64 and to β-strand I.

43, makes a further five hydrogen bonds across to strand III, which runs from Arg 47 to Ala 51. Strand III makes hydrogen bonds to the long loop between the cysteine pair and strand I, but none of these interactions are standard for antiparallel β-sheet.

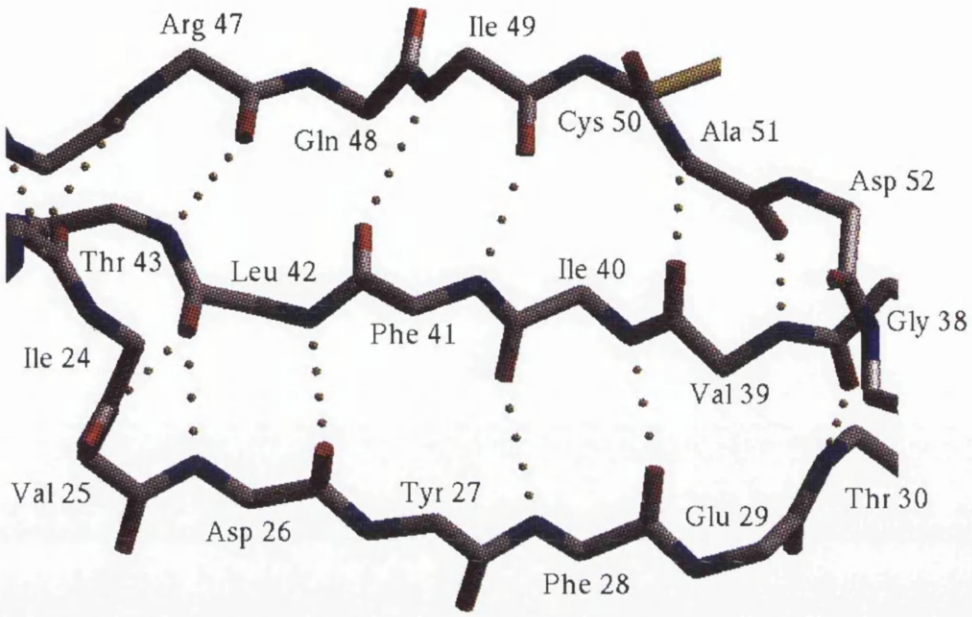


Figure 4.7: Hydrogen-bonding pattern in the antiparallel β-sheet.

The α-helix runs from Thr 56 to the C-terminus at Ala 69. Both α- and β-Chemokines tend to have amphipathic α-helices, and MIP-1α is no exception. The side of the helix which packs against the β-sheet is composed almost entirely of hydrophobic residues, including Trp 57, Asn 68, Tyr 61, Leu 65, Ala 69, Ile 62 and Val 58, whereas the solvent-exposed face of the helix consists mainly of residues which are hydrophilic, charged, or which would be charged in the native protein. Many of these hydrophobic residues are incorporated in the hydrophobic core of the monomer, with Trp 57 and Tyr 61 in particular interacting with other aromatics to hold the helix in place (See Figure 4.8).

The turn between strands I and II is anchored by the disulphide bond between Cys 11 and Cys 34 at the centre of the turn. The regions immediately to either side of Cys 34 can therefore be considered as separate entities. Residues Ser 31 to Cys 34 form a classic type I tight turn[142] with a hydrogen bond between Ser 31 O and Cys 34 N. On the other side of the Cys residue, the presence

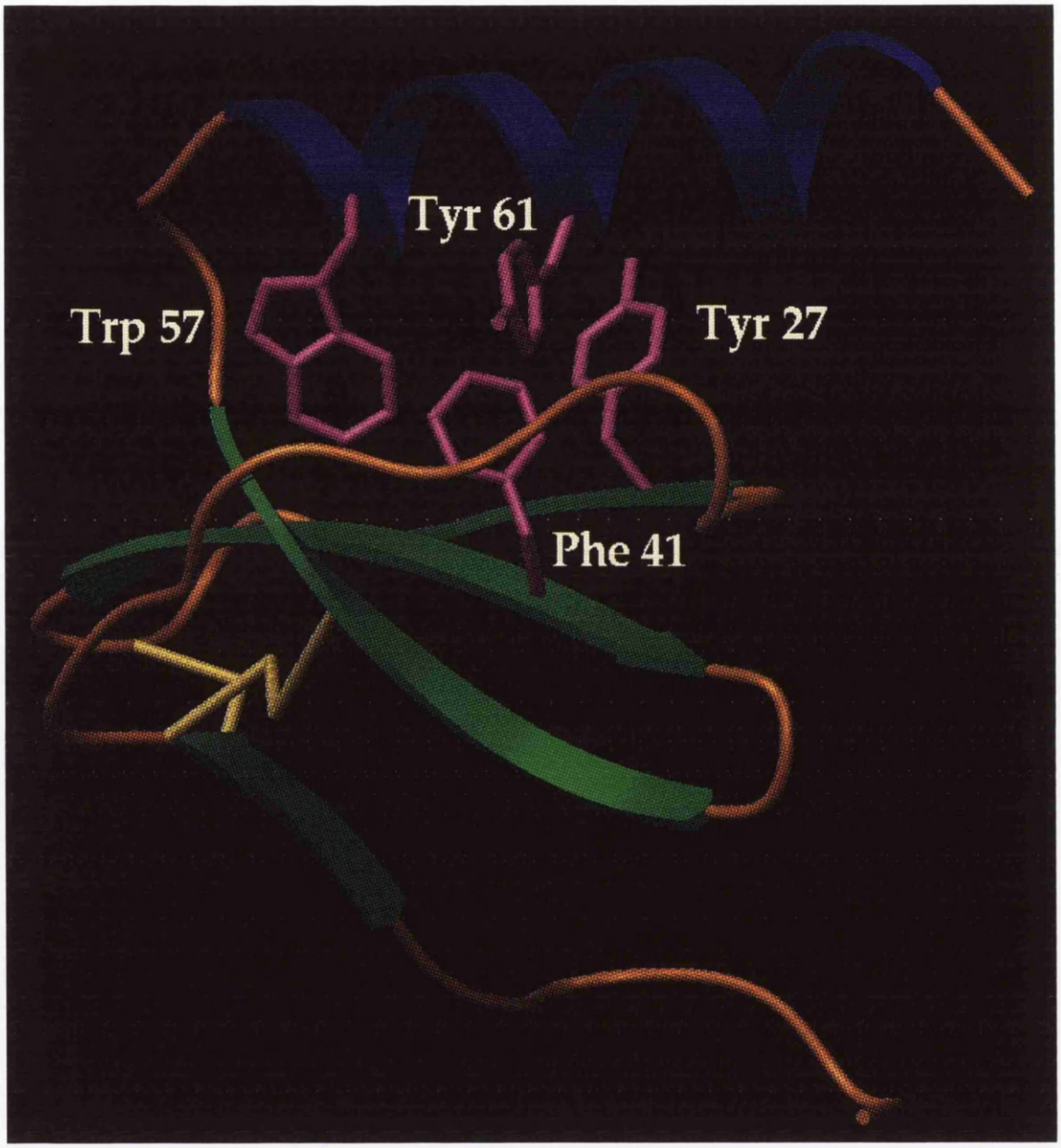


Figure 4.8: Aromatic residues in the hydrophobic core of MIP-1α.

of Pro 37 prevents any hydrogen bonding.

The turn between strands II and III is another classic type I turn. This region of the molecule is not well-defined, and has a high average temperature factor. In addition to the hydrogen bond between Thr 43 O and Asn 46 N which defines the turn, there is a main-chain-side-chain hydrogen bond between 43 Thr OG and Arg 47 N, which is almost a continuation of the β-sheet hydrogen-bonding between strands II and III.

The region between strand III and the helix also contains a type I tight turn between Asp 52 and Glu 55, with a hydrogen bond between Asp 52 O and Glu 55 N as well as interactions between the Glu 52 side chain and main chain atoms. The density for this loop is fairly good, although the side chain of Lys 54 is not visible; it is highly solvent-exposed and presumably is very flexible.

The final refined dimer model, at 2.3Å resolution, is missing 3 N-terminal residues and contains 49 solvent molecules. The final refined monomer model, at 2.9Å resolution, is missing 4 N-terminal residues and contains 17 solvent molecules. The average temperature factor for all atoms is 42Å^2 for the monomer and 60Å^2 for the dimer.

Overall then, the structure is almost exactly as would have been predicted from the structures of MIP-1β and RANTES. The greatest differences between MIP-1α and these molecules seems to be in the N-terminal region and in the loop anchored by Cys 34 between strands I and II. This is further evidence to support the proposal that this surface of the molecule must be involved in some way in receptor interaction. The α-helix is longest in the MIP-1α dimer, running all the way to the C-terminus, but this is possibly due to restraint of the last few residues by crystal contacts. In solution, these residues may be much more mobile.

4.2.2 Structure of Orthorhombic Form of Tetrameric Mutant.

In the orthorhombic crystal form of the tetrameric mutant, the asymmetric unit contains two independent monomers, which will be denoted $C222_1a$ and $C222_1b$. The dimer thus formed is related to another dimer by a crystallographic two-fold rotation in order to form the tetramer. It was initially difficult to identify the physiologically relevant tetramer, although eventual comparison with the monoclinic crystal form allowed unambiguous identification of the correct tetramer (See Figure 4.9. This process is discussed in more detail in Section 5.1.

As expected, the structures of each of the two independent monomers within the tetramer were very similar to the monomeric and dimeric mutant structures. The RMS deviations between the secondary structure elements of the dimer structure and those of $C222_1a$ and $C222_1b$ were 0.376Å and 0.218Å respectively. However, comparison of the entire structures gave higher figures

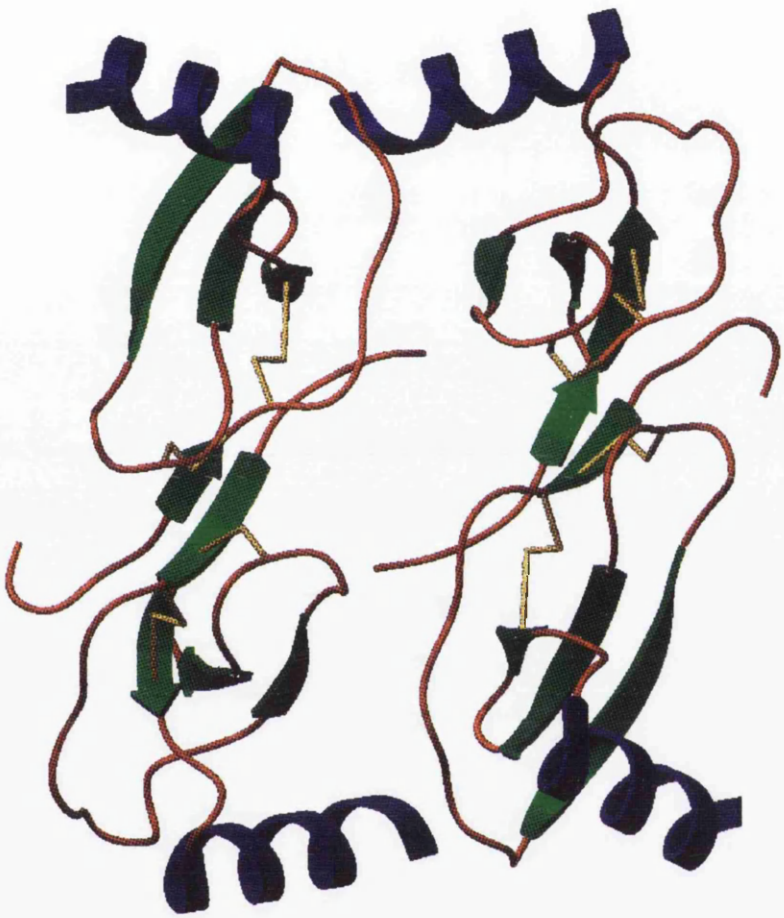


Figure 4.9: Structure of the MIP-1α tetramer.

of 0.444Å and 0.592Å (See Table 4.1). This suggested that the loop regions of both $C222_1a$ and $C222_1b$ were somewhat different to the corresponding regions in the dimer structure. While this could be construed as perhaps an artefact of the higher resolution of the tetramer structure, the comparison of $C222_1a$ and $C222_1b$ in the same way (See Table 4.1) proved that there are definite differences between the two independent monomers in the orthorhombic tetramer structure.

Each of the monomers exhibits the expected motif of three-stranded Greek-key and C-terminal α-helix. The three-residue β-strand on the N-terminal peptide is particularly well defined in both monomers. One difference between the monomers observed in the orthorhombic tetramer structure and those seen in the monomer and dimer structures was the presence of a very large electron density peak sitting on a crystallographic two-fold axis. The surrounding density suggested that whatever was responsible for the peak was being coordinated octahedrally by two crystallographically related molecules. The equatorial positions were occupied by main-chain carbonyls from the loop containing Cys 34, from two independent monomers. Solvent molecules occupied the other coordination positions. Consideration of the contents of the crystallisation mixture suggested that only a Calcium ion could have resulted in such a peak. The interatomic distances were consistent with a Ca^{2+} ion, and so one was included in the model and eventually refined to a reasonable temperature factor, consistent with those of its ligands and the protein atoms in the vicinity. The octahedral coordination of the Calcium ion is shown in Figure 4.10.

It is clear from inspection of the electron density maps obtained from this mutant that portions of this molecule are extremely well resolved, even for such a high resolution structure. In particular, the residues immediately surrounding the Calcium ion, as shown in Figures 4.10 and 4.11, have very low temperature factors, and consequently the quality of the density around these residues is very high. Comparing this region of the tetramer to the the corresponding residues in the monomer and dimer structures, it is clear that the loop is considerably more ordered in the tetramer. This can be attributed purely to the presence of the Ca^{2+} ion and the resulting crystal contacts.

Although the dimer observed within the orthorhombic tetramer crystals appeared to be exactly the same as that seen in both monomer and dimer, there were some slight differences. The angle between the two monomers was slightly different. This was probably a consequence of the presence of the Ca^{2+} ion in the tetramer, resulting in a modified conformation for the loop containing Cys 34. Any motion of this residue would be correlated to a motion of Cys 11, and would therefore result in movement of the entire N-terminal peptide relative to the remainder of the molecule. A different orientation for this part of the molecule would necessitate the angle between the two monomers being different.

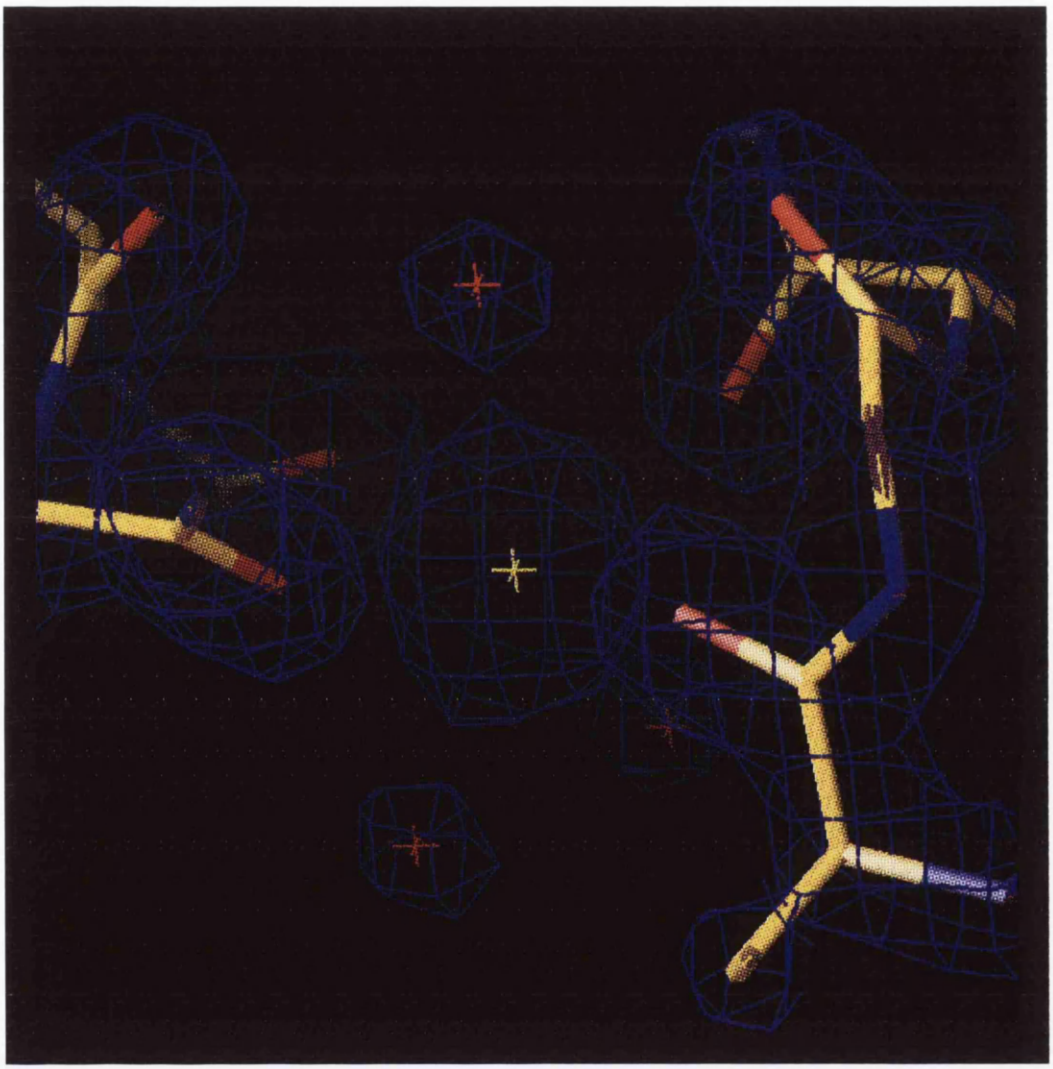


Figure 4.10: Electron density surrounding the Calcium ion.

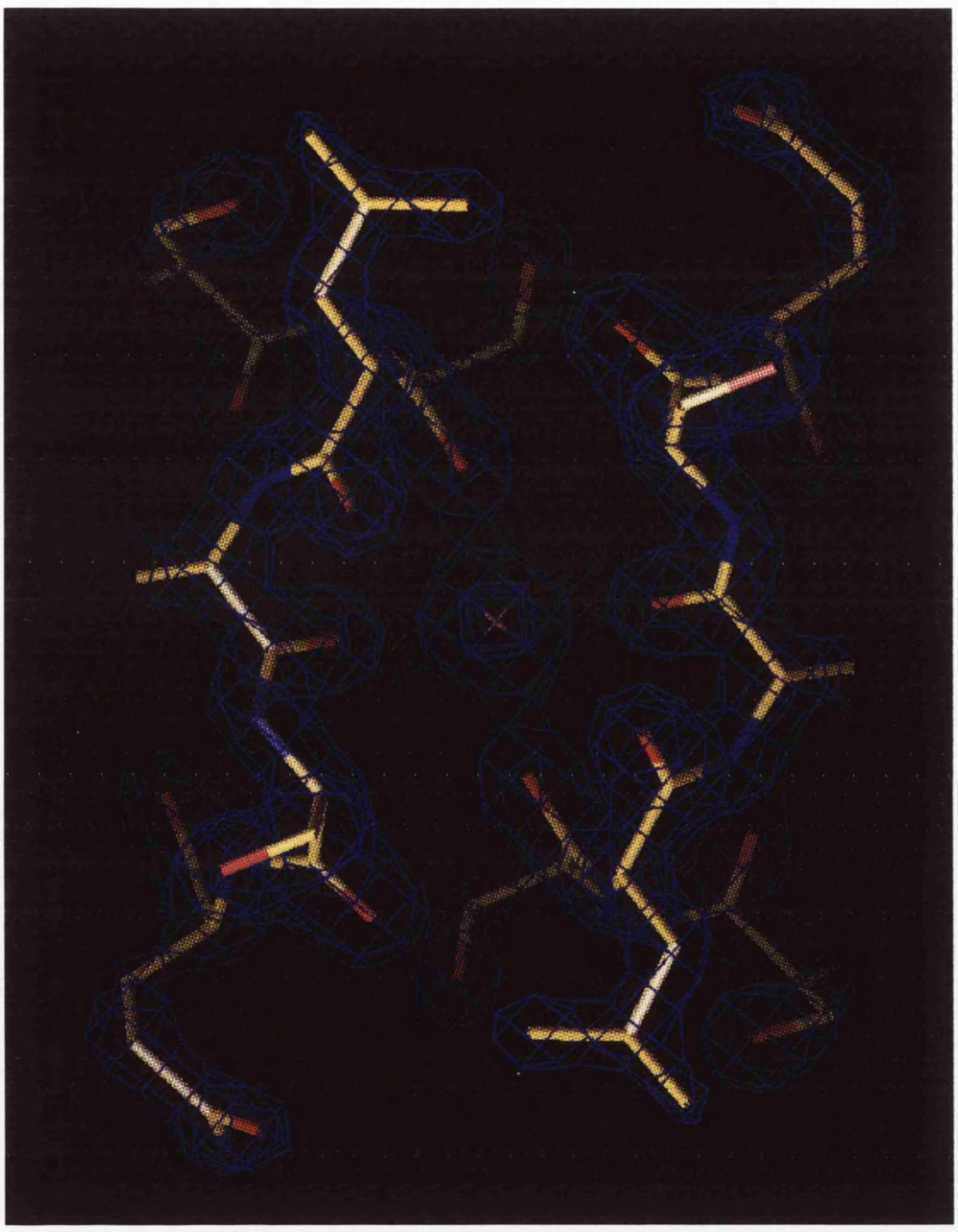


Figure 4.11: Calcium coordination: View down crystallographic two-fold axis.

The striking difference in the strength of diffraction obtained from dimeric and tetrameric crystals reflected the differing solvent contents in the two crystals. While the dimeric crystals showed diffraction to around 2.0Å, and had a fairly high solvent content of about 75%, the tetrameric crystals contained just 48% solvent. Since there are more crystal contacts in the tetrameric structure, the loop regions and exposed residues tend to be much better defined. For example, the basic loop between strands II and III is solvent-exposed and flexible in the dimeric structure, and consequently there is little density for the side chains. In the tetramer structure, this region is stabilised both by the dimer-dimer interaction that gives rise to the tetramer, and also by crystal contacts. The positions of the important side-chains of Lys 44, Arg 45 and Arg 47 are therefore well defined in the tetramer structure.

The refined model, at 1.65Å resolution, is missing 6 residues from each of the independent monomers, 3 from each N-terminus and 3 from each C-terminus. Two residues, Gln 22 from $C222_1a$, and Leu 42 from $C222_1b$, are modelled in dual conformations. The model contains a single Calcium ion, and 87 solvent molecules. The average temperature factor in the model is $41Å^2$.

4.2.3 Structure of Monoclinic Form of Tetrameric Mutant.

In the monoclinic crystal form of the tetrameric MIP-1α mutant, there were four independent monomers in the asymmetric unit. These will henceforth be referred to as $P2_1a$, $P2_1b$, $P2_1c$ and $P2_1d$. The tetrameric unit observed in this crystal form had identical topology to that seen in the orthorhombic form. As predicted, there was non-crystallographic symmetry, consisting of two independent non-crystallographic two-fold axes. A self-rotation function calculated using AMoRe[124] clearly showed that one of these axes was almost parallel to the c-axis of the crystal, but the signal for the second did not appear above the noise level.

It was immediately apparent from examination of the electron density for this crystal form that the pattern of crystal contacts for each of the monomers was quite different, and therefore there were certain regions in which there were significant differences between the monomers. For example, while the N-terminal regions of monomers $P2_1b$ and $P2_1c$ sat in the central cavity of the tetramer, those of monomers $P2_1a$ and $P2_1d$ were close to the edge of the tetramer and were consequently much more solvent-exposed. Similarly, while the C-terminal peptides of the $P2_1a$ and $P2_1b$ monomers followed α helical geometry, the $P2_1c$ monomer was α-helical only to residue Gln 66. The remaining residues adopted a non-helical conformation in order to facilitate the binding of a Ca^{2+} ion by Asp 64 and the carbonyl oxygen of Leu 67. The C-terminal region of monomer

$P2_1d$ pointed directly into a large solvent channel, as a result of which there was no interpretable density beyond residue Asp 64. In fact monomer $P2_1d$ was very highly solvent-exposed, and consequently most of the loop regions of this particular molecule appear to be disordered.

In addition to the Ca^{2+} ion bound to the helix of monomer $P2_1c$, another was observed bound on the loop containing Cys 34. This particular ion was involved in an interaction that was essentially identical to that observed in the orthorhombic crystal form, mediating contacts between two crystallographically-related molecules. Close inspection of the Ca^{2+} site on the $P2_1c$ helix showed that it too mediated crystal contacts between two independent molecules. In addition to these interactions, there was only one other tetramer-tetramer contact within the crystals. Therefore the situation is particularly interesting, with two thirds of all crystal contacts being mediated by Calcium ions.

4.3 Comparison of Chemokine Structures.

4.3.1 Comparison of the MIP-1α Monomers.

In order to assess and quantify the similarity between the eight independent MIP-1α monomers, the RMS deviations between each possible pair were measured. The deviations were defined on the basis of the overlap between the regions of the molecules with defined secondary structure.[I]. The RMS deviations for entire molecules are also given in each case. These results are shown in Table 4.1

These results show that the MIP-1α monomers are generally very similar. The analysis of the deviations between secondary structure elements indicates that each of them is essentially identical in those regions. However, there appear to be some slightly larger differences when the entire structure is considered.

This analysis could potentially indicate whether particular residues or loops change conformation on formation of the tetramer. The loop containing Arg 45 and Arg 47 is the obvious candidate. Monomers $P2_1a$ and $P2_1d$, in which this loop interacts with Glu 66 on another monomer, would be expected to have higher RMS deviations compared to the dimeric mutant than $P2_1b$ and $P2_1c$, in which this loop is solvent exposed. However, it is the RMS deviations of $P2_1b$ and $P2_1c$ relative to the dimeric mutant which are higher. It is also possible that this is a result of motion of the C-terminal helix upon tetramerisation, but that seems quite unlikely. Considering the tetramer as

[I] Secondary structure is defined as follows: β_4 - Thr 7 → Cys 12, β_1 - Asp 26 → Thr 30, β_2 - Gly 38 → Thr 43, β_3 - Arg 47 → Ala 51, α - Thr 56 → Leu 65. See also Figure 4.4

Molecule 1	Molecule 2	RMSD (Secondary structure)	RMSD (all)
Dimer	Monomer	0.274Å	0.345Å
Dimer	$C222_1a$	0.376Å	0.444Å
Dimer	$C222_1b$	0.218Å	0.592Å
Dimer	$P2_1a$	0.223Å	0.446Å
Dimer	$P2_1b$	0.266Å	0.729Å
Dimer	$P2_1c$	0.204Å	1.076Å
Dimer	$P2_1d$	0.460Å	0.607Å
$C222_1a$	$C222_1b$	0.431Å	0.756Å
$C222_1a$	$P2_1a$	0.359Å	0.496Å
$C222_1a$	$P2_1b$	0.672Å	0.914Å
$C222_1a$	$P2_1c$	0.473Å	0.854Å
$C222_1a$	$P2_1d$	0.418Å	0.527Å
$C222_1b$	$P2_1a$	0.405Å	0.755Å
$C222_1b$	$P2_1b$	0.256Å	0.372Å
$C222_1b$	$P2_1c$	0.261Å	0.464Å
$C222_1b$	$P2_1d$	0.478Å	0.722Å

Table 4.1: RMS deviations between MIP-1α monomers

a whole, it seems more feasible that the N-terminal regions of $P2_1b$ and $P2_1c$ are responsible for the higher RMS values, as their conformations are constrained due to their position internal to the tetramer.

Aside from this, the MIP-1α monomers all appear very similar in structure. It seems likely that the observed RMS deviations are primarily due to slight differences in the orientation of the N-terminal peptide in each of the monomers, and that there is no significant change to the structure of a MIP-1α dimer on formation of the tetramer.

4.3.2 Comparison between MIP-1α and β-Chemokine Structures.

The results in Table 4.2 show that, for the β-Chemokine structures known, the *i*-form of MCP-1 is the one which is most similar to MIP-1α. The *p*-form of MCP-1 is also very similar to MIP-1α, but the NMR structures of RANTES and MIP-1β are suprisingly the least similar of all.

It is not clear how much can be read into these comparisons, however. There are previous examples[137], even from within the Chemokine family, where the RMS differences between NMR and crystal structures of the same protein are greater than the differences between two crystals structures of different proteins. Therefore it is probably most useful to compare the MIP-1α structures with the coordinates of the two MCP-1 crystal forms. The RMS deviations for the secondary structure elements in MCP-1 are clearly comparable to those between the different MIP-1α crystal

Molecule 1	Molecule 2	RMSD (β)	RMSD (α + β)	RMSD (ss)	RMSD (all)	RMSD (dimer)
MIP-1α Dimer	RANTES	0.503Å	0.603Å	0.995Å	1.597Å	3.587Å
MIP-1α Dimer	MIP-1β	0.724Å	0.600Å	1.175Å	1.715Å	3.372Å
MIP-1α Dimer	MCP-1*p*	0.401Å	0.440Å	0.899Å	1.615Å	3.986Å
MIP-1α Dimer	MCP-1*i*	0.397Å	0.469Å	0.749Å	1.596Å	2.498Å

Table 4.2: RMS deviations between MIP-1α (dimer mutant) and other β-Chemokines

The regions of the molecules used to calculate these values were:

- β : Comparison of residues in strands β_1, β_2 and β_3
- α + β : Comparison of residues in β_1, β_2, β_3 and α.
- ss : Comparison of residues in β_1, β_2, β_3, α and β_4.
- all : Comparison over all residues present in both structures.

forms. However, the RMS deviations between MIP-1α and the other four known β-Chemokines are much larger when the N-terminal portions of these molecules are included in the calculations. This implies that the path of the residues between the N-terminus and the initial Cysteine pair is different in MIP-1α compared with the other β-Chemokines.

When the other members of the family are compared with each other (see Table 4.3), it becomes apparent not only that MIP-1α differs from the others, but that no pair actually have the same conformation in the region between the disulphide pair and the N-terminus. Since this is the region which forms the dimer interface, it would be expected that the dimers formed by each of the β-Chemokines would therefore be slightly different with respect to the relative orientations of the two dimers. This is reflected in the observed RMS deviations between the various dimers, all of which are significantly higher than for just the monomer.

Monomer 1	Monomer 2	RMSD (β)	RMSD (α + β)	RMSD (ss)	RMSD (all)	RMSD (dimer)
MIP-1β	RANTES	0.762Å	0.709Å	1.214Å	1.819Å	5.801Å
MIP-1β	MCP-1*p*	0.925Å	0.823Å	0.865Å	1.394Å	2.443Å
RANTES	MCP-1*p*	0.525Å	0.671Å	1.043Å	1.533Å	4.293Å
RANTES	MCP-1*i*	0.572Å	0.711Å	0.888Å	1.437Å	2.646Å
MIP-1β	MCP-1*i*	0.988Å	0.880Å	1.005Å	1.447Å	5.146Å
MCP-1*p*	MCP-1*i*	0.179Å	0.232Å	0.419Å	0.510Å	3.251Å

Table 4.3: RMS deviations between other β-Chemokine monomers

These observations may go some way towards explaining an aspect of the functionality of the Chemokines which has puzzled so many people for so long, namely the fact that their bio-

logical specificities can be so very different despite their structures being almost identical. Since the biological activities of the various members of the family overlap, and since each of the β-Chemokine receptors identified to date interacts with at least two and often several of them, it should be possible to begin to rationalise the specificities of the particular proteins by comparing their N-terminal conformations. This, however, is likely to be extremely difficult, given only a few example structures with which to work. It may therefore be more constructive to consider just the receptor CCR5, and its three ligands MIP-1α, MIP-1β and RANTES, as structures of all three ligands are available. By identifying aspects of the N-terminal regions of these specific proteins that are common to them but not to the other β-Chemokines, it may be possible to construct a basic model of the receptor. With the other β-Chemokine receptors, there may not yet be enough structural information available to perform the same type of analysis.

While the path of the N-terminal peptide is the most obvious difference between the various β-Chemokines, there are others which may be even more significant. The orientation of the N-terminus with respect to the β-strands seems to be controlled by the Cys 11 - Cys 34 disulphide and specifically by the conformation of the loop between strands I and II which contains Cys 34. This loop appears to be fairly different in MIP-1α structures compared to the other β-Chemokines. Tables 4.4 and 4.5 show the deviations in position of the sulphur atoms between MIP-1α monomer $C222_1b$ and all the other structures. $C222_1b$ was chosen as the template as it seemed to have the lowest average RMS deviations when comparing all MIP-1α structures with each other, and therefore can be considered as the most representative MIP-1α structure.

Molecule	δ-Cys 11	δ-Cys 34	δ-Cys 12	δ-Cys 50
Dimer	0.53Å	0.54Å	0.22Å	0.20Å
$C222_1a$	0.52Å	0.42Å	0.17Å	0.10Å
$P2_1a$	0.54Å	0.59Å	0.13Å	0.19Å
$P2_1b$	0.46Å	0.36Å	0.09Å	0.11Å
$P2_1c$	0.31Å	0.26Å	0.10Å	0.11Å
$P2_1d$	0.31Å	0.39Å	0.14Å	0.22Å
Average	0.445Å	0.427Å	0.14Å	0.155Å

Table 4.4: RMS deviations of Cysteine atoms in the various MIP-1α monomers
Monomer $C222_1b$ was the template for these comparisons

It is clear from Tables 4.4 & 4.5 that the positions of the sulphur atoms in the Cys 11 - Cys 34 disulphide are much more variable than those of the Cys 12 - Cys 50 disulphide. This implies that loop between strands I and II of each of these molecules which is anchored by the Cys 11 - Cys 34 disulphide is also flexible. It could be argued that this effect is mediated by Cys 11 and that the

Molecule	δ-Cys 11	δ-Cys 34	δ-Cys 12	δ-Cys 50
MIP-1β	1.35Å	2.55Å	0.65Å	0.30Å
RANTES	1.93Å	1.57Å	1.75Å	0.85Å
MCP-1*i*	1.33Å	1.51Å	0.75Å	0.68Å
MCP-1*p*	0.75Å	0.99Å	0.58Å	0.33Å
Average	1.34Å	1.66Å	0.93Å	0.54Å

Table 4.5: RMS deviations of Cysteine atoms in the β-Chemokine family.
Monomer $C222_1b$ was the template for these comparisons

movement of Cys 34 and the attached loop are a consequence of movement of Cys 11. However, the much lower deviations observed for Cys 12 and Cys 50 make this unlikely. Given that Cys 11 and Cys 12 are adjacent, any movement of Cys 11 would necessitate some displacement of Cys 12 also. The observed displacements for Cys 12 and Cys 50 are so much lower that the effect is more likely due to a displacement of Cys 34 for some reason, one consequence of which would be a correlated motion of Cys 11.)

It is clear from sequence alignment of β-Chemokines that there are also some significant differences between residues on β-strand I. Comparison of MIP-1α and MCP-1, in particular, suggests that the residues in positions 26, 27 and 29 (numbering relative to MIP-1α) are particularly important in determining the specific activities of β-Chemokines. MIP-1α and MCP-1 have essentially the same conformation in this region, but while MIP-1α has Asp 26, Tyr 27 and Glu 29, MCP-1 has Ser 26, Tyr 27 and Arg 29. In contrast to the disulphides, however, it seems to be just the relative distribution of electrostatic potential that is important in this region.

4.3.3 Comparison between β- and α-Chemokine Structures.

The Chemokine families are almost unique in the complexity and variability of their interactions, given the almost identical monomer units from which their structures are derived. The difference between α- and β-Chemokine dimers is currently the most evident example.

It is possible to rationalise most of these variations by examining the spatial distribution of charged and hydrophobic residues in the regions of the monomers which would be directly involved the interaction. Several groups have already done this[82] and their findings are very consistent with the structures known to date. However, this type of comparison has also resulted in some erroneous predictions in the past[139], and so it may be safer just to analyse the RMS deviations between the α- and β-Chemokines in order to search for the underlying pattern.

There are probably still too few examples for an experiment of this type to provide answers

that cannot be questioned. In addition, the apparent subdivision of the β-Chemokine family, into at least two and possibly more groups based on different dimer-dimer interactions, suggests that there may be even more structural distinctions within the Chemokine families than are currently appreciated.

Molecule 1	Molecule 2	RMSD (β)	RMSD (α + β)	RMSD (all)	RMSD (dimer)
MIP-1α	*h*PF4	0.833Å	1.246Å	4.442Å	15.182Å
MIP-1α	*b*PF4	6.905Å	6.574Å	4.459Å	15.123Å
MIP-1α	IL-8	0.734Å	2.073Å	6.143Å	14.355Å
MIP-1α	NAP-2	0.832Å	0.840Å	4.535Å	15.291Å
MIP-1α	*gro*-MGSA	0.742Å	0.810Å	4.543Å	14.259Å

Table 4.6: RMS deviations between α- and β-Chemokine monomers

Table 4.6 demonstrates that the MIP-1α monomer structure is comparable to the known α-Chemokine structures. While the low resolution structure of *b*PF4 appears to give an anomalous result, each of the others appears very similar to MIP-1α when comparing secondary structure elements. As expected the RMS deviation over the whole molecule is considerably higher, presumably as a result of the insertion between the first Cysteine pair changing the relative conformations of the N-termini and of β strands I and II. The huge RMS deviations between the dimers reflect the completely different dimer structures that were shown in Figure 4.2.

4.4 Conclusion

MIP-1α has been shown to form monomer and dimer structures which are effectively the same as those seen for the other β-Chemokines. While the close packing observed in the orthorhombic crystal form of the tetrameric mutant made it difficult to determine which pair of dimers gave rise to the tetramer, the solution of the monoclinic crystal form aided its identification.

Comparisons between the various independent MIP-1α monomers have shown that they do essentially all have the same structure. Similar comparisons with other β-Chemokines have highlighted certain regions of these molecules which may be responsible for the observed differences in aggregation potential and biochemical properties within the family.

Finally, comparisons between MIP-1α and the known α-Chemokine structures have shown that although the dimers formed by the two families are markedly different, the monomers are fairly similar.

5. IMPLICATIONS OF MIP-1α STRUCTURES

5.1 Aggregation Mechanism

As expected, the dimer identified in the MIP-1α structures corresponds very closely with those already identified for RANTES and MIP-1β. The MCP proteins MCP-1 and MCP-3 have also been shown to form dimeric structures, but these reflect the clear differences between these two branches of the β-Chemokine family which have recently become apparent. It was therefore expected that the structures of the MIP-1α and MCP-1 tetramers would be somewhat different, and might in some way account for the differences in their physiological properties. In particular, the large non-specific aggregates which MIP-1α, MIP-1β and RANTES all form under physiological conditions are not known to occur for the MCP proteins. Clearly MIP protein multimers beyond the dimer level will give snapshots of the building blocks from which these aggregates are formed.

The MCP-3 dimer is analogous to that of IL-8, with the two helices lying parallel and the three-stranded β-sheets coming together to form an extended β-structure. MCP-1 forms both this dimer (the MCP dimer) and that observed for MIP-1α, MIP-1β and RANTES (the MIP dimer) within its tetramer. The implication of the β-Chemokine structures known to date is that the dimers formed by the MCP and MIP branches are different. However, the structure of the MCP-1 tetramer shows that this distinction is by no means a definite one.

The two structures obtained from tetrameric MIP-1α mutants therefore provide an insight into the interactions which might be important in dimer formation and for the mechanism by which tetramers are then formed from dimers. The solution of both tetrameric crystal forms allowed unambiguous identification of the correct tetramer. The two independent examples of the tetramerisation interaction simplified the task of distinguishing between those interactions vital for tetramer formation and those which are merely crystal contacts.

It was apparent from the orthorhombic MIP-1α tetramer structure that the low solvent content and consequent tight packing of protein molecules within these crystals resulted in each monomer interacting with five others. While the structure of the dimer was known and could be readily identified within the crystal, the structure of the tetramer was not. Discounting the dimerisation interaction, there were then four different monomer-monomer interactions in the crystals which

could potentially be implicated in tetramer formation.

Identification of Tetramer

Tetramers within protein crystals most commonly occur in tetrahedral or pseudotetrahedral arrangements. In such an arrangement, each molecule has the same environment, and they are related by three orthogonal two-fold axes. Each of the three known Chemokine tetramers, MCP-1, PF4 and NAP-2 had been shown to have a pseudotetrahedral arrangement of monomers, and MIP-1α was expected to follow this trend.

It was surprising therefore that such an arrangement was not immediately obvious in the Orthorhombic crystal structure. After the location of the Ca^{2+} ion, it became clear that there were two dimer-dimer interactions which formed candidate tetramers, but neither of these had a tetrahedral arrangement of monomers. One potential tetramer (henceforth referred to as Tet1) consisted of two dimers lying side by side in very similar orientations and related by a rotation of around 30° (See Figure 4.9).

The second possible tetramer involved one monomer from each dimer coordinating a Calcium ion. The Calcium was octahedrally coordinated, with two adjacent equatorial coordination sites occupied by carbonyls from each monomer. The other coordination sites were filled by solvent. Again, the dimers were related by a two-fold axis, with the Calcium ion lying on this axis. In this mode of interaction (henceforth referred to as Tet2), only one monomer from each of the dimers was involved in dimer-dimer contacts - the other monomer only contacted its own dimer partner (See Figure 5.1). That particular aspect of Tet2 suggested that the dimer-dimer interaction via the Ca^{2+} ion was simply a crystal contact and was unlikely to have any physiological relevance.

However, when a near identical tetrameric orientation, likewise mediated by a Ca^{2+} ion, was also observed in the monoclinic crystal form, then the interaction had to be considered as potentially important. The RMS deviation between Cα atoms of all residues from each of these tetramers is just 1.09Å, demonstrating that the relative orientations of the two dimers in each case is almost identical. The Tet1 interaction was also observed in the monoclinic crystal form. The solvent content was higher and there were fewer intermolecular contacts overall, making it easier to distinguish between crystal contacts and those which were potentially involved in tetramerisation.

Both crystal forms therefore contained a near identical octameric assembly, one half of which was presumably the physiologically relevant tetramer, with the second half perhaps representing one of the interactions important in forming the higher aggregates.

It was assumed that the physiologically relevant tetramer would be the one which was most

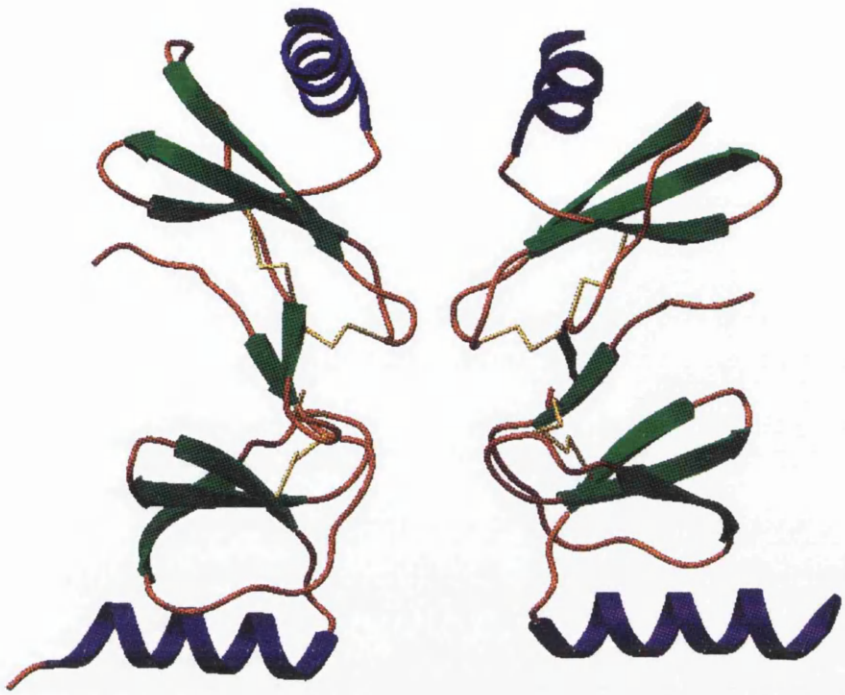


Figure 5.1: Tet2: View showing the single dimer-dimer contact.

stable. Therefore it was decided to examine factors which would stabilise the tetramer, such as interactions between specific residues and the amount of surface area buried on tetramerisation, and analyse them in order to determine which of the possible tetramers was likely to be more stable. These results are shown below in Tables 5.1 and 5.2, and their implications are discussed below.

All analysis of solvent accessible areas used the CCP4 program "SURFACE".

5.1.1 Analysis of Surface Accessibility Data

From these results the most obvious conclusion is that there is a vast difference in the amounts of surface area buried by the two tetramerisation modes. Whereas Tet1 buries around 1800Å^2 on each dimer, Tet2 only buries around 500Å^2. It is also interesting to note that on dimerisation around 1400Å^2 is buried on each monomer, and that this figure could potentially be higher if the missing N-terminal residues were also taken into account. In comparison to the total accessible

Crystal Form	Subunits	Solvent Accessible Area
$C222_1$	Monomer A	4581.4Å^2
$C222_1$	Monomer B	4725.3Å^2
$C222_1$	Monomer C	4568.9Å^2
$C222_1$	Monomer D	4710.2Å^2
$C222_1$	Dimer AB	7947.1Å^2
$C222_1$	Dimer CD	7944.6Å^2
$C222_1$	Tetramer (Tet1)	13927.0Å^2
$C222_1$	Tetramer (Tet2)	15400.5Å^2
$P2_1$	Monomer A	4756.6Å^2
$P2_1$	Monomer B	4871.2Å^2
$P2_1$	Monomer C	4855.2Å^2
$P2_1$	Monomer D	4338.5Å^2
$P2_1$	Dimer AB	8160.1Å^2
$P2_1$	Dimer CD	7827.5Å^2
$P2_1$	Tetramer (Tet1)	14285.1Å^2
$P2_1$	Tetramer (Tet2)	15498.2Å^2

Table 5.1: Solvent-Accessible Areas for Tetrameric Subunits of MIP-1α

Crystal Form	Relevant Area	Buried Surface Area
$C222_1$	Area buried on dimerisation of AB	1347.0Å^2
$C222_1$	Area buried on dimerisation of CD	1347.1Å^2
$C222_1$	Area buried on tetramerisation (Tet1)	1964.7Å^2
$C222_1$	Area buried on tetramerisation (Tet2)	581.2Å^2
$P2_1$	Area buried on dimerisation of AB	1451.7Å^2
$P2_1$	Area buried on dimerisation of CD	1382.2Å^2
$P2_1$	Area buried on tetramerisation (Tet1)	1702.5Å^2
$P2_1$	Area buried on tetramerisation (Tet2)	489.4Å^2

Table 5.2: Surface Areas buried by potential MIP-1α multimerisation interactions

surface area of each of the monomers, this figure appears high. It represents burial of just less than one third of the total surface area of each monomer upon dimerisation. This seems a fairly high proportion of the molecule to be buried by a single interaction, and implies that the two dimerisation surfaces are complementary and that the dimer thus produced is likely to be stable.

By contrast, Tet2 buries just 5% of the accessible surface area of each dimer, which is too low a proportion for this interaction to have any physiological significance. Tet1 buries around 15% of the available surface area of each dimer, and while this figure is higher than for Tet2 it does not represent an interaction as strong or as specific as the dimerisation of MIP-1α. The biochemical evidence for an equilibrium between monomer, dimer, tetramer and perhaps even other species is consistent with such a tetramer[84, 94], as is the crystallographic evidence presented in Section 2.6.

5.1.2 Comparison with other Chemokines

In comparison to the other Chemokines for which tetramer structures are known, the MIP-1α interaction appears much looser. An analysis similar to that above for MIP-1α was performed for the Chemokines MCP-1 and *h*PF4 and the results are shown below in Figures 5.3 and 5.4. Clearly the dimerisation interactions in each case are comparable, burying around 15% of the total accessible surface area of each monomer. However, both MCP-1 and *h*PF4 bury over 20% of the accessible surface of each dimer upon tetramer formation. This is higher than the figure for MIP-1α, and suggests this tetramerisation product is more stable, and that the tetramer will be a more significant aggregation state for these molecules.

The aggregation states of *h*PF4 have been studied in considerably more detail than most of the other Chemokines. It has been shown that the monomeric and tetrameric states of PF4 are more stable than the dimeric state, although an equilibrium between all three states exists under physiological conditions[143, 144]. In contrast to this, the structures of all β-Chemokines currently known demonstrate a highly stable dimeric state, although a disputed monomeric structure of MCP-3 has also been reported[141]. It would therefore appear that the association mode used by the β-Chemokines MIP-1α, MIP-1β and RANTES gives a particularly stable dimer, whereas the α-Chemokine dimer is somewhat less stable in some cases but can form a particularly stable tetrameric state. Both the stability, and the kinetics and thermodynamics of formation of the β-Chemokine tetramers would have to be carefully assessed before any conclusions could be drawn from their apparent lower stability relative to the α-Chemokines.

While analysis of buried surface areas provides much useful information regarding different association states, it should also be remembered that this method takes no account of specific inter-

subunit interactions such as salt-bridges or metal binding sites, which could have very significant stabilising effects. Examination of potential interactions of this type in the MIP-1α structures provides further evidence on the relative importance of the various potential aggregation states, and is discussed below.

Chemokine	Subunit	Accessible area
*h*PF4	Monomer A	4906.9Å^2
*h*PF4	Monomer B	5018.5Å^2
*h*PF4	Monomer C	4993.0Å^2
*h*PF4	Monomer D	4832.6Å^2
*h*PF4	Dimer AB	8342.4Å^2
*h*PF4	Dimer CD	8119.2Å^2
*h*PF4	Tetramer ABCD	12886.5Å^2
MCP-1	Monomer A	5415.0Å^2
MCP-1	Monomer B	5428.1Å^2
MCP-1	Monomer C	5415.0Å^2
MCP-1	Monomer D	5428.1Å^2
MCP-1	Dimer AB	9534.3Å^2
MCP-1	Dimer CD	9534.3Å^2
MCP-1	Tetramer ABCD	15191.5Å^2

Table 5.3: Solvent-Accessible Areas for subunits of other Chemokines

Chemokine	Relevant Area	Buried Surface Area
*h*PF4	Area buried on dimerisation of AB	1583.0Å^2
*h*PF4	Area buried on dimerisation of CD	1706.4Å^2
*h*PF4	Area buried on tetramerisation of ABCD	3575.1Å^2
MCP-1	Area buried on dimerisation of AB	1308.8Å^2
MCP-1	Area buried on dimerisation of CD	1308.8Å^2
MCP-1	Area buried on tetramerisation of ABCD	3877.1Å^2

Table 5.4: Areas buried on formation of multimers for other Chemokines

5.1.3 Shapes of Potential Tetramers

Considering just the relative shapes of the potential tetramers, Tet1 appears to be much more compact than Tet2, and is therefore more likely to be the physiological tetramer. However, the Tet1 tetramer is quite square and relatively flat, and does not appear at first glance to be a particularly stable arrangement. Unlike the α-Chemokine tetramers and even the MCP-1 tetramer, it is far from globular in shape. While the β-Chemokine dimers are also somewhat elongated, these have been shown to be stabilised by hydrophobic interactions, the burial of large amounts of surface area and

the presence of certain specific stabilising interactions, such as the formation of an antiparallel β-sheet between the N-terminal portions of the two monomers. As explained above, there is a fair degree of contact between the two dimers in the Tet1 tetramer, which would enhance stability. An examination of the importance of certain specific inter-dimer interactions and their implications for the stability of the two tetramerisation modes is given in Section 5.1.4.

The Tet2 tetramer has a much smaller area of contact between the two dimers and consequently appears more fragile (See Figure 5.1). Clearly this interaction depends critically on the Ca^{2+} ion, and its removal would certainly result in abrogation of this particular tetramer. If tetramerisation does indeed depend so critically on the presence of a particular ion, it would be expected that there be some biochemical evidence for its importance. The current literature does not contain any evidence to support a functional or structural role for Ca^{2+} ions with respect to MIP-1α or any other Chemokine.

5.1.4 Specific Interactions within Potential Tetramers

In both potential tetramerisation modes identified, there are surprisingly few specific interactions between dimers.

5.1.4.1 Tetramerisation Model : Tet1

The major specific interaction between the dimers in this tetramer model is a salt bridge between Asp 26 on one dimer and Arg 45 (and perhaps Arg 47) on another. The significant degree of buried surface area hints at there being many hydrophobic interactions as well. This is in agreement with a spectroscopic study of aggregation in the human form of MIP-1α[84], in which it is stated that the major stabilising force in tetramer formation is hydrophobic. It is almost certain that the human and murine forms of the protein will have identical properties with respect to mode and specificity of aggregation interactions, despite some charged residues not being conserved between them. Therefore, applying the spectroscopic evidence to the observed MIP-1α structure, the implication is that the single salt bridge observed makes little stabilising contribution in comparison with the many hydrophobic interactions. There are several other electrostatic interactions, but these all involve hydrogen-bonded interactions between main-chain and side-chain atoms, and are likely to be less significant than the Asp 26 - Arg 45 interaction.

5.1.4.2 *Tetramerisation Model : Tet2*

In this tetramer model there is only one direct interaction between dimers, a hydrogen bond between Ser 35 O on one and OG of Ser 32 on the other. The major interaction is the coordination of the Ca^{2+} ion by Ser 32 O and Cys 34 O on each of the dimers. There are also several electrostatic dimer-dimer interactions mediated by water molecules surrounding the Ca^{2+} ion. However, there are no hydrophobic interactions whatsoever. This fact alone effectively discounts Mode 2 from representing the physiologically relevant MIP-1α tetramer. Assuming again that both human and murine forms of MIP-1α will have identical aggregation properties, and applying the results of spectroscopic studies on human MIP-1α to the murine form, the major stabilising force of the tetramer must be hydrophobic. Human MIP-1α exists as a tetramer in conditions which disrupt ionic interactions, namely 500mM NaCl. A tetramer which depended on the presence of a calcium ion for its integrity would be disrupted under these conditions.

5.1.5 *Do the Mutations make sense?*

"No" would have to be the honest answer to that question. All the evidence points to the interactions giving rise to both dimer and tetramer being predominantly hydrophobic. This of course is at odds with the original assumption that these interactions were electrostatic, and implies that the rationale behind the production of the three MIP-1α mutants was flawed. Nevertheless, interactions between tetramers giving rise to larger aggregates do primarily involve electrostatic interactions, and hence the residue mutated in order to produce the tetrameric mutant is likely to be highly significant. However, the other mutations which give rise to the monomeric and dimeric mutants do not correspond with the spectroscopic data or with the observed structure, and it must be questioned what effects they really have on aggregation. The mutation E66Q which gives rise to the tetrameric mutant is discussed in more detail below, but the effects of the mutations D64N and E60Q will now be considered.

5.1.5.1 *Mutation D64N*

As can be seen in Figure 5.2, residue Asp 64 sits on the external face of the α-helix in both monoclinic and orthorhombic forms of the tetramer. Therefore it is some 10Å from the closest residue on the other dimer, and could not possibly have any direct effect on tetramerisation.

5.1.5.2 Mutation E60Q

As shown in Figure 5.2, in each of the four monomers composing the tetramer, residue Gln 60 sits at one end and points directly away from the rest of the tetramer. In the second potential tetramerisation mode, the situation is very similar, with each Gln 60 residue lying at one end of the tetramer and pointing away from the centre. Therefore it is clear that in no way could this residue have any influence on the monomer-monomer interactions giving rise to the dimer, or even on the formation of the tetramer. This of course is in complete contradiction to the gel filtration studies on the three mutants[86]. It is clear from both the dimer and tetramer structures that the monomer-monomer interactions involved in dimer formation are predominantly hydrophobic. The most significant electrostatic contribution appears to be the short antiparallel β-sheet formed by the N-terminal regions of the two monomers.

Each of the crystal structures therefore reinforces the biochemical and spectroscopic evidence suggesting predominantly hydrophobic interactions in dimerisation and tetramerisation. The most likely explanation for the molecular weights obtained for each of the mutants by gel filtration is that there is a dynamic equilibrium between monomer, dimer and tetramer. In such a case, the aggregation state of the molecules will depend critically on various environmental factors, such as ionic strength and pH. The mutation of one or two of the charged side-chains would modify the overall charge distribution of the monomer, which in turn could affect the equilibrium.

Therefore the tetramer model is consistent with the available biochemical and spectroscopic evidence[84, 94]. It is also consistent with the observation that the dimeric mutant was able to form tetramers or crystallise with the same morphology as the tetramer, under certain conditions. However, while this model appears to satisfy all the evidence with respect to the monomer, dimer and tetramer, it does not account for the formation of aggregates via electrostatic interactions.

Figure 5.2: Influence of mutated residues on aggregation

The above diagram shows the three residues which were mutated in the MIP-1α mutants. As this is the tetramer, only Glu 66 has been mutated. Clearly Asp 64 and Glu 60 have no direct influence on tetramer formation.

5.1.6 *Higher Aggregation States*

A tetramer has been identified in both PM1 crystal forms as being the physiologically relevant one. However, as explained in Section 5.1, the same Calcium-mediated dimer-dimer interaction was observed in both crystal forms. There is therefore a nearly identical octameric assembly within each of them. As MIP-1α is known to aggregate under a variety of conditions, this octamer must be considered as a possible example of an interaction giving rise to a higher aggregate. Before any possible physiological relevance of this octamer is assessed, the biochemical evidence for the mechanism of MIP-1α aggregation will be considered.

5.1.6.1 *Biochemical Evidence for Aggregation*

Spectroscopic studies[84] have suggested a potential molecular weight for the human MIP-1α aggregate of between 100 and 250*k*D. Importantly, the solubility of these MIP-1α aggregates is suggestive of some form of order being present in the self-association process. In other words, the aggregates are composed of discrete units, possibly tetramers, and probably do not involve non-specific interactions. The mass of the aggregate has also been assessed by various groups, and has been estimated at 650*k*D[85], 100-150*k*D[86] and 100*k*D[18] by different methods.

It has been suggested[86] that the aggregate may represent a discrete dodecamer, which would correspond to a molecular weight of around 95*k*D. However, the estimates from spectroscopic data and from gel filtration analysis are consistent with a population of aggregate molecules composed of at least twelve monomer units, and possibly many more. The highest molecular weight estimate[85], of 650*k*D, would correspond to an enormous aggregate of over 80 monomer units.

Biochemical evidence suggests that there is an equilibrium between various multimerisation states. In contrast to the most closely related β-Chemokines, MIP-1β and RANTES, it is proposed that the monomeric and tetrameric states predominate in MIP-1α. The dimeric state, which appears to be one of the most physiologically relevant to the other two molecules, was not observed in spectroscopic studies on the human form of MIP-1α[84]. However, different dimeric human MIP-1α molecules have been produced by mutagenesis[145]. Given the great similarity between MIP-1α, MIP-1β and RANTES, both in sequence and in monomer structure, it is likely that the aggregation mechanism of each will be the same.

It is therefore likely that, under any conditions, MIP-1α exists as an equilibrium mixture of multimers, from monomers up to aggregates. The multimerisation state of MIP-1α is then dependent only on how this equilibrium is affected by the local chemical conditions. Biochemical

studies on the MIP-1α mutant BB-10010[145], in which residue Asp 26 was changed to Ala, are consistent with this hypothesis, and suggest that even in this fairly homogeneous mutant, dimer, trimer and tetramer states are likely to be present.

5.1.6.2 Aggregation Models

As the packing in the orthorhombic crystal form is fairly tight, each monomer interacts with several others. It is quite possible that some of these interactions correspond to interactions relevant to the aggregation process. Certain assumptions can be made regarding the internal structure of the aggregate. Firstly, it is assumed that the aggregate will consist of discrete units, probably tetramers, arranged in a fairly ordered fashion. It can also be assumed that interactions between these building blocks will be similar, and therefore corresponding portions of each of these components will have the same environment. Aggregate models can therefore be constructed by considering tetramer-tetramer interactions observed in the orthorhombic crystal form, and using similar interactions to propagate successively larger models.

Naturally, the most interesting candidate models will be the ones involving electrostatic interactions of residue Glu 66. However, several other tetramer models will be possible, and all should be considered for the sake of completeness. It should also be remembered that as all interactions represent crystal contacts, none of them necessarily represents an interaction that is actually involved in forming the physiological aggregate.

5.1.6.3 Aggregation Model 1 (Agg1)

The most obvious way to generate a multimer from the MIP-1α tetramer is to use the relationship between two dimers to propagate an infinite chain of identically interacting MIP-1α dimers. As the two dimers within the tetramer differ in orientation by only around 30°, the multimer thus generated is helical, as shown in Figure 5.3. While such a model is very aesthetically pleasing, it seems inconsistent with the biochemical data on interactions within MIP-1α. The dimer-dimer interactions which form the tetramer are predominantly hydrophobic, and are not disrupted by the presence of 0.5M[84] or 1.0M[85] NaCl. Therefore, in Agg1, the aggregate is also stabilised by predominantly hydrophobic interactions. However, the interaction between Glu 66 and Arg45/Arg 47 is electrostatic, and although it does not appear to be a particularly strong interaction in the tetramer crystal structures, it seems feasible that a slight translation of one dimer relative to the other would put these residues in position to form a strong and more stable salt-bridge.

While various pieces of biochemical evidence point towards an aggregate that is formed by electrostatic interactions, different to the hydrophobic ones which are involved in tetramer formation, it is conceivable that the two interactions are in fact the same. The dimer-dimer interaction, while predominantly hydrophobic, does have a potential strong electrostatic contact in the Glu 66-Arg 45-Arg 47 interaction. While the hydrophobic interaction might be sufficient to hold together the tetramer, addition of further dimers or tetramers would produce a much more elongated arrangement. While the hydrophobic interactions alone might not be enough to stabilise this structure, the additional presence of two strong salt-bridge interactions could be enough to stabilise slightly larger aggregates.

This model would probably result in a population of aggregates of various molecular weights. It is quite possible that the aggregate mixture would represent an equilibrium centred around 100*k*D, and would therefore be fairly consistent with the various estimates of the mass of the aggregate. The average mass of such a distribution would be dependent, to some extent, on chemical environment. The considerable differences between the various estimates of aggregate mass could therefore be accounted for by differences in the environment within which the estimate was made.

Although the predominant aggregated species has been proposed to be a dodecamer, there is little evidence to indicate whether it is composed of tetramers or of other species. While the presence of a tetrameric state has been demonstrated for both human and murine forms of MIP-1α, there is no other evidence for the suggestion that the proposed dodecamer consist of a trimer of tetramers. As native MIP-1α consists of a range of species from monomer up to aggregate, there does not appear to be any reason why the aggregate could not be formed by interaction between tetramer and dimer species.

The major drawback of this model is the observed disaggregation to tetramers in 0.5M NaCl. While this has been interpreted as demonstrating that entirely different types of interaction are responsible for dimer-dimer and tetramer-tetramer interactions, this result can be viewed in another way. It may simply reflect the inability of the hydrophobic dimer-dimer interactions to stabilise any aggregate that becomes significantly elongated - in other words any aggregate that is larger than a tetramer. Naturally, this model would be more consistent with disaggregation to an equilibrium state where, although the predominant molecular weight corresponded to a tetramer, some smaller and larger species were also present. However, analytical ultracentrifuge and sedimentation equilibrium measurements[84] seem to indicate the formation of an almost completely monodisperse human MIP-1α tetramer in 0.5M NaCl, and are therefore inconsistent with this particular model.

Given the propensity for interaction via Ca^{2+} ions that both tetrameric mutant crystal structures have demonstrated, it is tempting to consider whether such interactions could have any bearing on aggregation.

The Ca^{2+} ion which is observed in both tetramer crystal forms would lie on the outer surface of the proposed helix, and would therefore be well-placed to interact with other molecules or even other helices. While it has previously been shown that such a tetramer-tetramer interaction would not confer very much stability, helices of moderate length could interact by this mechanism to produce aggregates of very high molecular weight, consistent with some observations[85]. Such aggregates would be considerably more globular, and hence presumably more stable, than single helical aggregates of similar molecular weight. It would therefore be extremely interesting to test the effects of Ca^{2+}, and of similar ions, on the aggregation of MIP-1α and related β-Chemokines.

There does not appear to be any steric or structural impediment to the addition of further dimers to the tetramer which has been identified within the two crystal forms of the tetrameric mutant. If, as seems likely, this tetramer is physiologically relevant, then interactions of these tetramers with further dimer or tetramer molecules by this mechanism would seem inevitable. Therefore, even if the proposed dodecameric aggregate is formed by another means, interactions of this type are still likely to occur to some extent.

5.1.6.4 Aggregation Model 2 (Agg2)

As mutant data suggested that residue Glu 66 is probably the most important in terms of aggregation interactions, a tetramer-tetramer interaction involving it was sought. Such an interaction was discovered, in which one monomer from the first tetramer was sitting in the central cavity of the second tetramer, and making interactions on either side (as seen in Figure 5.4). There are two specific electrostatic interactions of note; firstly the interaction between Glu 60 on the first tetramer and Arg 45/Arg 47 on the second, and secondly an interaction between Asp 26 and Glu 66 on the first tetramer and Lys 44, Arg 45 and Arg 47 on the second. In this second interaction, Asp 26 seems ideally placed to interact with either of the three basic side-chains, whereas Glu 66 is more remote but may interact with Arg 45 on the opposite tetramer. In addition there are several other hydrogen-bonded interactions, mainly between side-chain atoms on one tetramer and main-chain atoms on the other. It is also clear that there are significant hydrophobic interactions involved as well, specifically an aromatic stacking interaction involving Phe 23 on the first tetramer and Phe 28 on the second. Since this interaction is present within the crystal structure, there will of course be an infinite chain of tetramers related by this interaction within the crystal.

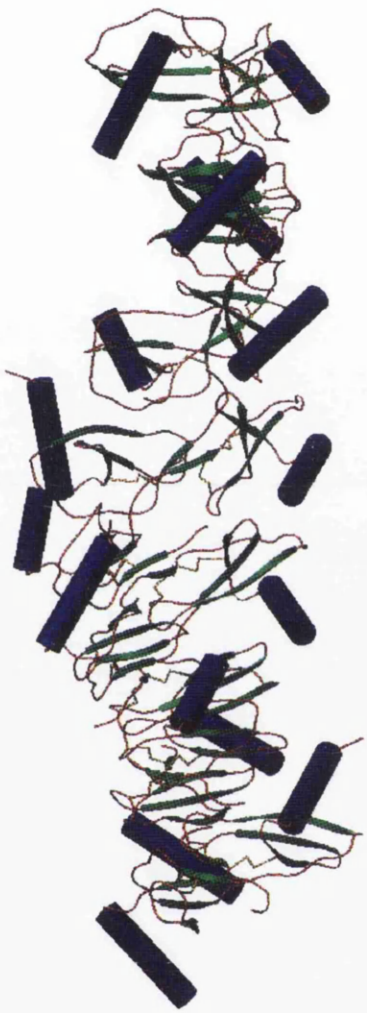


Figure 5.3: Structure of Aggregation Model Agg1

Therefore it is most unlikely that the interaction seen within the crystals is the same as is physiologically relevant for aggregation. However, it is possible that the interaction observed in the crystals is similar to the one of interest, and the interaction within the crystals represents an imperfect model of that interaction. It is clear that the fit between the two tetramers is not that good, and could probably be improved markedly by a minor motion of one tetramer relative to the other. Perhaps what is seen in the crystal is therefore a distorted representation of the true aggregation interaction. In particular it is clear that Glu 66 is not ideally placed to interact with Arg 45, but a minor motion of one tetramer would enable a stronger interaction between them.

It is possible to generate a tetramer of tetramers from this particular interaction, and the resultant "aggregate" looks reasonably compact and potentially stable (as seen in Figure 5.5). The molecular weight of this model, at 125*k*D, is consistent with most of the biochemical data, although it appears that it would not be possible to form a stable dodecamer using this interaction. Like Agg1, it is also consistent with mutagenesis evidence, with residues Asp 26, Arg 45 and Glu 66 all involved in the interaction.

5.1.7 Conclusion

The MIP-1α structures provide snapshots of two of the basic units used to build up the large aggregates observed for the proteins MIP-1α, MIP-1β and RANTES. They show that each of these proteins forms a dimer of identical topology, but also demonstrate that there are small but potentially significant differences between certain regions of these molecules. The tetrameric MIP-1α structures also show two potential tetramerisation modes. One of these buries a large amount of surface area, and thus is the one that is likely to be more significant to the aggregation process. The second of these is mediated by Ca^{2+} ions, although there is no evidence to suggest a role for Calcium in control of β-Chemokine aggregation. It is unlikely that both tetramerisation modes observed within the MIP-1α crystals are physiologically significant. It seems almost certain that tetramerisation mode 1 is in fact the correct tetramer.

In addition, some of the other crystal contacts observed within the tetramer crystals bury significant portions of surface area, and may represent interactions relevant to the formation of aggregates. Two tetramer models have been proposed, although others are of course possible. Both models are consistent with all the available mutagenesis evidence, but neither can completely account for all the biochemical data on MIP-1α aggregation.

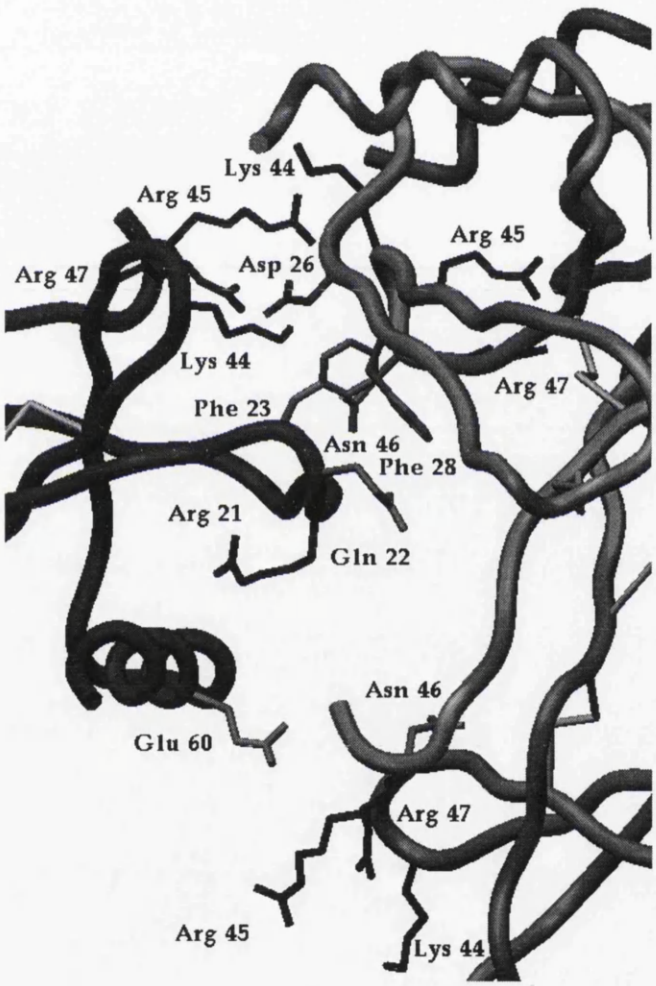


Figure 5.4: Interactions involved in formation of Aggregation Model Agg2

Figure 5.5: Structure of Aggregation Model Agg2

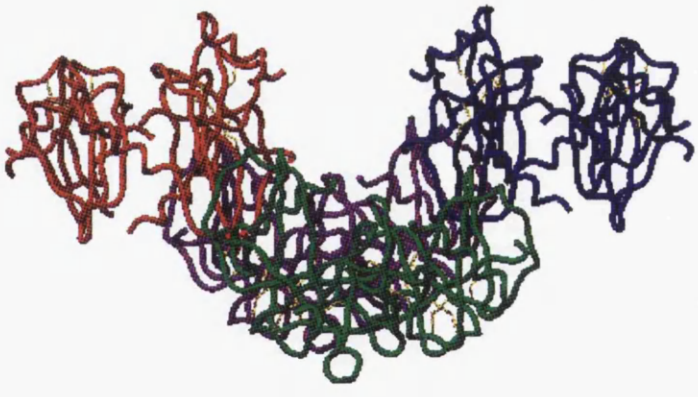


Figure 5.6: Structure of Aggregation Model Agg2

Figures 5.5 and 5.6 show two different views of Agg2

5.2 *The Potential Role of Ca^{2+} Ions.*

It is interesting that the two different tetrameric crystal forms were obtained from near identical crystallisation conditions. A similar phenomenon has been observed for other Chemokines[109]. Orthorhombic crystals were grown using a well solution containing 0.2M $CaCl_2$, 0.1M Tris pH 7.4 and 8% PEG 6000. This solution was diluted 1:1 with protein to give the drop solution. The monoclinic crystals were grown using a well solution containing 0.2M $CaCl_2$, 0.1M Tris pH 7.4, 8% PEG 6000 and 10% glycerol.

Glycerol had been added to this particular tray to reduce the rate of equilibration in an attempt to slow down crystal growth. It is possible that the glycerol had some additional precipitant effect, but the monoclinic crystals did take significantly longer to appear than the orthorhombic ones, suggesting that the glycerol may have had the desired effect.

The only other parameter that varied between the two crystallisation conditions was the protein concentration. The monoclinic crystals were grown using a protein solution of 1.4 mg/ml, whereas the optimal orthorhombic crystals came from a protein solution of 5.4 mg/ml. However, some of the wells containing monoclinic crystals also contained what appeared to be crystals of orthorhombic habit. While these crystals have not been diffraction tested, it seems likely that both crystal forms can be grown from the same initial crystallisation mixture.

Since the orthorhombic crystals appear within 2 days of being set up, and the monoclinic crystals take several weeks to appear, it is possible that the orthorhombic crystal form represents the "kinetic" product of the crystallisation mixture, whereas the monoclinic crystals represent the "thermodynamic" product. Nuclei that will generate orthorhombic crystals will be rapidly formed, and if the protein concentration is sufficiently high then these crystals will grow in several days. If the protein concentration is lower, these transient aggregates may not be able to add enough protein molecules to grow to a stable size, and will disperse. The presence of glycerol, slowing down the diffusion of protein molecules within the crystallisation drop, would also help to prevent crystals growing from such a mixture. As the mixture gets closer to equilibrium, monoclinic crystal nuclei could form, perhaps being stabilised by the extra cross-linking via the second Ca^{2+} ion. Possibly it is only the extra stability this interaction provides that allows the initial aggregate to survive long enough to become a viable crystal.

5.2.1 Are Ca^{2+} Ions involved in Aggregation?

It is possible that Ca^{2+} or similar ions are involved in some way in the aggregation phenomenon. The available evidence points to electrostatic interactions being predominantly responsible for the aggregation interaction, which would not discount the possibility of the involvement of ions of some kind.

There are several ways in which to test for presence of Ca^{2+} ions. Firstly, although most biochemical studies to date have used physiological conditions, and therefore have some salt present, it should be possible to perform similar experiments on samples from which all salt had been removed. A reagent such as EDTA could be used initially, and if there was some observed effect then reagents specific for particular ions could then be used in order to identify which were important. This would give some indication of the effects of ionic strength or the presence of particular ions on aggregation. Secondly, there are various methods which could be used in order to determine the presence of Ca^{2+} ions in the aggregate. For example, the aggregated MIP-1α molecules could be separated from the smaller multimers also present in solution by gel-filtration chromatography[85]. They could then be washed with non-Ca^{2+} containing buffer and the resulting solution analysed by Electrospray Mass Spectrometry to specifically identify any Calcium present.

The observed tetramer-tetramer interaction mediated by Ca^{2+} ions may be physiologically relevant. The fact that it occurs almost identically in both crystal forms suggests that it may be more than just a crystal contact. Indeed, the RMS deviation between the two potential Ca^{2+}-mediated tetramers, one from the orthorhombic structure and one from the monoclinic structure, was just 1.086Å for all C-α atoms (1.467Å for all atoms). While the analysis of buried surface area had shown that this particular "tetramer" is most unlikely to be physiologically relevant, it seems equally unlikely that the same crystal contact would appear in two entirely different crystal forms with an RMS deviation of around 1.0Å.

5.2.2 Are Ca^{2+} Ions necessary for Crystallisation?

Ca^{2+} ions seem to be necessary for crystallisation of the tetramer, but not for the monomer or the dimer. However, many of the originally identified conditions for those molecules did contain Ca^{2+}, as did those eventually identified as the optimal ones.

Following identification of the Ca^{2+} sites in the structures of the tetrameric MIP-1α mutants, crystals of the orthorhombic type were soaked in a solution containing $Lu_2(SO_4)_3$. It was expected that Lu^{3+} ions would enter the crystals and would be able to specifically replace some of

the Ca^{2+} ions. Lu^{3+} was chosen as it has an ionic radius of 0.98Å, compared to 1.00Å for an octahedrally coordinated Ca^{2+} ion[146]. The replacement of Ca^{2+} by Lu^{3+} has been described for Thermolysin[147], although the ionic radii quoted in this case were estimated differently[100].

Data collection parameters for soaked crystals are shown in Table 3.3. The Lu^{3+} ion was subsequently located using the program SHELX[120]. These results indicated a single-site derivative, of quality comparable to that of the Uranyl derivative obtained for the monomeric mutant. However, when refinement of the Lu^{3+} phasing was attempted[121], it was impossible to refine either the position or the temperature factor of the heavy atom, and the phasing did not seem to be self-consistent. The Lu^{3+} position obtained from SHELX was compared with the orthorhombic structure of the tetrameric mutant, and did not correspond with the expected Ca^{2+} position, even when taking into account the alternative origin in spacegroup $C222_1$. The observed position was in fact over 6Å from the closest protein atom, and could not therefore represent a bound derivative. The observed "signal" in the Patterson maps and in SHELX was probably be due to non-isomorphism or to lack of data completeness at low resolution.

5.2.3 Are Ca^{2+} Ions involved in MIP-1α Interactions with Receptor?

If it could be demonstrated that Ca^{2+} ions have any physiological effect on MIP-1α, particularly in receptor-binding, there would be scope for the investigation of the Ca^{2+} binding sites by mutagenesis. To date there is only one report of a mutant in this region of a β-Chemokine[148]. The insertion of a proline in the corresponding loop of MCP-1 caused an almost complete loss of activity. This mutation presumably had some effect on the conformation of the N-terminal peptide, and has been proposed to be the reason for the loss of activity.

However, it is interesting to consider the possibility of constructing similar mutants for MIP-1α, in order to see whether they could bind Ca^{2+}. Another possibility would be minor modification of the residues surrounding Calcium-binding sites in both crystal forms, in order to accommodate ions of other types.

5.3 Areas of the Structures involved in Receptor-Binding

5.3.1 Chemokine Receptors

Over the last few years many new Chemokine receptors have been identified, with the family now consisting of four distinct groups - Specific, Shared, Viral and Promiscuous. Distinction can also be made between α-Chemokine receptors and β-Chemokine receptors, although Promiscuous

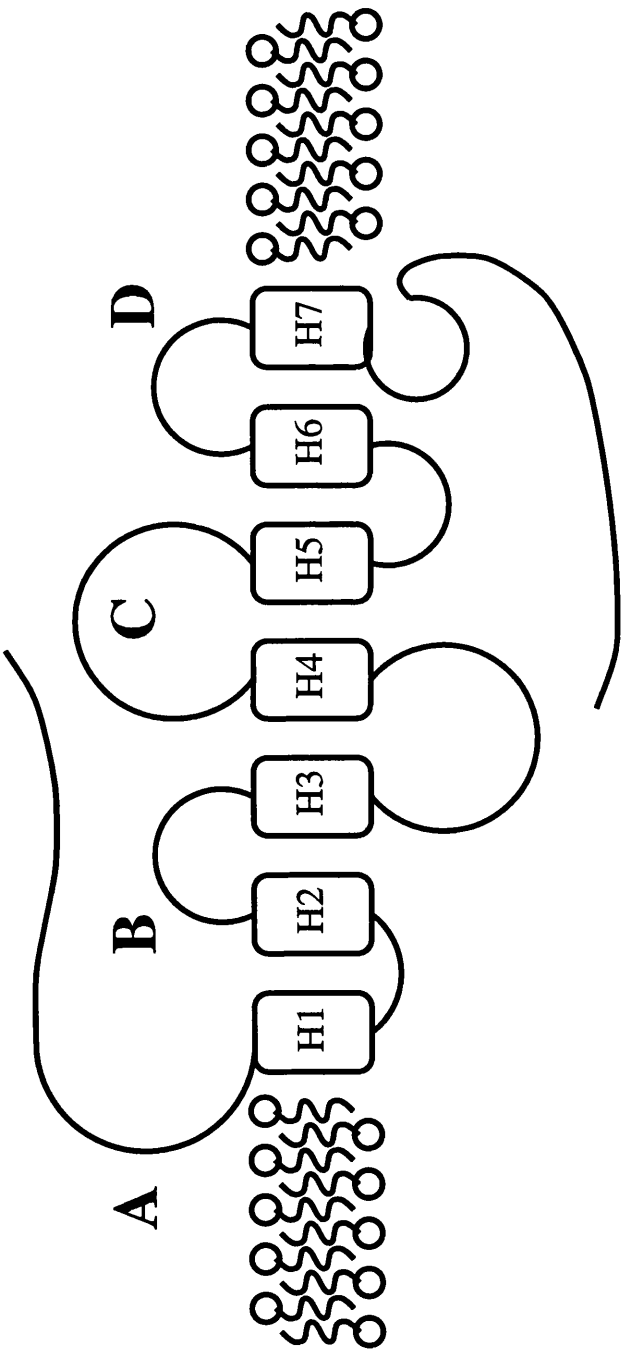


Figure 5.7: Schematic diagram of CC receptor

receptors by definition will bind proteins from either subfamily.

From a rather confusing beginning, the nomenclature for this receptor family is now fairly simple and very well established. The β-Chemokine receptors are numbered, with the prefix "CCR", and receptors CCR1 through to CCR8 are currently known. In addition, receptor CCR2 exists in two distinct forms, CCR2a and CCR2b. The α-Chemokine receptors follow a similar system, with the prefix "CXCR". α-Chemokine receptors CXCR1 to CXCR4 are known to date. The only known promiscuous receptor is the Duffy Blood Group Antigen (DBGA) which was first recognised as a receptor used by malaria parasites. The binding profile of each of the known human Chemokine receptors is shown in Table 1.8.

Each of these receptors is a seven-transmembrane helical receptor, and mediates its Chemotactic effects via interaction with G proteins. A schematic diagram of a Chemokine receptor is shown in Figure 5.7. There is a considerable degree of sequence identity between each of these receptors. As would be expected the putative transmembrane segments have the highest degree of identity, and the overall sequence identities range from 30% to around 65%.

5.3.2 β-Chemokine Receptors

As mentioned in Chapter 1, the receptor-binding profiles of many of the β-Chemokines are complex, and MIP-1α is no exception. Human MIP-1α is known to bind to the human forms of CCR1, CCR4 and CCR5[4]. The murine form has been shown to bind to the murine versions of CCR1, CCR4 and CCR5, in addition to a recently-identified receptor, D6[149].

Despite mutagenesis studies involving several β-Chemokines, including MIP-1α, little is known about which specific regions of these molecules are determinants of receptor-binding. It is quite feasible, and even likely, that different regions of these molecules are involved in binding to different receptors. The N-terminal region has been identified as important to the functionality of RANTES[150] and MCP-1[148]. However, in these studies residues close to the N-terminus were mutated and then the ability of the mutants to stimulate monocyte chemotaxis and intracellular calcium mobilisation was examined. This implied investigation of the ability of the mutants to interact with the receptors CCR1 (in the case of the RANTES mutants) and CCR2a and CCR2b (in the case of the MCP-1 mutants).

Similar mutagenesis studies on another region of both human and murine MIP-1α gave surprisingly different results. One study[94] showed that in murine MIP-1α, residues on the loop between β-strands II and III were essential for glycosaminoglycan (GAG)-binding, and were involved in binding to murine CCR1. The other study showed that mutation of a single residue

on the corresponding loop in human MIP-1α prevented glycosaminoglycan-binding, but had no effect on CCR1-binding.

Therefore it is likely that the N-terminal regions of the β-Chemokines are critical in determining whether they can bind to CCR1, CCR2a and CCR2b, although other parts of the molecules are also likely to be important. However, the other β-Chemokine receptors, most notably CCR5, may well interact with entirely different portions of their ligands.

It is puzzling that such similar mutations in human and murine forms of MIP-1α gave such different results. There are several possible explanations which, while not entirely compelling, may explain these observations.

- Human and murine forms of MIP-1α are quite different to each other when compared with human and murine forms of other β-Chemokines. There are also several charged residues that are unconserved between the two.
- Their receptor-binding properties are known to be different.
- In the human mutant, only Arg 45 was changed. In the murine mutant, both this residue and Lys 44 were altered.
- In the human mutant, Arg 45 was mutated to Ala. In the murine mutant, Lys 44 was mutated to Asn and Arg 45 to Ser - hence there are possible steric differences between the mutation products.

Two additional mutants of MIP-1α - R18A and R47A also abolished GAG-binding. They are dealt with in greater detail in Section 5.4.3.

In order to develop some kind of understanding of the interactions between the β-Chemokines and their receptors, the following Section will collect the available biochemical data for each receptor and attempt to discover any underlying connection between it and the known β-Chemokine structures.

5.3.3 CCR5

The β-Chemokine receptor CCR5 has been shown to bind the Chemokines MIP-1α, MIP-1β and RANTES. Many of the other β-Chemokines, most notably all of the Monocyte Chemoattractant Proteins, as well as all of the α-Chemokines, have been shown not to bind to or induce a signal from CCR5.

While it is clear from the analysis of β-Chemokine sequence similarities in Table 1.2 that these three proteins have quite a high degree of sequence identity, there are several other proteins, most notably HCC-1, DC-CK1 and MPIF-1, which are also very similar. Interestingly, HCC-1 and DC-CK1 also seem to be much more similar to MIP-1α than to RANTES, but there is currently no evidence to suggest that these proteins also bind to CCR5.

This makes it somewhat difficult to define the requirements for CCR5 binding. It would be expected that the high degree of sequence identity would result in three proteins with highly similar tertiary structures. It is likely though that HCC-1 and DC-CK1 have essentially the same tertiary structure as MIP-1α, MIP-1β and RANTES, but lack some specific residues or motifs that are absolutely required for CCR5 binding. While the structures of MCP-1 and MCP-3 have shown that the tertiary structure of the Monocyte Chemoattractant Proteins is very similar to that of the CCR5 ligands, these proteins are likewise unable to bind to CCR5. Again it is likely that the lack of an essential residue or structural motif prevents binding. The availability of the MCP-1 and MCP-3 structures, in addition to those of MIP-1α, MIP-1β and RANTES, allows investigation and potential identification of such residues.

5.3.3.1 Ligand Evidence

CCR5 binds MIP-1α, MIP-1β and RANTES, but none of the MCP proteins. Sequence identities in its ligands range from 45-68%. It is fairly simple to define a set of residues which are conserved in the CCR5 ligands but not in the MCP proteins. This set corresponds almost exactly with the residues previously identified as vital in MIP-1α aggregation (See Table 5.5). This may imply that one or more of the interactions made during the aggregation process mimics the interactions within multimers. Alternatively, Glu 29, which has been shown not to have a role in the formation of multimers, may be the residue which is vital to CCR5 binding, as it has never been demonstrated that MIP-1α, MIP-1β or RANTES are unable to bind to CCR5 as multimers.

A recent study on RANTES[151] has demonstrated that residues Phe 12 and Ile 15 are essential for CCR5 binding. In addition, residues Tyr 3 and Asp 6 are required for signal transduction. This suggests that the CCR5-binding epitope on RANTES and related β-Chemokines will involve residues close to the disulphide pair. There is also some evidence to suggest that the signal transduction process can be considered as separate from receptor-binding. This would imply a two-step interaction between ligand and receptor, each step involving a different set of residues.

From Table 5.5, it is clear that HCC-1 and DC-DK1, following this pattern of conserved residues, would be expected to also bind to CCR5. However, there has not yet been sufficient

MCP-1	...	S	**Y**	R	**R**	I	-	T	S	S	K	C	P	K	E	A	V	I	F	K	T
MCP-2	...	S	**Y**	T	**R**	I	-	T	N	I	Q	C	P	K	E	A	V	I	F	K	T
MCP-3	...	S	**Y**	T	**R**	T	-	T	S	S	H	C	P	R	E	A	V	I	F	K	T
MCP-5	...	S	**Y**	R	**R**	I	-	T	S	S	Q	C	P	R	E	A	V	I	F	R	T
Eotaxin	...	S	**Y**	R	**R**	I	-	T	S	G	K	C	P	Q	K	A	V	I	F	K	T
MCP-4	...	S	**Y**	V	-	I	-	T	T	S	R	C	P	Q	K	A	V	I	F	R	T
		26	*27*	*28*	*29*			*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	*39*	*40*	*41*	*42*	*43*
MIP-1α	...	**D**	Y	F	E	-	-	T	S	S	Q	C	S	K	P	G	V	I	F	L	T
MIP-1β	...	**D**	Y	Y	E	-	-	T	S	S	L	C	S	Q	P	A	V	V	F	Q	T
RANTES	...	**E**	Y	F	Y	-	-	T	S	G	K	C	S	N	P	A	V	V	F	V	T
HCC-1	...	**D**	Y	Y	E	-	-	T	N	S	Q	C	S	K	P	G	I	V	F	I	T
DC-CK-1	...	**D**	Y	S	E	-	-	T	S	P	Q	C	P	K	P	G	V	I	L	L	T

MCP-1	I	V	A	K	E	I	C	A	D	...
MCP-2	K	R	G	K	E	V	C	A	D	...
MCP-3	K	L	D	K	E	I	C	A	D	...
MCP-5	I	L	D	K	E	I	C	A	D	...
Eotaxin	K	L	A	K	D	I	C	A	D	...
MCP-4	K	L	G	K	E	I	C	A	D	...
	44	*45*	*46*	*47*	*48*	*49*	*50*	*51*	*52*	
MIP-1α	K	**R**	S	R	Q	V	C	A	D	...
MIP-1β	K	**R**	S	K	Q	V	C	A	D	...
RANTES	R	**K**	N	R	Q	V	C	A	N	...
HCC-1	K	**R**	G	H	S	V	C	T	N	...
DC-CK-1	K	**R**	G	R	Q	I	C	A	D	...

Table 5.5: Sequence alignment: Residues that may be involved in CCR5/CCR2 binding

investigation of these proteins to confirm or deny this proposal.

One other notable difference between the MIP and MCP proteins is the charge distribution at their C-termini. As the MIP-1α mutants have shown, modifications to charged residues in this region can have profound effects on the entire molecule, although it remains unclear precisely why. It is certainly possible that CCR5 binding requires the β-Chemokine helix to be proximal to a patch of positive potential, thereby favouring the MIP proteins over the MCPs.

5.3.3.2 Receptor Evidence

Several mutation studies on CCR5 have provided useful data on interactions with β-Chemokines. The N-terminal region has been shown not to be required for high affinity MIP-1α-binding[152]. This region has, however, been shown to be essential for signal transduction[153]. This implies an interaction mechanism for CCR5 and its ligands that is very different to the now partly characterised "two step" interaction between CCR2 and its ligands[152]. The third extracellular loop (Loop D in Figure 5.7) has been shown to be essential for both receptor-binding and signal transduction.

Naturally, the HIV-coreceptor activity of CCR5 has resulted in many other studies on the receptor. It has been shown that signal transduction and HIV-coreceptor activities are entirely uncoupled[153]. The cytoplasmic domains as a whole are unimportant to HIV-coreceptor activity.

5.3.3.3 Conclusion

- CCR5 Loop D may contain a Chemokine-binding motif - comparing the sequences of CCR5 and the other β-Chemokines in this region might allow some of the Chemokine-binding determinants to be elucidated.

- It is possible to construct a model of the interaction between CCR5 and β-Chemokines. (See Figure 5.8). It resembles the CCR2 receptor model, but has a different mechanism.

- HIV-binding requires a ternary interaction between CCR5, CD4 and gp120[154]. The interaction between MIP-1α and CCR5 may mirror the interaction between CCR5 and gp120 (or gp120/CD4) and explain HIV inhibition by β-Chemokines.

- Exposed residues close to the Cysteine pair are essential for receptor-binding.

- Residues closer to the N-terminus are implicated in Signal Transduction, suggesting a two-step mechanism for receptor binding and activation (See Figure 5.8).

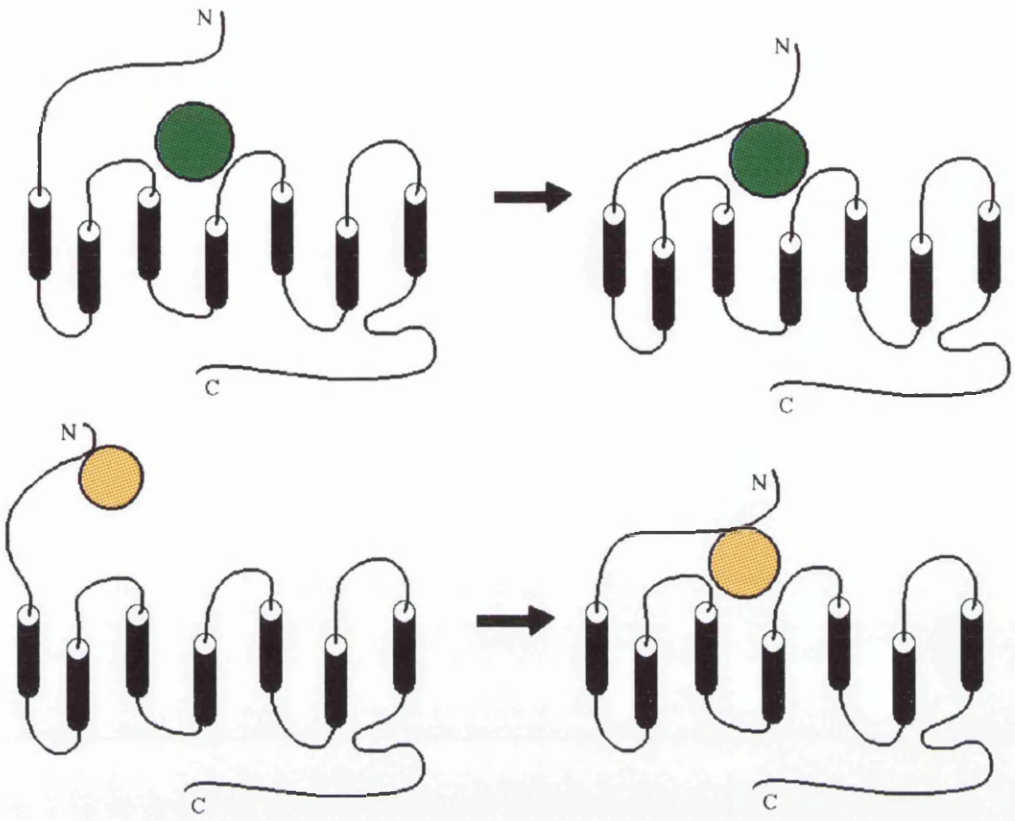


Figure 5.8: Models of MCP and MIP interactions with receptors

In the above diagram, MIP-1α is depicted as a green sphere and MCP-1 as a yellow sphere. The characterised two-step interaction between MCP-1 and CCR2 is shown. A proposed mechanism for binding of MIP-1α to CCR5 and subsequent signal transduction is also shown. In contrast to the MCP-1 mechanism, MIP-1α binds first to the extracellular loops of the receptor, and then the N-terminal region swings across to hold the ligand in place.

5.3.4 CCR1

CCR1 has not been studied to the same extent as CCR2 and CCR5, so the available evidence is a little limited.

It has been shown to bind MIP-1α, RANTES and MCP-3, but not MIP-1β, MCP-1 or MCP-2. Recent tentative evidence suggests that HCC-1 and MPIF-1 may also be viable ligands[6]. Unlike the ligands of CCR5, there is no clear pattern to suggest why only certain Chemokines can bind to CCR1. Sequence similarities between MIP-1α, RANTES, HCC-1 and MPIF-1 are high. However, MCP-3 does not fit this pattern and neither does MIP-1β, as it is very similar to the other ligands yet still does not bind.

In the particular case of MIP-1β there are only two residues in the entire sequence that might constitute significant differences between it and the related MIP proteins. There is an extra Tyr residue at position 63 (using MIP-1α numbering), but more significantly there is an extra proline residue in the N-terminal peptide, at residue 7. As can be seen from the MIP-1β structure[82], this results in the N-terminal residues of MIP-1β following a very different path to that seen for the other β-Chemokines. Since CCR1 binding has been shown[150] to be extremely dependent on the specific conformation of the N-terminal residues of its ligands, this observation is likely to explain the apparently anomalous behaviour of MIP-1β.

From the MCP-3 sequence, there is very little to separate it from the remaining MCP proteins. Looking specifically at the N-terminus, MCP-3 differs from all the other MCP proteins in lacking a proline residue at position 8. Again, it is possible that this residue is responsible for a significantly different N-terminal peptide conformation in MCP-3 relative to the other MCP proteins. However, it should also be noted that in common with all the other MCP proteins, MIP-1α and HCC-1 both possess Pro 8.

The interesting observations with respect to the non-GAG binding mutants of MIP-1α are relevant to the consideration of CCR1. While the murine mutant was reported not to bind to CCR1, the human mutant was reported to bind as normal. The implication is that the loop between strands II and III on the Chemokine ligands, with its high density of positive charges, may be important in CCR1 binding. However, the sequence comparisons point, admittedly vaguely, at the N-terminal regions of these molecules being the major determinant of CCR1 binding.

This assumption was recently partly confirmed in a mutagenesis study on RANTES[151]. It was demonstrated that Arg 17 was essential for CCR1 binding, and that Pro 2, Asp 6 and Thr 7 were all essential for CCR1 signal transduction by RANTES. This seems to imply a two-step ligand-receptor interaction, very similar to that proposed for CCR5. Although the same region appears to

be responsible for signal transduction in both cases, receptor-binding appears to be mediated by slightly different regions of the CCR1 and CCR5 ligands.

5.3.5 CCR2

CCR2 has been shown to bind all of the known MCP proteins - MCP-1, MCP-2, MCP-3, MCP-4 and MCP-5, but does not bind any of the MIP proteins. There is a considerable amount of evidence relating to the parts of the CCR2 molecule that are important in ligand-binding. The N-terminal portion of the receptor has been shown to be necessary for high affinity binding of MCP-1, and presumably for the other ligands as well. Each of the other three loops has been shown to be necessary for signal transduction, although the N-terminus is not. This data has resulted in a model of the interaction between CCR2 and its ligand, as described below.

The activation mechanism of CCR2 has been proposed to be a two step process[152] (See Figure 5.8). Step one involves a direct specific interaction between the N-terminal region of the receptor and the ligand. In step two, the N-terminal region of the receptor moves in order to place the bound ligand onto the surface of the other loops. Non-specific interactions between these loops and the ligand are then responsible for signal transduction.

There is some evidence identifying two regions of the MCP-1 molecule that are important in receptor binding. Both the N-terminus and the loop containing the third Cysteine residue have been shown to be important. The integrity of both regions has been shown, by mutagenesis, to be essential to Chemotactic function[148]. However, this study did not explicitly refer to CCR2 as the receptor involved, and therefore there is no firm evidence that these two regions of the MCP proteins are specific determinants of CCR2 binding.

5.3.6 CCR3

CCR3 has been shown to be specific for the Chemokines Eotaxin and Eotaxin-2, although certain other Chemokines will bind to a certain extent. RANTES, MCP-3 and MCP-4 have been shown to have some effect, although MIP-1α and the non-GAG-binding mutants of MIP-1α have been shown not to bind. The CCR3-binding epitope on RANTES has been proposed to lie close to the CCR5-binding epitope[151], but to be distinct. Residues close to the N-terminus have been identified as important to signal transduction, implying a ligand-receptor interaction similar to that expected for CCR1 and CCR5.

5.3.7 CCR4

CCR4 was known for a considerable time to be a receptor for MIP-1α and RANTES, but it has recently become clear that these two proteins have very minor effects when compared to TARC. While TARC is not strictly specific for CCR4, the affinity between the two is so much greater than that between CCR4 and any other β-Chemokine that they are effectively specific for one another.

5.3.8 CCR6, CCR7 and CCR8

These three β-Chemokine receptors have been identified very recently. Each of them currently appears to be specific for a single β-Chemokine, although that may be due purely to the lack of biochemical knowledge. CCR6 has been shown not to bind MIP-1α, MIP-1β, RANTES, MCP-3 or Eotaxin.

5.3.9 CCR9, CCR10 and other receptors

The most recently identified of the β-Chemokine receptors are CCR9 and CCR10. While CCR9 has been reported its ligands have not, and its role in the β-Chemokine receptor network has not yet been revealed. However, CCR10 has been shown to bind MCP-1 and MCP-3 particularly strongly, and also to be affected to a lesser extent by RANTES and MCP-4. It therefore has a similar ligand-binding pattern to CCR-1 and CCR-3, in that it binds both MIP and MCP proteins.

A variety of orphan receptors exist which are proposed to be Chemokine receptors on the basis of sequence similarity with portions of the existing β-Chemokine receptors. CCR6 and CCR8 are examples of what were originally orphan receptors of this type, but which were added to the β-Chemokine receptor family when their ligands were shown to be β-Chemokines. As a number of these orphan receptors currently exist and others are still being discovered, it is certain that many other β-Chemokine receptors remain to be identified and characterised.

The Duffy Blood Group Antigen (DBGA or DARC) is unusual in that it can bind both β- and α-Chemokines, despite having only 30% sequence identity with the appropriate receptors for those two families. DBGA binding studies involving the Chemokines RANTES, MCP-1, MGSA and IL-8 have provided a considerable body of information relating to the specific regions of the receptor important to this interaction.

The N-terminal region of DBGA has been shown to be essential, in much the same way as that of CCR2 was. In addition, it has been shown[155] that the formation of a disulphide bond between the N-terminus of the receptor and the third extracellular loop (D) forms a "pocket"

which accommodates the Chemokine ligand.

Some recently described chimeric mutant receptors showed Chemokine specificities somewhere between those of CCR2 and CCR5. By creating a receptor that contained the important determinants for MCP binding by CCR2 - the N-terminus and the first extracellular loop - and the important determinants for MIP binding by CCR5 - the second and third extracellular loops -they were able to combine the two activities. The chimeric receptor binds MIP-1α, MIP-1β, RANTES, and MCP-1, and appears to show that the ligand-binding mechanisms used by CCR2 and CCR5 are completely different.

A murine MIP-1α receptor that has no known human homologue has recently been described. It binds murine MIP-1α and also non-GAG binding variants of murine MIP-1α. It is known to be expressed on monocytes, therefore it is surprising that the human homologue, if it does exist, has not yet been identified.

5.4 Conclusion

5.4.1 Stem Cell Inhibition by MIP-1α

The three MIP-1α mutants are not more active than native protein. This suggests that they undergo spontaneous disaggregation in solution prior to functioning as monomer. This is consistent with aggregation model Agg1, where a dynamic equilibrium of aggregate states would be likely to exist. It is probably less consistent with aggregate model Agg2, which would constitute a probably fairly stable aggregate of 16 monomer units.

Stem Cell inhibition has been shown not to be affected by mutation of Lys 44 or Arg 45. Therefore Stem Cell inhibition may well use the same receptor-binding region on MIP-1α as interacts with CCR5. There is some evidence to suggest that interaction with CCR1 and Stem Cell Inhibition are mediated by different parts of the molecule.

5.4.2 HIV-1/CD4 Interactions of MIP-1α

HIV-1 glycoprotein120 (gp120) loop v3 interacts with CD4 and a Chemokine receptor in order to allow entry of HIV-1 to human cells. MIP-1α, as well as MIP-1β and RANTES, has been shown to prevent the infiltration the virus into some cells, presumably by inhibiting this interaction. It is possible that the viral protein mimics the way in which MIP-1α interacts with the receptor, and that any model of the interaction between MIP-1α and CCR5 would also have implications for the interaction between CCR5 and viral glycoproteins.

It is possible that the dimerisation of MIP-1α and the interaction between CCR5 and MIP-1α involve similar contacts. The precise conformation of MIP-1α in the dimer interface region might perhaps provide some indication of the way in which gp120 interacts with CCR5.

5.4.3 GAG-Binding by MIP-1α and β-Chemokines

While it is known that MIP-1α binds heparin and other glycosaminoglycans, the specific interactions involved and affinity for various GAG populations have not been studied. A recent investigation identified a probable heparin-binding site on human MIP-1α, which was consistent with previous mutagenesis data[156]. The residues Arg 18, Arg 46 and Arg 48 in human MIP-1α were each shown to be essential for heparin binding. In addition, residues Lys 45 and Lys 61 were shown not to be involved. A related murine MIP-1α mutant, called HepMut, in which Lys 44 and Arg 45 had been neutralised, did not bind to proteoglycans[94]. These two studies appear to be slightly at odds with one another, as the first identifies Lys 44 as unimportant but the second suggests that this residue is vital. Since the second case was a double mutant however, it is possible that all the biochemical properties ascribed to HepMut are due specifically to the neutralisation of Arg 45.

The observed biochemical properties of the various mutants appear also to be contradictory. While mutant R45A bound to CCR1 with the same efficacy as native protein[156] and was equally active, HepMut was inactive on human monocytes, which is likely to indicate a compromised ability to bind to human CCR1[94].

It is proposed that the residues Arg 17, Arg 45 and Arg 47 form what is known as a "cationic cradle" (as shown in Figure 5.9), and that Lys 18 may also contribute[157]. The MIP-1α structures are consistent with the formation of a short cleft lined by these residues, although Arg 21 seems to point away from the other basic residues and appears unlikely to be involved. What is interesting is that this proposed "cradle" is in fact the dimerisation site on MIP-1α. This implies that GAG binding would disrupt formation of MIP-1α multimers larger than dimers, and vice versa. It also suggests that at least some of the dimer-dimer interactions observed in the tetrameric mutant structures will mirror interactions that occur upon GAG binding by MIP-1α.

Although these results are apparently contradictory, they raise some very interesting points regarding the interactions of MIP-1α with proteoglycans. Firstly, both studies indicate that unlike members of the α-Chemokine family such as PF4, the heparin binding site on MIP-1α is a single patch of positive potential formed by, at most, three residues within a monomer. Secondly, the physiological role of such an interaction is unknown and there are several potential reasons for

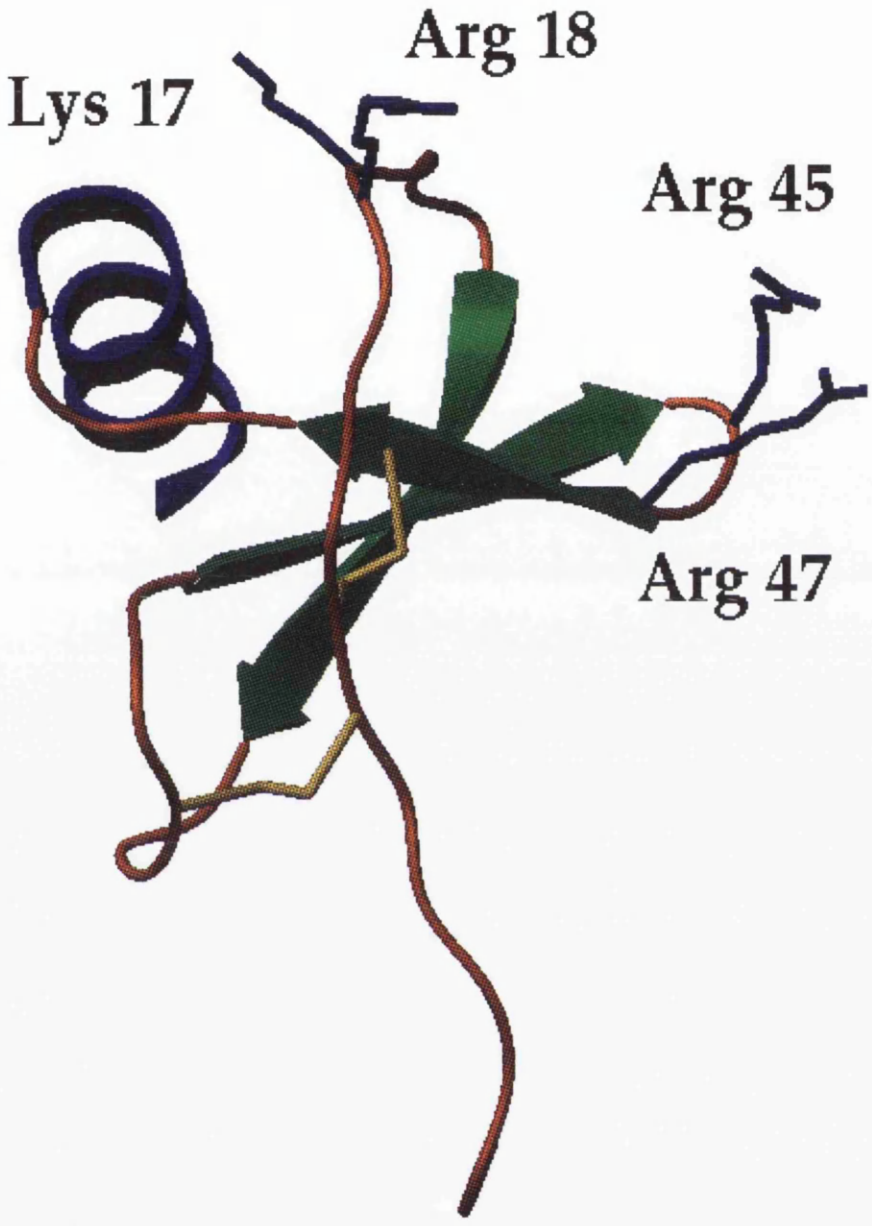


Figure 5.9: Residues involved in the proposed cationic cradle in MIP-1α

it. Glycosaminoglycans, bound to the surface of endothelial cells, may bind Chemokines in order to immobilise them and hence form a chemotactic gradient. It is also possible that Chemokines are bound in order to provide a method of presentation to target cells, or in order to increase local concentration of a particular Chemokine , perhaps to facilitate formation of multimers[158].

Previous mutagenesis studies on β-Chemokines identified the regions close to the first disulphide pair and the loop containing the third disulphide as being essential for receptor binding[148] on monocytes. Inspection of the MIP-1α structure indicates that in the monomer these regions are substantially removed from the region containing the heparin binding site. This makes it unlikely that both regions would be involved simultaneously in interactions with a common receptor, and suggests that the CCR1 and GAG receptor binding regions on MIP-1α are distinct.

The heparin-binding site in the MCP-1 structure shows a stripe of positive potential, similar to that seen for PF4 and the α-Chemokines. However, this band is nowhere near as pronounced, and does not encircle the molecule. Unlike the heparin-binding portion of PF4, this patch of positive potential appears not to involve the C-terminal helices at all. This suggests that the heparin-binding mechanisms of MIP and MCP β-Chemokines are distinct, and that both are different to that observed for the α-Chemokines. However, assuming that I-309 dimerises in the same way as the MCP proteins, the preponderance of basic residues in its probable C-terminal helix is very reminiscent of the corresponding region of the α-Chemokines, and suggests that this particular β-Chemokine may bind heparin in a way similar to that observed for PF4. In addition, mutagenesis studies have shown some of these basic residues to be essential for heparin binding, suggesting that there may be several different heparin interaction mechanisms, even within the β-Chemokine family.

Interestingly, although the β-Chemokines are all thought to function as monomers, there is evidence to suggest that MCP-1 can bind to its receptor in the dimeric form[158]. IL-8 has also been shown to be able to bind to its receptor as a dimer[159], and so it is tempting to speculate that there may be some role for the GAG-binding surface of the dimer in promoting such an interaction.

Another interesting observation is that C-terminally truncated IL-8 is unable to bind to heparin, and is not fully active as a chemoattractant. This implies that either the GAG-binding region of IL-8 is also involved in receptor-binding, or that the capacity to bind GAGs is required in order to bind to IL-8 receptors. Given that MCP-1 and MCP-3 form a similar dimer to IL-8, it is natural to wonder whether these two molecules should be able to cross-react with IL-8 receptors. However, the N-terminus of MCP-1 has been shown to be essential for binding to its receptor, CCR2. This suggests that the specific regions of IL-8 and MCP-1 required to bind to their respective

receptors occur at fairly distant parts of the molecules, despite their GAG-binding regions being comparitively close.

5.4.4 Redefinition of Chemokine Family Boundaries

Three residues, Tyr 28, Arg 30 and Asp 68 have been identified as being essential for the formation of the IL-8-like dimer by MCP-1. These residues are involved in a small hydrogen-bonded network which enables interaction between the helices on each monomer. In the MCP-1 tetramer which involves this dimerisation interaction, there also appear to be several important internal electrostatic interactions. Residues Arg 29 and Glu 50 interact via salt bridges, and Lys 44 is also internal to the tetramer and may also interact with Glu 50 either directly or via water molecules. In addition, two other residues, Thr 10 and Pro 8, point towards the internal cavity in the MCP-1 tetramer. Previous mutagenesis studies had replaced Thr 10 with Arg and generated an inactive MCP-1 mutant. Mutation of the residues Arg 30 and Asp 68 had significantly less effect on biological activity, while residue Tyr 28 appeared to be so vital to both activity and stability that when it was mutated the resultant peptide was improperly processed and was not secreted.

MCP-1 tetramerisation seems to be stabilised by electrostatic interactions involving residues Arg 29, Lys 44 and Glu 50. While the dimerisation and tetramerisation processes in MIP-1α have been shown to be predominantly hydrophobic, it is uncertain whether the same is true for MCP-1. Examination of sequence alignment for the β-Chemokines shows that these three residues are conserved within the MCP branch of the β-Chemokines and are completely unconserved within the MIP proteins (See Table 5.5). This suggests that these proteins will be members of a β-Chemokine subfamily, will be structurally and functionally similar, and will be the only members of the β-Chemokine family which can form multimers using the IL-8 dimer-like interaction. The implication therefore is that MCP-1, MCP-2, MCP-3, MCP-5 and Eotaxin, but perhaps not MCP-4, will be members of the MCP subgroup of the β-Chemokine family.

Conversely, it can also be seen from such an alignment that there are also certain residues conserved only in those proteins related to MIP-1α. More specifically, some of those residues already identified as being crucial to the formation of MIP-1α dimers, tetramers and higher aggregates are conserved in MIP-1β, RANTES, HCC-1, ILINCK and DC-CK-1 (See Tables 5.5 and 1.2). The sequence identity analysis of the various members of the β-Chemokine family has already provided some related clues, but this study seems to indicate that while MIP-1α, MIP-1β, HCC-1, ILINCK and DC-CK-1 are very closely related, RANTES is a little different and does not contain exactly the same pattern of conserved residues. This is reflected in the receptor-binding pattern

of RANTES in comparison to the other MIP-like proteins. RANTES, in binding to CCR3 and CCR10 appears to bear more similarity to the MCP proteins than the other members of the MIP subfamily. Residues Asp 26, Arg 45 and Arg 47, proposed as being important to tetramerisation, have been confirmed as such by the MIP-1α tetramer structures. Each of them is either conserved or conservatively mutated in MIP-1β, RANTES, HCC-1, ILINCK and DC-CK-1. While many other unrelated β-Chemokines have positively charged residues at the positions of the conserved motif KRXR, Asp 26 is completely unconserved throughout the remaining members of the β-Chemokine family. This suggests that the dimer-dimer interaction in these molecules requires a salt-bridge between a specific single negatively-charged residue and a patch of positively charged residues. The existing mutagenesis data for MIP-1α would concur with this hypothesis. Neutralisation of Asp 26 produces a molecule that cannot aggregate past the dimer state, whereas both Lys 44 and Arg 45 must be neutralised to produce a similar effect. In addition, there is still evidence for the presence of some larger multimers even when Lys 44 and Arg 45 have been neutralised. Glu 29 also appears to be conserved as an acidic amino acid in the MIP-like proteins, but as a basic amino acid in MCP-like proteins.

Mapping the MIP-1α structure onto that of MCP-3 allows investigation of the specific interactions within these molecules which determine the mode of dimerisation. Comparing the MIP-1α dimer to that of MCP-3, the following observations can be made:

- Tyr 28 (numbering according to MIP-1α) is conserved and has the same rotamer in both.
- Glu 66 replaces Asp 68 - the new side-chain is too long to allow H-bonding to Tyr 28.
- Asp 26 replaces Ser 27 - this produces a steric clash with Tyr 28' on the opposing monomer.
- Glu 29 replaces Arg 30 - the new side-chain is not long enough to allow interaction with Tyr 28. Arg 30 is presumably required to balance the charge on Asp 68 as well. In the MCP-3 dimer Asp 68 comes unfavourably close to Asp 68' on the other monomer - presumably the presence of Arg 30 alleviates this problem.
- Lys 44 and Arg 45 generate steric clashes with the N-terminus of the opposite monomer. In MCP-1, the corresponding residues are small hydrophobics. As Lys 44 is conserved in some MCP proteins, but Arg 45 is not, it is presumably the latter which is primarily responsible.

Although none of the examples given above represents a major clash, the presence of a few such adverse steric interactions would probably be enough to push the monomers slightly further

apart. Since the monomers in the MCP-3 dimer are already considerably more separated than in IL-8 and lack the stabilisation of an extended β-sheet, this extra loss of stability would be enough to make this dimer less favourable than that formed by the MIP proteins.

The problems encountered in modelling a MIP-1α monomer-monomer interaction via the MCP dimerisation method correspond very well with the comparison of the sequences of MIP and MCP proteins in Section 1.3.2. Most of the residues which are conserved differently within these two families are directly involved in monomer-monomer interactions forming one of the two dimers. The most notable exception is the observed difference in charged residue distribution in the C-terminal helices of the two families. While MIP-1α, MIP-1β and, to a lesser extent, RANTES, have a propensity for acidic residues in this region, most of the other β-Chemokines have a high proportion of basic residues. It is possible that the means to stabilise such a knot of positive charges exists in the MCP-3 dimer, but a similar knot of negative charges induces instability into this particular mode of interaction.

The idea that a single salt-bridge interaction between dimers is sufficient to explain the generation of the MIP-1α tetramer is at odds with spectroscopic evidence. As was discussed earlier, spectroscopic data indicate that the major stabilising force in the dimer-dimer interaction is hydrophobic. It seems likely that this is in fact true, given the extensive areas of surface on each dimer involved in the interaction. While such interactions may be favourable for all β-Chemokines, it is possible that only for MIP-1α and related proteins is there the possibility of the Asp 26 salt-bridge interaction to anchor the two dimers together. Without it, dimer-dimer interactions would be a transient phenomenon, and the dimer-tetramer equilibrium would favour the presence of the dimer.

It is possible that lack of either of the two motifs identified as important for tetramer formation precludes a Chemokine from self association to anything larger than a dimer. Certainly there are many novel members of the β-Chemokine family which do not fall into either of the newly defined groups, and are correspondingly different when sequence identity and chromosomal location are considered. Alternatively, there may be other Chemokine aggregation methods still to be discovered.

Three definite subfamilies, the MIPs, the MCPs and the Exodus proteins, appear to exist within the β-Chemokine family. Whether further subdivisions of the family will be necessary remains to be seen, although several of the known β-Chemokines do not seem to correspond to any of these three. It does seem likely that as more and more examples are discovered, the specific differences between the various proteins will become more apparent and their grouping and nomenclature will

have to be further rationalised.

Figure 5.10 is a summary of all relevant data for the characterised β-Chemokines. It is an attempt to subdivide the proteins based on their sequence similarities[26], chromosomal location and receptor-binding patterns. It seems clear from the sequence comparison in the right-hand portion of the diagram, and from the previous discussion in Section 1.3.2 that there are three distinct clusters - the MIP proteins, the MCP proteins and the Exodus proteins. The MIP proteins are coloured blue, the MCP proteins red, and the Exodus proteins green for clarity.

While the MIP and MCP groups occur at similar positions on the same chromosome, the other β-Chemokines have more remote locations. This observation, allied to the sequence comparison, suggests a more simple distinction between these subfamilies than would be expected from biochemical considerations.

However, the binding patterns of the various receptors reflect the true complexity of the β-Chemokine system. Clearly there is a subset of receptors which binds MIP proteins (CCR4, CCR5 and perhaps CCR9), and a subset which binds MCP proteins (CCR2). However, receptors CCR1, CCR3 and CCR10 all bind different subsets of both MIP and MCP proteins. Another receptor subset (CCR6, CCR7, CCR8) is specific for a single β-Chemokine. Even CCR4 is only weakly affected by RANTES and MIP-1α, and can be considered as a specific receptor for TARC.

At present, the underlying organisation of the β-Chemokine receptor family appears to be as follows:

- Proteins in the MIP-1α family bind to a subfamily of shared receptors.
- Proteins in the MCP family bind to a different subfamily of shared receptors.
- A third subfamily of receptors can bind proteins from both the MIP-1α and MCP branches.
- This receptor type appears to favour proteins which are intermediate in character between MIP and MCP families, most notably RANTES and MCP-4.
- A fourth subfamily of receptors is specific for a single β-Chemokine.

Clearly the network of β-Chemokines and their receptors in extremely complex, and is only beginning to be understood. However the pivotal role which the Chemokines occupy within the immune system has led to their exploitation by a variety of pathogenic organisms. Their importance in viral infections such as HIV is certain to promote the identification and characterisation of novel ligands and receptors, and should improve understanding of this fascinating and intricate

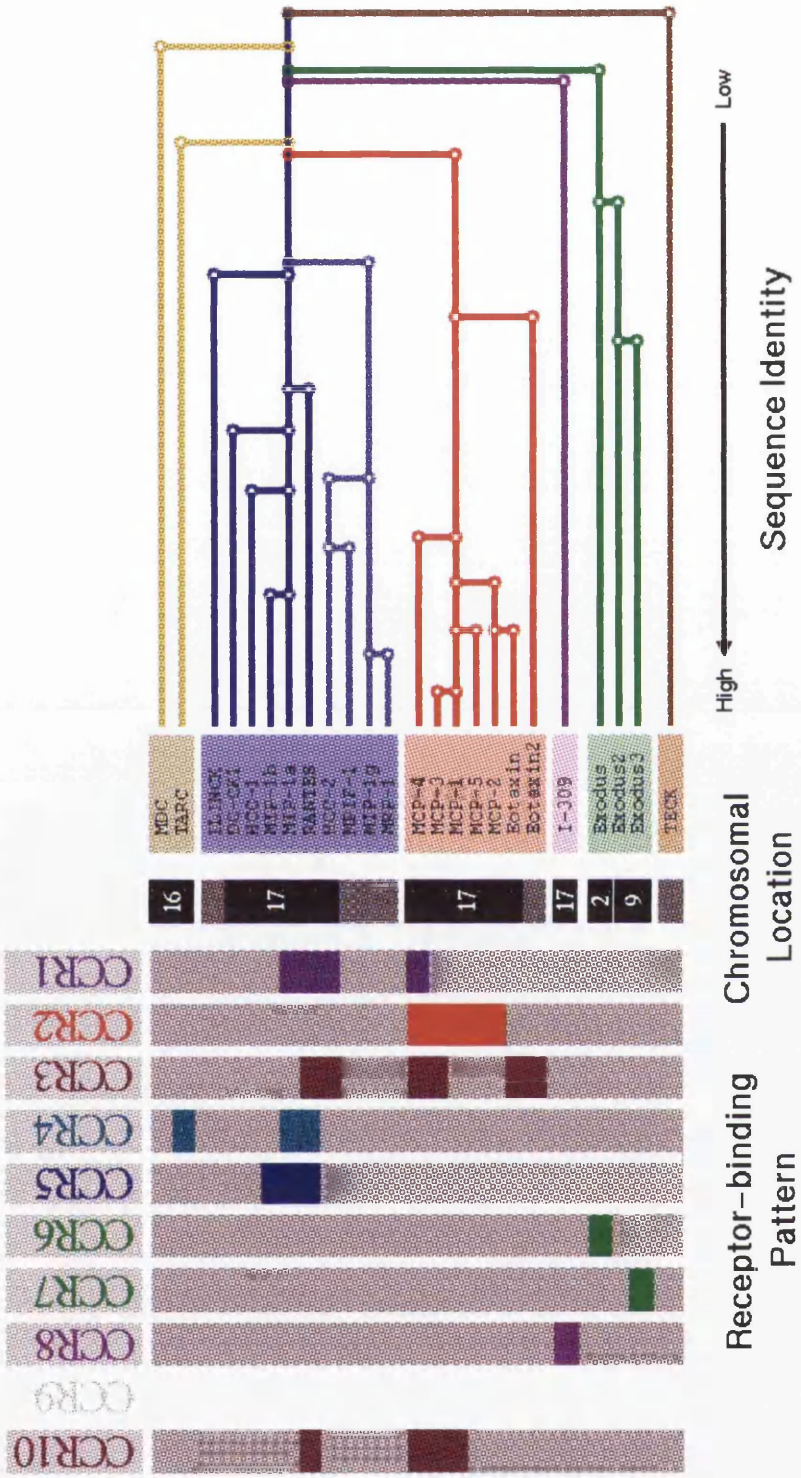


Figure 5.10: Summary of available data on β-Chemokines and their receptors

system. The great variation in receptor-binding specificities of the various β-Chemokine ligands, allied to the large number of receptors with overlapping but slightly differing functions, provides the Chemokine system with the flexibility and versatility required for such an important component of the immune response.

6. SUMMARY

The structures of three interesting structural mutants of the Chemokine MIP-1α have been solved by X-ray crystallography.

As MIP-1α tends to extensively self-aggregate under various conditions, non-aggregating mutants were generated by successive neutralisation of acidic amino acids in its carboxy-terminal region. The three mutants thus generated showed stable molecular weights corresponding to monomeric, dimeric and tetrameric variants.

These mutants were crystallised and X-ray diffraction data were collected from each of them. The phase problem was solved by SIRAS in the case of the monomeric mutant, and by Molecular Replacement in the case of the dimeric and tetrameric mutants.

The resulting structures have revealed the way in which MIP-1α forms multimers from monomeric units. The structures of monomer, dimer and tetramer were unambiguously determined, and models for the structures of larger aggregates have been proposed.

The high resolution structure of the MIP-1α tetramer will be extremely useful in investigating the interactions between β-Chemokines and their receptors. It is the highest resolution β-Chemokine structure known to date, and is also the first example of a novel method of Chemokine tetramerisation.

The existence of a second method of multimerisation within the β-Chemokine family has many implications. It is now clear that the β-Chemokine family consists of several subfamilies. While they were known to be functionally distinct, the MIP-1α structures have shown that there are also structural distinctions.

Greater understanding of the aggregation method of MIP-1α will direct the search for more effective and more clinically useful stem cell inhibitors. Knowledge of the precise conformation of receptor-binding portions of the molecule will aid the search for small molecules able to mimic the SCI and anti-HIV properties of MIP-1α.

As the amount of β-Chemokine structural data available increases, so too does the understanding of the complex way in which they interact with their receptors.

APPENDIX

A. ALTERNATE NAMES OF CHEMOKINES

Chemokine	Family	*Human* Name(s)	*Murine* Name(s)
9E3	α	*Only Chick form known - 9E3, EMF-1, CEF-4*	
ENA-78	α	ENA-78	*not known*
GCP-2	α	Granulocyte Chemotactic Protein	*not known*
γ-IP10	α	γ-IP10, INP-10	INP-10, CRG-2, IFI10, C7
GRO-α	α	GRO-α, MGSA, NAP-3, GRO, GRO-1	KC, Secretory Protein N51
IL-8	α	IL-8, MDNCF, TCCF, NAP-1 LDNAP, 3-10C, NAF, GCP-1, LUCT	*not known*
LIX	α	*not known*	LIX
MIG	α	MIG	mig, M119
MIP-2α	α	MIP-2α, GRP-β, Gro-β, Gro2,	MIP-2
MIP-2β	α	MIP-2β, GRP-γ, Gro-γ, Gro3	MIP-2
NCXC-1	α	NCXC-1	*not known*
PF4	α	Oncostatin-A, Iroplact	*not known*
SDF-1α	α	SDF-1α, SDF-1β, PBSF,	SDF-1, PBSF, TPAR-1, TLSF,
β-TG	α	β-TG, CTAP-III, LA-PF4, NAP-2,	*not known*
DC-CK1	β	DC-CK1, MIP-4, PARC, SCYA18, AMAC-1	*not known*
Eotaxin	β	Eotaxin, SCYA11	Eotaxin, Scya11
Eotaxin-2	β	Eotaxin-2, MPIF-2, Ckβ-6, SCYA24	*not known*
Exodus	β	Exodus, LARC, MIP-3α, SCYA20	*not known*
Exodus-2	β	Exodus-2, SLC, SCYA21, TCA4, 6CKINE	*not known*
Exodus-3	β	Exodus-3, EBI1-ligand, SCYA19, MIP-3β	*not known*
HCC-1	β	HCC-1, NCC-2, SCYA14, CCC-1, HCC-3	*not known*
HCC-2	β	HCC-2, MIP-5, NCC-3, CCC-2, SCYA15	*not known*
I-309	β	I-309, SCYA1	SIS-ε, P500, TCA-3, SIS-f
ILINCK	β	ILINCK, NCC-4, LEC, SCYA16	*not known*
MCP-1	β	MCP-1, MCAF, MIP-JE, HC11, SCYA2	JE, Scya2
MCP-2	β	MCP-2, HC14, SCYA8	*not known*
MCP-3	β	MCP-3, NC-28, SCYA7	MARC, FIC, Scya7
MCP-4	β	MCP-4, NCC-1, SCYA13, CKβ-10	*not known*
MCP-5	β	*not known*	MCP-5, Scya12
MDC	β	MDC, STCP-1, SCYA22	*not known*
MIP-1α	β	MIP-1α, LD78, SCYA3, G0S19-1 SIS-β, PAT 464.2	MIP-1α, Scya3, SIS-α, L2G25B, TY-5
MIP-1β	β	MIP-1β, TCAP-2, Act-2, Pat 744, H400 SIS-γ, LAG-1, HC21, SCYA4, G-26	MIP-1β, H400, SIS-γ, Scya4, L2G25C,
MIP-1γ	β	*not known*	MIP-1γ, P150-1, CCF-18, MRP-2, Scya9
MRP-1	β	*not known*	MRP-1, C10, Scya6, L2G75B
MPIF-1	β	MPIF-1, Ckβ-8, MIP-3, SCYA23	*not known*
RANTES	β	RANTES, SIS-δ, SCYA5	RANTES, SIS-δ, MURANTES, Scya5
TARC	β	TARC, SCYA17	*not known*
TECK	β	TECK, SCYA25	*not known*
Lymphotactin	γ	Lymphotactin, ATAC	Lymphotactin
Fractalkine	δ	Fractalkine, Neurotactin	Fractalkine, Neurotactin

B. CRYSTALLISATION THEORY

In a crystallography experiment, the aim is to derive the three-dimensional structure of a sample by examining the way in which it diffracts an X-ray beam of a particular wavelength. Crystals are used as they are a huge array of identically oriented objects, and therefore give a diffraction signal which is amplified with respect to that of a single object. The strength and quality of the diffraction signal obtained from a crystal will depend on its size and internal order. It is therefore desirable to be able to grow large, well-ordered single crystals of a protein to aid structure determination, although this is often far from straightforward. This Appendix gives an outline of the theory of protein crystallisation and describes some of the variables which are important in many crystallisation experiments.

B.1 Theory of Crystallisation

In any type of crystallisation experiment, the aim is to supersaturate the protein solution, and then to allow the solution to become less saturated in a controlled fashion, hopefully resulting in well-ordered crystals. This is most commonly done by vapour diffusion (see Section B.2.3), where diffusion of a solvent away from the protein solution results in supersaturation and often crystallisation.

The crystallisation of proteins is a very complex and poorly understood process. There are many parameters controlling the rate at which supersaturation is achieved and the rates of crystal nucleation and growth, and it is important to consider each of these when attempting to crystallise proteins. They include temperature, pH, protein concentration, presence and concentration of precipitants, salts, detergents or other additives. Given the large number of parameters that can be varied and the fact that only small amounts of protein are generally available, a full-matrix search of all of them is impossible. Therefore it is common to employ a sparse-matrix screening system in initial crystallisation trials, in order to sample as much of "crystallisation space" as possible[99]. Several sparse-matrix screens are now commercially available.

It will be thermodynamically favourable for a protein to come out of solution when it becomes supersaturated. However, this does not necessarily result in crystallisation. The strategy of a crys-

tallisation experiment is best explained by the use of a solubility diagram such as Figure B.1. At any point below the Solubility Curve in this diagram, the protein is soluble. At any point above the line, the protein solution is supersaturated. At any point above the Precipitation Curve, the protein will precipitate immediately, and crystallisation will not occur. These diagrams (Figures B.1 and B.2) are schematic representations of the relationship between protein concentration and crystallising agent concentration during crystallisation

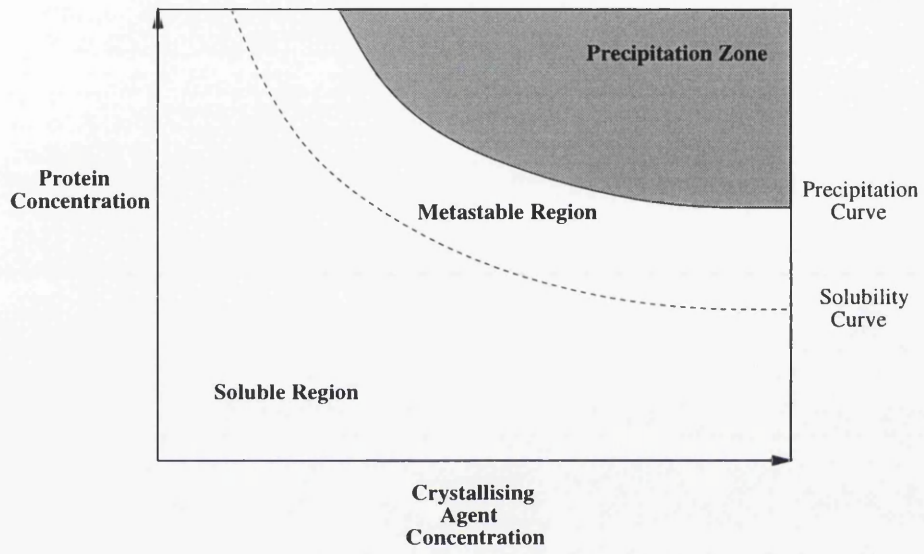
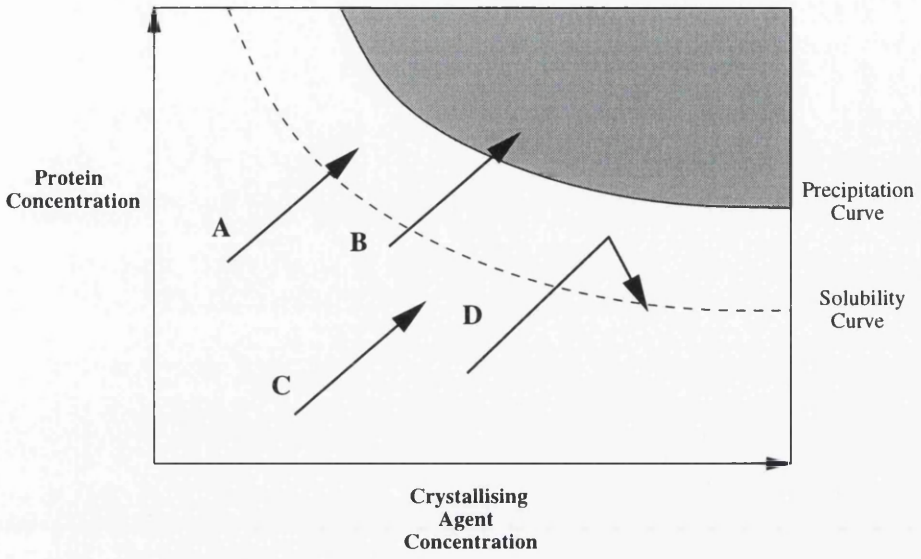


Figure B.1: Phase diagram for protein in solution.

In order for crystal nucleation to take place, the protein solution must lie within the Metastable Region. While here, the solution will be able to produce small clusters of protein molecules which can act as nucleation centres, and can grow larger by addition of more protein molecules. Recent theoretical studies suggest that the initial nuclei consist of as few as four molecules, and can be formed within microseconds of the crystallisation drop being prepared[160]. As more nucleation sites are formed, and as these proto-crystals grow larger, the protein concentration decreases. In order for large crystals to form, it is important that the rate at which nucleation sites are formed is low, or else many microcrystals will be formed and the solution will become depleted of protein before they can grow to reasonable size.

Therefore the ideal scenario is one in which the protein solution begins in the Metastable Region, and forms one or more nuclei. These crystals grow larger, and so the protein concentration decreases, bringing the solution close to the Solubility Curve again. Once at this point, equilibrium has been reached, and no more nucleation or crystal growth will occur. Figure B.2 depicts some

of the possible outcomes of crystallisation experiments.



A: Equilibrium reached before nucleation occurs.
No crystals formed.

B: Protein precipitates.

C: No crystals or precipitate formed.
Clear drop.

D: Crystals formed.

Figure B.2: Phase diagram for possible crystallisation experiments..

B.2 Factors affecting Crystallisation

B.2.1 Temperature

Protein crystals have been obtained from a wide variety of temperatures. It is advisable to maintain a constant temperature during crystallisation as even small changes in temperature can have significant effects. It is also possible to use a temperature gradient to bring a protein out of solution and form crystals, but this requires a very precise control of the temperature and its rate of change.

B.2.2 pH

The pH of the crystallisation medium is generally the parameter that can be most easily varied, but is often the most sensitive parameter as well. A low concentration of a buffering agent is usually present in crystallisation trials to maintain a constant pH. A variety of common buffers are used in protein crystallisation, covering a useful pH range of 2.0 to 12.0. A protein tends to be least soluble at or near its isoelectric point, so by choosing a buffering agent at a pH close to this it is possible to maximise the chances of the protein coming out of solution.

B.2.3 Precipitant

Although some proteins will naturally crystallise from solution given time, most must be forced out of solution using substances known as precipitants. These are mixed with the protein solution and equilibrated with a reservoir solution containing the precipitant at a higher concentration . As the vapour pressures of the two solutions equilibrate, water, or any volatile solvent, will diffuse from the protein solution to the reservoir, gradually increasing the concentrations of protein and the other components in the drop. This is the vapour diffusion technique.

For crystallisation to occur, the protein must reach a sufficient degree of supersaturation for nucleation to occur. In some cases this is not possible, but existing nuclei can be introduced to the crystallisation drop in the form of "seeds" which then grow normally.

One of the most commonly used precipitants is salt, which has a fairly significant effect on the ionic strength of a protein solution, and hence on the solubility of a protein in that solution. Protein solubility tends to decrease with increasing salt concentration. This property is exploited in the "salting out" procedure, a method of crystallising proteins by slowly increasing the salt concentration within the protein solution. As the salt concentration increases, so does the ionic strength of the protein solution. The salt ions mask the charges on the surface of the protein, with the result that the protein's solubility depends on hydrophobic effects rather than interactions of its charged residues with their environment. By coordinating the water surrounding the protein, sufficient salt concentrations begin to strip away water molecules involved in protein solvation. This forces the protein out of solution, hopefully resulting in crystallisation. Protein solubility also tends to be low at very low ionic strength, and in certain cases it is possible to crystallise proteins by "salting out". This involves decreasing the salt concentration in the protein drop to the point where the charges on the protein surface are no longer masked by salt ions, and the protein molecules begin to aggregate via salt bridges, potentially forming crystals. This method is

particularly useful for highly charged proteins.

Other commonly used precipitants include volatile compounds such as methanol, alcohols, sugars and ethylene glycols. Polyethylene Glycol (PEG), which is available as a polymer with a range of Molecular Weights from 200 to 20000, is the most commonly used precipitant. Each of these compounds performs a similar task to that of the salt in the "salting in" and "salting out" procedures; binding water molecules and eventually stripping off so many that the protein is no longer solvated. Unlike salts however, they have no effect on the ionic strength of the crystallisation solution. For this reason, and due to the variety of available molecular weights, PEG is an extremely useful and versatile reagent for protein crystallisation.

B.2.4 Protein Concentration

It is clear from Figure B.1 that the concentration of the protein sample used in a crystallisation experiment will have a profound effect on the kinetics of nucleation and crystal growth. Higher protein concentration will bring the protein closer to supersaturation. In general, protein concentrations of around 10mg ml^{-1} are used in initial trials, but there are examples where much higher concentrations were required for nucleation. At the other extreme, it is possible to crystallise proteins at relatively low concentrations (less than 1mg ml^{-1}) using a sufficiently high precipitant concentration. In this case, there is the advantage that equilibration occurs rapidly and crystals can therefore be obtained in a matter of hours.

B.2.5 Detergent

In the crystallisation of membrane proteins, it is necessary to simulate the lipidic environment of the exposed hydrophobic portions of the protein in order to maintain its structural integrity. This is commonly done by replacing the lipid molecules surrounding the crystal with molecules of a detergent. They generally consist of uncharged polar head groups with short aliphatic chains.

B.2.6 Additives

In some specific cases, most commonly for membrane proteins, the inclusion of a small molecule additive to the crystallisation mixture has been vital in obtaining high quality crystals. Amphiphilic molecules are generally used. Their mode of action is very poorly understood, but it is thought that they might interact in some way with detergent micelles to modify the parameters of a membrane protein crystallisation experiment.

C. MOLECULAR REPLACEMENT

C.1 Overview

Molecular Replacement (MR) is an increasingly common technique in which a protein of known structure is used to phase an unknown protein of similar structure. In order to calculate phases for the unknown structure, the orientation of the known protein relative to the unknown protein must be determined. The MR technique allows the determination of the six parameters, three rotational and three translational, that define this orientation.

C.2 Rotation Function

The rotational and translational searches can be performed separately, greatly simplifying the problem. The rotational search is conventionally performed in reciprocal space by making use of the Patterson function. As it is the self-vectors in the Patterson that are of interest in the rotation search, the Patterson is calculated only within a sphere of radius r from the origin. r is generally close to the largest intramolecular distance, and therefore will exclude most of the cross vectors and include the majority of the self vectors.

The success or failure of a rotation search is generally assessed by considering the "rotation function". By strict definition, this is a product function, composed of the original Patterson and the rotated Patterson[161]. Naturally the optimal Patterson overlap would be expected to correspond to the maximum of the product function.

C.3 Translation Function

For a successful rotation function solution, the second stage of the orientation procedure is known as the "translation function". In this case it is the intermolecular vectors which are useful, so many of the modern MR programs automatically subtract the intramolecular vectors during computation of the translation function.

There are a variety of possible translation function techniques, including both real and reciprocal space methods. Most commonly, the translation function is also treated as a Patterson overlap

function.

On successful completion of both rotation and translation steps, the search model is transformed by the rotational and translational portions of the calculated orientation matrix to obtain an initial model and hence initial phasing for the unknown structure.

C.4 Rigid-Body Refinement

It is increasingly common to improve the orientation matrix obtained from Molecular Replacement by performing rigid-body refinement of the oriented search model relative to the diffraction data. As well as improving the quality of the model and hence the initial phasing, this also makes it possible to deal with search models in which there is a domain motion relative to the unknown structure.

C.5 Modifications of the basic MR technique

Although the technique is now well-established, there are many recent and interesting variations. Patterson Correlation (PC) refinement is an analogue of the rigid-body refinement method, but is applied to a rotation function solution in order to improve the chances of obtaining a solution from the translation function. Phased translation searches can be performed when approximate phase information is available from another source. For example, if one part of a two-domain structure has been positioned, approximate phases from this part can be used to improve the translation search for the second component.

Although the number of proteins being solved by Molecular Replacement continues to rise, a variety of problems associated with the technique are also becoming apparent. Clearly it is necessary for the coordinates of a homologous structure to be available for use as a search model. This is becoming less and less of a problem, as the number of novel folds identified each year is declining. An increasing number of structures, particularly of smaller proteins, are being solved by NMR methods. It is generally thought (as discussed in Section 3.8.3) that such structures present some challenges when used in Molecular Replacement. Overcoming such difficulties will almost certainly be important in the near future as more and more NMR structures become available.

Finally, the majority of unsuccessful Molecular Replacement attempts are desribed as a failure of the MR technique. However, it is generally more correct to observe that among the many solutions that MR provided, one may well be correct. Unfortunately, the criteria for distinction between correct and incorrect solutions are unable to provide the correct one. It seems possible

that improving these criteria will have a more profound effect that other improvements to the MR technique.

D. GLOSSARY

- Basophil : A type of leukocyte or white blood cell.
- Chemotaxis : Directed movement of cells.
- Chemokinesis : Increased random movement of cells.
- Eosinophil : A type of leukocyte or white blood cell.
- Leukocyte : Generic term for white blood cell. Among the different types of leukocyte are basophils, eosinophils, lymphocytes (of which there are two types, T- and B-), macrophages, monocytes and neutrophils.
- Lymphocyte : A type of leukocyte. B-lymphocytes are responsible for antibody production, T-lymphocytes are responsible for cell-mediated immunity.
- Macrophage : A type of leukocyte. Found in mammalian tissue, derived from monocytes.
- Monocyte : A type of leukocyte. Found in blood, differentiate to form macrophages.
- Neutrophil : The most common type of leukocyte.

E. STRUCTURE VALIDATION

The accuracy of the derived structures was checked using the validation programs PROCHECK[162] and WHATCHECK[163]. Some of the statistics output by these programs are shown in Table E.1. Ramachandran plots for the dimer structure and for both crystal forms of the tetramer are presented in Figures E.1, E.2 and E.3

Structure	Dimer	*m*-Tetramer	*o*-Tetramer
RMSD (bond lengths)	0.025Å	0.016Å	0.015Å
ESU (atomic positions, Cruickshanks DPI)	0.189Å	0.226Å	0.129Å
RMSD (bond angles)	4.4°	2.9°	3.8°
Average Temperature Factor	60.7Å^2	44.4Å^2	32.0Å^2
ESU (temperature factor)	6.1Å^2	5.22Å^2	2.59Å^2
Correlation Coefficient (F_{obs} to F_{calc})	0.95	0.92	0.94
Free Correlation Coefficient (F_{obs} to F_{calc})	0.92	0.89	0.93

Table E.1: Geometric statistics for the refined dimer and tetramer structures.

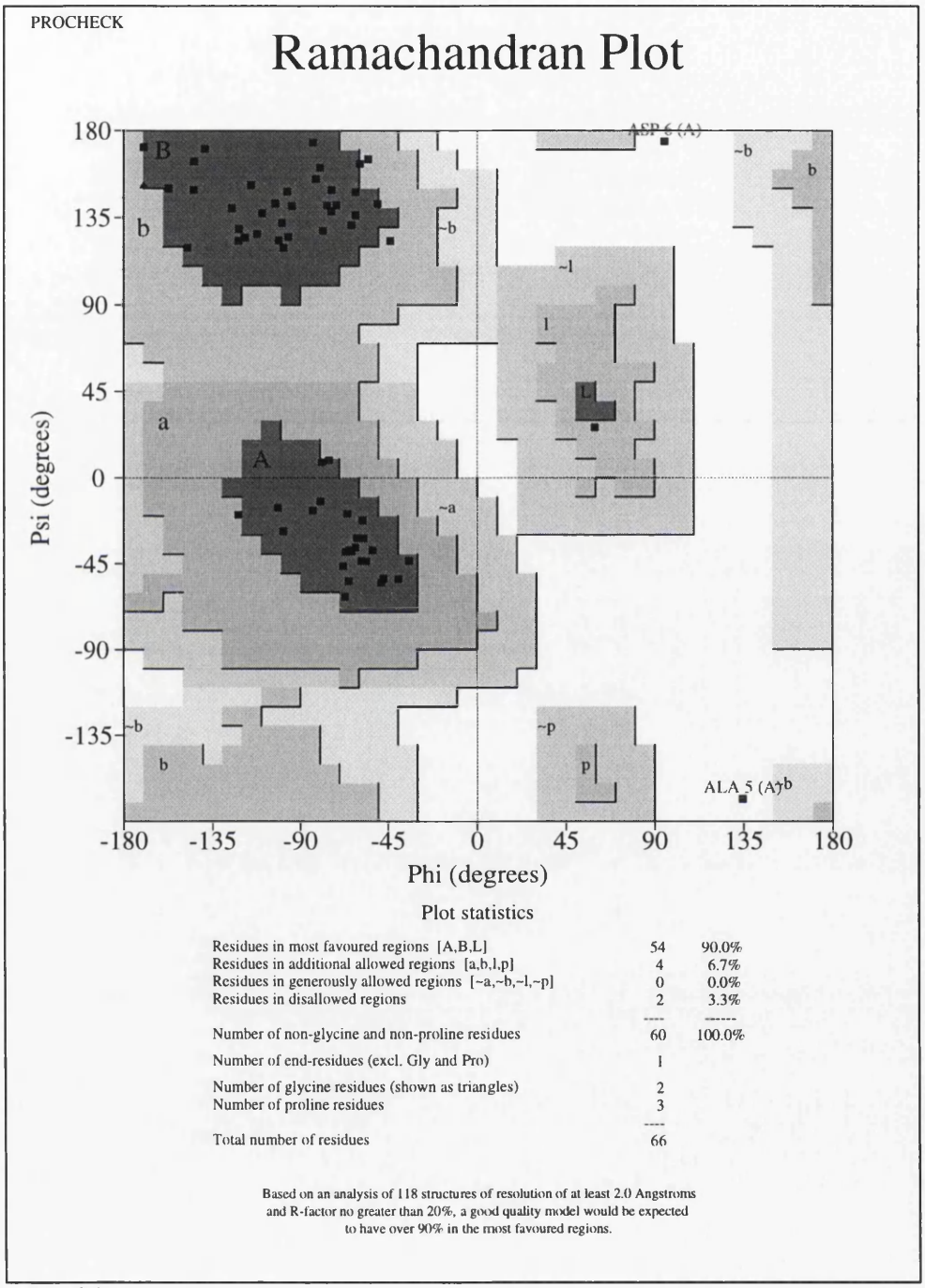
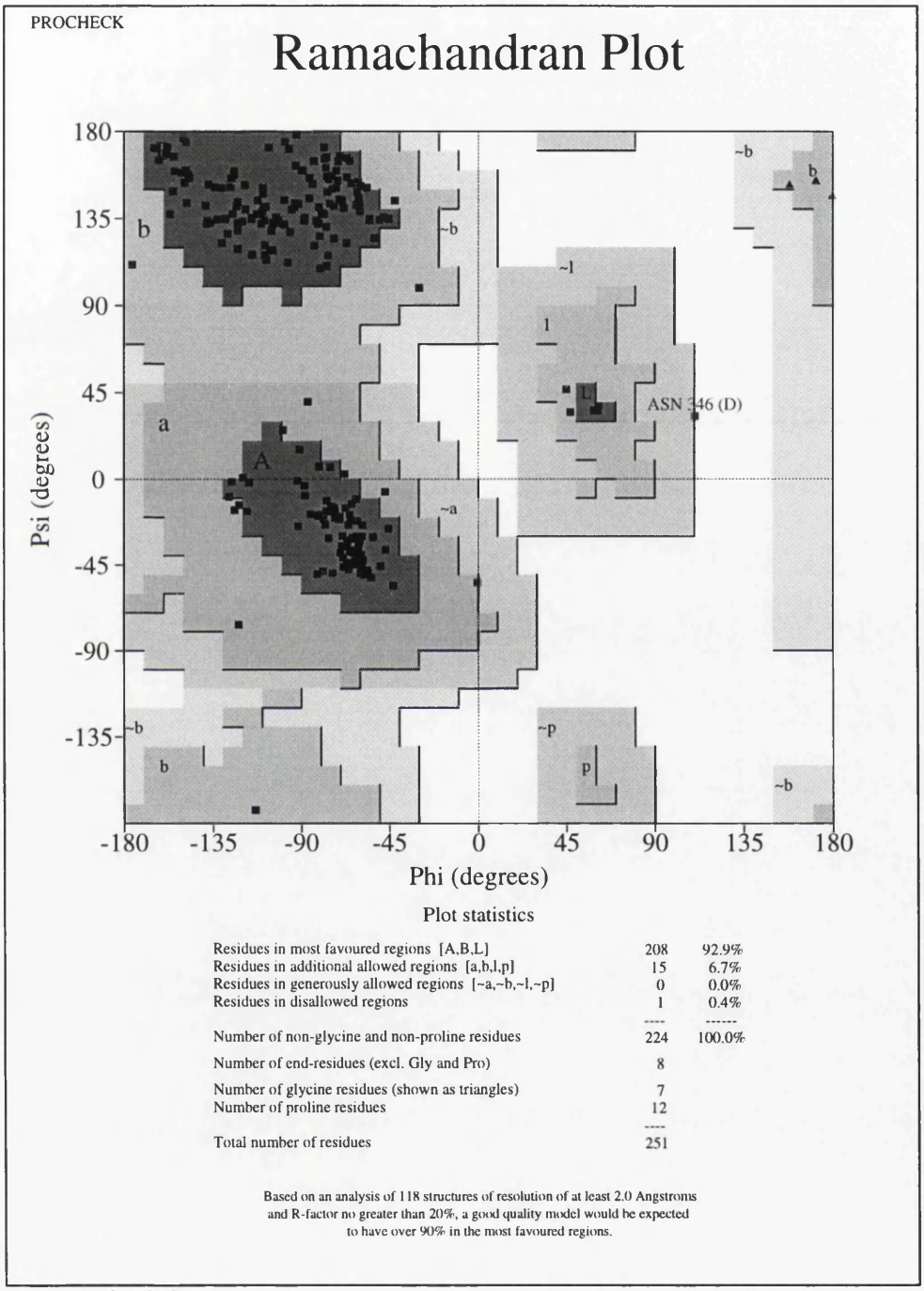
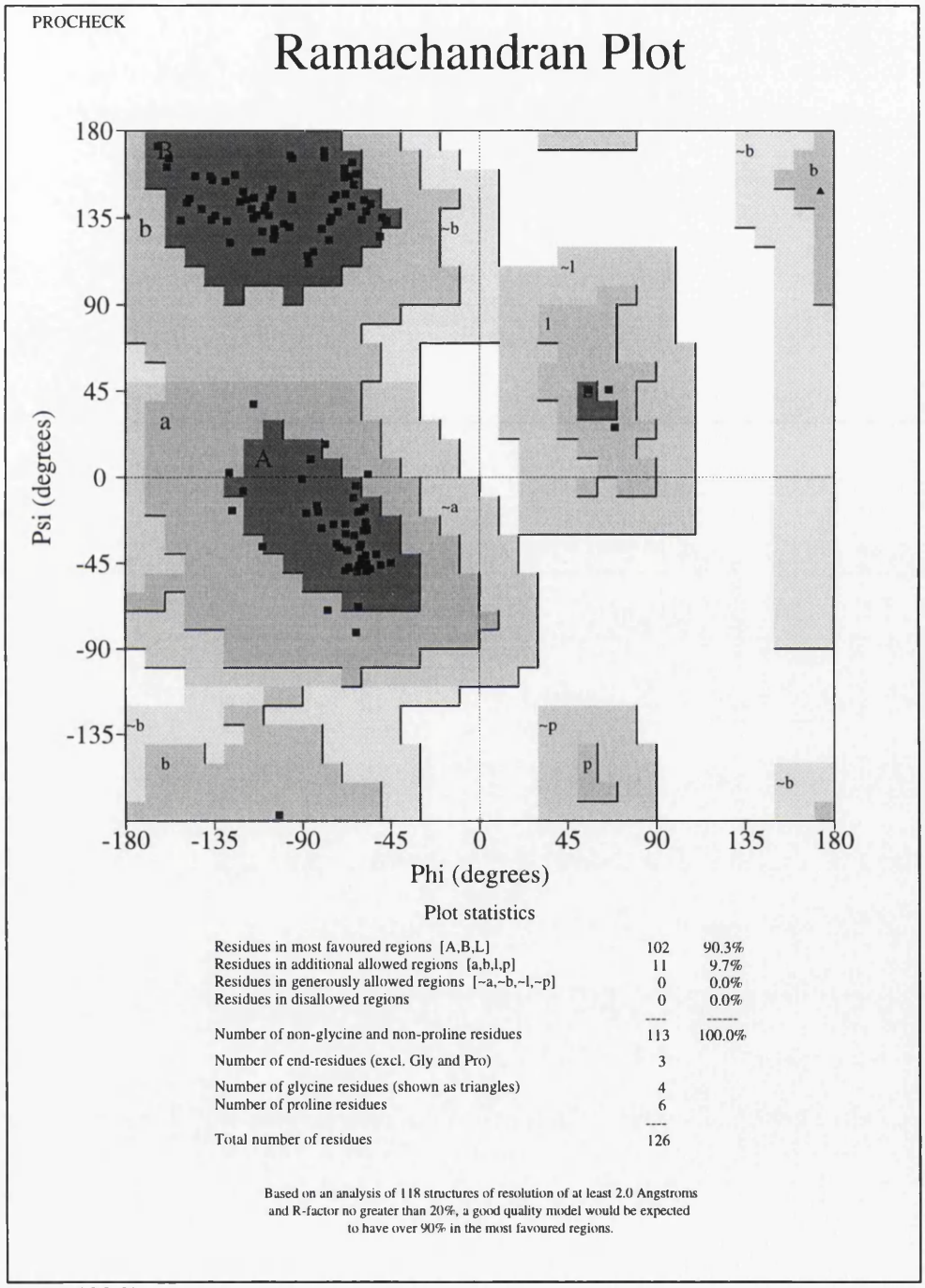


Figure E.1: Ramachandran plot: Dimeric mutant



pm1_mono_m3_o_fit_01.ps

Figure E.2: Ramachandran plot: Tetrameric mutant, monoclinic form



Plot statistics

Residues in most favoured regions [A,B,L]	102	90.3%
Residues in additional allowed regions [a,b,l,p]	11	9.7%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	113	100.0%
Number of end-residues (excl. Gly and Pro)	3	
Number of glycine residues (shown as triangles)	4	
Number of proline residues	6	
Total number of residues	126	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

pm1hr_m4.2.2_01.ps

Figure E.3: Ramachandran plot: Tetrameric mutant, orthorhombic form

References

1. Oppenheim, J. J., Zachariae, C. O. C., Mukaida, N. & Matsushima, K. *Annual Reviews in Immunology* **9**, 617–648 (1991).

2. Miller, M. D. & Krangel, M. S. *Critical Reviews in Immunology* **12**, 17–46 (1992).

3. Baggiolini, M., Ewald, B. & Moser, B. *Advances in Immunology* **55**, 97–179 (1994).

4. Baggiolini, M., Dewald, B. & Moser, B. *Annual Reviews in Immunology* **15**, 675–705 (1997).

5. Pan, Y. *et al. Nature* **387**, 611–617 (1997).

6. Mackay, C. R. *Current Biology* **7**, R384–R386 (1997).

7. Ibelgauft, H. (1994). http://www.lmb.uni-muenchen.de/groups/ibelgaufts/cytokines.html.

8. Davies, D. R. & Wlodawer, A. *FASEB Journal* **9**, 50–56 (1995).

9. Oefner, C., D'Arcy, A., Winkler, F. K., Eggimann, B. & Hosang, M. *EMBO Journal* **11**, 3921–3926 (1992).

10. Lapthorn, A. J. *et al. Nature* **369**, 455–461 (1994).

11. Wu, H., Lustbader, J. W., Liu, Y., Canfield, R. E. & Hendrickson, W. A. *Structure* **2**, 545–558 (1994).

12. Finzel, B. C. *et al. Journal of Molecular Biology* **209**, 779–791 (1989).

13. Zhu, X. *et al. Science* **251**, 90–93 (1991).

14. Eck, M. J. & Sprang, S. R. *Journal of Biological Chemistry* **264**, 17595–17605 (1989).

15. Acharya, K. R., Shapiro, R., Riordan, J. F. & Vallee, B. L. *Proceedings of the National Academy of Science of the USA* **92**, 2949–2953 (1995).

16. Clore, G. M. & Gronenborn, A. M. *FASEB Journal* **9**(1), 57–62 (1995).

17. Kartha, G., Bello, J. & Harker, D. *Nature* **213**, 862 (1967).

18. Wolpe, S. D. & Cerami, A. *FASEB Journal* **3**, 2565–2573 (1989).

19. Sun, P. D. & Davies, D. R. *Annual review of Biophysics and Biomolecular Structure* **24**, 269–291 (1995).

20. Sherry, B., Horii, Y., Manogue, K. R., Widmer, U. & Cerami, A. *Interleukin 8 (NAP-1) and Related Chemotactic Cytokines.*, volume 4, chapter Macrophage Inflammatory Proteins 1 and 2: An Overview, 117–130. Karger, Basel (1992).

21. Ryan, R. J. *et al. Progress in Hormone Research* **43**, 383–429 (1987).

22. Graham, G. J. & Pragnell, I. B. *Developmental Biology* **151**, 377–381 (1992).

23. Kelner, G. S. *et al. Science* **266**, 1395–1399 (1994).

24. Bazan, J. F. *et al. Nature* **385**, 640–646 (1997).

25. Hromas, R. *et al. Blood* **89**(9), 3315–3322 (1997).

26. Corpet, F. *Nucleic Acids Research* **16**(22), 10881–10890 (1988).

27. Henikoff, S. & Henikoff, J. G. *Proceedings of the National Academy of Science of the USA* **89**, 10915–10919 (1992).

28. Howard, O. M. Z., Ben-Baruch, A. & Oppenheim, J. J. *Trends in Biotechnology* **14**, 46–51 (1996).

29. Mohamadzadeh, M., Poltorak, A. N., Bergstresser, P. R., Beutler, B. & Takashima, A. *Journal of Immunology* **156**, 3102–3106 (1996).

30. Orlofsky, A., Berger, M. S. & Prystowsky, M. B. *Cell Regulation* **2**, 403 (1991).

31. Patel, V. P. *et al. Journal of Experimental Medicine* **185**(7), 1163–1172 (1997).

32. Wells, T. N. C. & Peitsch, M. C. *Journal of Leukocyte Biology* **61**, 545–550 (1997).

33. Kennedy, J. *et al. Journal of Immunology* **155**, 203–209 (1995).

34. Godiska, R. *et al. Journal of Experimental Medicine* **185**(9), 1595–1604 (1997).

35. Imai, T. *et al. Journal of Biological Chemistry* **272**, 15036–15042 (1997).

36. Hermodson, M., Schmer, G. & Kurachi, K. *Journal of Biological Chemistry* **252**, 6276–6279 (1977).

37. van Creveld, S. & Paulssen, M. M. *Lancet* **2**, 2242 (1951).

38. Begg, G. S., Pepper, D. S., Chesterman, C. N. & Morgan, F. J. *Biochemistry* **17**, 1739–1744 (1978).

39. Yoshimura, T. *et al. Proceedings of the National Academy of Science of the USA* **84**, 9233–9237 (1987).

40. Luster, A. D., Unkeless, J. C. & Ravetch, J. V. *Nature* **315**, 672–676 (1985).

41. Sugano, S., Stoeckle, M. Y. & Hanafasu, H. *Cell* **49**, 321–328 (1987).

42. Lindley, I. J. D., Westwick, J. & Kunkel, S. L. *Immunology Today* **14**, 24 (1993).

43. Castor, C. W., Miller, J. W. & Walz, D. A. *Proceedings of the National Academy of Science of the USA* **80**, 765–769 (1983).

44. Holt, J. C., Harris, M. E., Lange, E., Henschen, A. & Niewiarowski, S. *Biochemistry* **25**, 1988–1996 (1986).

45. Walz, D. A. & Baggiolini, M. *Biochemical and Biophysical Research Communications* **159**, 969–975 (1989).

46. Richmond, A., Lawson, D. H., Nixon, D. W. & Chawla, R. K. *Cancer Research* **45**, 6390–6394 (1985).

47. Koch, A. E. *et al. Journal of Clinical Investigation* **94**, 1012–1018 (1994).

48. Wuyts, A. *et al. Biochemistry* **36**, 2716–2723 (1997).

49. Smith, J. B. & Herschman, H. R. *Journal of Biological Chemistry* **270**, 16756–16765 (1995).

50. Oberlin, E. *et al. Nature* **382**, 833–835 (1996).

51. Farber, J. M. *Biochemical and Biophysical Research Communications* **192**, 223–230 (1993).

52. Obaru, K., Fukada, M., Maeda, S. & Shimada, K. *Journal of Biochemistry* **99**, 885–894 (1986).

53. Lipes, M. A., Napolitano, M., Jeang, K. T., Chang, N. T. & Leonard, W. J. *Proceedings of the National Academy of Science of the USA* **85**, 9704–9708 (1988).

54. Wolpe, S. D. *et al. Journal of Experimental Medicine* **167**, 570–581 (1988).

55. Sherry, B. *et al. Journal of Experimental Medicine* **168**, 2251–2259 (1988).

56. Yoshimura, T., Robinson, E. A., Tanaka, S., Appella, E. & Leonard, E. J. *Journal of Immunology* **142**, 1956–1962 (1989).

57. Schall, T. J. *et al. Journal of Immunology* **141**, 1018–1025 (1988).

58. van Damme, J., Proost, P., Lenearts, J.-P. & Opdenakker, G. *Journal of Experimental Medicine* **176**, 59–65 (1992).

59. Uguccioni, M. *et al. Journal of Experimental Medicine* **183**, 2379–2384 (1996).

60. Sarafi, M. N., Garcia-Zepeda, E. A., MacLean, J. A., Charo, I. F. & Luster, A. D. *Journal of Experimental Medicine* **185**, 99–109 (1997).

61. Miller, M. D., Hata, S., waal Melefyt, R. D. & Krangel, M. S. *Journal of Immunology* **143**, 2907–2916 (1989).

62. Ponath, P. D. *et al. Journal of Clinical Investigation* **97**, 604–612 (1996).

63. Forssmann, U. *et al. Journal of Experimental Medicine* **185**, 2171–2176 (1997).

64. Schulze-Knappe, P. *et al. Journal of Experimental Medicine* **183**, 295–299 (1996).

65. Nagira, M. *et al. Journal of Biological Chemistry* **272**, 19518–19524 (1997).

66. Yoshida, R. *et al. Journal of Biological Chemistry* **272**(21), 13803–13809 (1997).

67. Adema, G. J. *et al. Nature* **387**, 713–717 (1997).

68. Hedrick, J. A., Helms, A., Gorman, D. & Zlotnik, A. Submitted to EMBL/Genbank/DDJB data banks (Nov 1997).

69. Vicari, A. P. *et al. Immunity* **7**(2), 291–301 (1997).

70. Youn, B.-S. *et al. Journal of Immunology* **155**, 2661–2667 (1995).

71. Cocchi, F. *et al. Science* **270**, 1811–1815 (1995).

72. Doranz, B. J. *et al. Cell* **85**, 1149–1158 (1996).

73. Choe, H. *et al. Cell* **85**, 1135–1148 (1996).

74. Pantaleo, G. *Nature Medicine* **3**(5), 483–486 (1997).

75. Feng, Y., Broder, C. G., Kennedy, P. E. & Berger, E. A. *Science* **172**, 872–877 (1996).

76. Samson, M. *et al. Nature* **382**, 722–725 (1996).

77. Liu, R. *et al. Cell* **86**, 367–377 (1996).

78. Dean, M. *et al. Science* **273**, 1856–1862 (1996).

79. Huang, Y. *et al. Nature Medicine* **2**(11), 1240–1243 (1996).

80. Smith, O. *Nature Medicine* , 372–373 (1997).

81. Krell, T. Personal Communication.

82. Lodi, P. J. *et al. Science* **263**, 1762–1767 (1994).

83. Skelton, N. J., Aspiras, F., Ogez, J. & Schall, T. J. *Biochemistry* **34**, 5329–5342 (1995).

84. Patel, S. R. *et al. Biochemistry* **32**, 5466–5471 (1993).

85. Mantel, C., Kim, Y. J., Cooper, S., Kwon, B. & Broxmeyer, H. E. *Proceedings of the National Academy of Science (USA)* **90**, 2232–2236 (1993).

86. Graham, G. J. *et al. Journal of Biological Chemistry* **269**(7), 4974–4978 (1994).

87. Samson, M., Labbe, O., Mollereau, C., Vassart, G. & Parmentier, M. *Biochemistry* **35**, 3362–3367 (1996).

88. Alkhatib, G. *et al. Science* **272**, 1955–1958 (1996).

89. Deng, H. *et al. Nature* **381**, 661–666 (1996).

90. Dragic, T. *et al. Nature* **381**, 667–673 June (1996).

91. Graham, G. J. & Pragnell, I. B. *European Journal of Cancer* **27**, 952–953 (1991).

92. Chung, C. W., Cooke, R. M., Proudfoot, A. E. I. & Wells, T. N. C. *Biochemistry* **34**, 9307–9314 (1995).

93. Broxmeyer, H. E., Hague, N. L., Sledge, G. W., Rasmussen, H. & Gordon, M. S. *Blood* **86**, 37 (1995).

94. Graham, G. J. *et al. EMBO Journal* **15**(23), 6506–6515 (1996).

95. Laemmli, U. K. *Nature* **227**, 680 (1970).

96. Veesler, S., Marcq, S., Lafont, S., Astier, J. P. & Boistelle, R. *Acta Crystallographica* **D50**, 355–360 (1994).

97. Malkin, A. J. & McPherson, A. *Acta Crystallographica* **D50**, 385–395 (1994).

98. Cudney, B., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. *Acta Crystallographica D* **50**, 414–423 (1994).

99. Jancarik, J. & Kim, S.-H. *Journal of Applied Crystallography* **24**, 409–411 (1991).

100. Blundell, T. I. & Johnson, L. N. *Protein Crystallography.* Academic Press, fourth 1994 edition, (1976).

101. Stout, G. H. & Jensen, L. H. *X-ray Structure Determination : A practical guide.* Wiley-Interscience, (1989).

102. Glusker, J. P. & Trueblood, K. N. *Crystal Structure Analysis : A primer.* Oxford University Press, second edition, (1985).

103. Rhodes, G. *Crystallography made Crystal Clear.* Academic Press Inc., (1993).

104. Collaborative Computing Project, Number 4. *Acta Crystallographica D* **50**, 760–763 (1994).

105. Rossmann, M. G. & Blow, D. M. *Acta Crystallographica* **15**, 24 – 31 (1962).

106. Fairbrother, W. J., Reilly, D., Colby, T. J., Hesselgesser, J. & Horuk, R. *Journal of Molecular Biology* **242**, 252–270 (1994).

107. Charles, R. S., Walz, D. A. & Edwards, B. F. P. *Journal of Biological Chemistry* **264**, 2092–2099 (1989).

108. Zhang, X., Chen, L., Bancroft, D. P., Lai, C. K. & Maione, T. E. *Biochemistry* **33**, 8361–8366 (1994).

109. Lubkowski, J. *et al. Nature Structural Biology* **4**(1), 64–69 (1997).

110. Meunier, S., Bernassau, J.-M., Guillemot, J. C., Ferrara, P. & Darbon, H. *Biochemistry* **36**, 4412–4422 (1997).

111. Kabsch, W. *J. Appl. Crystallogr.* **21**, 67 – 71 (1988).

112. Kabsch, W. *J. Appl. Crystallogr.* **21**, 916 – 924 (1988).

113. Otwinowski, Z. & Minor, W. in *Methods in Enzymology,* (Carter, C. W. & Sweet, R. M., eds), volume 276. Academic Press (1996).

114. Otwinowski, Z. in *Proceedings of the CCP4 Study Weekend January 1992,* (Sawyer, L., Isaacs, N. W. & Bailey, S., eds), 56–62 (Science and Engineering Research Council, Daresbury Laboratory, Warrington, UK, 1992).

115. Leslie, A. G. W. in *Joint CCP4 and ESF-EACMD Newsletter on Protein Crystallography* (Daresbury Laboratory, Warrington, UK, 1992).

116. French, G. S. & Wilson, K. S. *Acta Crystallographica A* **34**, 517–525 (1978).

117. Matthews, B. W. *Journal of Molecular Biology* **33**, 491–497 (1968).

118. Wilmanns, M. & Nilges, M. *Acta Crystallographica D* **52**, 973–982 (1996).

119. Stewart, M. Personal Communication.

120. Sheldrick, G. M. *Acta Crystallographica A* **46**, 467–473 (1990).

121. Otwinowski, Z. in *Proceedings of the CCP4 Study Weekend January 1991,* (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds), 80–85 (Science and Engineering Research Council, Darebury Laboratory, Warrington, UK, 1991).

122. Cowtan, K. in *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34 – 38, (1994).

123. Navaza, J. *Acta Crystallographica A* **46**, 619–620 (1990).

124. Navaza, J. *Acta Crystallographica A* **50**, 157–163 (1994).

125. Kabsch, W. *Acta Crystallographica A* **32**, 522–523 (1976).

126. Turkenburg, J. P. & Dodson, E. J. *Current Opinion in Structural Biology* **6**, 604–610 (1996).

127. Anderson, D. H., Weiss, M. S. & Eisenberg, D. *Acta Crystallographica D* **52**, 469–480 (1996).

128. Murshudov, G. N., Dodson, E. J. & Vagin, A. A. in *Proceedings of the CCP4 Study Weekend January 1996,* (Dodson, E. J., Sawyer, L., Ralph, A. & Bailey, S., eds), 93–104 (Science and Engineering Research Council, Daresbury Laboratory, Warrington, UK, 1996).

129. Brunger, A. T. *Nature* **355**, 472–475 (1992).

130. Brunger, A. T. *Acta Crystallographica D* **49**, 24–36 (1993).

131. Brunger, A. T. *X-PLOR. A system for X-ray Crystallography and NMR.* Yale University Press, (1992).

132. Lamzin, V. S. & Wilson, K. S. *Acta Crystallographica D* **49**, 129–147 (1993).

133. Jones, T. A., Cowan, S., Zou, J.-Y. & Kjeldgaard, M. *Acta Crystallographica A* **47**, 110–119 (1991).

134. Charles, R. S., Ciaglowski, R. E., Walz, D. A. & Edwards, B. F. P. *Journal of Molecular Biology* **176**, 421–423 (1984).

135. Clore, G. M., Appella, E., Yamada, M., Matsushima, K. & Gronenborn, A. *Journal of Biological Chemistry* **264**, 18907–18911 (1989).

136. Clore, G. M., Appella, E., Yamada, M., Matsushima, K. & Gronenborn, A. *Biochemistry* **29**, 1689–1696 (1990).

137. Baldwin, E. T. *et al. Proceedings of the National Academy of Science (USA)* **88**, 502–506 (1991).

138. Malkowski, M. G., Wu, J. Y., Lazar, J. B., Johnson, P. H. & Edwards, B. F. P. *Journal of Biological Chemistry* **270**(13), 7077–7087 (1995).

139. Gronenborn, A. M. & Clore, G. M. *Protein Engineering* **4**(3), 263–269 (1991).

140. Shaw, J. P. *et al. Journal of Molecular Biology* **242**, 589–590 (1994).

141. Kim, K. S., Rajarathnam, K., Clark-Lewis, I. & Sykes, B. D. *FEBS Letters* **395**, 277–282 (1996).

142. Richardson, J. S. in *Advances in Protein Chemistry*, volume 34, 167–339. Academic Press Inc. (1981).

143. Mayo, K. H. & Chen, M.-J. *Biochemistry* **28**, 9469–9478 (1989).

144. Chen, M.-J. & Mayo, K. H. *Biochemistry* **30**, 6402–6411 (1991).

145. Hunter, M. G. *et al. Blood* **86**, 4400–4408 (1995).

146. Shannon, R. D. *Acta Crystallographica A* **32**, 751–767 (1976).

147. Colman, P. M., Weaver, L. H. & Matthews, B. W. *Biochemical and Biophysical Research Communications* **46**, 1999 (1972).

148. Beall, C. J., Mahajan, S., Kuhn, D. E. & Kolattukudy, E. *Biochemical Journal* **313**, 633–640 (1996).

149. Nibbs, R. J. B., Wylie, S. M., Pragnell, I. B. & Graham, G. J. *Journal of Biological Chemistry* **272**, 12495–12504 (1997).

150. Proudfoot, A. E. I. *et al. Journal of Biological Chemistry* **271**(5), 2599–2603 (1996).

151. Pakianathan, D. R., Kuta, E. G., Artis, D. R., Skelton, N. J. & Herbert, C. A. *Biochemistry* **36**, 9642–9648 (1997).

152. Monteclaro, F. S. & Charo, I. F. *Journal of Biological Chemistry* **271**(32), 19084–19092 (1996).

153. Gosling, J. *et al. Proceedings of the National Academy of Science of the USA* **94**, 5061–5066 (1997).

154. Picard, L. *et al. Journal of Virology* **71**(7), 5003–5011 (1997).

155. Tournamille, C. *et al. Journal of Biological Chemistry* **272**, 16274–16280 (1997).

156. Koopmann, W. & Krangel, M. S. *Journal of Biological Chemistry* **272**, 10103–10109 (1997).

157. Mann, D. M., Romm, E. & Mlgliorini, M. *Journal of Biological Chemistry* **269**, 23661–23667 (1994).

158. Zhang, Y. & Rollins, B. J. *Molecular and Cellular Biology* **15**(9), 4851–4855 (1995).

159. Schnitzel, W., Monschein, U. & Besemer, J. *Journal of Leukocyte Biology* **55**, 763–770 (1994).

160. Kierzek, A. M., Wolf, W. M. & Zielenkiewicz, P. *Biophysical Journal* **73**, 571–580 (1997).

161. Rossmann, M. G. & Blow, D. M. *Acta Crystallographica* **15**, 24 (1962).

162. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thorton, J. M. *Journal of Applied Crystallography* (26), 283–291 (1993).

163. Vriend, G. *Journal of Molecular Graphics* **8**, 52–56 (1990).