Towards a Total Chemical Synthesis of Human C-C Motif Chemokine Ligand 2

LEAR, SAM

School of Biological and Biomedical Sciences

Masters by Research (MRes) Thesis

Towards a Total Chemical Synthesis of Human C-C Motif Chemokine Ligand 2

Author:
Sam Lear

Supervisors:
Dr Ehmke Pohl
Dr Steven Cobb

February 2013

Abstract

Chemokines are cell signalling proteins that promote cellular migration. C-C motif chemokine ligand 2 (CCL2) plays an important role in the inflammatory response, and is involved in the pathogenesis of diseases such as asthma, atherosclerosis and cancer. The potential for modulation of inflammatory processes via inhibition of chemokine interactions makes CCL2 an attractive target for the development of novel anti-inflammatories. In order to characterise these interactions – and the more general role of chemokines within living systems – labelled chemokines are needed as tools for use in biomedical research. Total chemical synthesis has previously been used to produce site-specifically modified proteins (including chemokines) incorporating a wide range of unnatural alterations such as fluorophores and posttranslational modifications. Progress made towards the total chemical synthesis of human CCL2 via Fmoc solid-phase peptide synthesis (SPPS) and native chemical ligation (NCL) is described.

The microwave-assisted solid-phase synthesis of CCL2 as a single peptide chain was found to be problematic due to hydrophobic aggregation, and attempts to avoid this using the poly(ethylene glycol) (PEG)-based ChemMatrix® support (in combination with pseudoproline dipeptides) are reported. Progress has also been made towards a total chemical synthesis utilizing NCL, and syntheses for two of the required fragments – the C-terminal peptide acid CCL2(52-76) and a peptide thioester CCL2(12-35) – are presented. While SPPS of the thioester was achieved using the sulfonamide safety-catch linker, the use of the Dawson Dbz linker for the synthesis of peptide thioesters was also investigated, and the results given here represent the first reported use of the Dbz linker for microwave-assisted SPPS.

In addition, the structure determination of three analogues of a natural product found to inhibit CCL2-mediated monocyte chemotaxis has been carried out using single crystal X-ray diffraction data, thus contributing to ongoing research aiming to elucidate the role of chemokine interactions in biological systems.

Contents

Acknowledgements

First and foremost I would like to thank my supervisors Dr Ehmke Pohl and Dr Steven Cobb for their continued guidance and support. In addition I am grateful for assistance from Ian Edwards in the development of the peptide purification protocols given, as well as the invaluable advice of Dr Olga Chetina, without which the small molecule crystallization work described would not have been possible. Thanks also go to Natalie Tatum and Julie Thye for their help with X-ray data collection and peptide purification respectively, as well as Dr Oleg Dolomanov and Dr Dmitry Yufit for their assistance with the Olex2 software. The initial microwave synthesis trialling the use of the Dbz linker was carried out by Andrew Steer. In addition I would like to thank the mass spectrometry service in the Department of Chemistry for carrying out the characterisation of peptide samples. Finally thanks go to Dr Chris Mason of CEM for advice and assistance with the Liberty1 system, Dr Neil Colgin of Cambridge Research Biochemicals for advice regarding SPPS and microwave synthesis in general, and Dr Nathaniel Martin (Utrecht University) for useful discussions on the chemical synthesis of whole proteins.

Abbreviations

Acm	Acetamidomethyl
Alloc	Allyloxycarbonyl
AM	Aminomethyl
Boc	*tert*-Butoxycarbonyl
CAM	Cell adhesion molecule
CCL2	C-C motif chemokine ligand 2
CD	Circular dichroism
CM	ChemMatrix®
Dbz	3,4-Diaminobenzoic acid
DCM	Dichloromethane
DIPEA	*N*,*N*-Diisopropylethylamine
DKP	Diketopiperazine
DMF	*N*,*N*-Dimethylformamide
DSSP	Define Secondary Structure of Proteins
ERK1/2	Extracellular signal-regulated protein kinases 1 and 2
Fmoc	Fluorenylmethyloxycarbonyl
FPLC	Fast protein liquid chromatography
FRET	Förster resonance energy transfer
GAG	Glycosaminoglycan
Glp	Pyroglutamic acid
GPCR	G protein-coupled receptor
HATU	1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate
HBTU	*O*-(Benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate
HMPB	4-(4-Hydroxymethyl-3-methoxyphenoxy)-butyric acid
HPLC	High-performance liquid chromatography
IMAC	Immobilized metal ion affinity chromatography

ITC	Isothermal titration calorimetry
KCL	Kinetic chemical ligation
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
MeCN	Acetonitrile
MeOH	Methanol
MESNa	2-Mercaptoethanesulfonate sodium salt
MPAA	(4-Carboxymethyl)thiophenol
MCP-1	Monocyte chemoattractant protein-1
MS	Mass spectrometry
MW	Microwave
Nbz	*N*-acyl-benzimidazolinone
NCL	Native chemical ligation
NMP	1-Methyl-2-pyrrolidinone
NMR	Nuclear magnetic resonance
PA	Polyamide
Pbf	Pentamethyl-2,3-dihydrobenzofuran-5-sulfonyl
PEG	Poly(ethylene glycol)
PTFE	Polytetrafluoroethylene
PTM	Posttranslational modification
PyBOP®	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
RP	Reversed-phase
Rt	Room temperature
SEC	Size-exclusion chromatography
SPPS	Solid-phase peptide synthesis
SPS	Solid-phase synthesis
Tbfmoc	Tetrabenzo[a,c,g,i]fluorenyl-17-methoxycarbonyl
TBTU	*O*-(Benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium tetrafluoroborate
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
Thz	1,3-Thiazolidine-4-carboxyl
TIPS	Triisopropylsilane
TMS	Trimethylsilyl
TNBS	2,4,6-Trinitrobenzenesulfonic acid
Trt	Triphenylmethyl

Chapter 1

Introduction

A complete grasp of the structure and interactions of a particular protein within a biological system is key to understanding its function. A protein drug target cannot be truly exploited, for example, without considering the nature of its interactions with other elements of the signalling cascade or metabolic pathway. By elucidating the three-dimensional structure of a biomolecule and its sites of interaction with other proteins or ligands (using techniques such as X-ray diffraction and protein NMR), molecules can be designed to manipulate these interactions. Protein engineering is of crucial importance in the study of proteins, as often it is only when a protein can be synthesized artificially – thus modified or labelled versions can be produced – that its role can be thoroughly characterised, indeed at the very least this may provide large enough quantities for extensive biological assays or crystallographic analysis. Techniques such as site-directed mutagenesis are central to biotechnology, and organisms such as bacteria or fungi are classically used for the generation of modified proteins for research. In contrast, the production of proteins via *chemical* synthesis is an expanding field that provides a route to 'customized' proteins with greater control and the potential for a broader range of modifications.

The total chemical synthesis of whole proteins offers a number of advantages over conventional biosynthetic routes. While recombinant DNA-based techniques utilizing expression systems such as *Escherichia coli* offer some control over the primary sequence of the protein being produced,[1] a total chemical synthesis allows the site-specific incorporation of any number of unnatural amino acids, labels (such as fluorophores[2] or NMR probes[3]), or any other suitably reactive chemical group into the target protein. An application of particular interest is the selective generation of synthetic proteins

with posttranslational modifications (PTMs), such as phosphorylations[4] and glycosylations.[5] Furthermore, proteins of very high purity are accessible via chemical synthesis, circumventing issues with conventional methods such as endotoxin contamination.[6] Despite the area of therapeutic glycoproteins being the fastest growing field in the pharmaceuticals industry,[7] glycosylated proteins are not recombinantly expressed routinely in bacterial systems such as *E. coli*,[8] and the variability of glycosylation sites for recombinant protein can render the successful purification of a desired glycoform impossible.[7,9] Site-specifically glycosylated proteins can be obtained via chemical synthesis however. Total chemical synthesis also allows the generation of racemic proteins, potentially allowing more facile crystallization for X-ray diffraction studies.[10]

Solid-phase peptide synthesis (SPPS) is a mature field, and peptides with lengths of around 50 residues can now be routinely produced in high yield using microwave-assisted solid-phase techniques and highly efficient coupling reagents (*vide infra*). Linear syntheses of longer peptides can still be problematic however, and a number of ligation strategies have been proposed which allow access to larger peptides and proteins.[11,12] Perhaps the most influential of these is native chemical ligation, involving the condensation of a peptide thioester and a second peptide containing an N-terminal cysteine, with the subsequent S→N acyl shift generating a native amide bond (*vide infra*).[13]

The chemokine CCL2 is a target of interest due to the role it plays in a number of immune and inflammatory disease states including asthma, rheumatoid arthritis, atherosclerosis and cancer.[14] The 76 residue protein has previously been shown to be accessible via chemical synthesis,[15–19] allowing the creation of analogues bearing site-specific modifications.[20] The following sections will review the target protein and explore various general aspects of the chemical synthesis of whole proteins, concluding with a consideration of the structure of the CCL2 molecule and its bearing on a potential synthetic route.

1.1 Human C-C motif chemokine ligand 2 (CCL2)

Inflammation in response to damage of vascularized tissue is a complex process involving a large variety of cell types and signaling molecules.[21]

One particular reaction to injury (whether due to physical factors or infection with a pathogen) is the movement of leukocytes from blood vessels into the surrounding tissues in a process known as *extravasation*. The movement (or *chemotaxis*) of leukocytes is promoted by so-called chemotactic cytokines, or *chemokines*.

Chemokines, of which there are approximately 40 distinct types, act by binding to G protein-coupled receptors (GPCRs) expressed on the surface of leukocytes. A particular chemokine may bind multiple receptors – which in turn are capable of binding many different chemokines – leading to a highly complex code of chemotactic cues.[21]

1.1.1 CCL2 and extravasation of monocytes

One such chemokine – known as C-C motif chemokine ligand 2 (CCL2)* – (Figure 1.1) has a role in the extravasation of *monocytes* (a particular class of leukocyte). Monocytes can differentiate into macrophages or dendritic cells, and play an important role in the response at a site of infection.[21]

In order to move from the blood into the tissues, monocytes must first interact weakly with the apical face of epithelial cells comprising the blood vessel walls. This interaction is mediated by *selectins*, a family of cell adhesion molecules (CAMs) present on the endothelium and on the surface of leukocytes, and results in rolling of the cells along the endothelial layer.[25] The next stage involves chemokine binding, and it is this that triggers both firmer adhesion and the necessary cytoskeletal changes needed for extravasation to occur.[25] The process is illustrated in Figure 1.2.

In addition to receptor binding, CCL2 also binds to heavily sulfated (and thus negatively charged) polysaccharides present on the endothelium known as glycosaminoglycans (GAGs).[21] This is thought to be important in regulating local chemokine concentration, and for establishing a chemokine concentration gradient that directs monocyte movement.[27] It has been suggested that chemokine oligomerization is essential for this process to occur.[28]

*Also known as monocyte chemoattractant protein-1 (MCP-1). The nomenclature relates to the arrangement of two highly conserved Cys residues close to the protein N-terminus.[22]

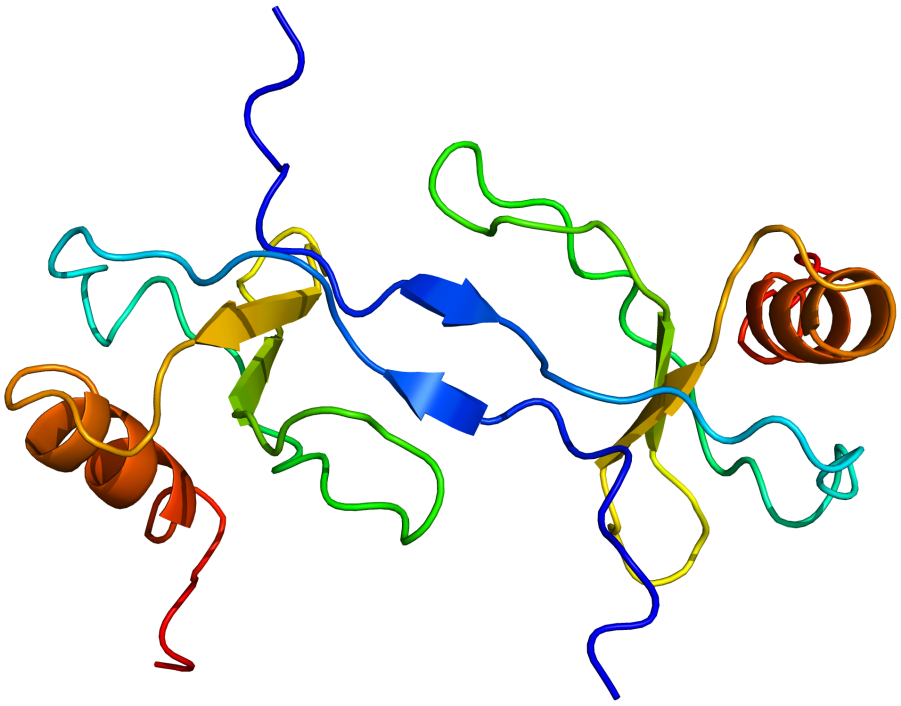


Figure 1.1: Ribbon representation of the CCL2 dimer (NMR solution structure).[23,24]

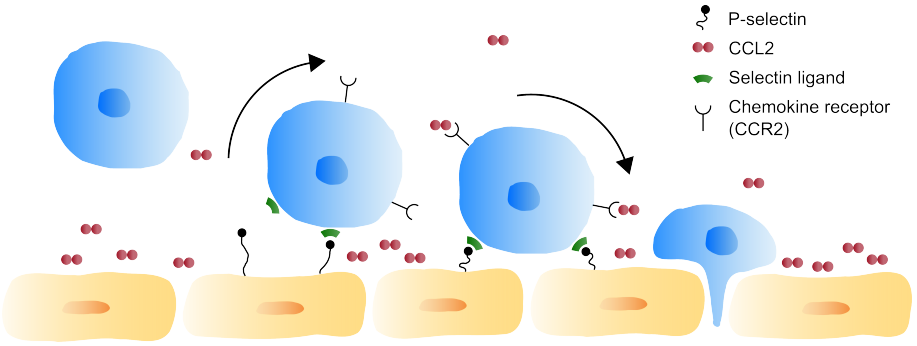


Figure 1.2: Stages of leukocyte extravasation: the moving cell initially rolls along the surface of the endothelium (a process mediated by selectin interactions) before adhering more strongly upon chemokine binding. Cell adhesion (resulting from integrin/CAM interactions, not shown) leads to transmigration of the leukocyte through the endothelium. The role of chemokine dimerization during this process has not yet been clearly elucidated.[26] Adapted from Lodish et al.[21]

1.1.2 Therapeutic targeting of chemokine interactions

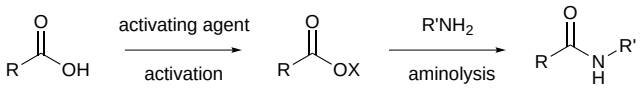
The myriad of chemokines and chemokine receptors – and their crucial role in various biological processes and disease states – makes chemokine interactions an attractive target for the development of new classes of drugs.[29] Variations in cell migratory behaviour can have dramatic effects on a living system – indeed a link has been proposed between cytokine-mediated chemotaxis and breast cancer metastasis[30,31] – and the inhibition of CCL2 protein-protein interactions (whether at the dimerization interface or at receptor/GAG interaction sites) may be used to modulate inflammatory processes for example. This would provide a new target for drug development, potentially offering a route towards a novel class of anti-inflammatories.[22] A number of small molecule chemotaxis inhibitors are the focus of Chapter 5, while the intervening chapters deal mainly with the techniques used to produce synthetic proteins (and hence tools with which chemokine interactions can be studied). The following sections introduce the area of peptide synthesis and its application to the total chemical synthesis of whole proteins.

1.2 Chemical synthesis of proteins

1.2.1 Solid-phase peptide synthesis (SPPS)

Peptide synthesis in general involves the formation of an amide bond between adjacent amino acids in a peptide sequence. The high temperatures needed to react an amine with an unmodified carboxylic acid are generally incompatible with the other reactive groups present on a peptide, and so a number of activation strategies have been developed for efficient, high-yielding amide bond formation.[32] Many of these involve the *in situ* formation of a reactive ester intermediate from the carboxyl terminus of the peptide, which subsequently undergoes aminolysis upon treatment with the next amino acid to form a new amide bond. This process is shown in Scheme 1.1.

Highly efficient amide bond formation is an essential aspect of peptide synthesis, as even small yield losses (0.5-1%) at each step result in increasingly low overall yields as the length of the target peptide increases. Despite the development of effective coupling reagents however, the solution-phase approach suffers from two major drawbacks. The first issue is the need to



Scheme 1.1: General scheme for amino acid coupling via active ester formation.[32]

isolate and purify the coupling products at each step, which can be time consuming. In addition, the inherent insolubility of the side chain-protected peptide intermediates is often problematic at higher sequence lengths.[13]

The development of functionalized polystyrene resins by Bruce Merrifield in the 1960s was a turning point for peptide synthesis.[33] These resins (in the form of insoluble beads decorated with reactive groups) act as a solid support upon which a peptide can be anchored and elongated, enhancing solubility and obviating the need for purification beyond simply washing the resin after each step. When used in conjunction with efficient coupling reagents, solid-phase strategies allow the synthesis of longer peptides than would be accessible in the solution-phase, and peptides of greater than 100 residues can be routinely produced using solid-phase peptide synthesis.[34]

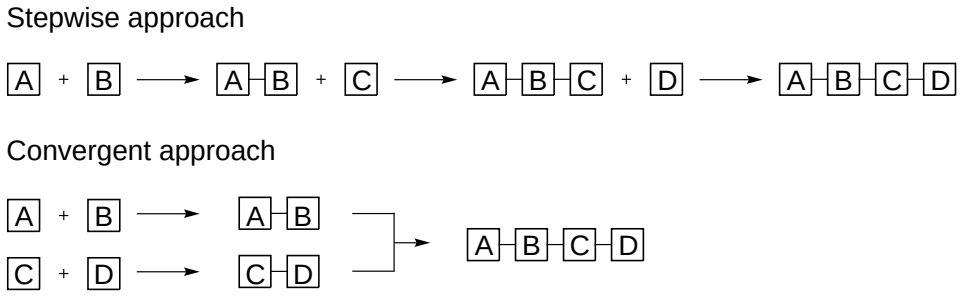
Microwave assisted SPPS has been demonstrated to produce peptides and proteins with dramatically reduced coupling times and improved crude product purity when compared to conventional room temperature syntheses.[35] The technique has proved especially useful for 'difficult' sequences, such as those liable to form β-sheet type structures during synthesis (*vide infra*). While it has been suggested that there may be some peptide-specific microwave effect that causes these beneficial effects – for example direct interaction of the amide bond dipole with the electric field, leading to disruption of hydrophobic aggregation (*vide infra*)[36] – it is generally accepted that they are simply due to the thermal effects also encountered during conventional heating.[37]

1.2.2 Convergent ligation: Accessing whole proteins

Despite becoming a commonplace technique for the synthesis of long peptides and small proteins, conventional stepwise SPPS also suffers limitations on target peptide length and can be problematic where long or particularly troublesome protein sequences are desired (*vide infra*). Poor solubility of the growing peptide chain is a major problem at higher sequence lengths, and can result in low yields due to decreased availability of the reacting peptide N-terminus.[38] Another issue is the presence of deletion sequences arising

from coupling steps that have proceeded in less than quantitative yield. While these inevitable side products are often unproblematic with shorter targets, the impurities generated increase with sequence length and can be difficult to separate at the end of the synthesis due to their similarity to the target peptide.

As an alternative to the stepwise attachment of single amino acids onto the growing peptide N-terminus, a so-called 'convergent approach' should allow the design of a synthesis that maximizes the difference between ligation product and impurities at each step (Scheme 1.2).[39] Furthermore, with the application of ligation chemistry that is tolerant of side chain groups, such an approach could make use of *unprotected* peptide intermediates which are freely soluble at millimolar concentrations in the presence of chaotropes[†].[13]



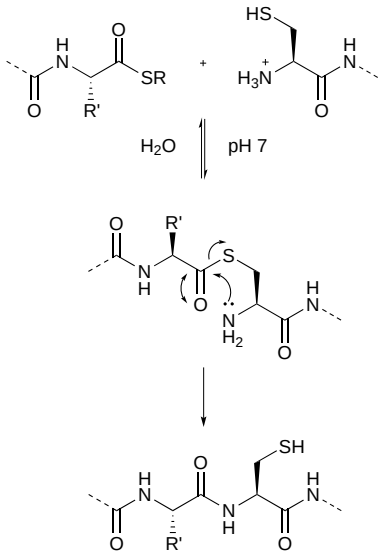
Scheme 1.2: A convergent approach maximizes the difference between product and impurities (unreacted starting materials) at each step, and allows reactions to be carried out in parallel.

1.2.3 Native chemical ligation (NCL)

A particularly important tool with regards to the synthesis of large peptides and proteins is 'native chemical ligation' (NCL). In this approach the side chain thiol of an N-terminal cysteine of one peptide segment is reacted with a C-terminal thioester formed on the other. A subsequent S→N acyl shift then generates a peptide bond between the segments that is indistinguishable from a 'native' amide linkage (Scheme 1.3). The sulfur chemistry utilized in this method is compatible with the presence of reactive groups present elsewhere

[†]Additives that increase the solubility of the peptide chain, for example aqueous 6 M guanidine·HCl.

on the peptide segments (thioesters are more stable to hydroxide-catalysed hydrolysis than their oxygen counterparts, while being more reactive towards thiolysis/aminolysis), and so fully deprotected peptide intermediates can be used.[13] Some specific considerations are detailed in the following sections.



Scheme 1.3: General mechanism for native chemical ligation.[13]

Thioester reactivity and catalysis

Conventional NCL typically makes use of alkyl thioesters which are relatively unreactive, therefore requiring the presence of a catalyst with which to form a more reactive thioester species in situ before ligation can occur. Commonly used catalysts include a 1% benzyl mercaptan (**1**) / 3% thiophenol (**2**) mix, or 2-mercaptoethanesulfonate sodium salt (MESNa, **3**) for recombinant peptide-thioesters (Figure 1.3).[40] Another catalyst, (4-carboxymethyl)thiophenol (MPAA, **4**), has also been proposed, being more effective and with a greater water solubility and less offensive odour than those currently used.[40]

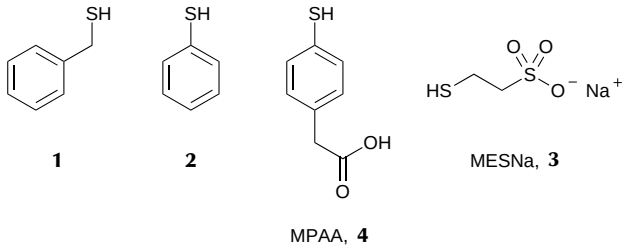


Figure 1.3: A selection of catalysts used in NCL.

Kinetic chemical ligation (KCL)

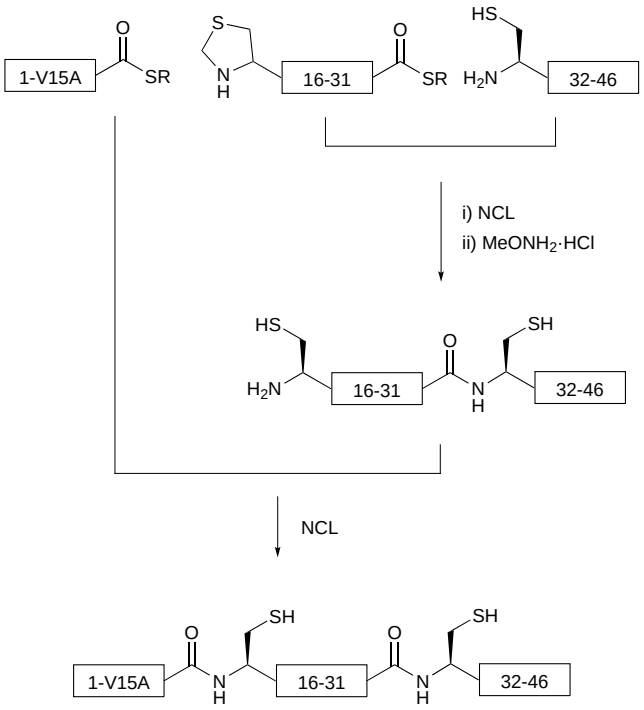
Another approach is to generate and purify reactive aryl thioesters separately, for subsequent use in a ligation. This can be achieved via treatment of alkyl precursors with an aryl thiol such as 4-mercaptophenylacetic acid.[41] These preformed 'active' thioesters will react preferentially over alkyl thioesters in a multi-component ligation where catalyst is *not* present, and can therefore be used for ligations in the N- to C-terminal direction.[42] This is known as 'kinetically controlled ligation' (KCL), and thioester segments with an unmasked N-terminal Cys can be used in this approach without head-to-tail cyclization occurring.

As an alternative to kinetic control, Shigenaga et al. used peptide thio*acids* as 'masked' thioesters in order to carry out NCL in the N- to C-terminal direction.[43] Conversion of the thioacids to the corresponding thioesters and subsequent NCL was accomplished in a one-pot matter.

N-terminal Cys thiol protecting groups

For stepwise ligations occurring in the C- to N-terminal direction, controlling thioester reactivity is not an issue. Instead N-terminal Cys thiols must be protected appropriately. Conventional approaches make use of the acetamidomethyl (Acm) protecting group.[44] The incorporation of a masked Cys residue in the form of a thiazolidine has been demonstrated however, with deprotection proving to be quicker, higher yielding and with fewer manipulations required.[45] The 1,3-thiazolidine-4-carboxyl (Thz) can be introduced easily during SPPS using Boc-L-thiazolidine-4-carboxylic acid and unmasked using 0.2 M methoxyamine hydrochloride (which has been used

without damage to the thioester).[42] Thz was utilized by Bang et al. in their synthesis of the protein crambin.[44,45] The synthesis is outlined in Scheme 1.4.



Scheme 1.4: Overview of the total chemical synthesis of [V15A]crambin.[45]

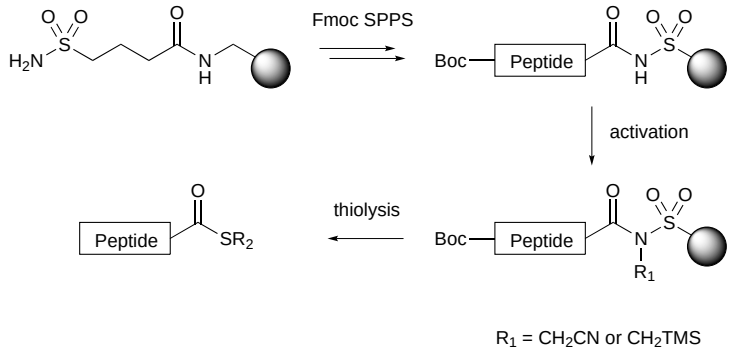
Photolabile protecting groups have also been used in native chemical ligations. A phenacyl-type protecting group was introduced by Ueda et al. in 2006, and was used successfully in a three-component one-pot NCL.[46]

Thioester formation

While a variety of methods exist for the formation of peptide C-terminal thioesters,[47] potentially the most convenient approach is to use a so-called 'safety-catch' linker compatible with Fmoc SPPS.‡ The *sulfonamide safety-catch*[48–50] is perhaps the best established technique for the solid-phase synthesis of C-terminal peptide thioesters whereby the required peptide

‡SPPS utilizing amino acids N-terminally protected with the base-labile fluorenylmethyloxycarbonyl (Fmoc) protecting group.

is synthesized on sulfamylbutyryl resin,[§] with subsequent alkylation (using iodoacetonitrile or trimethylsilyldiazomethane) and thiolysis resulting in release of the thioester product (Scheme 1.5). This linker has been used to produce peptide thioesters of up to 35 amino acids in length.[47]



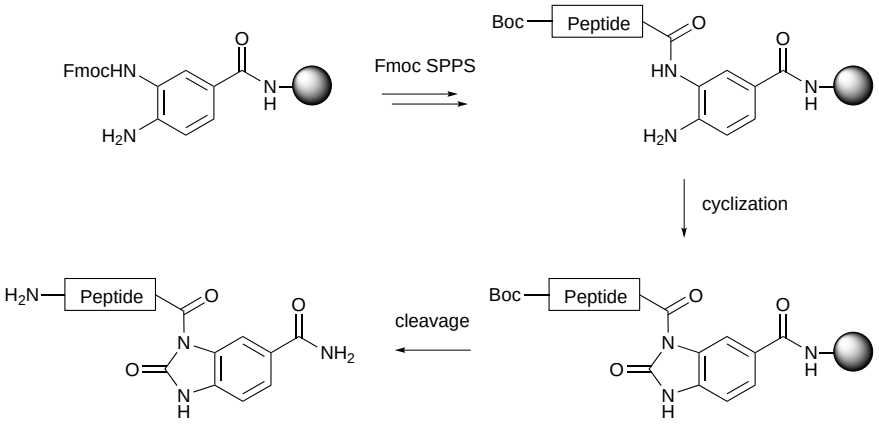
Scheme 1.5: Thioester formation using the sulfonamide safety-catch linker.[47]

Potential problems associated with this method include racemization of the C-terminal amino acid during lengthy loading reactions, aggregation of the protected peptide, and side reactions during alkylation (minimized by using TMS-CHN_2).[47] Ingenito et al. achieved high loadings with low levels of epimerization through the use of amino acid fluorides for the initial linker acylation.[51] A significant increase in the yield of isolated thioester was reported by Quaderer et al. when using 2 M LiBr/THF in place of DMF or THF as the solvent during the thioesterification step.[52] More recently the sulfonamide linker was attached to an acid-labile resin and used to prepare peptide *N*-alkylsulfonamides which could be used directly in ligation reactions to which thiol was added.[53]

As an alternative to the sulfonamide safety-catch, the Dawson Dbz linker can be used.[54] This is a safety-catch linker for Fmoc SPPS that generates *N*-acylurea peptides after treatment with *p*-nitrophenylchloroformate and cleavage from the resin (see Section 3.2). The process is summarized in Scheme 1.6. The *N*-acylurea peptides produced can be converted to a thioesters in solution via treatment with thiol, or used directly for NCL.[54] The *N*-acylurea moiety is also relatively stable to hydrolysis during storage,

[§]Available preloaded from Merk.

and the products can be subjected to the acidic conditions often used during HPLC purification (*vide infra*).[54]



Scheme 1.6: *N*-acylurea formation using the Dbz linker.[54]

1.3 CCL2: Synthetic considerations

1.3.1 Structure and residues of significance

With respect to probe incorporation, three major interactions should be considered:

- Receptor binding
- Glycosaminoglycan binding
- Chemokine oligomerization (and relevance to binding)

Although there are no receptor- or GAG-bound CCL2 crystal structures, a number of studies have shown the importance of certain residues for receptor binding and activity *in vitro*,[19,55–58] and GAG binding *in vitro*[28] and *in vivo*[28,59] (summarized in Table 1.1).

Figure 1.1 shows a ribbon diagram of the CCL2 homodimer in solution. A number of residues on the C-terminal α-helical domain are thought to be important for GAG binding,[28,59] while modification of the N-terminus has been shown to be detrimental receptor binding and activity.[19] Modification of the N-terminal β-sheet domain may also be detrimental to activity as this region is involved in H-bonding in the dimer (see Figure 1.1), therefore hindrance may inhibit oligomerization and hence GAG binding ability (*vide supra*).

A probe should ideally be incorporated at a residue on the chemokine surface, away from the proposed external binding and interaction sites involved in these processes and also avoiding internal folding interactions. The suitability of a potential site can be examined by consulting published crystal structures[15,61] or NMR solution structures[23] to confirm the spatial proximity of neighbouring residues. A representation of the CCL2 dimer is shown in Figure 1.4, with biologically significant residues (listed in Table 1.1) coloured red on one subunit (for the other the solvent-excluded surface is shown). The high number of important residues towards the N-terminus, and their role as part of the dimer interface, suggests that this would not be a suitable region for probe incorporation. A more appropriate position would be on the C-terminus – away from residues crucial for activity and those involved in dimerization.

The protein primary sequence is given in Figure 1.5. Cys residues are highlighted. A synthesis using NCL was developed by Grygiel et al. (*vide*

Table 1.1: Residues of biological significance in the CCL2 primary sequence.

Residue number	*Amino acid(s)*	*Biological significance* (*experiment*)
1	Gln/Glp[a]	Receptor binding and activity[b] *in vitro* (diminished upon deletion/extension)[19]
2-8	PDAINAP	Receptor binding and activity *in vitro*[19,55,56]
8	Pro	Oligomerization (some evidence) *in vivo*[28]
10-13	TCCY	Activity *in vitro*[57]
13	Tyr	Receptor binding and activity *in vitro*[28,55,58]
18	Arg	GAG binding *in vitro*[28]
19	Lys	GAG binding *in vitro*[28]
24	Arg	Receptor binding and activity *in vitro*,[56,58] GAG binding *in vitro*[28]
28	Tyr	Activity *in vitro*[56]
30	Arg	Activity *in vitro*[56]
34	Ser	Activity *in vitro*[57]
35	Lys	Receptor binding and activity *in vitro*[57,58]
38	Lys	Receptor binding[58]
49	Lys	Receptor binding,[58] GAG binding *in vitro*[28]
58	Lys	GAG binding ***in vivo***[c]
66	His	GAG binding ***in vivo***[c]
68	Asp	Activity *in vitro*[56]

[a] *Pyroglutamic acid*. The N-terminal Gln can cyclize to form Glp in the mature protein.[16,60] Grygiel et al. incorporate this as a Glp residue directly in their total chemical synthesis of CCL2 (*vide infra*).

[b] 'Activity' refers to *in vitro* cell chemotactic activity.

[c] Results of an *in vivo* study (intraperitoneal recruitment assay) - these residues may be more essential for function than *in vitro* binding/activity studies suggest.[28,59]

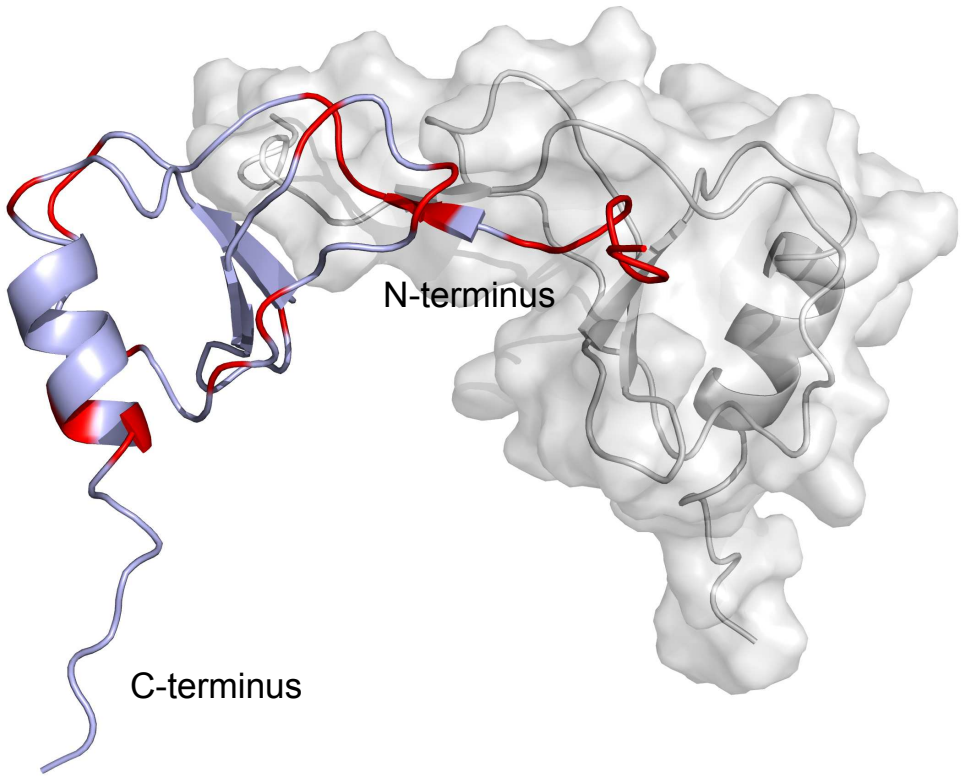


Figure 1.4: CCL2 dimer with biologically important residues listed in Table 1.1 coloured red on one subunit (the N- and C-termini are also labelled).[23,24] For the other subunit the solvent-excluded surface[62] is shown.

infra) involving ligation at the central cysteine. Development of a 3- or 4-segment NCL route may be useful as this would potentially allow the rapid generation of chemokines labeled with different probe combinations.

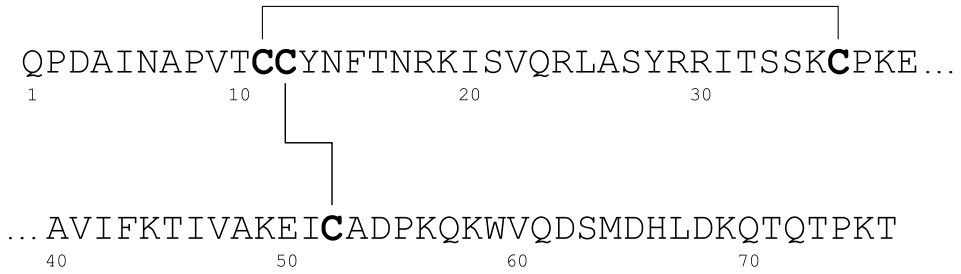


Figure 1.5: CCL2 sequence showing disulphide bond connectivity.[15]

1.3.2 Previous total syntheses

Some total syntheses of CCL2 have already been published and are summarized in Table 1.2. Although stepwise Fmoc SPPS approaches are described, none of the syntheses published to date have used microwave assistance. The general trend away from the use of Boc SPPS as a method of chemically synthesizing whole proteins is due to the tendency of the harsh conditions involved to degrade peptides (repeated TFA treatment), and the risks associated with handling liquid hydrogen fluoride.[63] Of the papers which utilize stepwise Fmoc SPPS, only Kruszynski et al. give a yield for synthesis of the linear peptide (30%). The group used affinity chromatography during purification however, which required the use of a monoclonal antibody and is hence a relatively expensive technique when compared to more commonly used peptide purification methods such as RP-HPLC (Section 4.2). It was decided that the approach explored for chemokine generation should be as general as possible and with a low cost, therefore affinity chromatography was deemed to be unsuitable for this purpose. The highest yield (60%) was reported by Grygiel et al., who employed a two-segment NCL approach utilizing pseudoproline dipeptides (Section 2.3.2). A linear stepwise synthesis is conceptually simpler however, allowing the more rapid generation of chemokines and chemokine analogues. Advances in microwave SPPS means

that attempting this is now feasible, and it is for this reason that a stepwise approach was explored initially (Chapter 2).

Table 1.2: Previous total syntheses of CCL2.

Authors	*Year*	*Approach*	*Yield*[a] /%	*Additional details*
Grygiel et al.[15]	2010	NCL (segments synthesized via Fmoc SPPS)	60	Pseudoprolines[b] were incorporated to disrupt β-sheet structure and improve solvation (leading to increased yield).
Kruszynski et al.[16]	2006	Stepwise Fmoc SPPS[c]	30	Affinity chromatography was used for purification (in addition to HPLC). The method was used to produce site-specific biotinylated analogues.[20]
Reid et al.[17]	2006	Stepwise Fmoc SPPS	-	-
Brown et al.[18]	1996	Stepwise Fmoc SPPS[d]	-	The Tbfmoc protecting group was used to aid HPLC purification (high hydrophobicity and characteristic absorption).
Gong and Clark-Lewis[19]	1995	Stepwise Boc SPPS[e]	-	-

[a] Linear peptide after purification (before oxidation and folding).
[b] Dimethyloxazolidine dipeptides can be used in place of Ser or Thr and are best incorporated at intervals of approximately six residues.[64] (Acid-mediated ring opening regenerates Ser and Thr during the TFA resin cleavage step.)
[c] Significantly decreased product quality was reported with double coupling/capping protocol over single coupling.
[d] Utilizing double coupling/capping.
[e] SPPS utilizing amino acids N-terminally protected with the acid-labile *tert*-butoxycarbonyl (Boc) protecting group.

1.4 Summary

It can be seen that CCL2 is a biologically relevant and synthetically accessible target protein. A range of techniques in peptide synthesis (such as SPPS and NCL) have been described and will be utilized in attempts to synthesize CCL2 chemically, enabling the generation of CCL2 analogues bearing site-specific modifications. These will be useful as tools for investigating the role and biomolecular interactions of the chemokine in biological systems, and to evaluate its potential as a drug target.

Chapter 2

Linear stepwise synthesis

When considering the application of total chemical synthesis to a protein of biological interest, two major factors should guide the choice of approach – namely the speed and simplicity of the synthesis, and the ease with which it can be used to produce site-specifically modified analogues of the target molecule. A linear stepwise synthesis – that is, one in which the target is built one amino acid at a time as a single peptide strand – is conceptually simpler than a strategy involving native chemical ligation (*vide infra*). The stepwise approach nonetheless affords comparable control over the incorporation of modifications, while circumventing many of the difficulties associated with NCL (thioester generation, optimization of ligation conditions etc).

With regards to CCL2, a synthetic route should ideally allow the rapid generation of labelled chemokines, and for this reason a linear stepwise approach was initially adopted.

2.1 Initial considerations

The complete human CCL2 target sequence is shown in Figure 2.1. A schematic indicating secondary structure assigned from crystal structure data[15] using Kabsch and Sander's DSSP algorithm[65] is also shown. Information of this type is especially relevant as formation of β-sheet secondary structure during a synthesis can lead to aggregation of the peptide chain resulting in poor coupling efficiencies.[37] For this reason it was anticipated that difficulties would be encountered upon reaching the β-sheet regions shown.

The software utility Peptide Companion – a powerful tool used for the prediction of sequence difficulties – was also utilized.[66] Difficult coupling steps predicted for the CCL2 primary sequence (based on calculated aggregation potentials) are summarized in the graph shown in 2.1. A potentially problematic section can be identified between Lys-49 and Lys-38, which coincides with the start of the first region of β-sheet secondary structure.

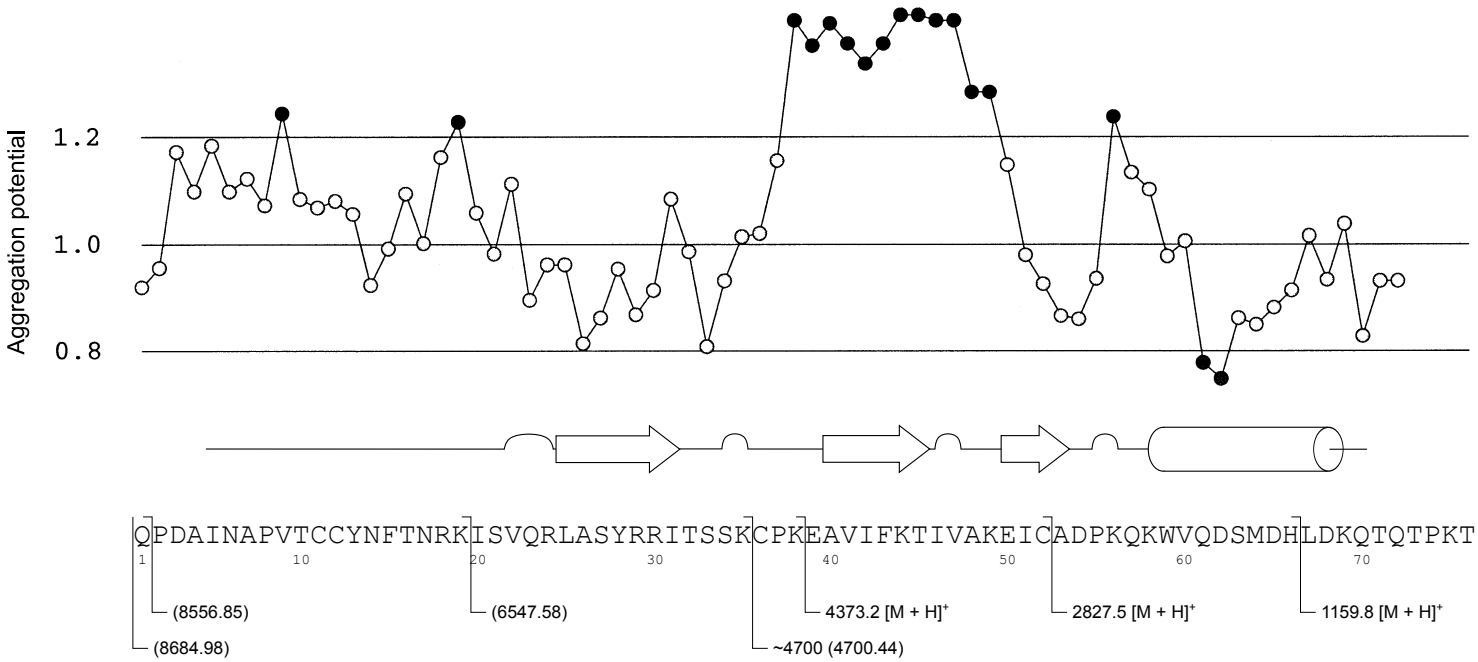


Figure 2.1: CCL2 primary sequence and secondary structure assignment.[15,65] α-Helical regions and β-sheets are represented by cylinders and arrows respectively. Experimental *m/z* values are also shown for points in the synthesis where intermediate MALDI-TOF MS analysis resulted in meaningful data (see Figure 2.3, Section 2.2). Calculated values are given in brackets, including those expected for test cleavages where no molecular ion peak was observed in the spectrum (*vide infra*). A representation of 'sequence difficulty' is also shown in the form of a plot of calculated aggregation potential against residue number obtained using the Peptide Companion software.[66] Difficulties are predicted in the synthesis for those residues with potentials above 1.2 (shown in black) – a problematic region (coinciding with β-sheet secondary structure) is clearly predicted between Lys-49 and Lys-38.

2.2 Microwave synthesis using polystyrene resin

A stepwise approach utilizing solid-phase Fmoc chemistry was first adopted. It was thought that microwave assistance would increase the efficiency of the coupling and deprotection steps, decreasing reaction times and improving on the crude product yield and purity attained in previous literature syntheses.[15–18] It was hoped that this type of approach would allow the development of a rapid, easily reproducible synthesis in accordance with the aims discussed.

2.2.1 The solid support: Handling, polymer choice and linker types

Initial MW-SPPS runs were attempted utilizing polystyrene Wang resin, a support traditionally used for the solid-phase synthesis of peptide acids. The structure of the Wang linker is shown in Figure 2.2 alongside some other linkers routinely used for Fmoc SPPS. Here it is important to emphasize the distinction between the linker type (Wang) and the resin polymer (polystyrene) – careful selection of both is crucial for an efficient synthesis. Commonly used polymers include polystyrene, poly(ethylene glycol) (PEG), and polyamide (PA), each with different swelling properties in various solvents.[67] For the linkers shown in Figure 2.2 either a hydroxyl group (for a so-called 'acid resin') or amine group ('amide resin') on the linker is used to anchor the C-terminal amino acid onto the resin via an esterification or amide bond formation respectively during the initial resin loading. After elongation the peptide is cleaved from the resin using 95% trifluoroacetic acid (TFA) to give either a peptide acid or peptide amide (depending on whether an acid or amide resin is used).*

As diffusion rates of reagents into the polymer matrix largely dictate reaction kinetics in solid-phase synthesis, only light agitation of the reaction is required.[67] The use of magnetic stirrers should be avoided so as not to crush the resin beads – gentle shaking of the reaction vessel or bubbling using a stream of nitrogen is recommended. Effective swelling of the polymer support in the reaction solvent is also vital, as the resin bead surface represents only 1% of coupling sites on the resin.[67] For optimum results pre-swelling in

*Many other linker types exist giving different C-terminal functionalities upon cleavage in various conditions.[68]

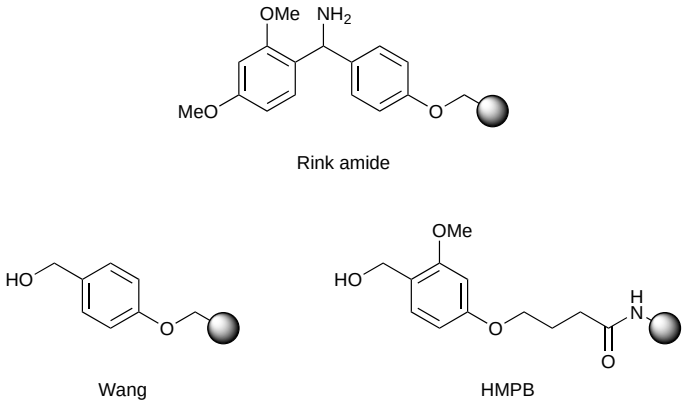


Figure 2.2: Commonly used linkers for Fmoc SPPS.[67] For Rink amide the C-terminal amino acid is attached via coupling to an amine, for Wang and HMPB a hydroxyl group is used. Peptide cleavage from the resin is carried out using 95% TFA for all of the above linkers to give peptide amides or peptide acids depending on the linker type (see text).

dichloromethane (DCM) overnight is preferred, although swelling for 1 h beforehand in DCM, DMF or a mixture of both was found to be adequate for most applications. DMF was used as the reaction solvent for all SPPS carried out due to its favourable swelling properties and suitability for MW heating (low volatility and high relative permittivity).[69]

2.2.2 Automated SPPS

Wang resin was purchased from Novabiochem preloaded with the amino acid threonine, the C-terminal residue of CCL2. Preloaded resin was used in order to circumvent problems typically associated with the esterification required for resin loading, such as racemization.[67] A resin with a low substitution (0.27 mmol/g) was chosen as it was anticipated that effects from steric hindrance and aggregation would be more significant the larger the target peptide.

Stepwise syntheses were run on an CEM Liberty1 automated peptide synthesizer equipped with a Discover microwave. Single couplings were carried out with activation (*vide infra*) using 5 equivalents of Fmoc amino acid under the default conditions (75 °C, 10 min). Removal of the Fmoc group was effected via treatment with 20% piperidine for 3 min with microwave assistance (75 °C).

Characterisation of peptide intermediates cleaved from the resin at various points during the synthesis was carried out using MALDI-TOF mass spectrometry (see Section 2.2.3). Mass spectra showing the products of a number of test cleavages are shown in Figure 2.3. While synthesis up to Ala-53 is clearly achievable (24-mer), there is an apparent reduction in the purity of the intermediate obtained on reaching Glu-39 (38-mer). Continuation of the synthesis past Cys-36 did not yield crude products of the correct molecular weight observable by MALDI-TOF MS. Isolation of the complete chemokine from the crude product mixture was attempted using high resolution FPLC (Section 4.2) but no species above ~6 kDa was observed in any of the column fractions. Syntheses were performed using three different coupling reagents: PyBOP®,[70] HBTU[71] and TBTU[72] (Figure 2.4). No significant difference in reaction efficiency was observed between reagents.

It was suggested that the difficulties encountered upon reaching Glu-39 may result from aggregation of the peptide during synthesis, leading to a reduction in solubility of the growing peptide and increased steric hindrance at the N-terminal amine.[37,73] It is therefore perhaps unsurprising that the problems described arise after Ala-53, where the first region of β-sheet secondary structure begins (Figure 2.1). In light of this it was decided that an alternative resin polymer should be sought, which would be more suitable for the synthesis of long peptides via SPPS (*vide infra*).

2.2.3 Synthesis monitoring and characterisation

The most commonly used technique for confirming the presence of the desired peptide sequence during SPPS is MALDI-TOF mass spectrometry (see Section 4.1 for more detail). In addition to the characterisation of the final product after cleavage from the resin, MALDI-TOF MS analyses can be run on small-scale 'test' cleavages carried out on samples of peptide-resin taken at intermediate points during the synthesis. Samples for test cleavage be taken after any step in a manual synthesis, and the Liberty1 software allows modification of reaction cycles to incorporate a pause at any reaction step (allowing access to the reaction vessel during an automated SPPS run).

For manual syntheses (*vide infra*), qualitative colourimetric tests provide a cruder (but far less time-consuming) method of confirming whether a particular coupling or deprotection step has been successful. The tests

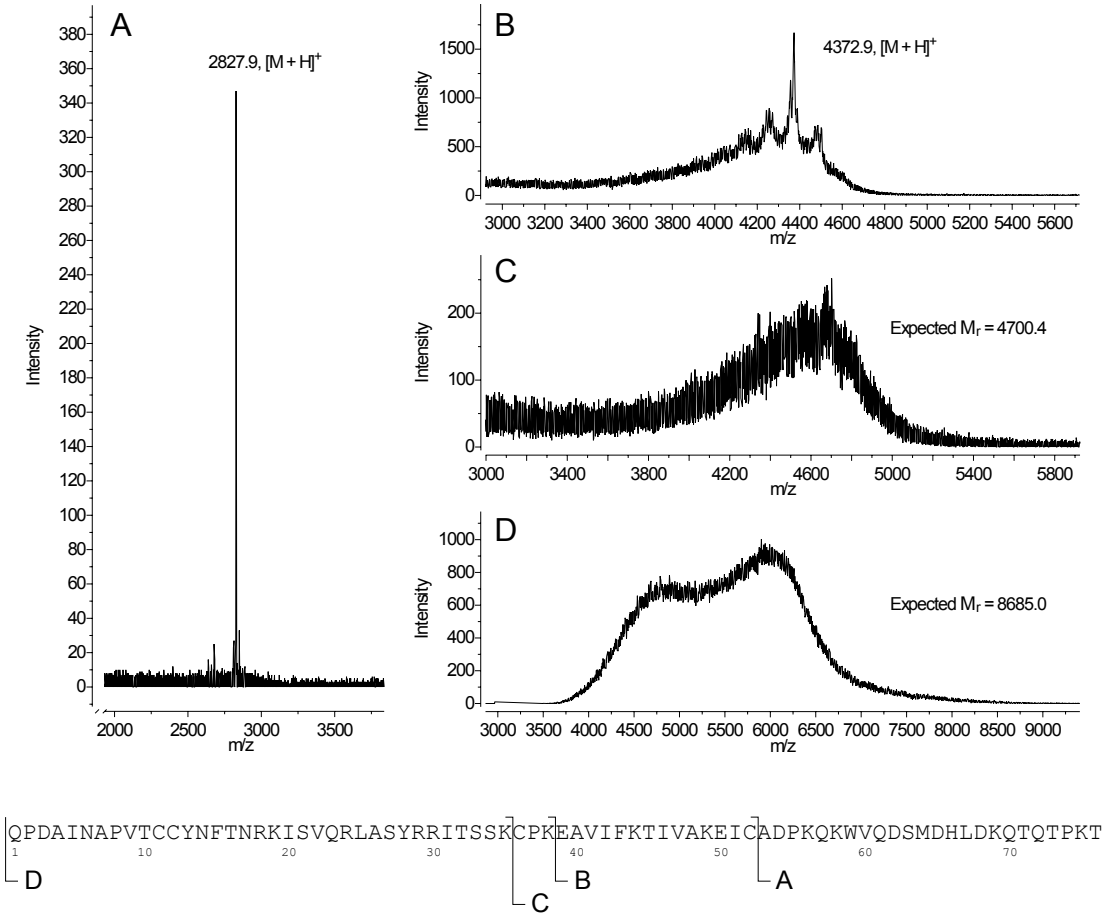


Figure 2.3: MALDI-TOF mass spectra resulting from analysis of test cleavage products obtained at various points during an attempted stepwise synthesis of CCL2(1-76) – **5** (full sequence) – via MW-SPPS. Synthesis up to Ala-53 (fragment **6**) is feasible using polystyrene Wang resin (spectrum **A**), although difficulties are encountered upon reaching Glu-39 (**B**). A broad molecular weight distribution due the presence of deletion peptides (indicating synthesis inefficiency) is observed at Cys-36 (**C**). Continuation of the synthesis did not yield any products of the correct molecular weight upon completion (**D**).

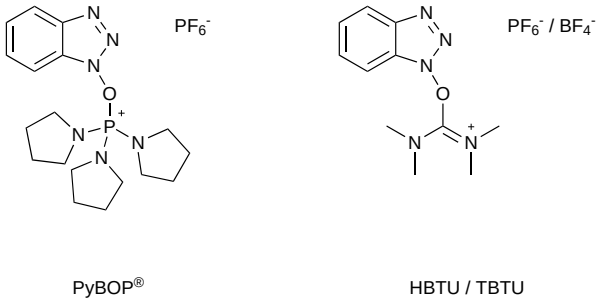


Figure 2.4: The coupling reagents PyBOP®, HBTU and TBTU. The last two differ only in their counterion, which has no influence on the reactivity.[32]

summarized in Table 2.1 can be carried out on resin-bound peptide directly, without the need for cleavage from the support.

Table 2.1: Qualitative tests for the presence of free amines.[63]

	TNBS Test[74]	Acetaldehyde/Chloranil Test[75]
Procedure[a]	1% TNBS in DMF and 10% DIPEA in DMF	2% acetaldehyde in DMF and 2% chloranil in DMF
Positive result[b]	Red	Dark blue/green
Negative result[b]	Grey/yellow	Colourless/yellow
Additional details	Primary amines only	For the detection of secondary (and primary) amines

[a] Reagents to be added in equal amounts to a sample of the resin (see Experimental section).
[b] Colouration of the *resin beads* themselves. Observation under a microscope may be required in ambiguous cases.

2.3 ChemMatrix® resin: A PEG-based support for SPPS

The significant loss of coupling efficiency observed during attempts thus far to synthesize CCL2 on-resin in a single run (Section 2.2) was suspected to be

the result of aggregation of the peptide or steric hindrance at the N-terminus during coupling and deprotection steps. It was therefore decided that an alternative resin polymer should be sought which would be more suited to the assembly of long peptides and proteins via SPPS.

ChemMatrix® is a PEG-based solid support[73] that has been utilized for the synthesis of long and/or difficult sequences – including β-amyloid peptide,[76] HIV-1 protease[77] and the chemokine CCL5[78†] – via SPPS. The amphipathic nature of PEG – due to its ability to form a number of helical structures of varying hydrophobicity – ensures the resin is well-solvated (and hence swells effectively) in a range of polar and nonpolar solvents.[73] A low molecular weight PEG was even demonstrated to act as the solvent when utilized as the solid support and melted during a MW-assisted reaction where no additional solvent was present.[79] Crucially, ChemMatrix® swells better than polystyrene in DCM and DMF, the most effective solvents for swelling polystyrene-based resins.[73]

2.3.1 MW-SPPS using ChemMatrix® resin

An automated microwave synthesis was attempted using ChemMatrix® HMPB‡ resin purchased from Sigma-Aldrich. As with the previous synthesis, resin preloaded with the C-terminal Thr residue was used. A test cleavage was carried out upon reaching Ala-53, yielding peptide **6** which was previously shown to be easily accessible using polystyrene Wang resin (Figure 2.3). The mass spectrum of the cleavage mixture is given in Figure 2.5. The crude product from the ChemMatrix® synthesis appears to be of poorer quality, and the major peaks present can be assigned to deletion sequences.

The apparent failure of the synthesis of such a short peptide was hypothesized to be at least partly due to degradation of the PEG support under microwave conditions.[79,80] Before attempting a room temperature synthesis however, it was deemed necessary to address the problem of hydrophobic aggregation.

†The synthesis of CCL5 also made use of pseudoproline dipeptides (*vide infra*).

‡For linker structure see Figure 2.2.

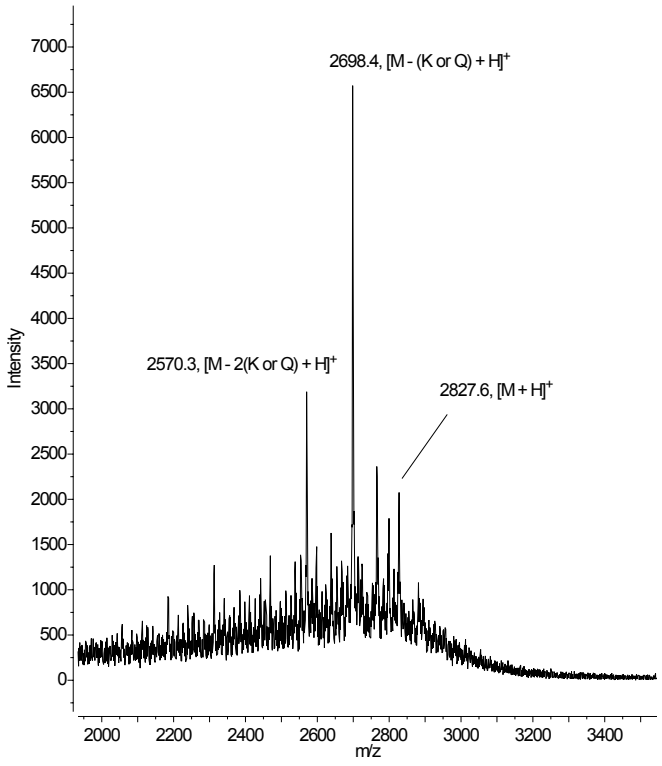


Figure 2.5: MALDI-TOF mass spectrum of test cleavage mixture from MW synthesis of CCL2 up to Ala-53 using ChemMatrix® resin. While target peptide **6** (ADPKQKWVQDSMDHLDKQTQTPKT) is present, the purity of the crude product appears poorer than that of the corresponding intermediate in the previous synthesis (Figure 2.3, spectrum **A**).

2.3.2 Disrupting aggregation using pseudoproline dipeptides

The problems encountered during the attempts described thus far to synthesize CCL2 on polystyrene Wang resin were largely attributed to hydrophobic aggregation exacerbated by the tendency of the peptide to form regions of insoluble secondary structure. 'Pseudoproline' dipeptides (or oxazolidine dipeptides) have been shown to be a useful tool to increase synthesis yields for difficult β-sheet-rich sequences where aggregation is an issue.[64,81] These preformed cyclized dipeptides of Ser/Thr and another amino acid (Scheme 2.1) are available commercially, and can be incorporated during a synthesis using standard Fmoc SPPS conditions. The cyclic oxazolidine moiety is isosteric with proline, hence the incorporation of pseudoprolines can disrupt the formation of secondary structure such as β-sheets. In some cases pseudoprolines have enabled the synthesis of proteins where SPPS runs have otherwise failed completely to generate any product.[82]

For optimal effect pseudoprolines should be separated from each other and from Pro residues by a distance of 5 or more amino acids, with a minimum separation of 2 amino acids.[82] On inspection of the CCL2 primary sequence it was found that the spacing of Pro, Ser and Thr residues was ideal for the incorporation of pseudoproline dipeptides.[15§]

Microwave-assisted synthesis

A manual MW-assisted synthesis utilizing preloaded ChemMatrix® resin and pseudoproline dipeptides was attempted. Reactions were run under identical conditions to the automated syntheses (except where detailed in the Experimental section), with microwave heating carried out using a Discover microwave. Test cleavage mass spectra from various points in the sequence are shown in Figure 2.6. Sites of pseudoproline inclusion are also indicated. Successful incorporation of the first pseudoproline was demonstrated (spectrum **B**), as well as synthesis up to Gln-61 (16-mer, using the second pseudoproline) – although the product peaks for the intermediate reached at this stage are very small (spectrum **C**).

[§]Positioning as follows: Q**P**DAINA**P**VTCCYN**FT**NRK**IS**VQRL**AS**YRR**IT**SSKC**P**K EAVIF**KT**IVAKEICAD**P**KQKWVQ**DS**MDHLDK**QT**QT**P**KT.

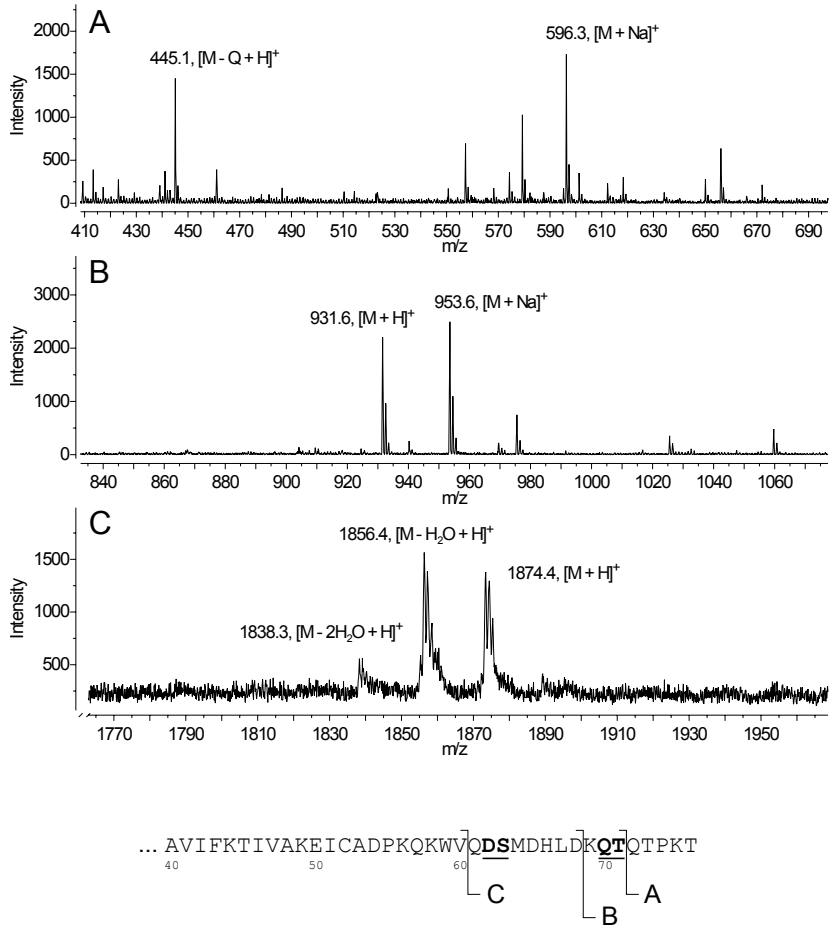
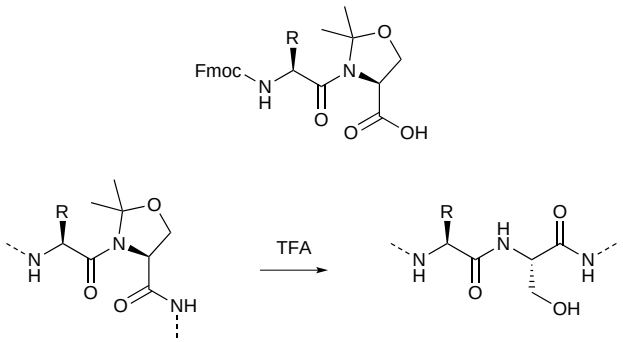


Figure 2.6: MALDI-TOF mass spectra for test cleavages at various points in the partial MW synthesis of CCL2 utilizing pseudoproline dipeptides (on CM resin). Bold underlined residues represent sites where pseudoproline dipeptides were included. Both pseudoprolines are incorporated successfully, although the product peaks observed upon reaching Gln-61 are very small (**C**).



Scheme 2.1: Pseudoproline dipeptides (of the form XS or XT, where X is another amino acid) can be incorporated via Fmoc SPPS and undergo acidolysis during the TFA cleavage step to generate the uncyclized dipeptide.[64] The Fmoc-protected serine dipeptide building block is shown (side chain of X represented by R group).

Room temperature synthesis

A room temperature synthesis (utilizing pseudoprolines) was also carried out to investigate whether microwave irradiation may be playing a part in the apparent loss of product quality observed with CM resin. Automated couplings were run at room temperature on the Liberty1/Discover system, with single couplings lasting 1 h and deprotections performed in two steps (5 min then 10 min – see Experimental section for full details).

Test cleavage mass spectra are shown in Figure 2.7. When compared with corresponding analysis from the MW synthesis, the spectrum obtained upon reaching Gln-61 (**D**) contains sharper, more clearly defined peaks of higher intensity – although only deletion sequences are observed in the cleavage mixture. From this result it can perhaps be suggested that while MW assistance does help to improve reaction efficiency (thereby decreasing the occurrence of deletions), the apparent yield of the crude product suffers when CM resin is subjected to microwave conditions (as tentatively inferred from MALDI-TOF MS data). It is therefore recommended that ChemMatrix® acid resins should not be used for SPPS with microwave heating.

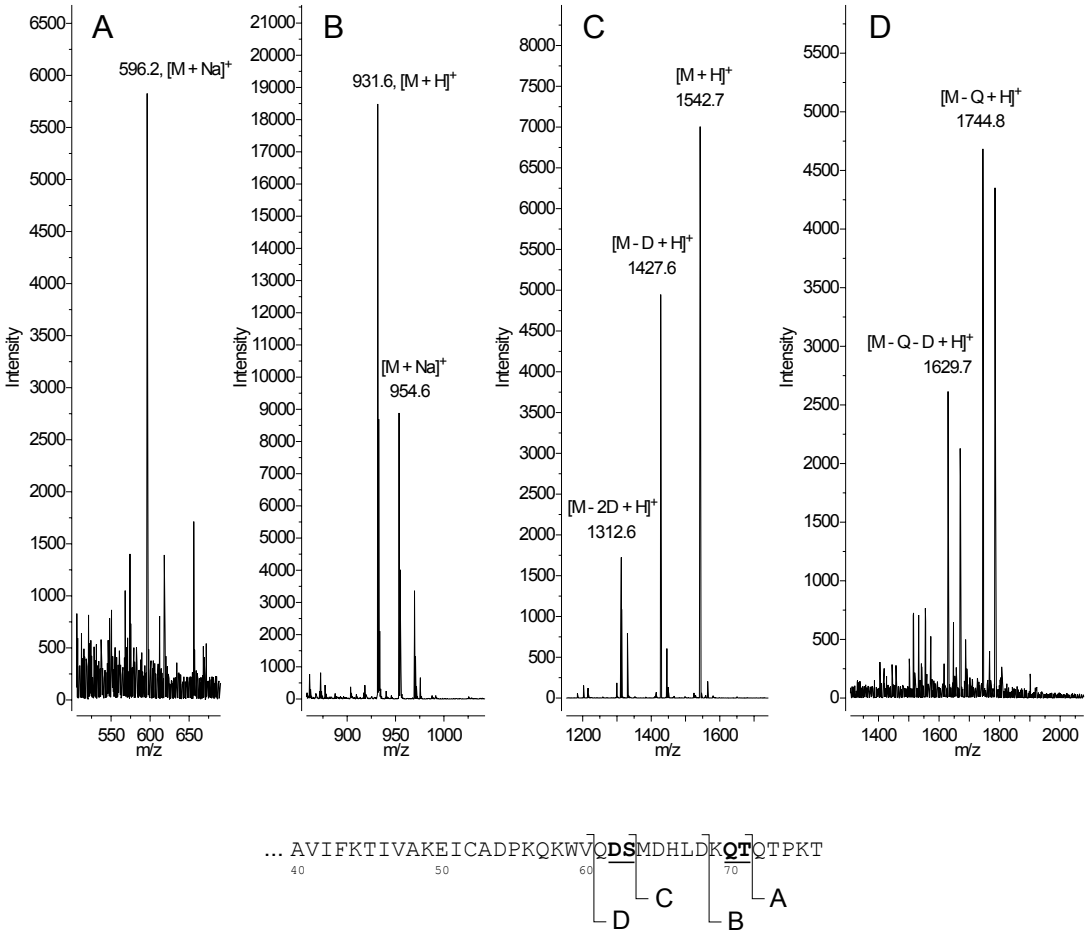


Figure 2.7: MALDI-TOF mass spectra for test cleavages during synthesis of part of CCL2 at room temperature using pseudoproline dipeptides (on CM resin). Pseudoproline incorporation at the points shown (bold underlined) was successful, although only peaks belonging to deletion peptides are observed in the spectrum obtained upon reaching Gln-61 (**D**).

2.4 Summary

The stepwise synthesis of CCL2 as a single peptide has been attempted using polystyrene Wang resin with microwave assistance. While elongation past Glu-39 proved problematic (potentially due to hydrophobic aggregation of the peptide on-resin), synthesis up to Ala-53 was shown to be achievable under these conditions. The PEG-based ChemMatrix® resin was investigated as an alternative support that may be better suited to the synthesis of the (relatively large) target protein, although initial MW-SPPS attempts using CM resin were disappointing. The use of pseudoproline dipeptides appeared to increase crude product quality however, suggesting that the 'sequence difficulties' experienced were at least in part due to hydrophobic aggregation. It was found that room temperature synthesis conditions were perhaps more suitable when using CM resin, although the lack of microwave assistance resulted in deletions that need to be eliminated via the use of a double coupling protocol.

Chapter 3

Towards native chemical ligation

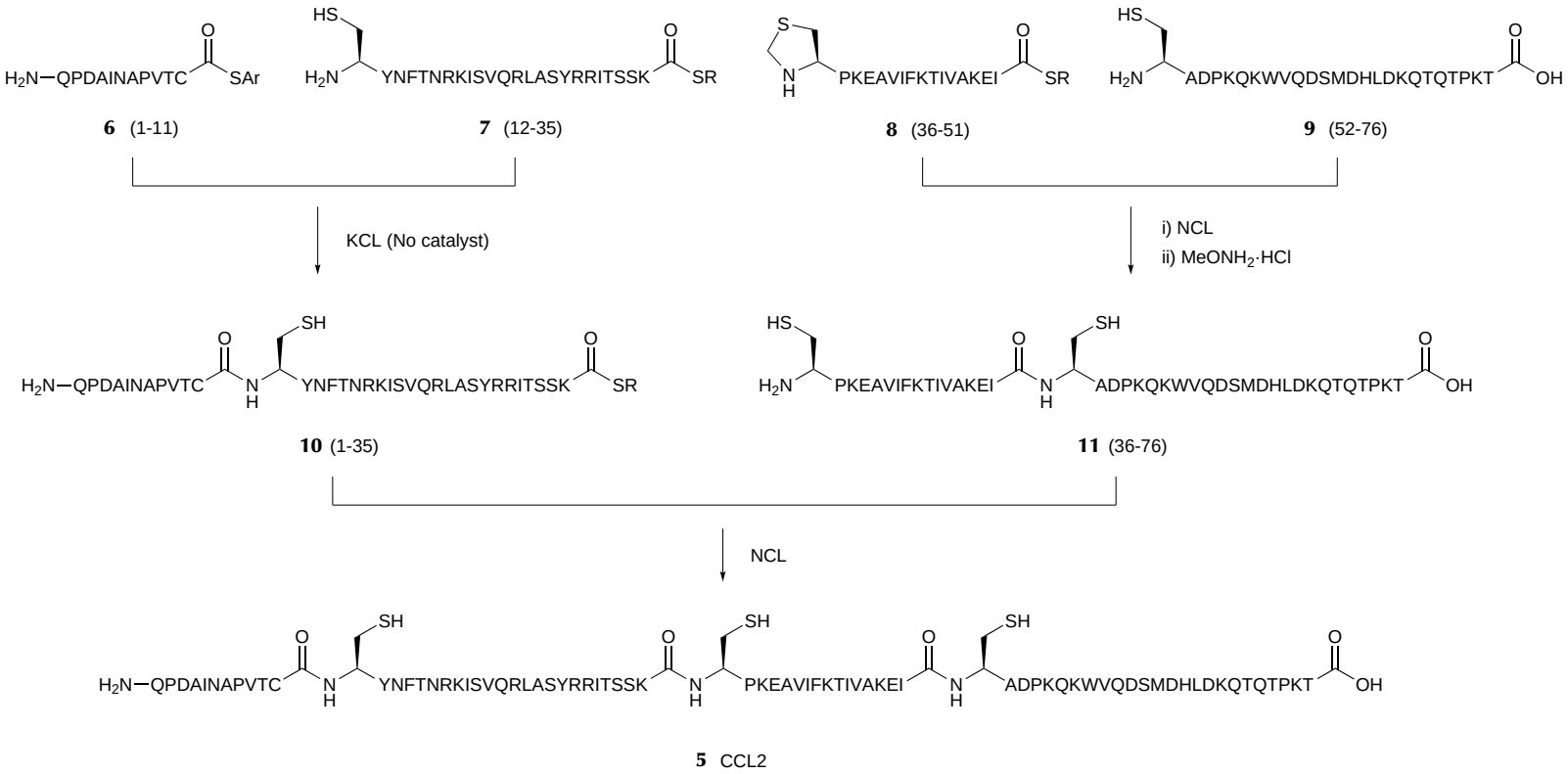
The attempts at a stepwise linear synthesis described in the previous section highlight the obstacles often encountered during the solid-phase synthesis of long peptides. Of these, two problems – namely aggregation on-resin (leading to loss of coupling efficiency) and difficulties separating the end product from similar deletion sequences (see Chapter 4) – can be avoided by adopting a strategy in which shorter segments of the target are synthesized and ligated in solution. One such ligation technique, 'native chemical ligation' (Section 1.2.3), utilizes the transthioesterification of a peptide thioester with a thiol ligation partner (followed by an S→N acyl shift) to form an amide bond between two fully deprotected peptides in aqueous solution.

Due to the lack of success with attempts to assemble the whole of CCL2 on-resin as a single peptide, it was thought that a convergent approach utilizing NCL would be promising as an alternative strategy for the total synthesis of CCL2. An evaluation of two methods of thioester generation, as well the synthesis of a number of thioester segments and a Cys ligation partner are described in the following sections.

3.1 Proposed synthesis using NCL

Scheme 3.1 outlines a proposed (fully convergent) strategy for the total chemical synthesis of CCL2 (**5**) via NCL in 4 segments. The ligation of fragments **6** and **7** is kinetically controlled (the aryl thioester of **6** should react preferentially over the alkyl thioester of **7**), and the N-terminal Cys of fragment **8** is masked as a thiazolidine during the the 'conventional' ligation of

8 and **9**. The two ligations described can be performed in parallel, potentially allowing rapid generation of CCL2 analogues (cf. a stepwise approach with ligations occurring in the C- to N-terminal direction). Ligation of the resulting segments (**10** and **11**) affords the full chemokine.



Scheme 3.1: Fully convergent 4-segment total chemical synthesis of CCL2 utilizing NCL. The ligation of fragments **6** and **7** is kinetically controlled (see text). **6** is an aryl thioester, and 'Ar' represents an electron withdrawing aromatic group such as phenyl. The N-terminal Cys of fragment **8** is masked as a thiazolidine.

3.2 The *N*-acylurea approach for the synthesis of peptide thioesters

Although the number of Fmoc SPPS strategies for producing peptide thioesters has grown dramatically in recent years,[47] none are as established as conventional methods for the generation of peptide acids and amides, and hence no single approach has emerged as the dominant technique for peptide thioester generation. Strategies utilizing the so-called 'safety-catch' principle (Section 1.2.3) offer convenient methods for obtaining peptide thioesters directly using SPPS, and it was therefore decided that these would be the best option for producing the thioester segments required for the synthesis of CCL2 via NCL.

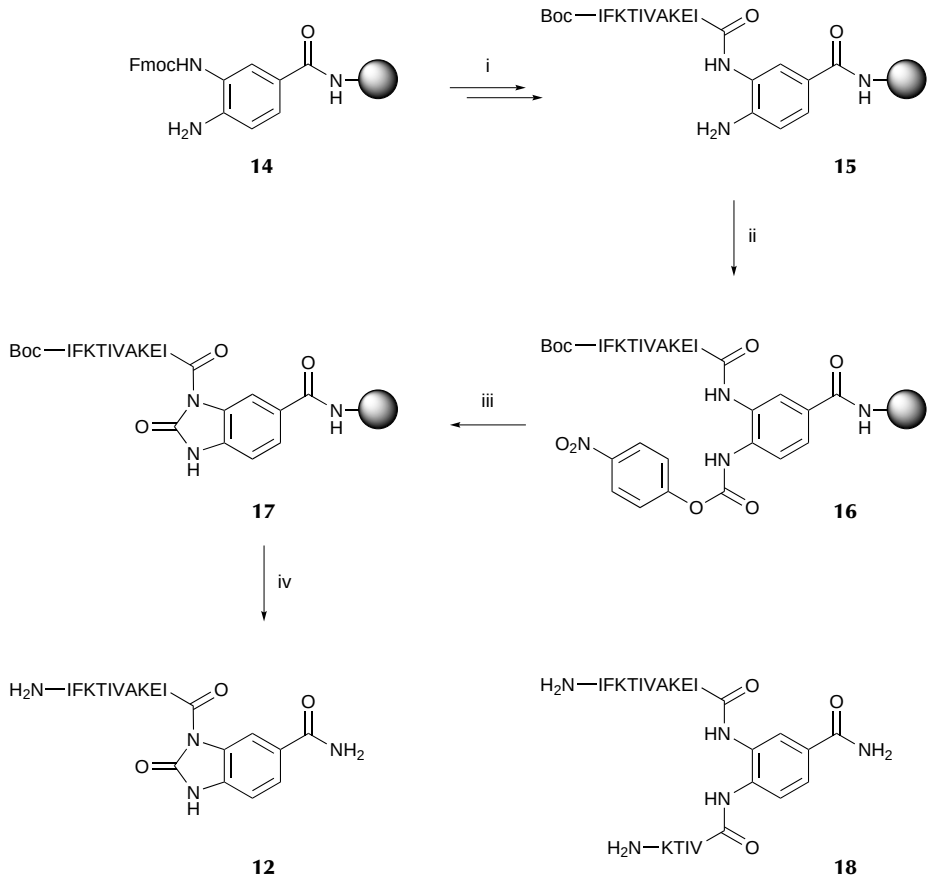
The *N*-acylurea approach introduced in 2008 by Blanco-Canosa and Dawson utilizes an acid-labile linker modified with 3,4-diaminobenzoic acid (Dbz) onto which a peptide chain can be assembled.[54] Treatment with *p*-nitrophenylchloroformate forms the cyclic *N*-acylurea moiety (also known as an *N*-acyl-benzimidazolinone, Nbz), which can be cleaved under standard conditions to give peptides incorporating a C-terminal Nbz group. These Nbz-peptides can be converted to thioesters in solution or used directly in ligation reactions.[54]

3.2.1 Suitability of the Dbz linker for MW-SPPS

The synthesis of initial test fragment **12** – CCL2(42-51) *N*-acylurea (Scheme 3.2) – was attempted using MW-SPPS and Dawson Dbz AM* resin available from Novabiochem. The 10-residue sequence was assembled using standard automated MW conditions (see Experimental section), with Boc-protected isoleucine coupled as the final amino acid to protect the N-terminus during the subsequent linker activation and cyclization. Synthesis monitoring via cyclization of a sample of peptide-resin followed by cleavage of *N*-acylurea intermediates is unfeasible using this approach, as the resin-bound peptide cannot be elongated once the necessary N-terminal Boc group has been incorporated. Nevertheless, cleavage of the Rink amide linker without prior activation/cyclization releases diaminobenzamide-peptides that can be

*Aminomethyl. AM resin (effectively polymer-bound benzylamine) is a support used in solid-phase synthesis. Dawson Dbz AM resin comprises Dbz linked via the Rink amide linker to an aminomethylpolystyrene support (one Dbz amine is Fmoc-protected).

observed (for example **13**, Scheme 3.4), hence allowing chain elongation to be followed easily by small-scale cleavage of resin-bound peptides *without* N-terminal Boc protection. This is an advantage over the sulfonamide linker, which is not acid-labile and must be cleaved via attack by a thiol nucleophile (*vide infra*).



Scheme 3.2: Synthesis of Nbz-peptide **12** using Dawson's *N*-acylurea approach. Side chain protecting groups are not shown. Structure **18** is a putative branched product produced as a result of linker over-acylation (see text). Reagents and conditions: (i) Fmoc MW-SPPS (see Experimental section); (ii) *p*-nitrophenylchloroformate (5 equiv), DCM, rt, 1 h; (iii) DIPEA, DMF, rt, 45 min; (iv) TFA (95%, water), TIPS, rt, 4 h.

Activation and cyclization of resin-bound product **15** were carried out via treatment with 5 equivalents of *p*-nitrophenylchloroformate for 1 h, followed by 0.5 M DIPEA for 45 min. The resultant peptide was cleaved from the resin

using standard TFA cleavage conditions and analysed via MALDI-TOF MS. No peak corresponding to **12** (M_r = 1320.6) was observed in the mass spectrum, and the major peaks (Figure 3.1, spectrum **A**) correspond to protonated and sodiated versions of a species that is ~400 Da greater than the mass of the target peptide (m/z = 1736.2 and 1757.8 respectively).

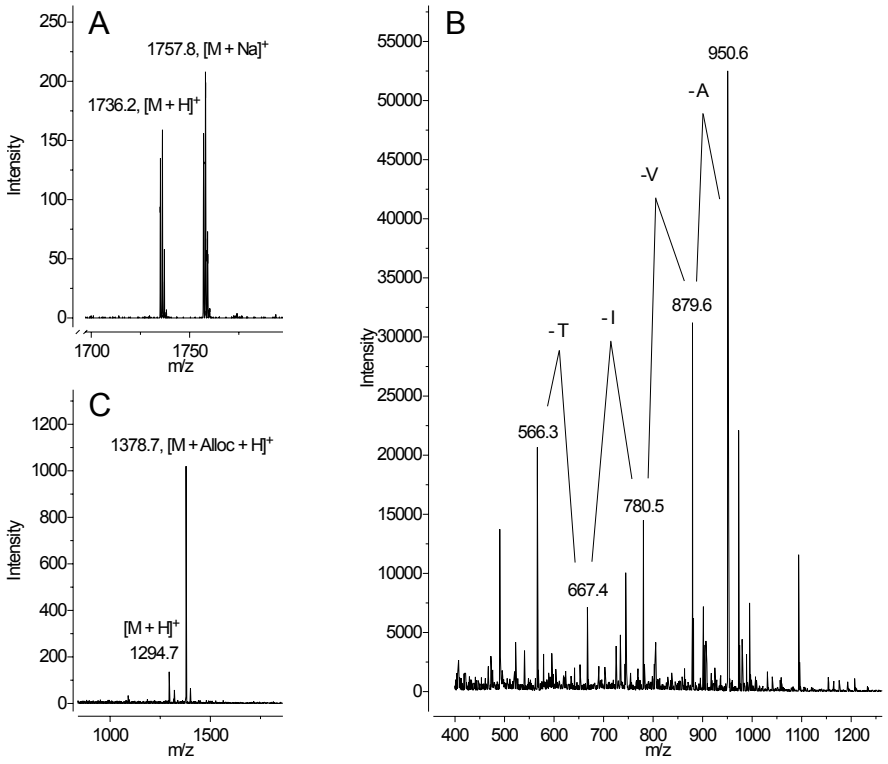


Figure 3.1: MALDI-TOF mass spectra for attempts to synthesize CCL2(42-51) *N*-acylurea (**12**) using Dawson Dbz AM resin under various conditions. Spectrum **A** shows the products of the initial MW-assisted synthesis: the target sequence is not present, although peaks corresponding to putative structure **18** (Scheme 3.2) are observed (resulting from 'over-acylation', see text). For the subsequent room temperature synthesis (**B**) only a series of deletion peptides are present. MW-SPPS using an Alloc-protected linker (spectrum **C**) produces the correct peptide sequence with no branching (although the desired *N*-acylurea species is not observed, see text).

SPPS using the Dbz linker relies on the fact that after coupling of the C-terminal amino acid to the first linker amine group, steric and electronic factors generally prevent reaction at the second.[54] Reaction at both groups

is nevertheless possible, and so-called 'over-acylation' has been observed for sequences rich in sterically unhindered residues such as glycine.[83] This results in the formation of branched products that cannot be cyclized to form the required *N*-acylurea, and it was suspected that the major peaks observed in the mass spectrum were due to the presence of a branched product such as **18** (m/z = 1736.2), shown in Scheme 3.2.

Despite tentatively assigning the observed signals, it was nonetheless surprising that over-acylation should be observed for a sequence comprising only sterically *hindered* residues. Given the success of previous syntheses utilizing the Dbz linker carried out at room temperature,[84–86] it was suggested that MW conditions may promote reaction at the second linker amine even for hindered amino acids. To test this, a second automated synthesis was run at room temperature – the MALDI-TOF mass spectrum is given in Figure 3.1 (spectrum **B**).

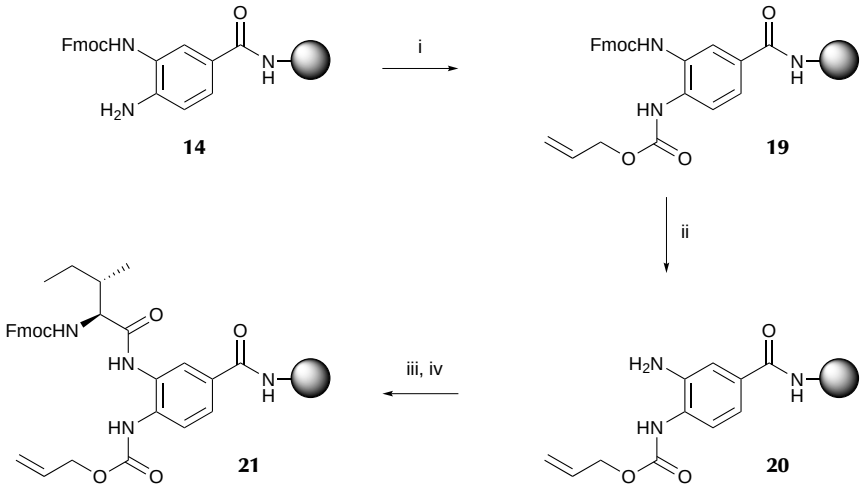
Peptide **12** is once again absent from the spectrum, although in this case the observed signals have m/z values *lower* than expected, suggesting that they perhaps correspond to deletion sequences rather than branched products. Furthermore, a series of deletions (shown on spectrum) can clearly be linked to the peak with m/z = 950.6, indicating that the major signal corresponds to a peptide containing those residues. These results suggest that while MW conditions promote over-acylation of the Dbz linker, a protocol utilizing room temperature single couplings is not sufficient for the synthesis of this sequence. In light of the benefits afforded by MW assistance, it was decided that developing a modified Dbz strategy for use in MW-SPPS would be preferable to optimizing the room temperature synthesis.

3.2.2 Preventing over-acylation

As over-acylation is caused by unwanted coupling of amino acids onto both linker amines, it was suggested that protection of one of the amine groups would prevent the formation of branched products, hence allowing use of the Dbz linker for MW-SPPS. Mahto et al. previously demonstrated the use of the allyloxycarbonyl (Alloc) protecting group for this purpose, with elimination of over-acylation when an Alloc-protected resin was used.[83]

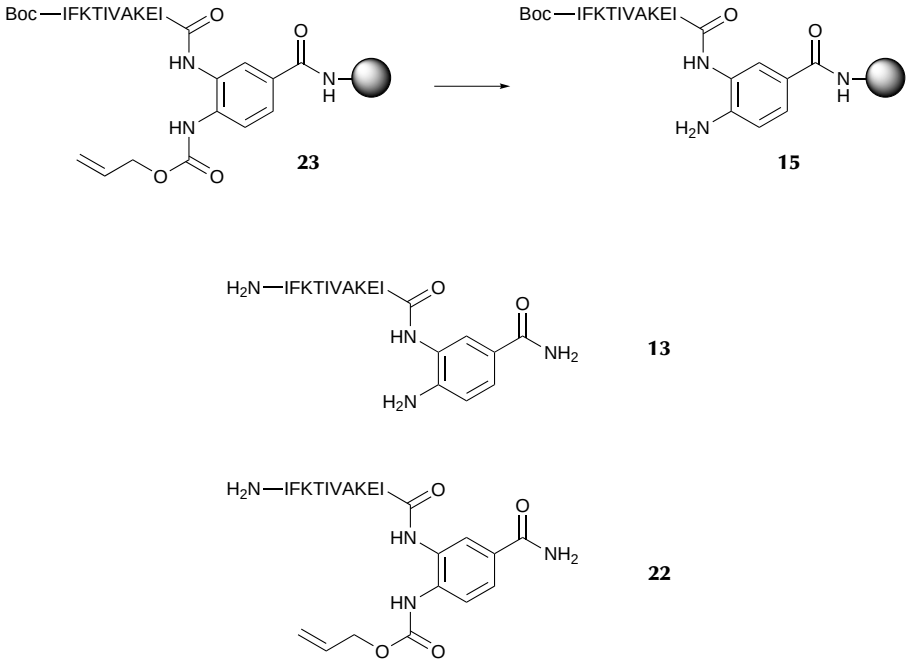
The free amino group of Fmoc-protected Dbz resin was first Alloc-protected by treatment with allyl chloroformate and DIPEA for 24 h (Scheme 3.3). Complete protection was monitored qualitatively using the tests

described in Section 2.2.3. After removal of the Fmoc group, the C-terminal isoleucine residue was double coupled using the coupling reagent HATU and 15 equivalents of Fmoc-Ile-OH. These conditions were required as coupling onto the Alloc-Dbz linker was found by Mahto et al. to be inefficient for the β-branched hydrophobic amino acids Ile and Val.[83] Unreacted amine groups were then capped via acetylation with acetic anhydride and the sequence CCL2(42-51) assembled on the resin via MW-SPPS.



Scheme 3.3: Alloc protection and loading of the Dbz linker. Reagents and conditions: (i) allyl chloroformate (10 equiv), DIPEA (1.0 equiv), DCM, rt, 24 h; (ii) 20% piperidine, DMF, rt, 2 × 10 min; (iii) Fmoc-Ile-OH (15 equiv), HATU (13.5 equiv) and DIPEA (27 equiv), DMF, rt, 2 × 1 h; (iv) 20% acetic anhydride, DMF, rt, 30 min.

Removal of the Alloc group was carried out via treatment with Pd(0) and $PhSiH_3$ (Scheme 3.4), and linker activation and cyclization were carried out as described previously. The mass spectrum resulting from analysis of the TFA cleavage mixture is shown in Figure 3.1 (spectrum **C**). While no peak corresponding to an *N*-acylurea species is observed, the uncyclized species present – deprotected **13** (m/z = 1294.7) and Alloc-protected **22** (m/z = 1378.7), Scheme 3.4 – are unbranched (no over-acylation) and complete (no deletions), demonstrating that MW-assisted synthesis of this sequence is feasible using the Dbz linker. Although the Alloc deprotection and cyclization issues need to be addressed, to the best of our knowledge this represents the first reported use of the Dawson Dbz linker for microwave-assisted SPPS.



Scheme 3.4: Removal of the Alloc group from the Dbz linker. Reagents and conditions: $PhSiH_3$ (20 equiv), $Pd(PPh_3)_4$ (0.35 equiv), DCM, rt, 30 min. Uncyclized products **13** and **22** (m/z = 1294.7 and 1378.7 respectively - spectrum **C**, Figure 3.1) observed following TFA cleavage from the resin are also shown.

3.3 Thioester generation using the sulfonamide safety-catch linker

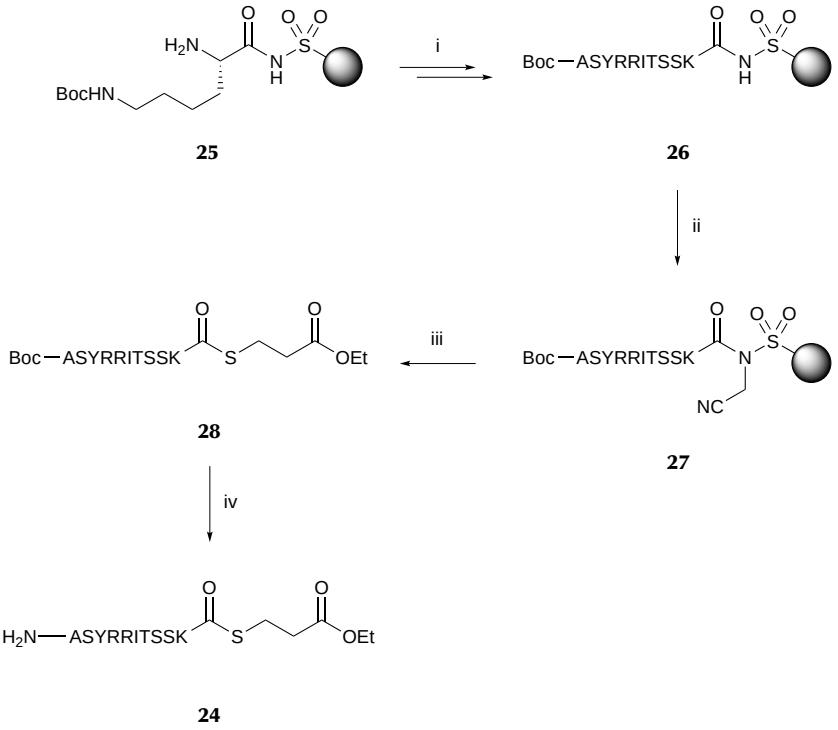
Alongside attempts to synthesize peptide thioesters using the Dbz linker, the well-established 'sulfonamide safety-catch' approach was also investigated. This method circumvents problems associated with over-acylation, but does not form intermediate *N*-acylurea peptides (as the thioesters are produced directly). In addition to thiols, the linker can be attacked after activation with a variety of other nucleophiles[50] to produce peptides with various C-terminal functionalities (such as amides[87] and esters[88]) or the N-terminus of the peptide itself[89] (head-to-tail cyclization). The sulfonamide linker was used by Grygiel et al. to generate peptide thioesters used in their synthesis of CCL2 via NCL.[15]

3.3.1 Trial synthesis and optimization of deprotection conditions

An automated room temperature synthesis of fragment **24** – CCL2(26-35) thioester (Scheme 3.5) – was attempted in order to investigate the usefulness of the sulfonamide linker for producing peptide thioesters. The peptide was assembled using standard room temperature conditions (see Experimental section) on sulfamylbutyryl resin, with Boc-Ala-OH coupled as the N-terminal amino acid. As with the Dbz linker (Section 3.2.1), the Boc group is essential to prevent alkylation of the N-terminus during the subsequent activation step. In contrast to the *N*-acylurea approach however – which utilizes the acid-labile Rink amide linker facilitating rapid analysis of peptide intermediates – peptides anchored to the sulfonamide linker can only be released via treatment with an alkylating agent followed by attack with a thiol nucleophile (each step ~24 h). This limits the feasibility of convenient synthesis monitoring via small-scale test cleavages when using the sulfonamide safety-catch approach.[†]

Linker activation was carried out via alkylation of the *N*-acylsulfonamide with iodoacetonitrile (24 h), followed by treatment with thiol (ethyl 3-mercaptopropionate) and sodium thiophenolate catalyst (24 h) in order to

[†]A solution to this problem is to anchor the safety-catch to the resin via an acid-labile linkage such as Rink amide (see Section 1.2.3). Preloaded sulfamylbutyryl resin (without Rink amide linker) is commercially available and was used due to the difficulties associated with sulfonamide loading.[51,53] During the writing of this report however, a number of preloaded sulfamylbutyryl Rink amide resins were made available to order from Novabiochem.



Scheme 3.5: Synthesis of peptide thioester **24** using the sulfonamide safety-catch approach. Side chain protecting groups are not shown. Reagents and conditions: (i) Fmoc MW-SPPS (see Experimental section); (ii) ICH_2CN (10 equiv), DIPEA (10 equiv), DMF, rt, 24 h; (iii) ethyl 3-mercaptopropionate (50 equiv), PhSNa (0.5 equiv), DMF, rt, 24 h; (iv) TFA (95%, water), TIPS, rt, 4 h.

generate protected peptide thioester **28** (Scheme 3.5). The standard global deprotection conditions (95% TFA/TIPS, 4 h) were found to be insufficient for **28**, and species with Pbf-protected arginine residues were observed in the mass spectrum (Figure 3.2, left-hand spectrum). Treatment of the partially protected crude peptide with the cleavage cocktail for a further 3 h gave fully deprotected peptide thioester **24** (right-hand spectrum), alongside a single serine deletion peptide.

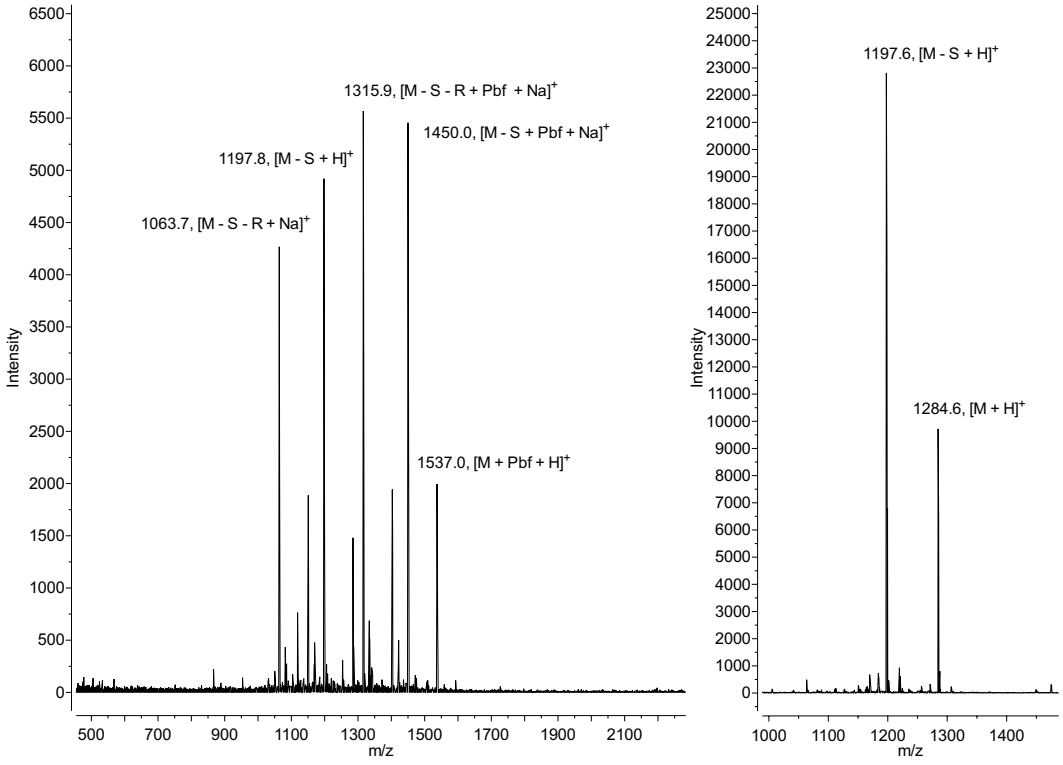


Figure 3.2: MALDI-TOF mass spectra of cleavage mixture for TFA deprotection of test thioester **28** after 4 h (left-hand spectrum) and 7 h (right-hand spectrum). Various partially deprotected peptides and/or deletion sequences are present in the first spectrum, while after extended TFA treatment only a single serine deletion peptide remains alongside the intended product (**24**, ASYRRITSSK-thioester) in the second spectrum.

3.3.2 Synthesis of thioester fragments for NCL

Having demonstrated the effectiveness of thioester generation using the sulfonamide safety-catch linker, the next objective was to complete the synthesis of fragment **7** (Scheme 3.1). Synthesis up to Ala-26 had previously been achieved (and Arg deprotections optimized), therefore completion of the fragment was expected to be unproblematic. Due to the occurrence of deletions in the previous synthesis however, it was decided that double couplings should be employed for all residues in the sequence. In addition, selected residues were triple coupled[‡] in anticipation of potential sequence difficulty predicted using the Peptide Companion software (Section 2.1) and through consideration of the occurrence of β-sheet regions and the deletions observed previously.

An automated room temperature synthesis of CCL2(12-35) thioester (**7**) was carried out as described, with Boc-Cys(Trt)-OH coupled as the N-terminal amino acid. Linker alkylation, thiolysis and global deprotection were performed as described. A second thioester, **29**, was also synthesized with Boc-Thz-OH coupled at the N-terminus in place of the Cys derivative. This was done so that a version with masked N-terminal Cys would be available if it was decided that a sequential (rather than convergent) ligation strategy[§] should be adopted (*vide infra*).

Mass spectra showing thioesters **7** and **29** are given in Figure 3.3. A small by-product peak is present in both corresponding to a single Gln or Lys deletion – potentially indicative of difficulty at Gln-23, and which may be eliminated by triple coupling at this residue.

3.4 Synthesis of C-terminal segment

As a result of the efforts to assemble CCL2 in a single SPPS run described previously (Chapter 2), it was shown that synthesis up to Ala-53 (24-mer) is achievable. The addition of a cysteine residue gives C-terminal fragment **9** (Scheme 3.1), hence the synthesis was repeated (using the same conditions: automated MW-SPPS, polystyrene Wang resin) and extended to Cys-52, and all of the peptide was cleaved from the resin. A mass spectrum confirming the

[‡]Triple couplings were used for the following residues: **Y**NF**T**NR**KI**S**V**QRLA**S**YRRITSSK.
[§]Occurring in the C- to N-terminal direction (see Section 1.2.3).

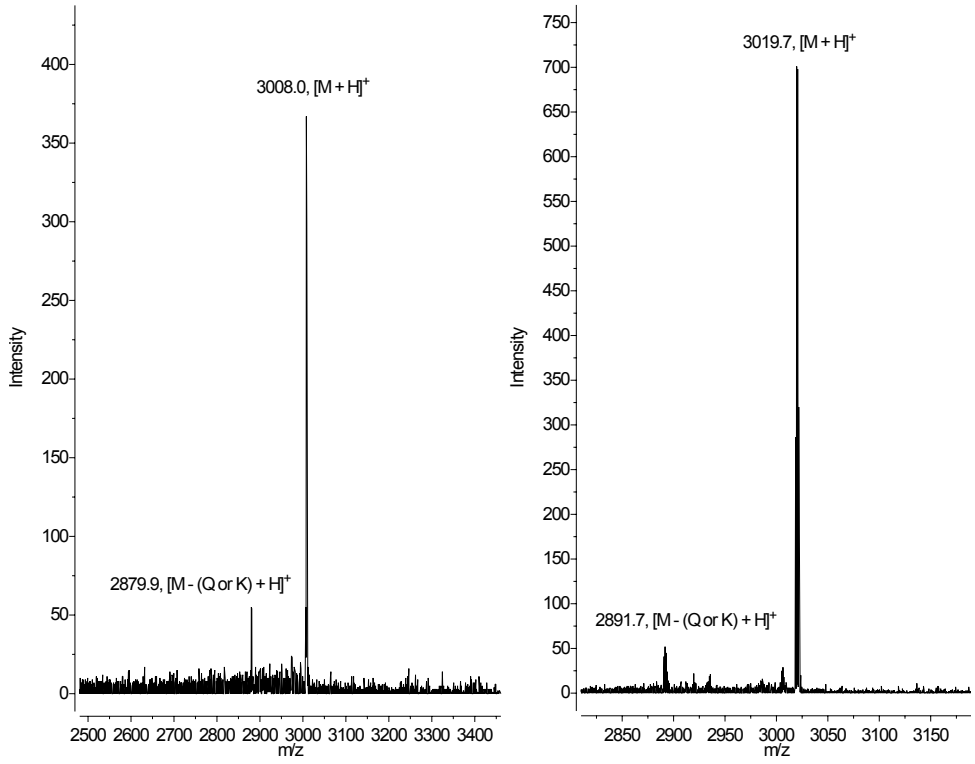


Figure 3.3: MALDI-TOF mass spectra of thioester fragment CCL2(12-35) (Xaa-YNFTNRKISVQRLASYRRITSSK-thioester) incorporating N-terminal Cys (**7**, left-hand spectrum) or Thz (**29**, right-hand spectrum). A small peak corresponding to a Gln or Lys deletion is present in each spectrum.

successful generation of peptide **9** is shown in Figure 3.4. The peptide was purified via RP-HPLC and FPLC (see Section 4.2 and Experimental section).

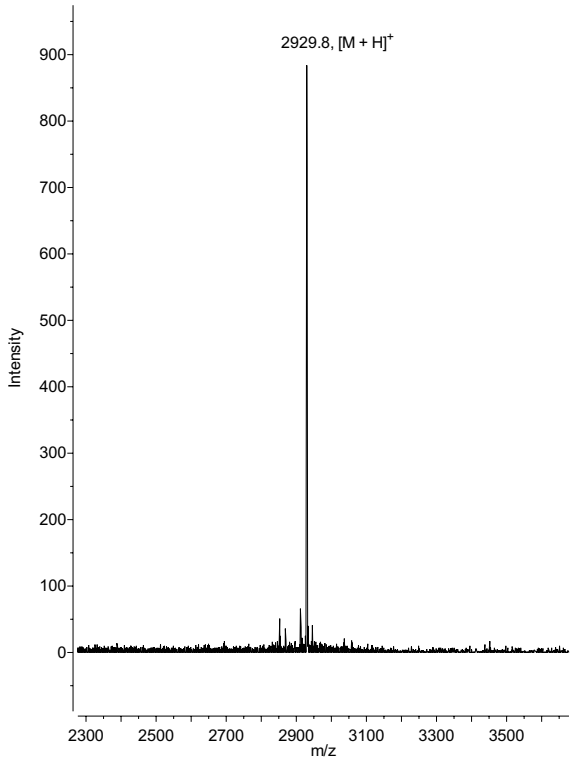


Figure 3.4: MALDI-TOF mass spectrum of C-terminal fragment **9** – CCL2(52-76), CADPKQKWVQDSMDHLDKQTQTPKT.

3.5 Summary

Significant progress has been made towards a total chemical synthesis of CCL2 using native chemical ligation. Two solid-phase methods of peptide thioester synthesis have been investigated, and thioester fragment **7** (Scheme 3.1) was successfully synthesized – along with Thz-masked counterpart **29** – using the sulfonamide safety-catch approach. The alternative *N*-acylurea method – utilizing Dawson's Dbz linker – was successfully adapted for MW-SPPS using the Alloc protection strategy developed by Mahto et al. While issues with Alloc deprotection and linker cyclization need to be addressed, the technique was used for the partial synthesis of fragment **8**, representing the first reported use of the Dawson Dbz linker for microwave-assisted SPPS. In addition, C-terminal fragment **9** was synthesized on a suitable scale (~160 mg) and purified for use in the total synthesis.

Chapter 4

Peptide purification and characterisation

The characterisation and purification of peptides carries with it potential difficulties beyond those experienced for small organic molecules. Mass spectrometric observation of peptide species can become less trivial with increasing molecular weight, requiring the use of more suitable techniques such as matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry and greater attention to sample preparation. In addition, SPPS can generate numerous side products alongside the target peptide in the form of chemically similar deletion sequences.[37] The chromatographic separation of these from the target is often based solely on subtle differences in hydrophobicity, and therefore can also prove challenging.

The following sections outline some of the considerations that need to be taken into account when using MALDI-TOF MS for peptide characterisation, along with a description of the steps taken to purify the products of SPPS and the optimization of parameters used for chromatographic separation.

4.1 The characterisation of synthetic peptides using MALDI-TOF MS

MALDI-TOF mass spectrometry facilitates the analysis of heterogeneous mixtures of large molecules[90,91] and is therefore well suited to the analysis of crude reaction products from SPPS. The MALDI ion source generates primarily singly charged ions with no fragmentation, hence simplifying the

interpretation of spectra obtained from samples containing complex mixtures of different peptide sequences. Care must be taken when interpreting MALDI-TOF mass spectra: while the technique has been shown to be useful for reaction monitoring (*vide supra*), the intensity of a mass spectral peak is highly dependent on the ionization efficiency of the corresponding peptide – which in turn is affected by peptide composition and chemical environment.[92] Therefore MALDI-TOF MS should only ever be used as a very rough indicator of product purity.

The two mass spectra given in Figure 4.1 correspond to Fmoc-protected (right-hand spectrum) and deprotected (left-hand spectrum) versions of the peptide KQTQTPKT. As well as the dramatic difference in relative intensity of the sodiated/unsodiated peaks between spectra, the observed by-products are much less prominent for the Fmoc-protected form, suggesting that in this case the protected peptide is perhaps more easily ionized. This example highlights the fact that the apparent absence of impurities from a MALDI-TOF MS spectrum can often give a false impression of crude product purity. Issues relating to sample preparation and how this can affect the resulting spectrum are discussed in the following section.

4.1.1 Sample preparation

The conditions used for sample preparation can dramatically affect the quality of MALDI-TOF MS spectra. The choice of solvent can change which peaks are observed in a spectrum, while the hydrophobicity of an analyte – and its solubility in a particular solvent – can dictate whether the desired peptide is observed in the spectrum at all. Care must be taken to ensure the peptide dissolves adequately when preparing the sample solution – dissolving the peptide in water should be the first step, with subsequent addition of a solvent such as acetonitrile to make sure that any hydrophobic species present are also dissolved. For particularly hydrophobic peptides a higher proportion of organic solvent can be used, although solvents with a polarity lower than water should ideally make up no more than 50% of the sample solution by volume.[93] Other organic solvents such as methanol or THF can also be tried if results are not satisfactory with MeCN. The addition of 0.1% TFA was also found to improve solvation for particularly insoluble peptides, and has been shown to increase mass resolution and accuracy for MALDI-TOF MS.[93]

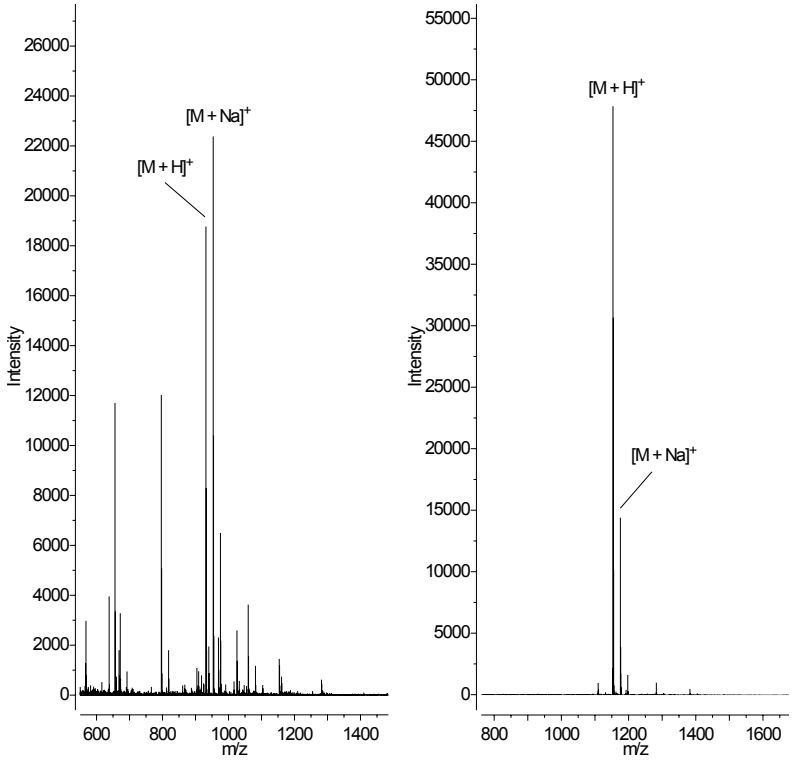


Figure 4.1: MALDI-TOF mass spectra of the crude peptide KQTQTPKT, with N-terminal Fmoc protection (right-hand spectrum) and without (left-hand spectrum). A clear difference in the relative peak intensities can be seen between spectra.

Figure 4.2 illustrates the effect solvent choice can have on the resulting mass spectrum. The left-hand spectrum (taken from Figure 3.1, Section 3.2.1) was collected from a sample prepared using MeCN as the organic solvent – peaks corresponding to peptides **13** and **22** are observed. The sample used for the right-hand spectrum was prepared using MeOH instead of MeCN. Not only is the relative intensity of the peak corresponding to **13** much higher, but the intensities for *all* peaks are at least 20 times higher than in the first spectrum – potentially indicating that the corresponding peptides are more soluble in methanol. The apparent increased prominence of impurities in the second spectrum should also be noted. This comparison serves to demonstrate the huge effect solvent choice and species solubility can have on the prominence of peaks in MALDI-TOF mass spectra, while highlighting the importance of trialling different sample preparation conditions before even tentative conclusions about the success of a particular synthesis are inferred from mass spectral data.

If no peaks (other than matrix/baseline noise) are observed, an amount of solid larger than the standard 1 mg/mL should be used for sample preparation – particularly for higher molecular weight targets, where the proportion of target to impurities (and the amount of one peptide in a given sample mass) becomes increasingly small. It is also vitally important to remove all small organic by-products resulting from protecting group cleavage during TFA treatment. These impurities can affect the quality of resulting mass spectrum and therefore the crude product should be washed thoroughly in diethyl ether to ensure only peptide species remain before MS sample preparation (see Experimental section). A poor signal to noise ratio may also be improved by desalting the crude peptide prior to analysis – special desalting pipette tips or microcentrifuge tubes are commercially available and can be used for this purpose.[94]

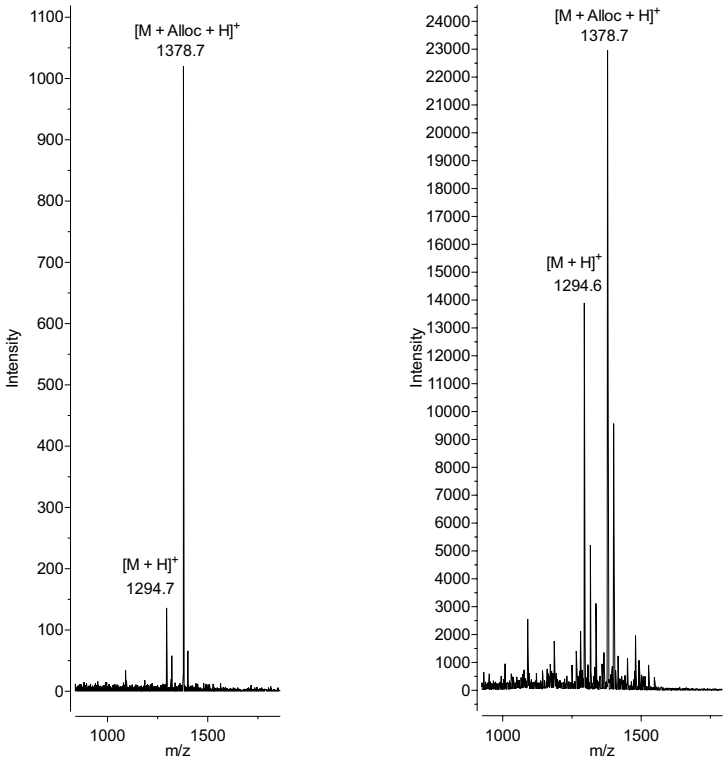


Figure 4.2: MALDI-TOF mass spectra of the crude mixture of peptides **13** (without Alloc group) and **22** (with Alloc). For the left-hand spectrum the sample was prepared using MeCN/water, whereas for the right-hand spectrum MeOH/water was used. The clear difference in peak intensities demonstrates the effect solvent choice during sample preparation can have on the resulting spectra.

4.2 Chromatographic purification of synthetic peptides

Reversed-phase high-performance* liquid chromatography (RP-HPLC) is the technique most commonly used for the routine purification of synthetic peptides and proteins. In most cases the principle impurities are deletion peptides - this is in contrast to recombinantly expressed protein, where the crude product will often be present alongside a variety of (potentially toxic[6]) cell-derived proteins of varying molecular weights in addition to other unwanted cellular matter. Chromatographic methods such as immobilized metal ion affinity chromatography (IMAC) and size-exclusion chromatography (SEC) may be better suited to the purification of expressed protein, with RP techniques used as a final 'polishing step' in the separation protocol. Thierry et al. however reported the use of SEC as an initial purification step in their attempts to obtain chemokines via total chemical synthesis, where preliminary RP-HPLC separation failed.[95] It was reasoned that heavy contamination of the crude product with small organic by-products of protecting group cleavage was to blame, thus SEC was used to remove these before HPLC purification.

The initial chromatographic separation of a number of crude peptides is described in the following sections alongside a discussion of subsequent high resolution purification steps and the optimization of conditions necessary for effective separations.

4.2.1 Initial RP-HPLC separation of crude peptides

Chromatographic separation of crude synthesis products was performed using a semi-preparative SB Analytical C-18 reversed-phase column (5.0 μm, 10.0 × 250 mm). A 30 minute gradient of 0-100% of solvent B was utilized for initial analytical runs. It was found that these small-scale runs (~1 mg loading) facilitated peak assignment (using MALDI-TOF MS) and the optimization of separation conditions prior to scale-up. Once the retention times of any peaks of interest had been ascertained (and conditions optimized), only the gradient region in which the peaks in question eluted was deemed important for scale-

*'High-performance' (or high-pressure) refers to the fact that separations are carried out at a higher pressure than with standard liquid chromatography. This enables the use of shorter columns with a smaller stationary phase particle diameter, therefore allowing faster separations to be a achieved with a higher resolution than normal.

up, and a much steeper gradient was used to reach the point where fraction collection could begin. An illustrative chromatograph – for the separation of peptide **9** (Section 3.4) – is shown in Figure 4.3.

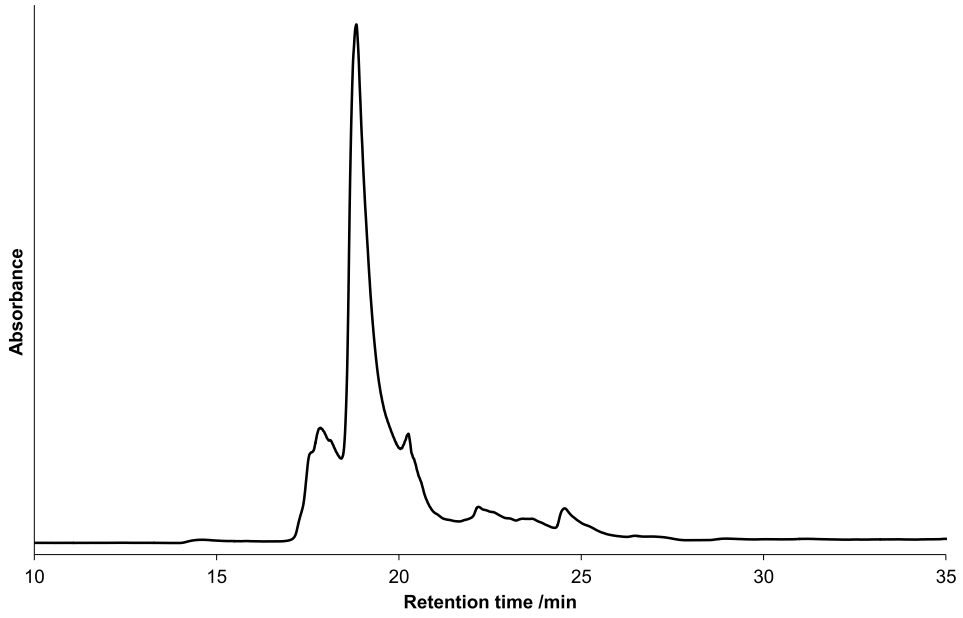


Figure 4.3: Chromatogram for HPLC separation of peptide **9** (CADPKQK-WVQDSMDHLDKQTQTPKT) using a gradient of 0-100% B at a flow rate of 2 mL/min. The solvent composition was as follows: A = 5% MeCN in 0.1% TFA, B = 95% MeCN in 0.1% TFA (further details given in Experimental section). Using a small amount of organic modifier in the A solvent improves wetting of the hydrophobic stationary phase.[96] Absorbance at 280 nm is plotted – the major peak observed corresponds to **9** (m/z = 2929.9). While baseline resolution is not achieved, these conditions were useful as a preliminary purification step prior to high resolution purification on a smaller scale (*vide infra*).

4.2.2 High resolution FPLC purification

High resolution separations were carried out using an ÄKTAexplorer FPLC system equipped with a high resolution SOURCE 15RPC ST 4.6/100 column (reversed-phase). For analytical runs ~0.5 mg in 0.5 mL of solvent A was loaded onto the column per injection, and the separation was run with a gradient of 0-100% solvent B in 20 column volumes at a flow rate of 2 mL/min.

A standard H_2O/MeCN/TFA solvent system was utilized initially, although it was found that a basic buffer system was more suitable for the purification of chemokine-derived sequences (*vide infra*).

While pH was found to be the parameter having the most significant effect on selectivity, other conditions that were optimized in order to improve resolution were flow rate and gradient duration. In the case of a poor separation the gradient slope was first decreased, followed by the flow rate.

An FPLC chromatogram for the separation of crude peptide **9** is shown in Figure 4.4, the peak corresponding to the target peptide (as confirmed by MALDI-TOF MS) is marked with an asterisk (*). It can be seen that the peaks are sharper with a better separation than in Figure 4.3.

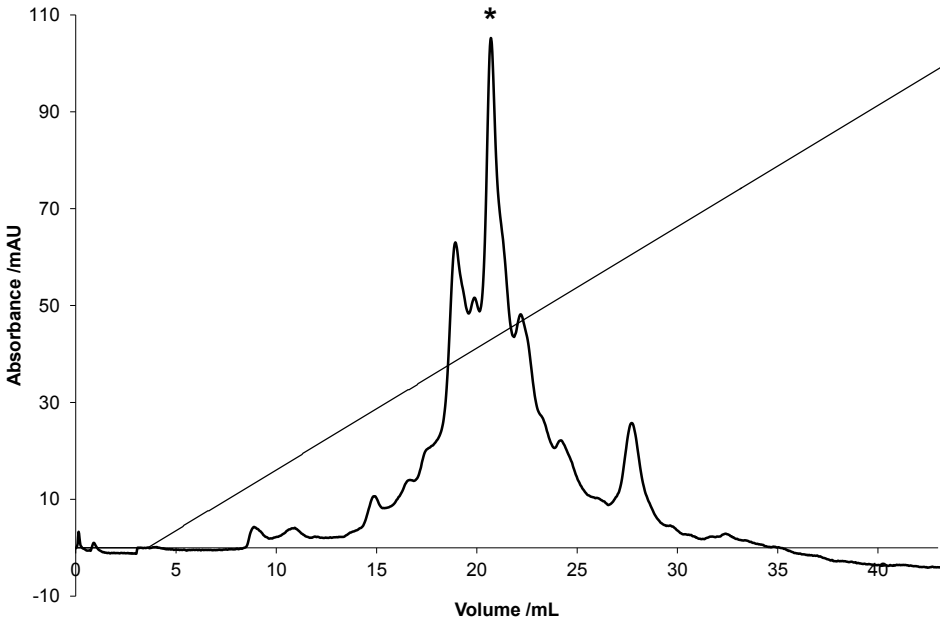


Figure 4.4: Chromatogram for FPLC separation of ligation fragment **9** in basic conditions (*vide infra*). A gradient of 0-80% was used (shown) at a flow rate of 1 mL/min. Absorbance at 280 nm is plotted and the peak corresponding to peptide **9** (m/z = 2929.7) is marked with an asterisk (*). It can be seen that the resolution achieved is significantly higher than for the previous separation (Figure 4.3).

4.2.3 Optimization of solvent pH

Optimum buffer pH is one of the most important parameters to establish, and pH optimization should be one of the first steps in FPLC method development.[96] Figure 4.5 shows an FPLC chromatogram for the C-terminally amidated fragment CCL2(63-76) – SMDHLDKQTQTPKT-amide (**30**) – run in an acidic solvent system (details given in figure caption). **30** was synthesized via automated MW-SPPS on Rink amide AM resin (for linker structure see Figure 2.2), and the crude product subjected to an preliminary HPLC purification as detailed in the Experimental section. It can be seen that the attempt at FPLC purification of the resulting material in acidic conditions is unsatisfactory, as almost no peak separation is achieved (despite using a high resolution column).

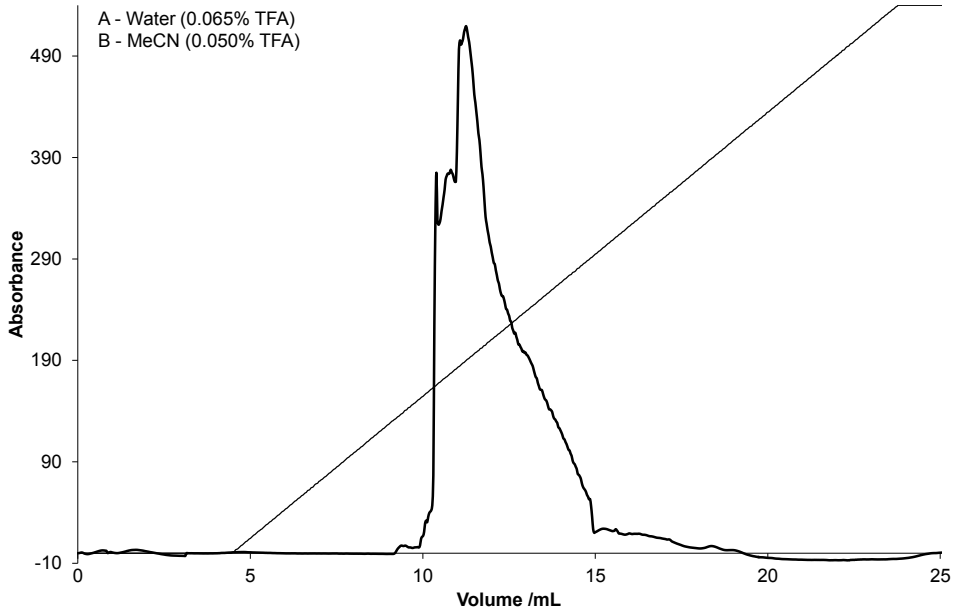


Figure 4.5: Chromatogram for FPLC separation of peptide **30** in acidic conditions. Solvent composition according to the manufacturer's instructions: A = H_2O with 0.065% TFA, B = MeCN with 0.050% TFA (further details given in Experimental section). A gradient of 0-100% was used (shown) at a flow rate of 2 mL/min. Absorbance at 280 nm is plotted. It can be seen that the conditions utilized are not sufficient for the resolution of individual peaks.

The poor resolution achieved using the standard acidic buffer system was thought to be due to the fact that chemokines (and chemokine-derived fragments) are basic species, and as such they are liable to elute as tailing peaks from reversed-phase columns at low pH.[96] SOURCE 15RPC – being a polymeric (polystyrene/divinyl benzene) medium – can be used for separations performed at a pH well above neutral, and so it was hypothesized that a separation under basic conditions may be more effective. A pH 9.0 Tris hydrochloride buffer system was trialled which dramatically improved separations, with almost baseline separation being attained for peptide **30** (Figure 4.6).

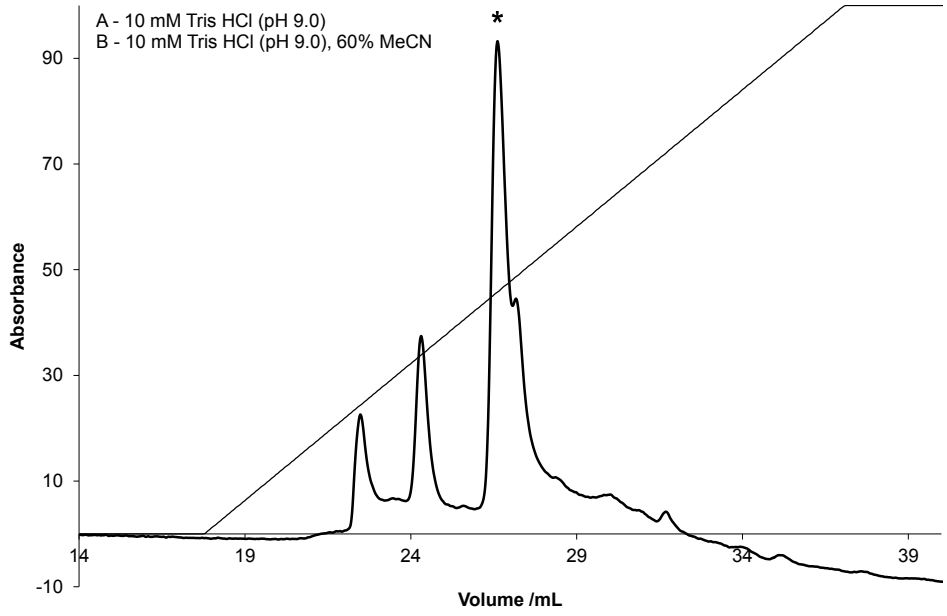


Figure 4.6: Chromatogram for FPLC separation of fragment **30** utilizing a basic buffer system. Buffer composition according to the manufacturer's instructions: A = 10 mM Tris HCl (pH 9.0), B = 60% MeCN in 10 mM Tris HCl (pH 9.0). A gradient of 0-100% was used (shown) at a flow rate of 2 mL/min. Absorbance at 280 nm is plotted and the major peak (corresponding to a Lys or Gln deletion, m/z = 1500.9) is marked with an asterisk (*). The resolution is greatly improved when compared with the separation achieved in acidic conditions (Figure 4.5).

As a result of these findings it is recommended that the FPLC purification of chemokines and chemokine-derived peptide fragments should be carried out using a basic buffer system.

4.3 Summary

Protocols for the chromatographic purification of synthetic chemokine fragments have been established. Semi-preparative RP-HPLC was found to be useful for preliminary purification of the crude material on a large scale, followed by high resolution FPLC purification. While the standard acidic buffer system was initially trialled for FPLC separations, a basic (pH 9.0) Tris hydrochloride buffer system proved to be superior for achieving decent peak separation with the chemokine fragments described. Tris hydrochloride is a non-volatile buffer component and hence must be removed using dialysis from the final product – this is impractical for individual column fractions, and the presence of Tris was found to reduce the sensitivity of MALDI-TOF MS analysis during peak assignment. For this reason a volatile basic buffer component such as ammonium acetate[96] should be investigated for basic separations.

Chapter 5

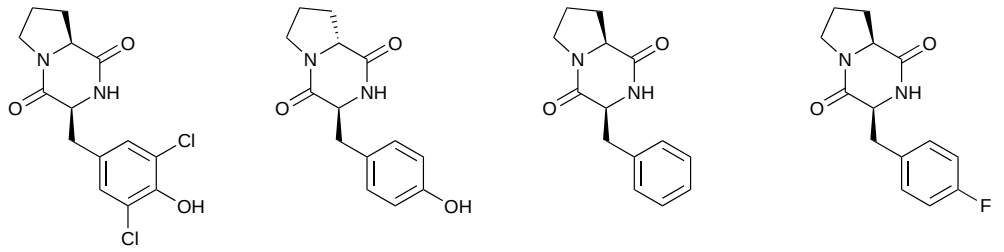
Structure elucidation of small molecule chemotaxis inhibitors

Structural data is crucial for understanding the mechanism of action of any small molecule ligand found to modulate the activity of a cell signalling protein or enzyme. In the search for potential anti-inflammatories targeting CCL2-mediated processes, the three-dimensional structure and conformation of a bioactive lead compound is needed in order to fully understand the interactions underpinning its activity. The most powerful technique for structure elucidation is single crystal X-ray diffraction – although NMR spectroscopy can also provide a wealth of structural information for a particular ligand. Once 3D structures have been obtained for the protein and ligand, so-called 'molecular docking' can be used to investigate protein-ligand interactions computationally.[97–99] During the docking process the optimum 'pose' of the ligand in a binding site on the protein is sought, which can then be used as a model of the conformation of the ligand in the real binding site. The 'correctness' of a particular pose can be evaluated *in silico* by running molecular dynamics simulations or analysing how well the ligand and protein 'fit' together from a purely geometric perspective (the method used depends on the docking approach taken). The end result is a computationally-generated model of how the lead compound interacts with its binding site on the protein, and using this information structural features of the ligand can be proposed that are particularly important for binding.

The structure elucidation of three small molecule inhibitors of CCL2-mediated chemotaxis – with a view to investigating CCL2-inhibitor interactions using molecular docking – is described in the following sections.

5.1 Background

A number of synthetic analogues (**31**, **32** and **33**) of natural product **34** (Figure 5.1) – isolated from the fungus *Leptoxyphium* sp.[100] – were assayed for inhibitory activity against CCL2-mediated monocyte chemotaxis *in vitro* (Mahsa Saleki and Simi Ali, Newcastle University).[101] While compounds **32** and **33** were found to be more potent inhibitors than the natural product (**34**), Western blots showed that stimulation with **32** and **33** did not reduce ERK1/2 phosphorylation – a marker for receptor activation – in the model cell line. This suggests that the inhibitors are not receptor antagonists, and that the mechanism of inhibition does not involve disruption of the interaction between CCL2 and its receptor directly. It was therefore proposed by Saleki et al. that the inhibitory activity of the dipeptides described may be due to an interaction with CCL2 itself, perhaps resulting in the disruption of chemokine dimerization. The exact role of CC chemokine oligomerization in the regulation of leukocyte trafficking is still not fully understood,[26] and it is hoped that a clearer understanding of chemotaxis inhibition – with reference to dimer disruption – may lead to a more precise description of how chemokine dimerization plays a part in this regulatory process. Molecular docking can be used to investigate the nature of the CCL2-inhibitor interactions proposed, although a three-dimensional structure for each ligand is required before this can be carried out. Dipeptide **31** did not show inhibitory activity despite being closely related in structure to the other compounds.



34 **31** **32** **33**

Figure 5.1: Natural product **34** and synthetic analogues **31**, **32** and **33**. Compound **34** has previously been shown to inhibit CCL2-mediated monocyte chemotaxis.[100,101]

5.2 Crystallization

In order to characterise a molecule via single-crystal X-ray diffraction, ordered crystals of the compound of interest must first be obtained. The simplest method involves crystallization from a supersaturated solution of the compound in a solvent in which it has intermediate solubility. To obtain good quality crystals the conditions should be optimized so that crystal formation happens as slowly as possible. This ensures only a limited number of crystals form (nucleation events are rare/ growth only occurs from a small number of seed crystals), resulting in larger crystals with fewer defects or satellites.

If single crystals or crystalline material cannot be obtained from a single solvent, the addition of an *antisolvent* (a solvent in which the compound is insoluble/ has low solubility) may encourage an amorphous-crystalline phase transition. An initial solvent screen – which can be carried out on a microscope slide using < 1 mg of material – is essential in order to assess the solubility of the compound in a range of solvents. Crystalline and amorphous material can be distinguished via observation under a microscope fitted with a pair of crossed polarizing filters due the fact that a crystalline sample will appear bright against a dark background. The tendency of a particular solvent to promote a transition to the crystalline phase is of crucial importance, and it is these solvents which should be used in subsequent crystallization experiments.

Crystalline material was initially obtained via recrystallization of dipeptide **31** from ethanol. Addition of a small amount of ethanol to a saturated solution of **31** in acetonitrile – with gentle heating to ensure complete dissolution followed by seeding – yielded diffraction-quality crystals after 1-2 days. Crystals of **32** were obtained from toluene using the same method.

Compound **33** proved more difficult to crystallize, remaining amorphous in a wide range of solvents. Eventually crystalline material was obtained via recrystallization from a saturated ethanol solution using diethyl ether as an antisolvent, although the crystals were not single. In order to allow crystallization to occur more gradually, a *vapour diffusion* setup was utilized in which an open vial containing a saturated solution of **33** was placed within a larger (sealed) vessel containing (more volatile) diethyl ether, allowing slow diffusion of the antisolvent into the compound solution. A small number of crystals suitable for testing were obtained using this method.

5.3 Data collection

X-ray diffraction data were collected at 100 K on a Bruker MicroStar diffractometer using a Cu-rotating anode. A photograph of a crystal of **33** mounted on the goniometer head of the diffractometer is shown in Figure 5.2. Crystallographic data for compounds **31**, **32** and **33** are summarized in Table 5.1. The absolute configuration was confirmed by the Flack parameter (given in the table).[102]

Unfortunately a complete data set could not be obtained for compounds **32** and **33** as data collection had to be aborted due to technical problems (cryosystem failure). In addition, only a small number of reflections beyond 1 Å resolution were collected for **32** and **33** due to limited crystal quality. For these reasons the overall completeness (∞ to 0.85 Å) reached only 70% (Friedel pairs not merged).

A complete data set should be collected for each compound, and further crystallization attempts are ongoing. Nevertheless the data were sufficient to solve both structures and assign absolute configuration.

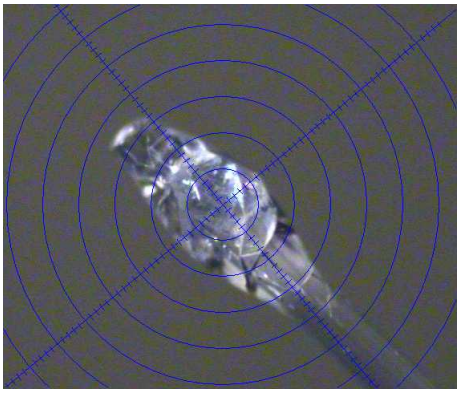


Figure 5.2: Crystal of compound **33** mounted on the diffractometer goniometer head (glass fibre in Fomblin oil).

Table 5.1: Crystallographic data for **31**, **32** and **33**.

	Dipeptide **31**	Dipeptide **32**	Dipeptide **33**
Empirical formula	$C_{14}H_{16}N_2O_3$	$C_{14}H_{16}N_2O_2$	$C_{14}H_{15}FN_2O_2$
M_r	260.29	244.29	262.28
T [K]	100	100	100
Crystal system	Hexagonal	Monoclinic	Monoclinic
Space group	$P6_3$	$P2_1$	$P2_1$
a [Å]	16.3840(3)	5.5875(7)	11.7856(6)
b [Å]	16.3840(3)	10.0300(14)	9.0051(4)
c [Å]	8.9875(2)	10.6734(15)	12.4478(6)
α [°]	90.00	90.00	90.00
β [°]	90.00	92.272(7)	109.1380(10)
γ [°]	120.00	90.00	90.00
Resolution [Å]	0.843	0.849	0.868
V [Å^3]	2089.34(7)	597.70(14)	1248.08(10)
Z	8	2	4
Z'	1	1	2
ρ_{calc} [mg mm^{-3}]	1.209	1.271	1.061
μ [mm^{-1}]	0.712	0.740	0.754
F(0 0 0)	800.0	229.0	397.0
Reflections collected	12958	2832	6500
Independent reflections	2456	1361	2794
Data/restraints/parameters	2456/1/183	1361/1/163	2794/16/343
Completeness (∞ to 0.85 Å) [%]	96.0	69.7	70.1
R_1, wR_2 [$I > 2\sigma(I)$][a]	0.0228, 0.0669	0.0499, 0.1266	0.0428, 0.1267
R_1, wR_2 [all data][a]	0.0228, 0.0670	0.0503, 0.1276	0.0429, 0.1273
Largest diff. peak, hole [e Å^{-3}]	0.12, -0.14	0.28, -0.27	0.28, -0.19
Flack parameter[102]	0.04(14)	0.3(3)	0.16(16)

[a] R-factors calculated using

$$R_1 = \frac{\sum ||F_0| - |F_C||}{\sum |F_0|} \qquad \text{and} \qquad wR_2 = \sqrt{\frac{\sum w(F_0^2 - F_C^2)^2}{\sum w \cdot F_0^4}}$$

where w is a weighting term.

5.4 Molecular structure and packing

5.4.1 *Cyclo*(L-Tyr-D-Pro)

The molecular structure of dipeptide **31** is shown in Figure 5.3. The 6-membered diketopiperazine ring adopts a shallow boat conformation, while the bicycle as a whole exists as the twist-boat conformer. The boat conformation of the 6-membered ring is characteristic of proline-containing 2,5-DKP compounds. In contrast, the 6-membered ring in monocyclic DKPs is highly flexible and can adopt a twist-boat or planar conformation.[103–105] The phenol side chain of **31** is folded towards the DKP ring system – this is in contrast to the conformers observed for dipeptides **32** and **33**, for which the side chain is extended (*vide infra*).

A molecular dynamics simulation of the enantiomer of **31** carried out by Li et al.[106] gave an optimized geometry almost identical to the one presented, suggesting that the folded conformation observed is not simply an artefact of intermolecular packing interactions within the crystal. It has been suggested that nonsteric factors play a role in dictating the conformational distribution for such systems, and that the folded conformer may predominate due to interaction of the aromatic ring with the π system of the amide bond.[107]

An X-ray diffraction study of *cyclo*(L-Tyr-L-Pro) found that the diastereomer also exists as the folded conformer, although the DKP ring adopts a flattened chair conformation in this case.[108] This suggests that it is not simply the difference in proline configuration between **31** and the other synthetic dipeptides that causes it to exist uniquely in the folded conformation.

The crystal packing arrangement is shown in Figure 5.4. Interestingly, the structure is composed of helices which span the length of the crystal and associate via O19–H19···O11 hydrogen bond contacts (side chain hydroxyl and ring carbonyl). Molecules of the solvent (MeCN) are interspersed amongst these channels, which are arranged in parallel within the crystal structure. Within each helix the individual molecules interact by N4–H4···O10 hydrogen bonds. A single helix along with the solvent of crystallization is shown in Figure 5.5 – the helix is left-handed and has 6 molecules per turn. Electron density occurring within the channels themselves (and which could not be assigned) was thought to be due to the presence of further residual disordered solvent.* 2,5-Diketopiperazines are molecules of interest

*Solvent masking was used during structure refinement (see Experimental section).

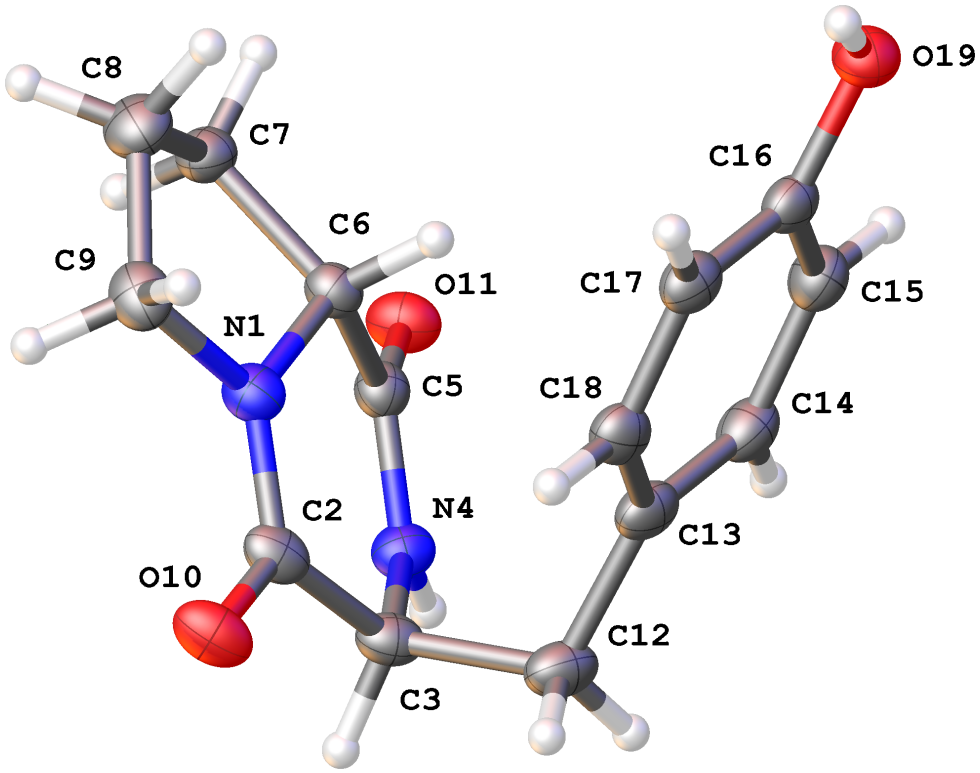


Figure 5.3: Molecular structure of compound **31** showing 50% probability anisotropic displacement ellipsoids.

in crystal engineering,[109,110] and the X-ray diffraction data presented here reveal a novel, highly ordered supermolecular structure of potential use in the development of materials that exhibit nonlinear optical properties and electrical superconductivity.[105]

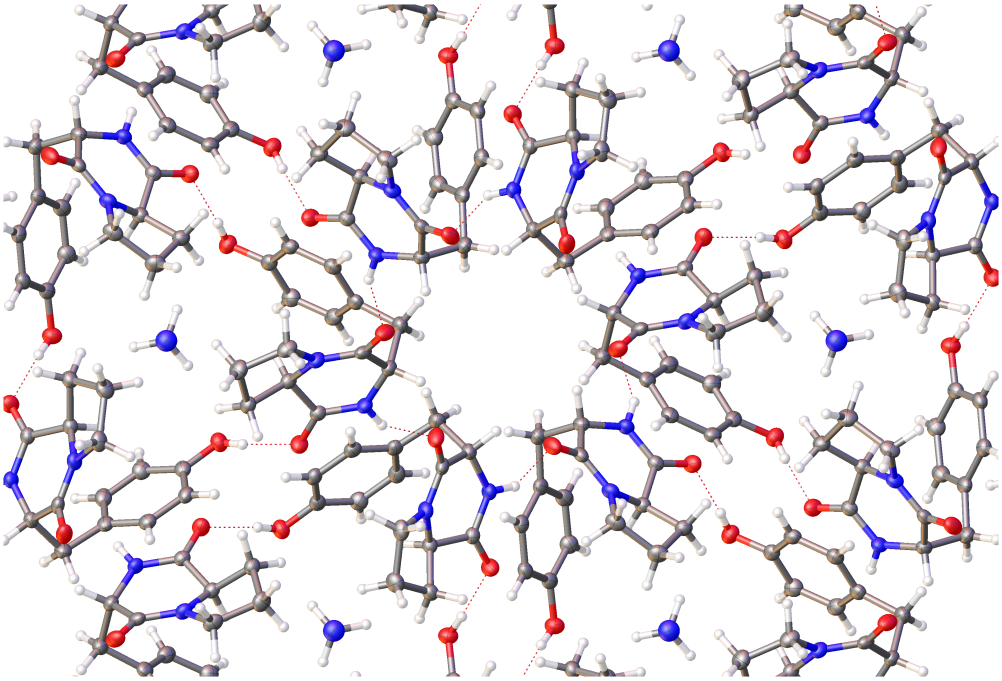


Figure 5.4: Crystal packing arrangement for compound **31**. The structure is comprised of hexagonal channels surrounded by ordered molecules of acetonitrile. Residual disordered solvent (MeCN) appears to be present inside the channel (see text).

5.4.2 *Cyclo*(L-Phe-L-Pro)

The molecular structure of **32** is shown in Figure 5.6. The 6-membered DKP ring adopts a shallow boat conformation as previously observed, although the phenylalanine side chain is extended unlike the tyrosine side chain in the folded conformer observed for compound **31**. The packing arrangement is stabilized by N4–H4···O10 hydrogen bonding.

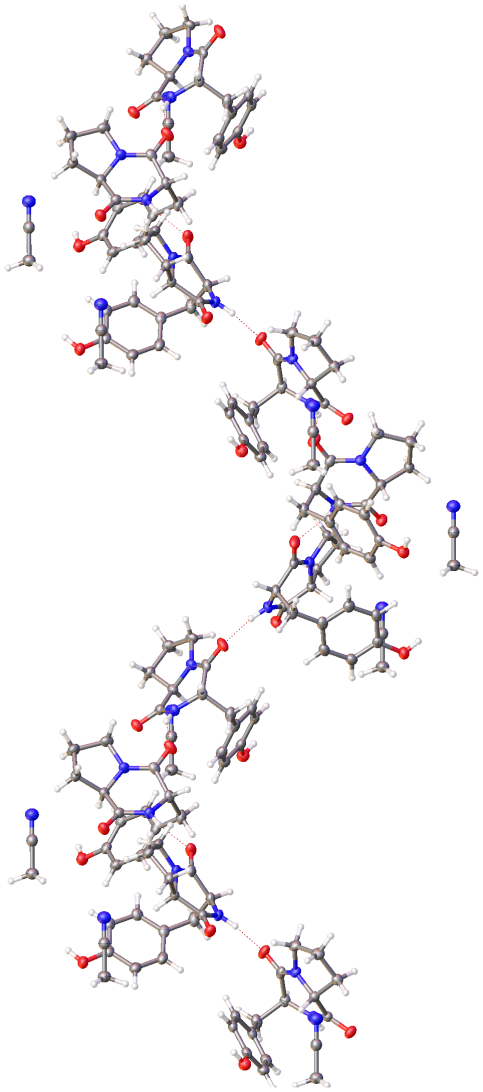


Figure 5.5: Helical arrangement of molecules in the crystal packing observed for compound **31** (ordered solvent molecules are shown). A N4−H4···O10 hydrogen bond stabilizes the helix.

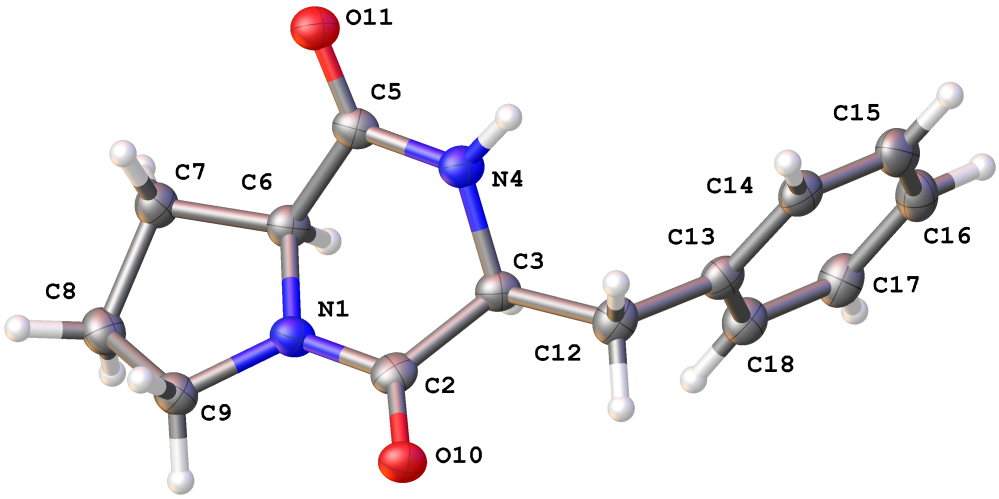


Figure 5.6: Molecular structure of compound **32** showing 50% probability anisotropic displacement ellipsoids.

5.4.3 *Cyclo*(L-*p*-FPhe-L-Pro)

The diffraction data for **33** reveal that the dipeptide crystallizes with two molecules in the asymmetric unit ($Z' = 2$). Crystal structures with $Z' > 1$ are rare, and the phenomenon often occurs for chiral molecules that pack awkwardly with each other in the solid state.[111,112] Current opinion is divided as to whether these structures are thermodynamic minima or simply metastable forms obtained under kinetic conditions, and discussions of the nature and reasons for their formation are ongoing.[113] The asymmetric unit is shown in (Figure 5.7). For both molecules the DKP ring adopts a shallow boat conformation with the side chain extended. The 5-membered ring incorporating C27 is disordered, and consequently the bicyclic ring system exists as both the chair (C27A) and boat (C27B) conformer.

A similar 2,5-DKP compound was found to have $Z' > 1$ in an X-ray crystallographic study by Budesinsky et al.[114] Data collection was carried out at both 150 K and room temperature, with reversible changes in crystal lattice parameters and packing geometry occurring during the transition. In light of this result the authors suggested that the reason for $Z' > 1$ in the low-temperature structure was likely to be kinetic. Structural data for compound

33 were collected at ~100 K and therefore it is conjectured that the structure obtained is a metastable form.

The apparent 'packing frustration' and inability of the side chain fluorine to act as a hydrogen bond donor may account for the relatively low density[113] of **33** (see Table 5.1), and the difficulty experienced during attempts to crystallize it in comparison with the tyrosine derivative (**32**).

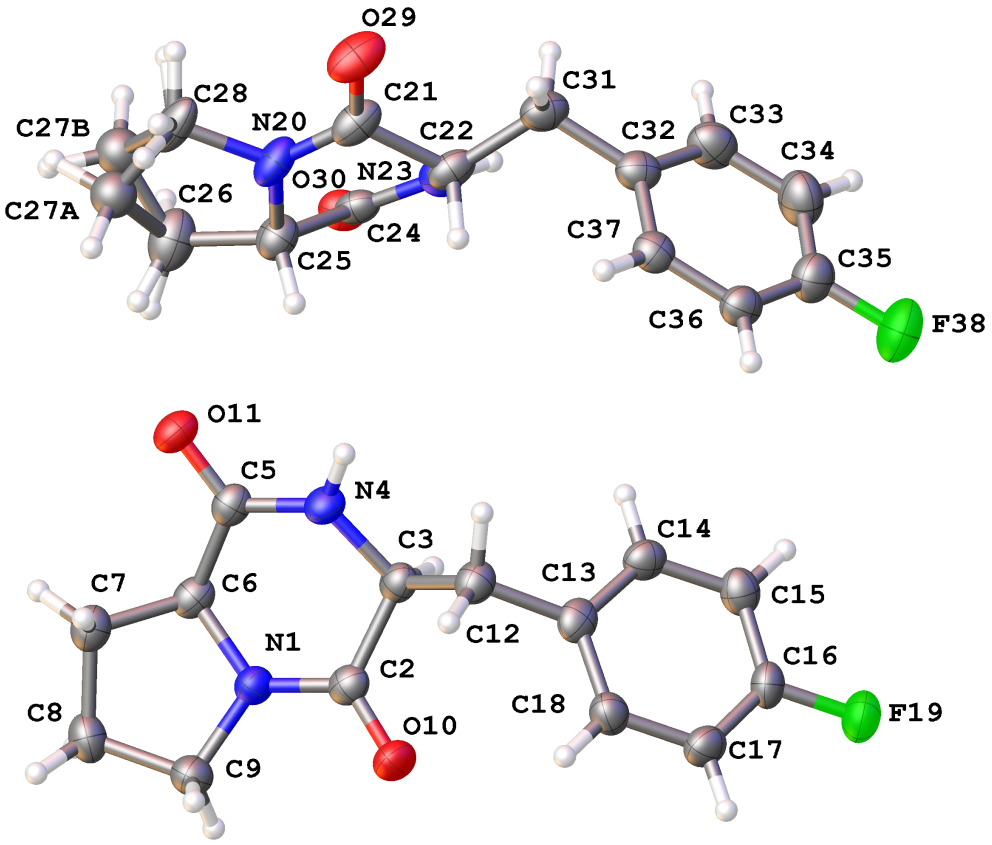


Figure 5.7: Molecular structure of compound **33** showing 50% probability anisotropic displacement ellipsoids (with the exception of C27, which could not be refined anisotropically).

5.5 Summary

Diffraction-quality single crystals were obtained for two inhibitors of *in vitro* monocyte chemotaxis (**32** and **33**), in addition to a third structurally similar compound (**31**) that exhibited no inhibitory activity. X-ray diffraction data were collected successfully for all three compounds and their structures solved. It is suggested that the lack of inhibitory activity for **31** is due to the dipeptide existing in a folded conformation, while the inhibitors in question both have an extended aromatic side chain. This may be due to the presence of the tyrosine hydroxyl (absent from the inhibitor structures), which may play a role in the purported nonsteric interactions previously discussed. The fact that the natural product exhibits inhibitory activity (despite possessing a hydroxyl group) can potentially be accounted for by considering the effect of steric hindrance from the ring chlorines, which may prevent occurrence of the folded conformation – although a crystal structure of this molecule is required in order to further elucidate the nature of this interaction.

Ligand structural data are required in order to carry out molecular docking studies, and to establish whether chemokine-DKP interactions may play a part in chemotaxis inhibition. These results will therefore not only be relevant to investigations of the 2,5-DKP scaffold for anti-inflammatory drug discovery, but will also find use in continued efforts to elucidate the role of chemokine-chemokine and chemokine-receptor interactions in CCL2-mediated monocyte chemotaxis.

Chapter 6

Conclusions and future work

Significant progress has been made towards the total chemical synthesis of the chemokine CCL2. In addition, the structures of a number of inhibitors of CCL2-mediated chemotaxis have been obtained via single crystal X-ray diffraction.

Attempts to synthesize the whole chemokine in one piece using Fmoc SPPS were hindered by the on-resin hydrophobic aggregation of the peptide during elongation. In order to avoid aggregation, the PEG-based ChemMatrix® polymer support was investigated in combination with pseudoproline dipeptides as an alternative to conventional polystyrene-based resin. It was decided however that a synthesis involving the ligation of smaller fragments of CCL2 in solution might be a more promising route, and the work already completed showed that microwave-assisted SPPS using conventional polystyrene Wang resin was the most effective method for the generation of acid fragments of up to ~40 residues in length.

A total chemical synthesis involving the native chemical ligation of four fragments of CCL2 was proposed (Section 3.1), and the synthesis of two of the required fragments – the C-terminal peptide acid and a peptide thioester – has been completed. Room temperature SPPS using the sulfonamide safety-catch linker was found to be effective for producing the thioester, while the peptide acid was synthesized via MW-SPPS using polystyrene Wang resin. Versions of the thioester with and without the N-terminal cysteine masked as a thiazolidine have been synthesized for potential use in a sequential or fully convergent ligation strategy respectively (should the former be required). The Dawson Dbz linker was also investigated as a method of generating peptide thioesters, and progress towards a third fragment has also been made. While

'over-acylation' of the Dbz moiety proved to be an issue for microwave-assisted syntheses initially, orthogonal protection of the linker eliminated this problem resulting in the first reported use of the Dawson Dbz linker for MW-SPPS.

Protocols for the purification of synthetic CCL2-derived sequences using RP-HPLC and high resolution FPLC have been developed, and the peptide acid fragment described has been successfully purified. It was found that separations of basic chemokine fragments were more effective when carried out using a Tris buffer at high pH. Tris hydrochloride is a non-volatile buffer component however, and the use of a more volatile compound such as ammonium acetate may allow more facile removal from column fractions, increasing the sensitivity of MALDI-TOF MS analysis. The development of purification protocols for the remaining peptide thioesters is also necessary.

Once all the required ligation fragments have been synthesized and purified, suitable conditions for ligations, correct folding and oxidation[15] of the chemokine will be investigated. It is anticipated that a combination of RP-HPLC, FPLC and SEC will be utilized in the purification of the complete protein. Upon obtaining purified CCL2, biological testing (chemotaxis and calcium mobilization assays[15,101]) will be used to verify that the protein obtained is the desired product and that it is folded correctly. Circular dichroism (CD) spectroscopy will also be used to confirm correct folding.

Alongside the synthetic work, single crystals of three synthetic analogues of a naturally occurring CCL2 inhibitor were obtained. X-ray diffraction data were collected and molecular structures for the three compounds successfully determined. These will be useful for future molecular docking studies, with the next step being the use of a pocket finding algorithm[115] to identify potential ligand binding sites on the surface of CCL2 using existing crystal structure data for the protein.[15,61] The inhibitor crystal structures can then be docked with each of the candidate sites on the surface of CCL2 in order to determine for which of these the binding affinity would be highest. Isothermal titration calorimetry (ITC) can be used to quantify the strength of this interaction. A binding site occurring at the dimerization interface would lend support to the hypothesis that the inhibitory activity of the compounds described is due to CCL2 dimer disruption. Cocrystallization of the CCL2 molecule with the inhibitors discussed would allow detailed examination of the protein-ligand interactions involved via X-ray diffraction experiments.

When complete, the total chemical synthesis of CCL2 will enable the generation of site-specifically labelled versions of the chemokine which will be extremely valuable for use in biomedical research. As well as enabling the facile production of fluorescently labelled and/or mutant chemokine for use in biological assays, a disulphide-bonded obligate dimer[26] can be chemically synthesized, which could be used in cell-based studies with the aim of further elucidating the role of CCL2 dimerization with regards to chemotactic activity. The formation of a covalent dimer may also be possible using Cu(I)-catalysed azide-alkyne 'click' cycloaddition to link azide- and alkyne-modified CCL2 monomers,[116] allowing nonsymmetric labelling[41] of the dimer. Further characterisation of the monomer-dimer dynamic equilibrium may be accomplished through a Förster resonance energy transfer (FRET) experiment[117,118] – this requires two sets of monomer each labelled with a different chromophore, which will be readily accessible via chemical synthesis. Finally, a synthetic route will enable the site-specific incorporation of PTMs, which will be utilized to investigate CCL2 chemokine nitration, a process which regulates leukocyte infiltration in tumours and is therefore relevant to drug research in the field of cancer immunotherapy.[119]

It can be seen that the protein tools accessible using total chemical synthesis – as well as the small molecule ligands characterised – will be extremely useful for research in the field of chemokine immunobiology. Not only will these tools facilitate the investigation of the CCL2 oligomerization process, but also the complex role of PTMs in the regulation of the immune response. More generally, the techniques described will find broad application in the field of chemical protein synthesis as a whole, allowing access to a larger range of proteins, thus enabling the detailed characterisation of their interactions in living systems.

Chapter 7

Experimental

7.1 Solid-phase peptide synthesis

All reagents were purchased from Sigma-Aldrich unless otherwise specified. Peptide synthesis grade DMF was purchased from AGTC Bioproducts (Hessle, UK) and amino acid derivatives were purchased from CEM, Novabiochem (Merck) or AGTC. HBTU was purchased from CEM, TBTU from AGTC and PyBOP® from Apollo Scientific (Stockport, UK). All resins were purchased from Novabiochem with the exception of ChemMatrix®, which was purchased from Sigma-Aldrich.

Fmoc SPPS procedures are detailed in the following sections. Fritted polypropylene reaction vessels were used for all manual reactions and resin swelling, with DMF as the reaction solvent. For peptide couplings HBTU, TBTU or PyBOP® was used as the activator. Amino acid side chain functionality was protected as follows: Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(*t*Bu)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(*t*Bu)-OH, Fmoc-His(Trt)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Ser(*t*Bu)-OH, Fmoc-Thr(*t*Bu)-OH, Fmoc-Trp(Boc)-OH and Fmoc-Tyr(*t*Bu)-OH. Fmoc deprotections were carried out using a 20% (v/v) solution of piperidine in DMF. Pre-swelling of the resin was carried out in DCM/DMF (1:1) for a minimum of 1 h (overnight preferred), followed by washing with DMF. The resin was shrunk in diethyl ether before storage and to remove DMF in preparation for TFA cleavage.* Peptide-resin was stored at −18 °C.

*It should be noted that ChemMatrix® resin *swells* in MeOH, which is conventionally used for shrinking polystyrene-based supports.[73]

7.1.1 Automated Fmoc SPPS

Automated SPPS was carried out on a CEM Liberty1 single-channel microwave peptide synthesizer equipped with a Discover microwave unit. All reactions were carried out using the 30 mL PTFE reaction vessel, with microwave heating and agitation by bubbling nitrogen. Couplings were carried out using a 5-fold excess of Fmoc-protected amino acid, activator (10 equiv) and DIPEA (20 equiv, 2 M solution in NMP) unless otherwise stated.[†] For double and triple couplings the reaction vessel was drained after each cycle and fresh reagents were added.

Microwave-assisted

Microwave couplings at 0.05 mmol scale were carried out for 10 min at 75 °C and 20 W power unless otherwise stated. Cys and His residues were coupled at low temperature: 10 min at room temperature followed by 10 min at 50 °C (20 W). Arg residues were double coupled, with the first coupling carried out for 45 min at rt followed by 5 min at 75 °C (20 W), and the second using the standard microwave conditions given. The Fmoc group was removed by a single treatment with piperidine solution at 75 °C (35 W) for 3 min. For 0.1 mmol scale syntheses the power was increased to 25 W for couplings and 45 W for deprotections.

Room temperature

Couplings were carried out at room temperature for 60 min. The Fmoc group was removed by two successive treatments with piperidine solution (5 min then 10 min).

7.1.2 Manual microwave-assisted Fmoc SPPS

Manual SPPS was carried out in fritted polypropylene reaction vessels using the Discover SPS system. Couplings and deprotections were carried out using the conditions for automated MW-SPPS already described (*vide supra*).

[†]N.B. For syntheses at 0.05 mmol scale the mass of activator calculated by the Liberty1 software should be halved in order to give the correct number of equivalents when making up the activator solution. This is especially important for HBTU and TBTU, as uronium type coupling reagents can cap the peptide N-terminus if used in excess.

Efficient mixing during the synthesis was achieved using a magnetic stirrer bar (without nitrogen agitation). The success of each coupling/deprotection step was confirmed using the qualitative monitoring protocols described below.

7.1.3 Alloc protection and loading of Dbz resin

The Dawson Dbz linker was Alloc protected and loaded using the procedure described by Mahto et al.[83] 3-(Fmoc-amino)-4-aminobenzoyl (Dawson Dbz) AM resin was first treated with allyl chloroformate (10 equiv) and DIPEA (1.0 equiv) in DCM (3 mL) for 24 h with agitation. The resin was washed with DCM (3×5 mL) and DMF, and the linker deprotected manually with two piperidine treatments at room temperature (as described previously in procedure 7.1.1). After thorough washing with DMF, the resin was treated with Fmoc-Ile-OH (15 equiv), HATU (13.5 equiv) and DIPEA (27 equiv) (1:0.9:1.8) in DMF (3 mL), and agitated for 1 h at room temperature. The solvent was removed and the coupling step repeated. The resin was washed with DMF and the remaining unreacted amine groups were capped via treatment with a 20% (v/v) solution of acetic anhydride in DMF for 30 min.

7.1.4 Qualitative tests for the presence of free amines

TNBS test

The TNBS test was used to confirm the presence of primary amines during SPPS.[74] Approximately 10 resin beads were added to a microcentrifuge tube containing 3 drops of a 1% (w/v) solution of 2,4,6-trinitrobenzenesulfonic acid (TNBS) in DMF, and 3 drops of a 10% (v/v) solution of DIPEA in DMF. A positive result (presence of primary amine) was indicated by a red colouration of the beads, whereas a grey/yellow colouration remained in the case of a negative result. Observation of the beads under a microscope was used to identify colouration of the bead interior that could not be detected with the naked eye.

Acetaldehyde/chloranil test

The acetaldehyde/chloranil test was used to confirm the presence of secondary amines during SPPS.[75] Approximately 10 resin beads were added to a microcentrifuge tube containing 3 drops of a 2% (w/v) solution of chloranil

in DMF, and 3 drops of a 2% (v/v) solution of acetaldehyde in DMF. A positive result (presence of primary or secondary amine) was indicated by a dark blue/green colouration of the beads, whereas the resin remained colourless/yellow in the case of a negative result. Observation of the beads under a microscope was used to identify colouration of the bead interior that could not be detected with the naked eye. Reagent solutions were not kept any longer than one month.

7.1.5 Cleavage of peptides from acid-labile resins

Peptide-resin was treated with 0.95 mL TFA and 0.05 mL deionized water, with 0.05 mL TIPS as a scavenger for 3-4 h at room temperature (values were tripled when the mass of resin was greater than 100 mg). Longer cleavage times were used for arginine-containing peptides. The resin was then removed by filtration and the filtrate split into two portions which were each added dropwise to 45 mL of ice-cold diethyl ether. After centrifugation the supernatant was removed and the pellets resuspended in fresh ether and centrifuged again. Following further centrifugation, the supernatant was discarded. For cleavages where the mass of peptide was deemed to be too small for precipitation/centrifugation to be feasible the TFA was first removed under vacuum before precipitation using ether and decanting of the liquid (followed by subsequent ether washes). The resulting solid peptide was dissolved in deionized water and lyophilized. For hydrophobic sequences a small amount of acetonitrile was added ($< 50\%$ v/v) in the event of poor solubility (water and MeCN containing 0.1% TFA were also used to improve solubility for particularly insoluble peptides).

7.1.6 Cleavage of peptides from Dbz resin

Activation and cyclization[83]

Alloc-protected peptide-resin was first deprotected using the method detailed below. The resin was then treated with *p*-nitrophenylchloroformate (5 equiv) in DCM (2.5 mL) for 1 h with nutation at room temperature. The reaction vessel was drained, the resin washed with DCM (3×5 mL) and DMF (3×5 mL), and DIPEA was added in DMF (0.5 M). After agitation for a further 45 min the solvent was removed and the resin washed in DMF.

Alloc deprotection[83]

Alloc-protected peptide-resin was swollen in DCM for 1 h and sparged with argon for 5-10 s. $PhSiH_3$ (20 equiv) and $Pd(PPh_3)_4$ (0.35 equiv) were added in a small amount of DCM and the reaction was agitated for 30 min at room temperature. The resin was then washed with DCM and DMF.

7.1.7 Cleavage of peptides from sulfonamide resin

Peptide-resin was treated with iodoacetonitrile (10 equiv) and DIPEA (10 equiv) in DMF (1 mL) for 24 h with agitation in a foil-wrapped reaction vessel.[87,120] The solvent was removed and the resin washed with DMF (3 × 5 mL), THF (3 × 5 mL) and DMF again. Ethyl 3-mercaptopropionate (50 equiv) and sodium thiophenolate (0.5 equiv) were then added in DMF (2 mL) and the reaction agitated for a further 24 h. After removal of the resin beads via filtration, the resulting protected peptide was precipitated in excess ice-cold diethyl ether (200 mL). The solution was chilled for 2.5 h and the fine precipitate was isolated via filtration of the suspension through a glass fibre filter. The filtrand was washed with a large amount of ice-cold ether and global deprotection of the product was achieved via TFA treatment of the filter/filtrand (procedure 7.1.5).

7.1.8 CCL2 fragments

CCL2(53-76), ADPKQKWVQDSMDHLDKQTQTPKT (6)

Synthesized via automated MW-SPPS (procedure 7.1.1) on low loaded Fmoc-Thr(*t*Bu)-Wang polystyrene resin (0.27 mmol/g substitution, 0.05 mmol scale). Resin cleavage and deprotection using TFA was carried out according to procedure 7.1.5 to afford fragment **6**, MALDI-TOF mass: found *m*/*z* 2827.9 $(M + H)^+$, calcd 2827.1 $(M + H)^+$.

CCL2(52-76), CADPKQKWVQDSMDHLDKQTQTPKT (9)

Synthesized via automated MW-SPPS (procedure 7.1.1) on low loaded Fmoc-Thr(*t*Bu)-Wang polystyrene resin (0.27 mmol/g substitution, 0.10 mmol scale). Resin cleavage and deprotection using TFA was carried out according to procedure 7.1.5 to afford fragment **9** (158 mg, 54%), MALDI-TOF mass:

found *m/z* 2929.8 $(M + H)^+$, calcd 2930.3 $(M + H)^+$. The crude product was purified via HPLC (procedure 7.2.1) and FPLC (procedure 7.2.2), using a flow rate of 1 mL/min over 30 column volumes for the latter.

CCL2(39-76), EAVIFKTIVAKEICADPKQKWVQDSMDHLDKQTQTPKT (35)

Synthesized via automated MW-SPPS (procedure 7.1.1) on low loaded Fmoc-Thr(*t*Bu)-Wang polystyrene resin (0.27 mmol/g substitution, 0.05 mmol scale). Resin cleavage and deprotection using TFA was carried out according to procedure 7.1.5 to afford fragment **35**, MALDI-TOF mass: found *m/z* 4372.9 $(M + H)^+$, calcd 4372.0 $(M + H)^+$.

CCL2(42-51), IFKTIVAKEI-diaminobenzamide (22)

3-(Fmoc-amino)-4-aminobenzoyl (Dawson Dbz) AM resin (0.42 mmol/g substitution, 0.03 mmol scale) was first Alloc protected, loaded with Fmoc-Ile-OH and capped according to procedure 7.1.3. The peptide was synthesized on the Alloc-protected resin via automated MW-SPPS (procedure 7.1.1), with Boc-Ile-OH coupled as the N-terminal amino acid. Alloc deprotection and cyclization of the linker were then carried out according to procedure 7.1.6, and subsequent TFA treatment (procedure 7.1.5) resulted in cleavage from the resin and global deprotection to afford Alloc-protected diaminobenzamide **22** (3 mg, 9%), MALDI-TOF mass: found *m/z* 1294.7 $(M + H)^+$, calcd 1295.6 $(M + H)^+$.

CCL2(26-35), ASYRRITSSK-thioester (24)

Synthesized via automated SPPS at room temperature (procedure 7.1.1) on H-Lys(Boc)-Sulfamylbutyryl NovaSyn® TG resin (0.2 mmol/g substitution, 0.03 mmol scale), with Boc-Ala-OH coupled as the N-terminal amino acid. Linker activation and thiolysis were carried out according to procedure 7.1.7, and global deprotection of the product (procedure 7.1.5) with an extended cleavage time of 7 h afforded thioester **24**, MALDI-TOF mass: found *m/z* 1284.6 $(M + H)^+$, calcd 1285.5 $(M + H)^+$.

CCL2(12-35), Cys-YNFTNRKISVQRLASYRRITSSK-thioester (7)

Synthesized via automated SPPS at room temperature (procedure 7.1.1) on H-Lys(Boc)-Sulfamylbutyryl NovaSyn® TG resin (0.2 mmol/g substitution, 0.03 mmol scale). Double couplings were used for all residues, except for those which were triple coupled (given in Section 3.3.2) in anticipation of difficult couplings predicted by the Peptide Companion software.[66] Boc-Cys(Trt)-OH was triple coupled manually as the N-terminal amino acid under the conditions used for previous couplings. Linker activation and thiolysis were carried out according to procedure 7.1.7, and global deprotection of the product (procedure 7.1.5) with an extended cleavage time of 7 h afforded thioester **7**, MALDI-TOF mass: found *m/z* 3008.0 $(M + H)^+$, calcd 3009.5 $(M + H)^+$.

CCL2(12-35), Thz-YNFTNRKISVQRLASYRRITSSK-thioester (29)

Synthesized via automated SPPS at room temperature (procedure 7.1.1) on H-Lys(Boc)-Sulfamylbutyryl NovaSyn® TG resin (0.2 mmol/g substitution, 0.03 mmol scale). Double couplings were used for all residues, except for those which were triple coupled (given in Section 3.3.2) in anticipation of difficult couplings predicted by the Peptide Companion software.[66] Boc-Thz-OH was triple coupled manually as the N-terminal amino acid under the conditions used for previous couplings. Linker activation and thiolysis were carried out according to procedure 7.1.7, and global deprotection of the product (procedure 7.1.5) with an extended cleavage time of 7 h afforded thioester **29**, MALDI-TOF mass: found *m/z* 3019.7 $(M + H)^+$, calcd 3020.5 $(M + H)^+$.

7.2 Peptide purification and characterisation

7.2.1 Reversed-phase high-performance liquid chromatography (RP-HPLC)

Crude peptides were purified using a Speck and Burke Analytical C-18 column (5.0 μm, 10.0 × 250 mm) attached to a PerkinElmer Series 200 LC Pump and 785A UV/Vis Detector. A linear gradient of 0-100% B (solvent A = 5/95/0.1 MeCN/H_2O/TFA; B = 95/5/0.1 MeCN/H_2O/TFA) over 30 min with a flow rate of 2 mL/min was used for separations unless otherwise specified. Absorbance data were collected at 280 nm. 0.5 mL of a 1 mg/mL solution was injected for analytical runs, from which selected peak fractions were lyophilized and a mass assigned using MALDI-TOF MS. For preparative runs, concentrations of up to 20 mg/mL (0.5 mL) were injected and peak fractions of interest pooled and lyophilized. The peptide solution was centrifuged before injection to remove undissolved material.

7.2.2 Fast protein liquid chromatography (FPLC)

High resolution purification was carried out using an ÄKTAexplorer FPLC system equipped with a SOURCE 15RPC ST 4.6/100 column (15 μm, 4.6 × 100 mm) (GE Healthcare). A linear gradient of 0-100% B (solvent A = 10 mM Tris HCl, pH 9.0; B = 60% MeCN in 10 mM Tris HCl, pH 9.0) over 20 column volumes with a flow rate of 2 mL/min was used for separations unless otherwise specified. The fraction size was set at 1 mL and absorbance data were collected at both 220 and 280 nm. 0.5 mL of a 1 mg/mL solution was injected per run. The peptide solution was centrifuged before injection to remove undissolved material.

7.2.3 Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)

MALDI-TOF mass spectra we collected using an Autoflex II ToF/ToF mass spectrometer (Bruker Daltonik GmBH) equipped with a 337 nm nitrogen laser. Peptides were dissolved in 1:1 deionized water/MeCN for MS analysis (solvents containing 0.1% TFA were used for particularly insoluble analytes). Sample solution (1 mg/mL) was mixed with matrix solution (α-cyano-4-

hydroxy-cinnamic acid, ~50 mg/mL) in a ratio of 1:9, and 1 μL of the resulting solution spotted onto a metal target and placed into the MALDI ion source. Reflectron mode was used for molecules with $m/z > 4000$. MS data was processed using FlexAnalysis 2.0 (Bruker Daltonik GmBH).

7.3 Crystallography

7.3.1 Crystallization

Cyclo(L-Tyr-D-Pro) (31)

A saturated solution of dipeptide **31** was prepared in acetonitrile with gentle heating. A small amount of ethanol was added to ensure complete dissolution of the solid, and the solution was seeded with additional crystalline material. Diffraction-quality crystals of **31** were obtained overnight.

Cyclo(L-Phe-L-Pro) (32)

Crystals of **32** were obtained using the same method as for dipeptide **31**, but using toluene as the solvent.

Cyclo(L-*p*-FPhe-L-Pro) (33)

A vapour diffusion setup was utilized comprising a open vial containing a saturated solution of **33** in ethanol, placed inside a larger sealed vial containing the antisolvent (diethyl ether). Crystals were obtained within 1 week.

7.3.2 Data collection

Crystallographic data for both compounds were collected on a Bruker MicroStar diffractometer using a Cu-rotating anode, equipped with Helios mirror optics as a monochromator and a Platinum 135 CCD detector. The crystals were flash-cooled to 100 K during data collection using a cold nitrogen gas stream. Measurements were performed using Cu Kα radiation (λ = 1.5417). All data were integrated and scaled using SAINT (Bruker, 2000).

7.3.3 Structure solution and refinement

The structures were solved by direct methods (SHELXS-97) and refined by least squares methods against F2 (SHELXL-97).[121] All non-hydrogen atoms were refined with anisotropic displacement parameters and all hydrogen atoms were unambiguously identified and refined isotropically on calculated

positions using a riding model. Molecular graphics were created using Olex2 software.[122]

An electron density peak (corresponding to approximately 4 electrons) present after structure refinement for compound **31** could not be assigned an atom type, and was thought to be a result of residual disordered solvent occupying the channel formed by packing of the dipeptide around a central axis. Refinement was improved by using Olex2 solvent masking.[123]

Conferences

Portions of the work described here were presented at the British Biophysical Society 2012 Conference, Durham University (poster and oral presentation), and the 2012 RSC Bioorganic Group Postgraduate Symposium, Cardiff University (poster presentation).

References

1. T. S. Young and P. G. Schultz, *J. Biol. Chem.*, 2010, **285**, 11039.
2. M. Sainlos and B. Imperiali, *Nat. Protoc.*, 2007, **2**, 3201.
3. S. L. Cobb and C. D. Murphy, *J. Fluorine Chem.*, 2009, **130**, 132.
4. P. W. R. Harris, G. M. Williams, P. Shepherd and M. A. Brimble, *Int. J. Pept. Res. Ther.*, 2008, **14**, 387.
5. S. Mezzato, M. Schaffrath and C. Unverzagt, *Angew. Chem. Int. Ed.*, 2005, **44**, 1650.
6. D. Petsch and F. B. Anspach, *J. Biotechnol*, 2000, **76**, 97.
7. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, in *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, New York, 2nd edn., 2009.
8. F. Schwarz, W. Huang, C. Li, B. L. Schulz, C. Lizak, A. Palumbo, S. Numao, D. Neri, M. Aebi and L. -X. Wang, *Nat. Chem. Biol.*, 2010, **6**, 264.
9. D. P. Gamblin, E. M. Scanlan and B. G. Davis, *Chem. Rev.*, 2009, **109**, 131.
10. K. Mandal, B. L. Pentelute, V. Tereshko, A. A. Kossiakoff and S. B. H. Kent, *J. Am. Chem. Soc.*, 2009, **131**, 1362.
11. H. P. Hemantha, N. Narendra and V. V. Sureshbabu, *Tetrahedron*, 2012, **68**, 9491.
12. J. M. Chalker, *Chem. Biol. Drug Des.*, 2013, **81**, 122.

13. S. B. H. Kent, *Chem. Soc. Rev.*, 2009, **38**, 338.

14. A. Yadav, V. Saini and S. Arora, *Clin. Chim. Acta*, 2010, **411**, 1570.

15. T. L. R. Grygiel, A. Teplyakov, G. Obmolova, N. Stowell, R. Holland, J. F. Nemeth, S. C. Pomerantz, M. Kruszynski and G. L. Gilliland, *Biopolymers*, 2010, **94**, 350.

16. M. Kruszynski, N. Stowell, A. Das, J. Seideman, P. Tsui, M. Brigham-Burke, J. F. Nemeth, R. Sweet and G. A. Heavner, *J. Peptide Sci.*, 2006, **12**, 25.

17. C. Reid, M. Rushe, M. Jarpe, H. van Vlijmen, B. Dolinski, F. Qian, T. G. Cachero, H. Cuervo, M. Yanachkova, C. Nwankwo, X. Wang, N. Etienne, E. Garber, V. Bailly, A. de Fougerolles and P. A. Boriack-Sjodin, *Protein Eng. Des. Sel.*, 2006, **19**, 317.

18. A. R. Brown, M. Covington, R. C. Newton, R. Ramage and P. Welch, *J. Peptide Sci.*, 1996, **2**, 40.

19. J. -H. Gong and I. Clark-Lewis, *J. Exp. Med.*, 1995, **181**, 631.

20. M. Kruszynski, P. Tsui, N. Stowell, J. Luo, J. F. Nemeth, A. M. Das, R. Sweet and G. A. Heavner, *J. Peptide Sci.*, 2006, **12**, 354.

21. H. Lodish, A. Berk, C. A. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh and P. Matsudaira, in *Molecular Cell Biology*, W. H. Freeman and Company, New York, 6th edn., 2008.

22. S. Ali, G. O'Boyle, P. Mellor and J. A. Kirby, *Mol. Immunol.*, 2007, **44**, 1477.

23. T. M. Handel and P. J. Domaille, *Biochemistry*, 1996, **35**, 6569.

24. The PyMOL Molecular Graphics System, Version 1.2r1, Schrödinger, LLC.

25. T. A. Springer, *Cell*, 1994, **76**, 301.

26. J. H. Y. Tan, M. Canals, J. P. Ludeman, J. Wedderburn, C. Boston, S. J. Butler, A. M. Carrick, T. R. Parody, D. Taleski, A. Christopoulos, R. J. Payne and M. J. Stone, *J. Biol. Chem.*, 2012, **287**, 14692.

27. A. E. I. Proudfoot, T. M. Handel, Z. Johnson, E. K. Lau, P. LiWang, I. Clark-Lewis, F. Borlat, T. N. C. Wells and M. H. Kosco-Vilbois, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 1885.

28. E. K. Lau, C. D. Paavola, Z. Johnson, J. -P. Gaudry, E. Geretti, F. Borlat, A. J. Kungl, A. E. Proudfoot and T. M. Handel, *J. Biol. Chem.*, 2004, **279**, 22294.

29. T. Jin, X. Xu and D. Hereld, *Cytokine*, 2008, **44**, 1.

30. A. Müller, B. Homey, H. Soto, N. Ge, D. Catron, M. E. Buchanan, T. McClanahan, E. Murphy, W. Yuan, S. N. Wagner, J. L. Barrera, A. Mohar, E. Verástegui and A. Zlotnik, *Nature*, 2001, **410**, 50.

31. K. Koizumi, S. Hojo, T. Akashi, K. Yasumoto and I. Saiki, *Cancer Sci.*, 2007, **98**, 1652.

32. C. A. G. N. Montalbetti and V. Falque, *Tetrahedron*, 2005, **61**, 10827.

33. G. R. Marshall, *J. Peptide Sci.*, 2003, **9**, 534.

34. B. Gutte and R. B. Merrifield, *J. Am. Chem. Soc.*, 1969, **91**, 501.

35. S. L. Pedersen, A. P. Tofteng, L. Malik and K. J. Jensen, *Chem. Soc. Rev.*, 2012, **41**, 1826.

36. J. M. Collins and N. E. Leadbeater, *Org. Biomol. Chem.*, 2007, **5**, 1141.

37. B. Bacsa, K. Horváti, S. Bõsze, F. Andrea and C. O. Kappe, *J. Org. Chem.*, 2008, **73**, 7532.

38. B. L. Nilsson, M. B. Soellner and R. T. Raines, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 91.

39. J. Y. Lee and D. Bang, *Biopolymers*, 2010, **94**, 441.

40. E. C. B. Johnson and S. B. H. Kent, *J. Am. Chem. Soc.*, 2006, **128**, 6640.

41. V. Y. Torbeev and S. B. H. Kent, *Angew. Chem. Int. Ed.*, 2007, **46**, 1667.

42. D. Bang, B. L. Pentelute and S. B. H. Kent, *Angew. Chem. Int. Ed.*, 2006, **45**, 3985.

43. A. Shigenaga, Y. Sumikawa, S. Tsuda, K. Sato and A. Otaka, *Tetrahedron*, 2010, **66**, 3290.

44. D. Bang, N. Chopra and S. B. H. Kent, *J. Am. Chem. Soc.*, 2004, **126**, 1377.

45. D. Bang and S. B. H. Kent, *Angew. Chem. Int. Ed.*, 2004, **43**, 2534.

46. S. Ueda, M. Fujita, H. Tamamura, N. Fujii and A. Otaka, *ChemBioChem*, 2011, **6**, 1983.

47. F. Mende and O. Seitz, *Angew. Chem. Int. Ed.*, 2011, **50**, 1232.

48. B. J. Backes and J. A. Ellman, *J. Org. Chem.*, 1999, **64**, 2322.

49. R. Ingenito, E. Bianchi, D. Fattori and A. Pessi, *J. Am. Chem. Soc.*, 1999, **121**, 11369.

50. P. Heidler and A. Link, *Bioorg. Med. Chem.*, 2005, **13**, 585.

51. R. Ingenito, D. Drežnjak, S. Guffler and H. Wenschuh, *Org. Lett.*, 2002, **4**, 1187.

52. R. Quaderer and D. Hilvert, *Org. Lett.*, 2001, **3**, 3181.

53. F. Burlina, C. Morris, R. Behrendt, P. White and J. Offer, *Chem. Commun.*, 2012, **48**, 2579.

54. J. B. Blanco-Canosa and P. E. Dawson, *Angew. Chem. Int. Ed.*, 2008, **47**, 6851.

55. K. Jarnagin, D. Grunberger, M. Mulkins, B. Wong, S. Hemmerich, C. Paavola, A. Bloom, S. Bhakta, F. Diehl, R. Freedman, D. McCarley, I. Polsky, A. Ping-Tsou, A. Kosaka and T. M. Handel, *Biochemistry*, 1999, **38**, 16167.

56. Y. J. Zhang, B. J. Rutledge and B. J. Rollins, *J. Biol. Chem.*, 1994, **269**, 15918.

57. C. J. Beall, S. Mahajan, D. E. Kuhn and P. E. Kolattukudy, *Biochem. J.*, 1996, **313**, 633.

58. S. Hemmerich, C. Paavola, A. Bloom, S. Bhakta, R. Freedman, D. Grunberger, J. Krstenansky, S. Lee, D. McCarley, M. Mulkins, B. Wong, J. Pease, L. Mizoue, T. Mirzadegan, I. Polsky, K. Thompson, T. M. Handel and K. Jarnagin, *Biochemistry*, 1999, **38**, 13013.

59. L. Chakravarty, L. Rogers, T. Quach, S. Breckenridge and P. E. Kolattukudy, *J. Biol. Chem.*, 1998, **273**, 29641.

60. E. A. Robinson, T. Yoshimura, E. J. Leonard, S. Tanaka, P. R. Griffin, J. Shabanowitz, D. F. Hunt and E. Appella, *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 1850.

61. J. Lubkowski, G. Bujacz, L. Boqué, P. J. Domaille, T. M. Handel and A. Wlodawer, *Nat. Struct. Biol.*, 1997, **4**, 64. Note this structure includes an additional N-terminal methionine residue which may have an effect on the conformation of the crystallized dimer (Grygiel et al., 2010); in addition this modification may enhance GAG binding (Lau et al., 2004).

62. M. L. Connolly, *J. Appl. Cryst.*, 1983, **16**, 548.

63. M. Mergler and J. P. Durieux, in *The Bachem practise of SPPS*, Bachem AG, Bubendorf - Switzerland, 2nd edn., 2005.

64. P. White, J. W. Keyte, K. Bailey and G. Bloomberg, *J. Peptide Sci.*, 2004, **10**, 18.

65. W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577.

66. Peptide Companion, M. Lebl, G. Lebl and V. Krchnak, Version 1.25, CSPS Pharmaceuticals Inc., 1994.

67. M. Amblard, J. -A. Fehrentz, J. Martinez and G. Subra, *Mol. Biotechnol.*, 2006, **33**, 239.

68. F. Guillier, D. Orain and M. Bradley, *Chem. Rev.*, 2000, **100**, 2091.

69. P. Lidström, J. Tierney, B. Wathey and J. Westman, *Tetrahedron*, 2001, **57**, 9225.

70. J. Coste, D. Le-Nguyen and B. Castro, *Tetrahedron Lett.*, 1990, **31**, 205.

71. V. Dourtoglou, J. -C. Ziegler and B. Gross, *Tetrahedron Lett.*, 1978, **19**, 1269.

72. R. Knorr, A. Trzeciak, W. Bannwarth and D. Gillessen, *Tetrahedron Lett.*, 1989, **30**, 1927.

73. F. García-Martín, M. Quintanar-Audelo, Y. García-Ramos, L. J. Cruz, C. Gravel, R. Furic, S. Côté, J. Tulla-Puche and F. Albericio, *J. Comb. Chem.*, 2006, **8**, 213.

74. W. S. Hancock and J. E. Battersby, *Anal. Biochem.*, 1976, **71**, 260.

75. T. Vojkovsky, *Pept. Res.*, 1995, **8**, 236.

76. B. Bacsa, S. Bősze and C. O. Kappe, *J. Org. Chem.*, 2010, **75**, 2103.

77. S. Frutos, J. Tulla-Puche, F. Albericio and E. Giralt, *Int. J. Pept. Res. Ther.*, 2007, **13**, 221.

78. F. García-Martín, P. White, R. Steinauer, S. Côté, J. Tulla-Puche and F. Albericio, *Biopolymers*, 2006, **84**, 566.

79. B. Sauvagnat, F. Lamaty, R. Lazaro and J. Martinez, *Tetrahedron Lett.*, 2000, **41**, 6371.

80. M. Larhed, G. Lindeberg and A. Hallberg, *Tetrahedron Lett.*, 1996, **37**, 8219.

81. T. Haack and M. Mutter, *Tetrahedron Lett.*, 1992, **33**, 1589.

82. F. El Oualid, R. Merkx, R. Ekkebus, D. S. Hameed, J. J. Smit, A. de Jong, H. Hilkmann, T. K. Sixma and H. Ovaa, *Angew. Chem. Int. Ed.*, 2010, **49**, 10149.

83. S. K. Mahto, C. J. Howard, J. C. Shimko and J. J. Ottesen, *ChemBioChem*, 2011, **12**, 2488.

84. P. Siman, O. Blatt, T. Moyal, T. Danieli, M. Lebendiker, H. A. Lashuel, A. Friedler and A. Brik, *ChemBioChem*, 2011, **12**, 1097.

85. K. S. A. Kumar, L. Spasser, L. A. Erlich, S. N. Bavikar and A. Brik, *Angew. Chem. Int. Ed.*, 2010, **49**, 9126.

86. K. P. Chiang, M. S. Jensen, R. K. McGinty and T. W. Muir, *ChemBioChem*, 2009, **10**, 2182.

87. J. -w. Park and K. -H. Lee, *Bull. Korean Chem. Soc.*, 2009, **30**, 2475.

88. G. Triola, M. Gerauer, K. Görmer, L. Brunsveld and H. Waldmann, *Chem. Eur. J.*, 2010, **16**, 9585.

89. L. Yang and G. Morriello, *Tetrahedron Lett.*, 1999, **40**, 8197.

90. J. K. Lewis, J. Wei and G. Siuzdak, *Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis*, in *Encyclopedia of Analytical Chemistry*, ed. R. A. Meyers, John Wiley & Sons Ltd, Chichester, 2000, pp. 5880-5894.

91. K. Strupat, *Molecular Weight Determination of Peptides and Proteins by ESI and MALDI*, in *Methods in Enzymology*, ed. A. L. Burlingame, Academic Press, 2005, vol. 405, pp. 1-36.

92. W. Timm, A. Scherbart, S. Böcker, O. Kohlbacher and T. W. Nattkemper, *BMC Bioinformatics*, 2008, **9**, 443.

93. F. Dubois, R. Knochenmuss, R. J. J. M. Steenvoorden, K. Breuker and R. Zenobi, *Eur. Mass Spectrom.*, 1996, **2**, 167.

94. T. Baçzek, *J. Pharmaceut. Biomed.*, 2004, **34**, 851.

95. A. -C. Thierry, S. Pinaud, N. Bigler, G. Perrenoud, B. Denis, M. A. Roggero, N. Fasel, C. Moulon and S. Demotz, *Biologicals*, 2001, **29**, 259.

96. *Hydrophobic Interaction and Reversed Phase Chromatography*, GE Healthcare Bio-Sciences AB, Uppsala, 2006.

97. S. F. Sousa, P. A. Fernandes and M. J. Ramos, *Proteins*, 2006, **65**, 15.

98. I. Halperin, B. Ma, H. Wolfson and R. Nussinov, *Proteins*, 2002, **47**, 409.

99. T. Lengauer and M. Rarey, *Curr. Opin. Struc. Biol.*, 1996, **6**, 402.

100. P. Klausmeyer, O. M. Z. Howard, S. M. Shipley and T. G. McCloud, *J. Nat. Prod.*, 2009, **72**, 1369.

101. M. Saleki, MRes Thesis, Newcastle University, 2011.

102. H. D. Flack, *Acta Crystallogr. A*, 1983, **A39**, 876.

103. V. Madison, P. E. Young and E. R. Blout, *J. Am. Chem. Soc.*, 1976, **98**, 5358.

104. I. L. Karle, H. C. J. Ottenheym and B. Witkop, *J. Am. Chem. Soc.*, 1974, **96**, 539.

105. A. D. Borthwick, *Chem. Rev.*, 2012, **112**, 3641.

106. Z. Li and S. Mukamel, *J. Phys. Chem. A*, 2007, **111**, 11579.

107. P. E. Young, V. Madison and E. R. Blout, *J. Am. Chem. Soc.*, 1976, **98**, 5365.

108. P. J. Milne, D. W. Oliver and H. M. Roos, *J. Cryst. Spectrosc.*, 1992, **22**, 643.

109. G. T. R. Palmore and M. T. McBride, *Chem. Commun.*, 1998, 145.

110. S. Palacin, D. N. Chin, E. E. Simanek, J. C. MacDonald, G. M. Whitesides, M. T. McBride and G. T. R. Palmore, *J. Am. Chem. Soc.*, 1997, **119**, 11807.

111. R. Bishop and M. L. Scudder, *Cryst. Growth Des.*, 2009, **9**, 2890.

112. E. Pidcock, *Acta Crystallogr. B*, 2006, **B62**, 268.

113. K. M. Anderson and J. W. Steed, *CrystEngComm*, 2007, **9**, 328.

114. M. Budesinsky, I. Cisarova, J. Podlaha, F. Borremans, J. C. Martins, M. Waroquier and E. Pauwels, *Acta Crystallogr. B*, 2010, **B66**, 662.

115. J. C. Fuller, N. J. Burgoyne and R. M. Jackson, *Drug Discov. Today*, 2009, **14**, 155.

116. J. Xiao and T. J. Tolbert, *Org. Lett.*, 2009, **11**, 4144.

117. E. A. Morrison, G. T. DeKoster, S. Dutta, R. Vafabakhsh, M. W. Clarkson, A. Bahl, D. Kern, T. Ha and K. A. Henzler-Wildman, *Nature*, 2012, **481**, 45.

118. A. N. Wein, B. N. Williams, S. Liu, B. Ermolinsky, D. Provenzano, R. Abagyan, A. Orry, S. H. Leppla and M. Peredelchuk, *J. Med. Chem.*, 2012, **55**, 7998.

119. B. Molon, S. Ugel, F. Del Pozzo, C. Soldani, S. Zilio, D. Avella, A. De Palma, P. Mauri, A. Monegal, M. Rescigno, B. Savino, P. Colombo, N. Jonjic, S. Pecanic, L. Lazzarato, R. Fruttero, A. Gasco, V. Bronte and A. Viola, *J. Exp. Med.*, 2011, **208**, 1949.

120. Q. Li, M. Moutiez, J. -B. Charbonnier, K. Vaudry, A. Ménez, E. Quéméneur and C. Dugave, *J. Med. Chem.*, 2000, **43**, 1770.

121. G. M. Sheldrick, *Acta Crystallogr. A*, 2008, **A64**, 112.

122. O. Dolomanov, L. Bourhis, R. Gildea, J. A. K. Howard and H. Puschmann, *J. Appl. Cryst.*, 2009, **42**, 339.

123. P. van der Sluis and A. L. Spek, *Acta Crystallogr. A*, 1990, **A46**, 194.