

Research School
of Chemistry



Australian
National
University

CHLORINATED AMINO ACIDS IN PEPTIDE PRODUCTION

A thesis submitted for the admission to the degree of

Doctor of Philosophy

By

Mantas Liutkus

Canberra, Australia

April 2016

Declaration

This is to declare that the research presented in this Thesis represents original work that I carried out during my PhD candidature at the Research School of Chemistry at the Australian National University from 2011 to 2015, except for contributions to multi-author papers incorporated in the Thesis where my contributions are specified in the following Statement of Authorship.

To the best of my knowledge the work presented in this Thesis does not contain material that has been submitted for a degree or diploma in any university or any other tertiary institution. Established results and methodologies published or written by another person have been acknowledged by citation of the original work throughout the text.

I give consent to this copy of my thesis, when deposited in the University library, to be made available for loan and photocopying.

Mantas Liutkus

April 2016

Statement of Authorship

This is to delineate the contributions of individual authors to the multi-authored papers and manuscripts that are included into this Thesis:

Aspects of the work described in Chapter 2 have been published in the paper “In Situ Deprotection and Incorporation of Unnatural Amino Acids during Cell-Free Protein Synthesis”, I. N. Arthur, J. E. Hennessy, D. Padmakshan, D. J. Stigers, S. Lesturgez, S. A. Fraser, M. Liutkus, G. Otting, J. G. Oakeshott and C. J. Easton, *Chem. Eur. J.* **2013**, *19*, 6824-6830. My contribution to the paper was all work involving 4-chlorovaline and 4-chlorovaline methyl ester, including the synthesis, purification and use in protein expression.

The work described in Chapter 3 has been published in the paper “Peptide Synthesis through Cell-Free Expression of Fusion Proteins Incorporating Modified Amino Acids as Latent Cleavage Sites for Peptide Release”, M. Liutkus, S. A. Fraser, K. Caron, D. J. Stigers and C. J. Easton, *ChemBioChem* **2016**, doi: 10.1002/cbic.201600091. I am responsible for the identification and exploration of the latency of γ -chlorinated amino acids for peptide bond cleavage and the development of the method for peptide production through the expression of heat-labile fusion proteins; Mr. Fraser and Dr. Stigers are responsible for the incorporation studies of unsaturated amino acids and the adaptation of iodo-lactonisation of unsaturated amino acids to peptide production through the expression of iodine-labile fusion proteins; Ms. Caron carried out

Chlorinated Amino Acids in Peptide Production

comparative peptide production studies utilising enzymatic protein digestion; and Prof. Easton supervised various aspects of the project.

Mantas Liutkus

April 2016

I hereby certify that the above statement is correct.

Prof. Christopher J. Easton

April 2016

Acknowledgements

I would like to thank my supervisor Prof. Chris J. Easton for his guidance during my PhD research and his ceaseless and restless support with the preparation of this Thesis. I would also like to thank Dr. Dannon J. Stigers for affording me the opportunity to discern the elusive properties of chlorinated amino acids.

I would like to thank all members of the Easton group, both current and those who've left, for their help with my work that I could not have managed without and for making my research years an enjoyable experience, for making our group feel like a family.

I would like to thank Dr. Paul D. Carr for his help with protein crystallography and Dr. Max Keniry for his help with NOE analysis.

A thank you to the RSC and everyone in it for making my research possible.

A thank you to my family for continuous support during my academic journey.

And thank you Australia for the amazing weather.

Abstract

A new method for the production of peptides through biological expression was developed, utilising lactonisation-prone chlorinated amino acids for latent peptide bond digestion. Incorporation of halogenated amino acids into proteins is possible due to the inherent inability of the biological synthetic machinery to discriminate against compounds structurally similar to the natural substrates. As isoleucine, leucine and valine are primarily recognised by size exclusion, all isosteres with a methyl group replaced by similarly-sized chlorine are mistaken as substrates, and the ability of halogens to mimic the bonding interactions of sulfur enabled the design of a chlorinated and a brominated analogue of methionine.

Amino acids with chlorine at the 4-position, incorporated into proteins in place of isoleucine, leucine or methionine during cell-free protein expression, were found to trigger cleavage of the proximal peptide bond on the C-terminal side at elevated temperatures. The reaction mechanism is similar to that of cyanogen bromide induced cleavage, driven by the formation of a highly favourable 5-membered ring. This presents the first case of heat induced proteolysis, resulting in almost instantaneous peptide bond cleavage through exposure to 100°C in water, and avoids the need for toxic, expensive or sensitive external agents. When encoded in a fusion protein, the chlorinated residues can be used for rapid separation of the fusion partners.

The utility of the method was demonstrated through the preparation of small peptides human gastrin releasing peptide prohormone, cholecystokinin prohormone and

oxytocin, lacking isoleucine, leucine and methionine, respectively, expressed as fusion proteins. Through simultaneous replacement of two amino acids, deuterated and fluorinated analogues of the peptides were also prepared. The method was then expanded to prepare more complicated targets. Homologous substitution of leucine with isoleucine, and *vice versa*, enabled the preparation of a small protein aprotinin. A crystal structure of the mutated aprotinin demonstrated that the protein structure was not affected by the substitution, thus establishing that interchange of similar amino acids can be used to overcome sequence limitations.

In stark contrast, amino acids with chlorine at the 3-position are resistant to high temperature. Through the ability to substitute valine or isoleucine during protein expression, the chlorides were incorporated into aprotinin, thus establishing a method for the preparation of proteins of essentially unlimited size containing 3-chloro amino acids. Crystal structures of the chlorinated aprotinin showed that the amino acid analogues are not inherently detrimental to protein structure and can lead to stable proteins, while the preparation through the expression of heat-labile fusion proteins juxtaposed the differences in reactivity between amino acids halogenated at the 3- and 4-positions.

Table of Contents

Declaration	i
Statement of Authorship	iii
Acknowledgements	v
Abstract	vii
Table of Contents	ix
Abbreviations	xi
Chapter 1. Introduction	1
1.1. Production of proteins	1
1.2. Proteolytic methods	5
1.3. Unnatural amino acids in proteins	22
1.4. Analogues of the branched-chain amino acids	30
1.5. Substitution of methionine	35
Chapter 2. Substrate Selectivity of Isoleucyl-tRNA Synthetase	41
2.1. Alternate approaches to γ -chlorovaline methyl ester	108
2.2. Incorporation of 3',3',3' – trifluoroisoleucine	111
2.3. Synthesis of 5,5,5 – trifluoroisoleucine	115
2.4. Incorporation of amino acids for 'click' chemistry	121
2.5. Conclusions	126
Chapter 3. Cleavage of Peptide Bonds	129

Chapter 4. Engineering with Chlorinated Amino Acids without Sequence Restrictions	185
4.1. Expression of aprotinin	186
4.2. Crystallisation and crystal analysis	189
4.3. HPLC purification of aprotinin	198
4.4. Use of trypsin column	200
4.5. Conclusions	203
Chapter 5. Labile Analogues of Methionine	205
5.1. Preparation of the analogues of methionine	208
5.2. Fragmentation analysis of modified His ₆ PpiB	218
5.3. Production of oxytocin	227
5.4. Conclusions	236
Chapter 6. Conclusions and Future Directions	239
Chapter 7. Experimental	243
7.1. General experimental	243
7.2. Cell-free protein expression	245
7.3. Synthesis of compounds from Chapter 2.	246
7.4. Additional experimental data for Chapter 3.	259
7.5. Experimental procedures of Chapter 4.	267
7.6. Synthesis of compounds from Chapter 5.	269
References	277
Appendix	285
A1. Spectroscopic characterisation of new compounds	285
A2. X-Ray data and structure validation reports	291

Abbreviations

AARS	amino acyl tRNA synthetase
BCAA	branched chain amino acids
^t Boc	<i>tert</i> -butoxycarbonyl
br	broad
CRS	cysteinyl tRNA synthetase
DIPEA	diisopropylethylamine
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
EI	electron impact
ESI	electrospray ionisation
FRS	phenylalanyl tRNA synthetase
HPLC	high performance liquid chromatography
IRS	isoleucyl tRNA synthetase
LRS	leucyl tRNA synthetase
mRNA	messenger RNA
MRS	methionyl tRNA synthetase
MS	mass spectrometry
<i>m/z</i>	mass-to-charge ratio
NBS	<i>N</i> -bromosuccinimide

Chlorinated Amino Acids in Peptide Production

NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
obs	obstructed
PpiB	peptidyl-prolyl <i>cis-trans</i> isomerase B
RNA	ribonucleic acid
SDS	sodium dodecylsulfate
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
THF	tetrahydrofuran
tRNA	transfer RNA
VRS	valyl tRNA synthetase
UV	ultraviolet
YRS	tyrosyl tRNA synthetase

Chapter 1. Introduction

Peptides are critical in biological systems^[1] and are becoming increasingly important in medicinal applications.^[2-4] Due to greater potency and specificity as compared to small molecule agents as well as non-toxic metabolic break-down products (amino acids), synthetically produced peptides and proteins are being approved for therapeutic use with increasing rate, with around a hundred peptide drugs currently in use and more than five hundred in various stages of clinical trials,^[5] steadily replacing small molecule chemotherapeutics, antivirals, cardiovascular medication and antimicrobial drugs as well being increasingly used as vaccines and diagnostic tools instead of attenuated pathogens. In addition, over the last two decades the peptides developed for pharmaceutical use have been increasing in size, and the trend is likely to continue.^[4] The precisely defined sequence^[6] and the highly specific interactions make peptides valuable tools in nanotechnology^[7] as well. As the demand for peptides continues to grow, new methods for peptide production are continuously developed.

1.1. Production of proteins

Short amino acid sequences are referred to as peptides, while longer chains are known as proteins, but the boundary between the two is artificial and not well defined. The size of the target generally determines the production method. For large proteins

over-expression in biological systems is the only currently feasible option.^[8] Ever since organisms were demonstrated to translate foreign genetic material into proteins in the 1970s,^[9] the easy-to-handle bacterium *Escherichia coli* became a routine host for protein expression.^[10,11] The procedures have been perfected over the years,^[8,11-13] with the development of modified cell strains, optimised growth media, customised expression vectors and highly refined protocols. Despite all the advances, however, recombinant protein production is not always trivial,^[11] as protein yields vary significantly and unpredictably, and even when satisfactory amounts of protein are obtained, protein may lack activity.

While many strategies to overcome the difficulties of recombinant protein expression in cells have been devised,^[13] many of the problems can be overcome by the use of cell-free systems. The earliest cell-free protein expression systems were explored in the 1950s,^[14] and the first practical application of the system was achieved in 1961 with the deciphering of the genetic code.^[15] Since then these systems have become significantly more sophisticated and their use for protein production has become routine.^[14,16,17] Cell-free systems hold an advantage over whole-cell systems in that cellular extracts (or reconstituted mixtures) can be stored and handled as chemicals and can be directed to produce a single protein, as native genetic material is absent; multiple proteins can be co-expressed simultaneously without the need to consider plasmid complementarity^[18] (an issue affecting the use of related vectors in living cells). Expressed protein can then be extracted without further treatment of the mixture. As all (or most) metabolic activity is uncoupled from protein production, the background processes do not interfere with protein production and most proteases are neutralised beforehand. On the other hand, due to the uncoupling of protein expression from the catabolic machinery, energy for

protein synthesis could no longer be generated from cheap and stable nutrients like sugars or protein hydrolysates. Energy in the form of ATP (or immediate sources of activated phosphate) has to be supplied and constantly replenished throughout the reaction. This led to the development of continuous-flow (CF) and continuous-exchange cell-free (CECF) systems where the sensitive nutrients were constantly resupplied and reaction by-products removed, allowing the system to function for longer (Figure 1.1). Regardless, scale-up using cell-free systems has remained to date problematic.

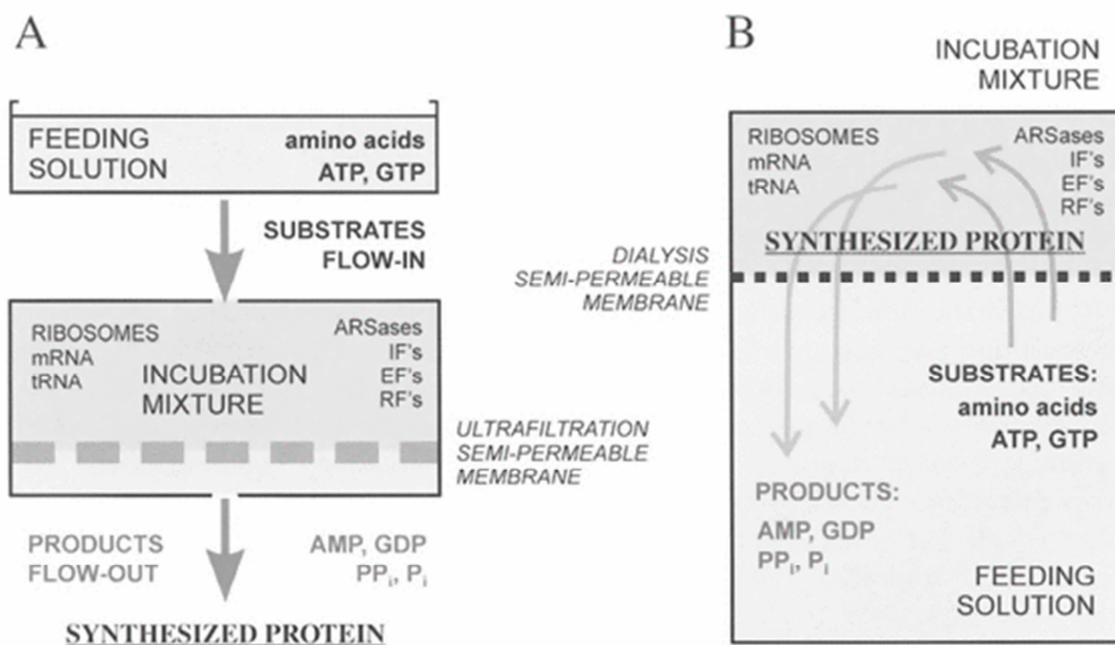


Figure 1.1. The principles of continuous-flow (A) and continuous-exchange (B) cell-free protein expression systems. Reproduced with permission from Springer.^[14]

Another problem associated with heterologous protein production is protein instability. The lack of native machinery required to process and facilitate correct protein folding may lead to non-functional or completely insoluble protein. On occasion, homologous chaperone proteins are co-expressed in efforts to stabilise the target protein.^[19] A more routine approach, however, is to express the target protein as a part of a larger fusion protein.^[11,20] In 1977, the first target to be produced from synthetically generated DNA, 14-residue long hormone somatostatin, was attached to the C-terminus of β -galactosidase,^[21] which has become a common strategy with small peptides.^[22,23] Since then, other proteins like SUMO (small ubiquitin-like modifier), MBP (maltose binding protein) and GST (glutathione *S*-transferase) were demonstrated to stabilise sensitive proteins and have become routine fusion partners. Attached protein tags, like His₆-tag, c-Myc or FLAG-tag, also facilitate protein purification. Some fusion partners, like MBP or GST, serve both as stabilisers and purification tags.

The ability to artificially increase the target size enables biological production of smaller peptides.^[24] More routinely, however, chemical synthesis is employed for short targets.^[25-27] This severs the reliance on biological machinery and the need for handling DNA. The difficulty, though, lies in the size and the linear nature of proteins and peptides, requiring sequential addition of monomers. As the polymer is extended, incomplete coupling and side reactions result in product heterogeneity and introduce purification challenges.^[28] Introduction of solid phase peptide synthesis in 1963^[29] and the subsequent developments^[25,30] in automation, racemisation suppressants and the protection group chemistry made the sequential build-up of peptide more efficient and less time-consuming. Still, yields and purity drop off with increasing chain length, and the success of the synthesis is somewhat sequence dependent, with some nascent chains

failing to extend due to aggregation. Long peptides (up to 100 amino acids) have been prepared by using harsh denaturing conditions^[31] or conformationally restricted amino acid derivatives^[32] to disrupt chain aggregation, but more routinely assembly of smaller fragments, some obtained through biological means, is required.^[26,33,34] These modifications, however, significantly extend the preparation time and costs, making the biological approaches more favourable. Additionally, biological methods are significantly easier to scale up and are more amenable to continuous as opposed to batch synthesis. As a result, several hundreds of therapeutic proteins are still produced through recombinant expression,^[3] and the treatment costs of tens of thousands of dollars per patient annually associated with the use the celebrated 36 residue long synthetic HIV fusion inhibitor enfuvirtide^[35] prompted the development of biological routes of production.^[23]

A continuous exchange cell-free (CECF) protein expression system derived from *E. coli* BL21 Star (DE3) strain^[16] was used to conduct the research described in this Thesis.

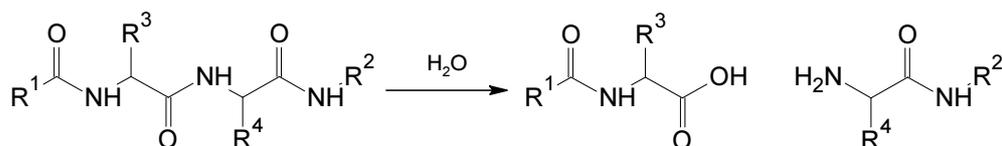
1.2. Proteolytic methods

As mentioned above, proteins are routinely expressed as larger constructs to facilitate proper folding or purification. When the tag is considered a potential hindrance in the following experiments, the added sequence is removed. Proteases are routinely used for protein digestion,^[36] as enzymes function under conditions that do not adversely affect other proteins. Many proteases with specific target sequences, like enterokinase, factor Xa and tobacco etch virus (TEV) protease, have been characterised and adapted in routine laboratory work. The enzymes, however, are generally introduced in catalytic

amounts and take up to 24 hours to complete the reaction. In this prolonged reaction, most enzymes are prone to cleave additional non-specific sequences. Additionally, enzymes require specific buffers, pH and temperature to function properly, as well as soluble substrate and exposed proteolytic sites. Thus, the use of enzymes can be problematic.

Chemical reagents can perform under significantly more diverse conditions than enzymes and the small molecule nature of these agents makes them easy to remove after the reaction is complete. Therefore, chemical proteolysis methods that can cleave peptides in a sequence dependent manner (also known as limited proteolysis) are an attractive alternative to enzymatic cleavage.

While proteases are divided into classes based on the mechanism of action^[37] (serine proteases like enterokinase or trypsin and cysteine proteases like papain use an internal nucleophile, whereas aspartate proteases like pepsin and HIV protease position a water molecule for direct bond hydrolysis), they all split the substrate at the peptide bond (Scheme 1.1). Chemical agents exhibit a far more diverse array of cleavage mechanisms and tend to irreversibly modify the targeted amino acids introducing chemistry required to facilitate the chain cleavage.



Scheme 1.1. Enzymatic hydrolysis of peptide bonds.

Metal based cleavage reagents do hydrolyse amide bonds as metals have Lewis acid characteristics and peptide break-down proceeds *via* enzyme-like transition states. Early examples^[38] of metal-catalysed proteolysis were more position than sequence dependent (i.e. N-terminal or C-terminal amino acids were cleaved). However, metal ions do have different affinities for different amino acids and can hydrolyse peptide bonds in the vicinity of their recognised sequences.^[39,40] They are proposed to catalyse the bond cleavage by forming a multivalent complex with the protein, although the exact mechanistic details remain to be confirmed (Figure 1.2).

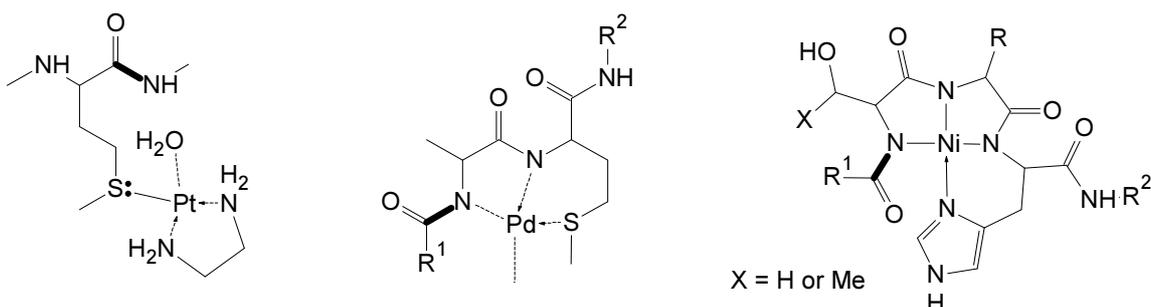


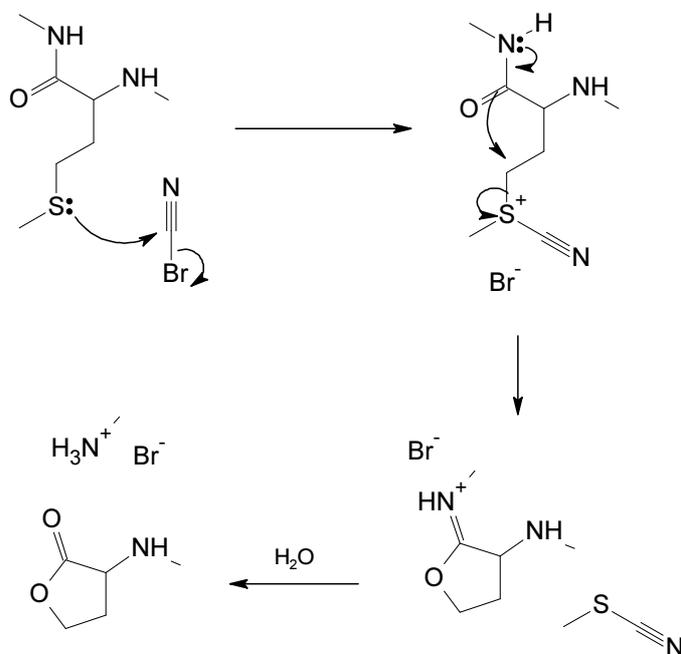
Figure 1.2. Proposed hydrolytic protein – metal complexes.^[39] The scissile bond is highlighted.

One of the earliest and the most wide-spread chemical proteolysis reagent is cyanogen bromide, giving some of the highest yields and generating the least amount of side reactions. Cyanogen bromide was used for the cleavage of the aforementioned β -galactosidase-somatostatin fusion protein.^[21]

Cyanogen bromide was found to selectively cyanylate the residue of methionine on the sulfur atom, thus predisposing the side chain of methionine for a nucleophilic

substitution.^[41,42] In a peptide chain the closest available nucleophile is the amide bond at the C-terminal side of the modified residue that can displace the cyanosulfonium group with concomitant formation of a 5-membered ring; aqueous hydrolysis of a newly formed imine bond leads to cleavage of the peptide bond (Scheme 1.2). The reaction takes place at room temperature and, reportedly, leads to near quantitative breakage of susceptible bonds.

Cyanogen bromide was identified as a peptide cleavage reagent as a part of a larger screen of related alkylating compounds, like methyl iodide and iodoacetamide, all of which were capable of promoting protein fragmentation.^[43] While any alkylation of the side-chain of methionine can promote the reaction, cyanogen bromide gave the highest yields at the mildest conditions.

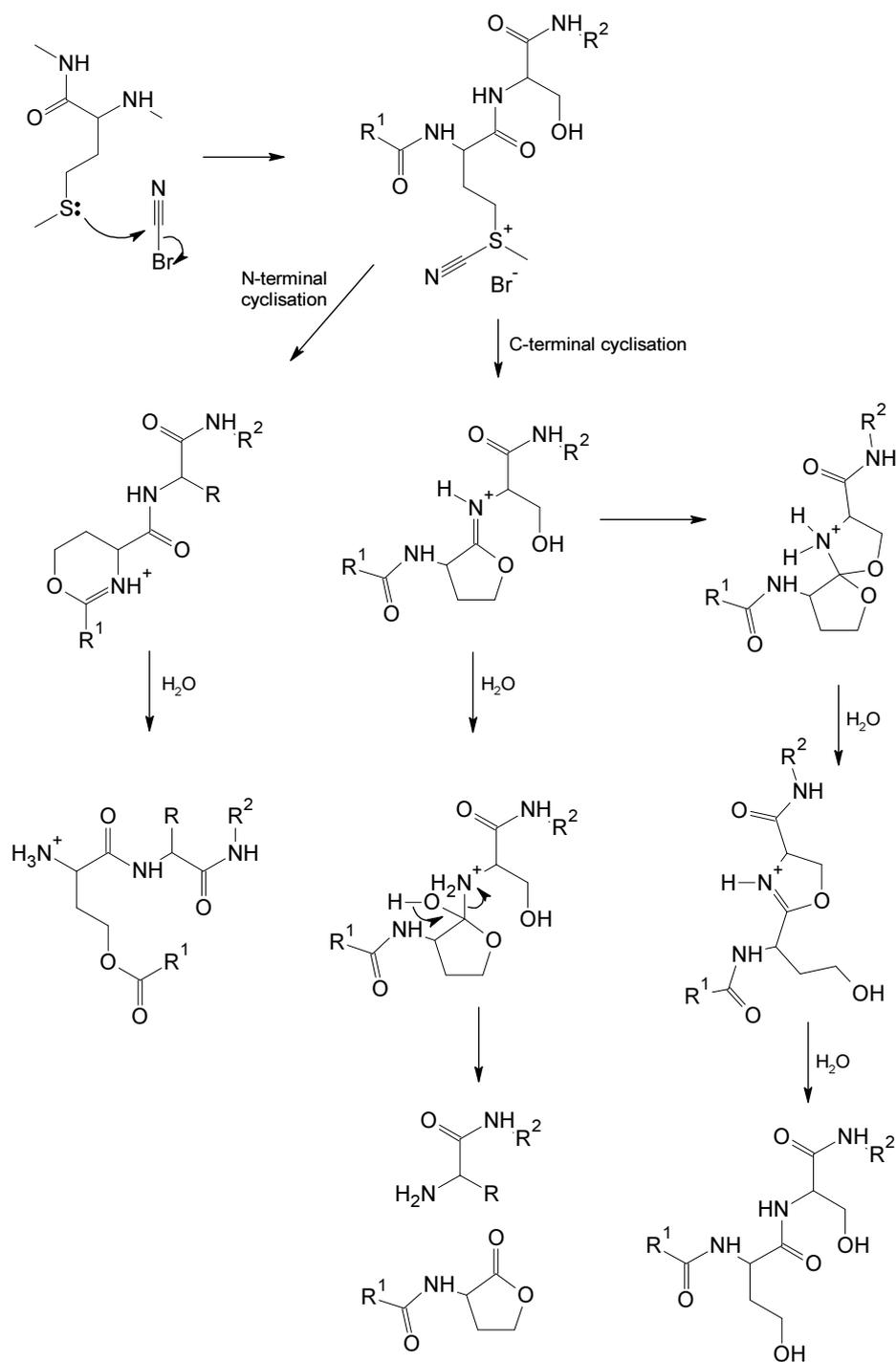


Scheme 1.2. Original proposed mechanism of peptide cleavage with cyanogen bromide.^[41,42]

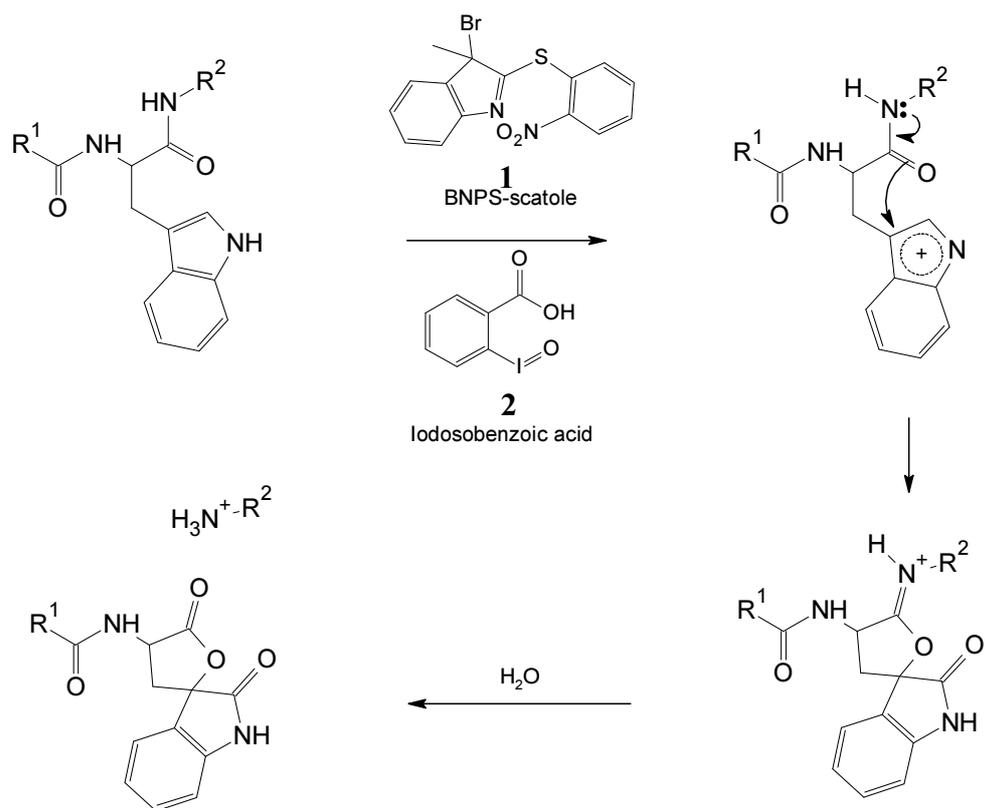
The viability of the method was first demonstrated on a small dipeptide ethyl *N*-benzoyl-DL-methionylglycinate that enabled unambiguous establishment of the reaction mechanism by characterisation of the products,^[41] followed by more thorough mechanistic analysis in subsequent years.^[44] It was then tested on ribonuclease,^[42] a protein routinely used for protein folding experiments. On the very first set of experiments cyanogen bromide fragmentation allowed for a correction of previously mis-identified sequence of ribonuclease.

Subsequent work uncovered several alternative pathways cyanylation of methionine can follow. In cases where methionine is followed by a serine or threonine, the close hydroxyl group can participate in a second intramolecular ring formation (with opening of the first ring) and thus ‘rescue’ the peptide chain (Scheme 1.3, right).^[45] Over the past few decades the procedure has been optimised to reduce the likelihood of this pathway and up to 80% of alkylated methionine-serine/threonine bonds result in cleavage.^[46] The reaction is routinely conducted in 70% aqueous formic acid to minimise participation of the hydroxyl groups as well as to limit the oxidation^[47] of methionine that renders it immune to cyanylation.

Another alternative is the participation of the peptidic bond at the N-terminal side of methionine (Scheme 1.3, left).^[48] In this case a 6-membered ring is formed and does not immediately result in spontaneous chain cleavage. The amide bond is instead replaced by a new ester bond that can be selectively hydrolysed under mild alkaline conditions. This pathway has not been observed in long peptide chains, possibly because, all things being equal, 5-membered rings form more favourably. Studies on small molecules and N-terminal acetylated peptides, however, revealed the pathway is possible.

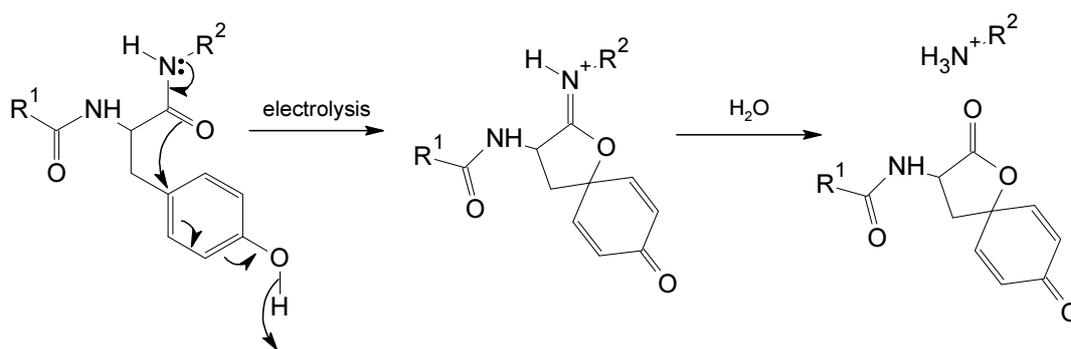


Scheme 1.3. The possible outcomes of cyanylation of methionine. The classical pathway resulting in peptide cleavage (middle)^[41,42,44] can be precluded if methionine is followed by a serine or threonine (right).^[45,46] Alternatively, the N-terminal peptide bond can be replaced by an ester bond (left).^[48]



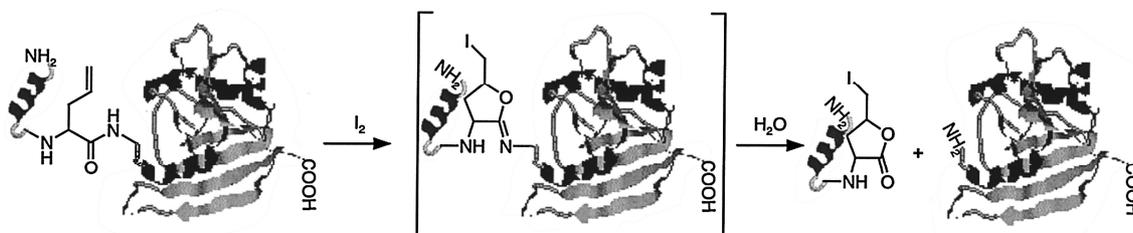
Scheme 1.4. Cleavage of peptide bond resulting from oxidation of tryptophan.^[49,50]

Oxidation of aromatic amino acids results in peptide bond hydrolysis following an analogous mechanism. After initial attempts with harsh oxidants like *N*-bromosuccinimide (NBS),^[38] that can substantially damage the protein in untargeted regions, more sensitive and select reagents have been developed. BNPS-scatole **1**^[49] and iodosobenzoic acid **2**,^[50] when prepared with great purity, can oxidise tryptophan with little or no damage to tyrosine, the other highly susceptible amino acid (Scheme 1.4). Conversely, electrolysis has been used to oxidise tyrosine (Scheme 1.5).^[51] Prolonged exposure to the oxidants leads to oxidation of histidine, which, again, leads to bond cleavage at the modified residue.^[38] After Birch reduction of the protein, even peptide bond cleavage near phenylalanine becomes mechanistically possible.^[38]



Scheme 1.5. Mechanism of peptide bond cleavage as a result of oxidation (by electrolysis or otherwise) of tyrosine.^[51]

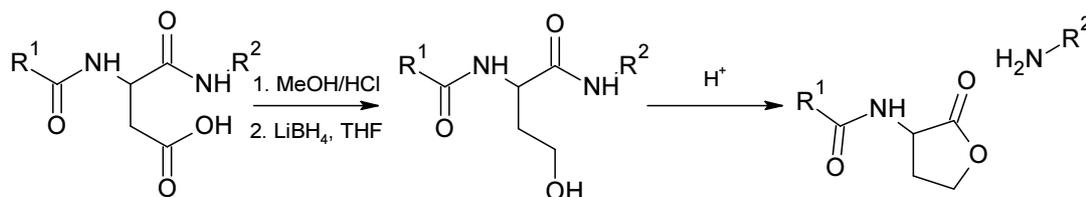
In a similar fashion, iodolactonisation of allylglycine that was incorporated into proteins through the use of chemically acylated tRNA (*vide infra*) led to peptide bond cleavage (Scheme 1.6).^[52,53]



Scheme 1.6. The mechanism of iodine-mediated peptide bond cleavage next to allylglycine. Reproduced with permission from ACS.^[53]

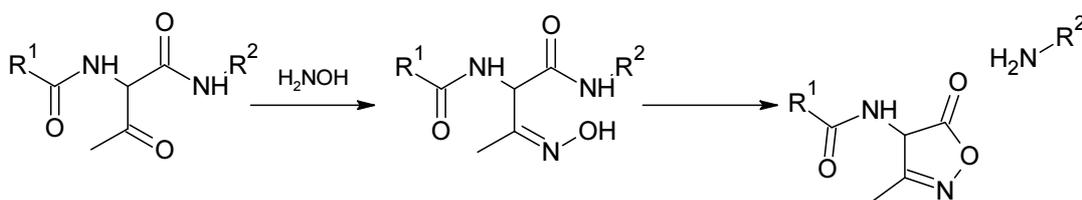
The chemical treatments described above lead to the formation of 5-membered rings by promoting a nucleophilic attack by the oxygen of the proximal amide bond at the C-terminal side of the reacted residue; the peptide bond breaks as a result of the ring

formation. The amide can act as an electrophile when the side-chain of aspartate or glutamate is reduced to a nucleophilic OH. The highly favourable lactone replaces the otherwise more thermodynamically stable amide (Scheme 1.7).^[38]



Scheme 1.7. Protein fragmentation caused by reduced aspartyl residue.^[38]

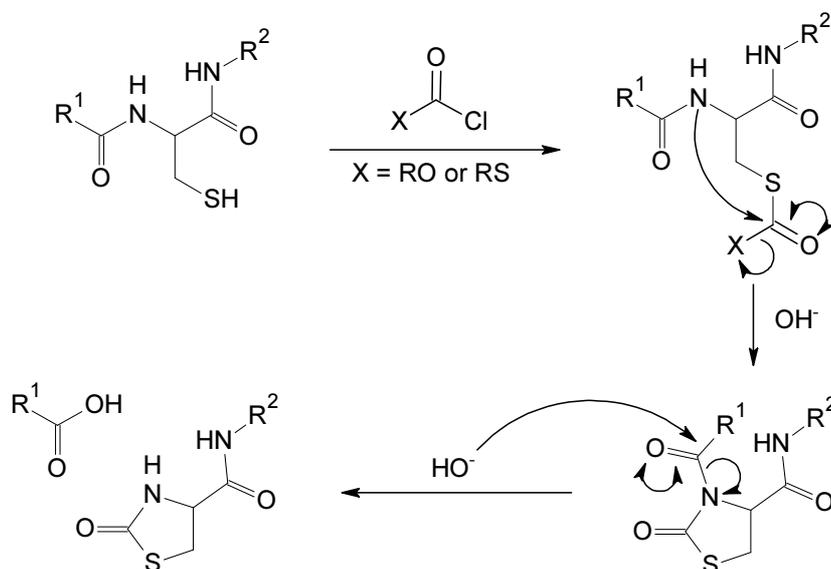
Yet another example of peptide bond cleavage on the C-terminal side of the residue involves hydroxylamine treatment of threonine or serine after oxidation with DCC (Scheme 1.8).^[38] Due to side reactions the newly oxidised residues can undergo, however, this method is not applicable to large proteins.



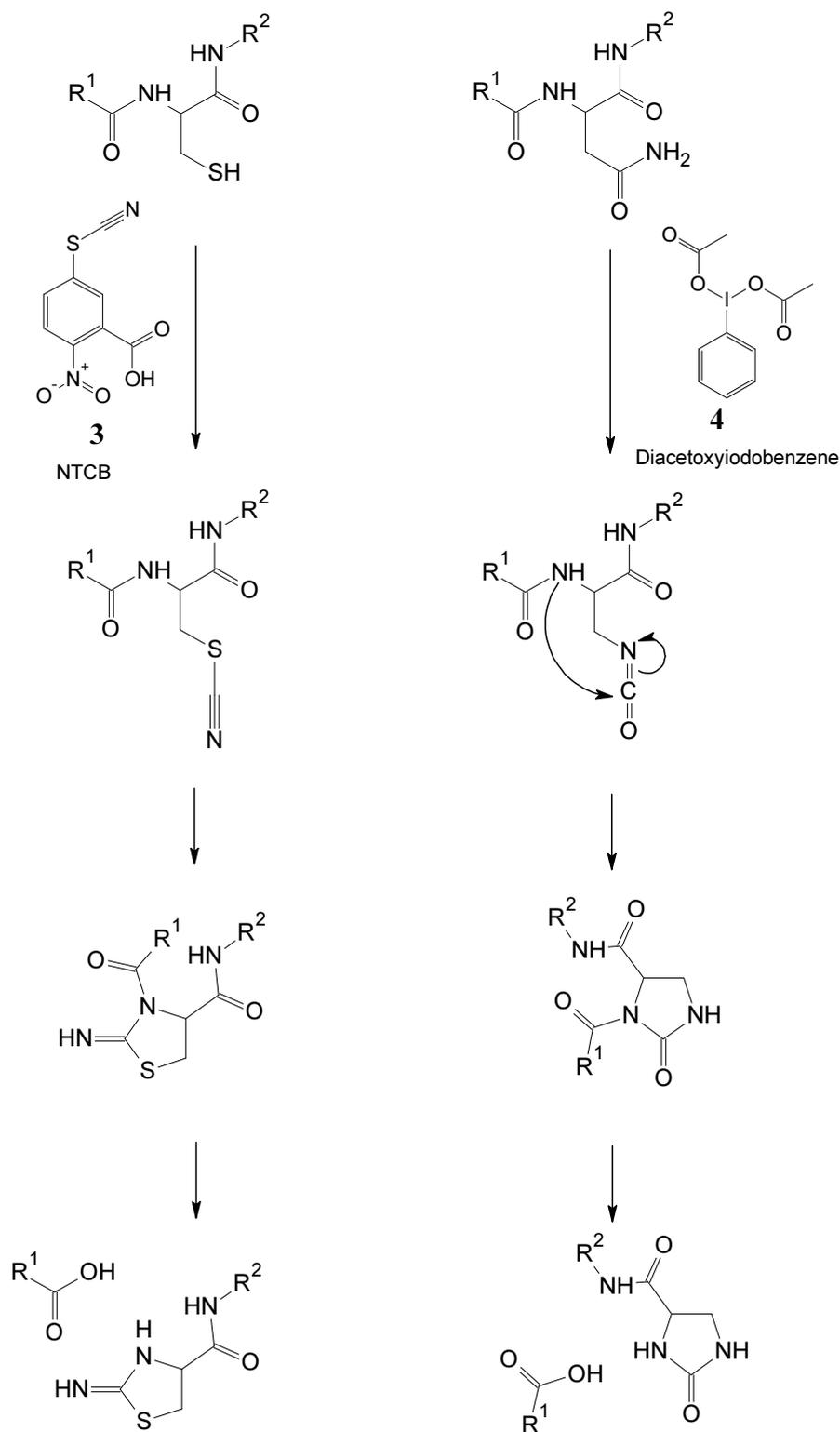
Scheme 1.8. Bond cleavage induced by treatment of oxidised threonyl residue with hydroxylamine.^[38]

All of the bond cleavage reactions described above are driven by the formation of a 5-membered ring (with the exception of reduced glutamyl residue) at the C-terminal side of the modified residue. If, however, a shorter residue is modified or the side chain is trimmed by the treatment, the favourable ring forms in the opposite direction, resulting in bond cleavage on the N-terminal side of the residue.

Reaction of cysteine residues with chloroformates or chlorothioformates is a prototypical method of peptide bond cleavage, with yields of cleavage up to 70% (Scheme 1.9).^[38] Similar results are obtainable by reaction of serine and threonine with phosgene, although higher temperatures are required for the reaction to proceed.^[38]



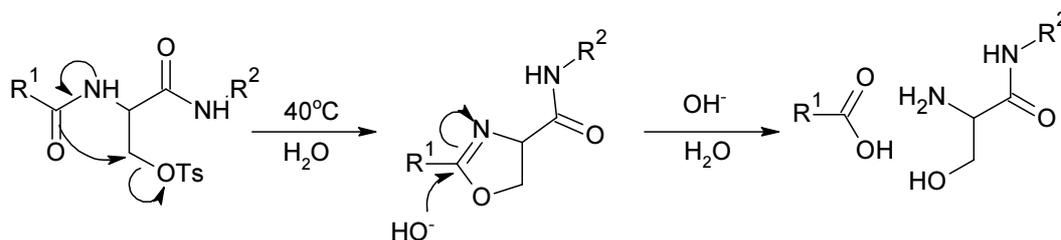
Scheme 1.9. Fragmentation of X-cysteinyl peptide bond due to treatment with chloroformates or chlorothioformates.^[38]



Scheme 1.10. The mechanisms of peptide cleavage by cyanylation of cysteine by NTCB **3** (left)^[54] and oxidation of asparagine by diacetoxyiodobenzene **4** (right).^[55]

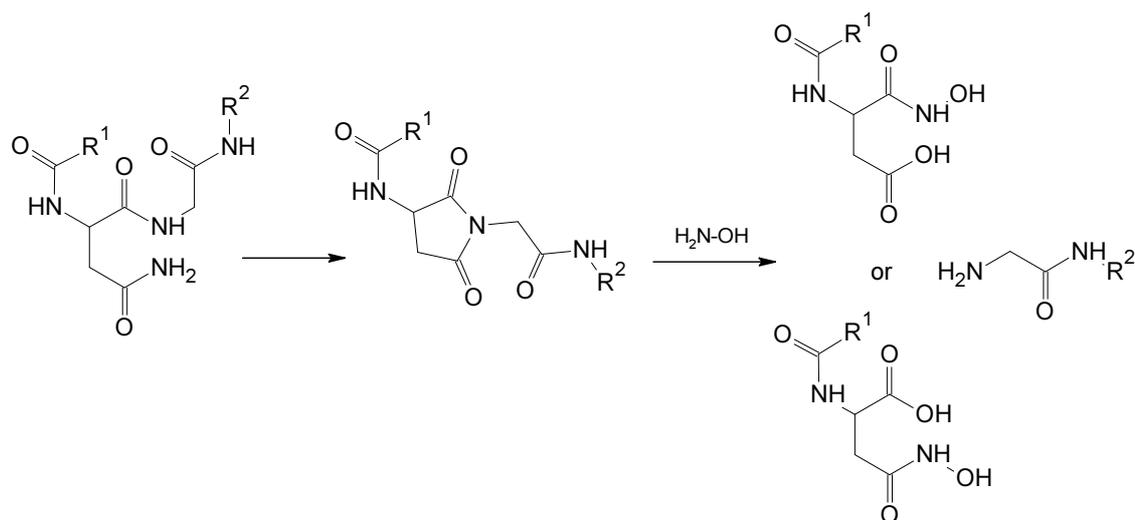
More sensitive cysteine modifying reagents 2-nitro-5-thiobenzoic acid (NTCB) **3**^[54] and Cyssor-I^[56] were later introduced to make the reaction cleaner. The modified residue is able to form a ring with the participation of the amide bond on the N-terminal side, forming an imide-like intermediate that facilitates peptide bond cleavage. Cyanylation of cysteine is also achieved by reacting disulfide bonds directly with inorganic cyanide.^[38] Years later oxidation of asparagine with diacetoxyiodobenzene **4** was found to lead to a Hofmann rearrangement and peptide bond cleavage, by a virtually isosteric mechanism (Scheme 1.10).^[55]

When the electrophilic site is set up closer to the α -centre, the N-terminal peptide bond can react through the oxygen, as was the case with *N*-acetylmethionine (Scheme 1.3, left). Tosylated (or otherwise sulfonylated) serine presents a good leaving group for a substitution reaction by the neighbouring carbonyl group.^[38] The transformation resolves in an oxazolinium ring that leads to cleavage of the peptide chain (Scheme 1.11).



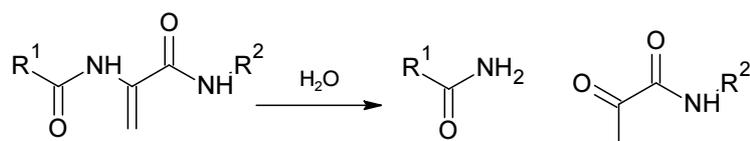
Scheme 1.11. Peptide bond cleavage as a consequence of tosylation of serine.^[38]

More conventional laboratory chemicals were also used to achieve partial protein hydrolysis. Hydroxylamine was found to preferentially hydrolyse asparagine-glycine bonds (Scheme 1.12), proceeding once again through a cyclic intermediate.^[57]



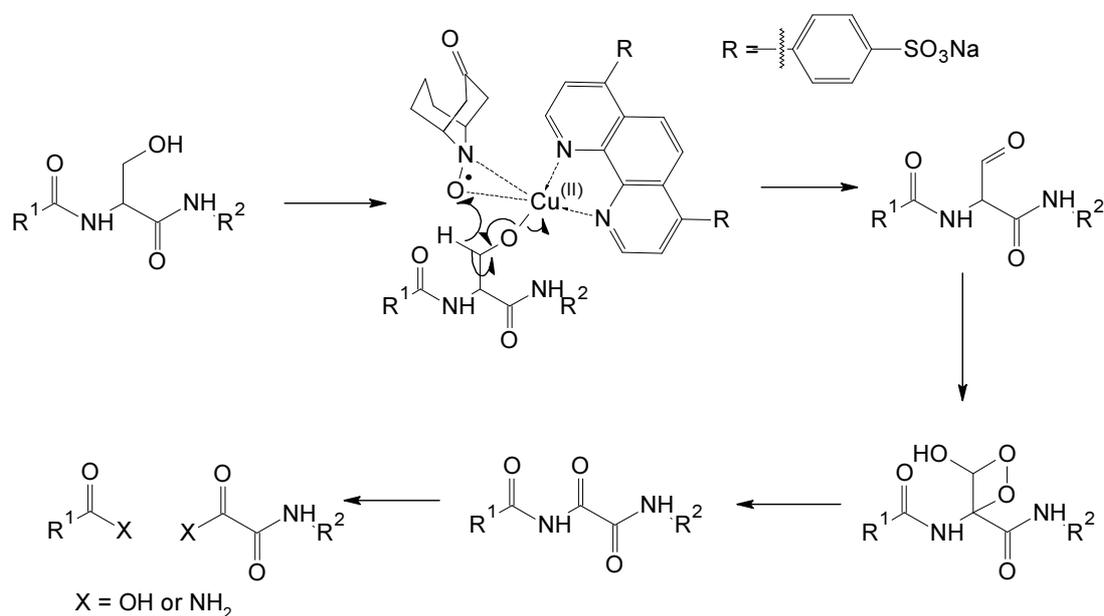
Scheme 1.12. Hydroxylamine-induced break-down of asparaginylglycine bond.^[57]

Peptide chain can also fragment a result of an introduced α - β unsaturated amino acid, generally prepared from serine or cysteine by β -elimination. The enamine hydrolysis has been achieved under mild conditions (Scheme 1.13),^[58] once the dehydroalanine has been produced.



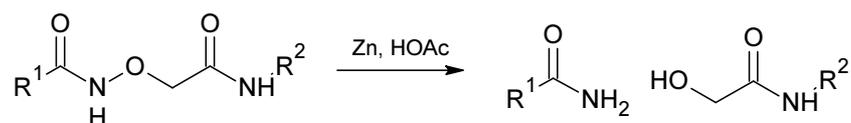
Scheme 1.13. Hydrolysis of the dehydroalanine enamine.^[58]

A more elaborate mechanism has been proposed for a recently introduced copper reagent that hydrolyses the peptide bond at the N-terminal side of serine.^[59] While the reaction intermediates have not been detected and the reaction mechanism has not been verified, this aerobic method completely strips off the side-chain of serine converting the amide bond into a more labile imide (Scheme 1.14).



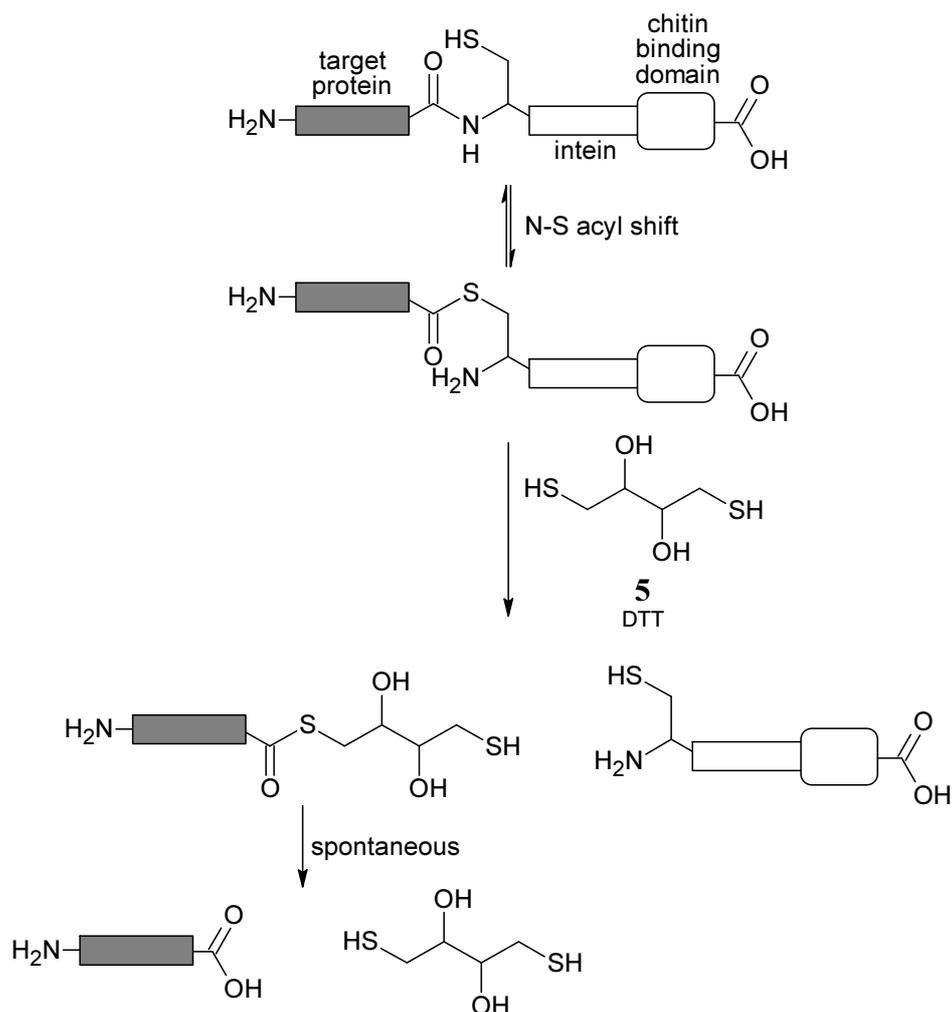
Scheme 1.14. Mechanism proposed for the peptide cleavage at serine.^[59]

A labile nitrogen-oxygen bond of (aminoxy)acetic acid, another analogue incorporated into proteins through chemical acylation of mutated tRNA (*vide infra*), is cleaved under mild reducing conditions (Scheme 1.15).^[60]



Scheme 1.15. Reductive cleavage of (aminooxy)acetic acid.^[60]

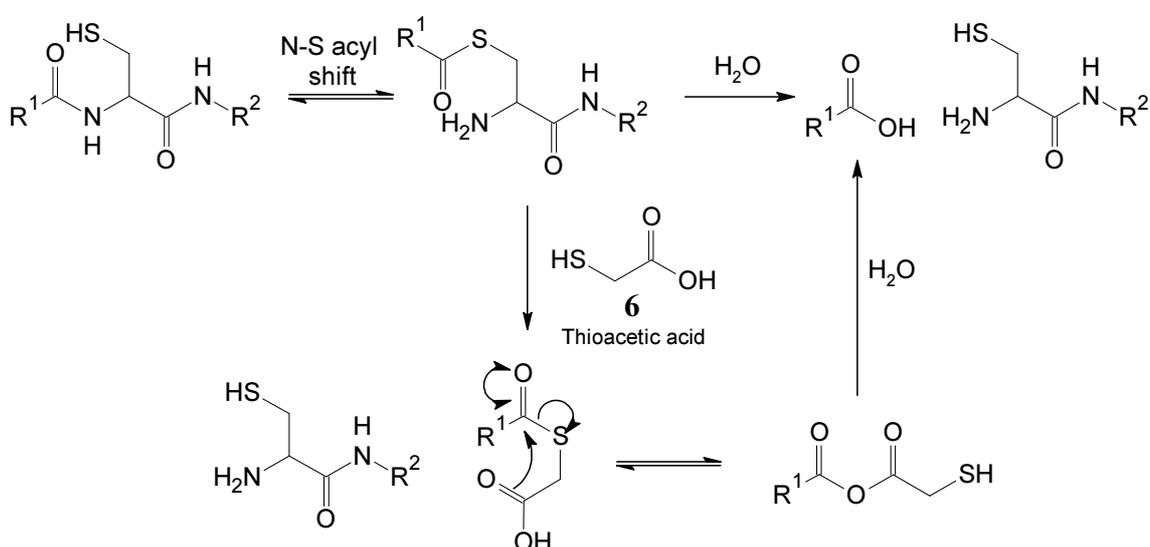
A self-digesting protein purification system has also been developed. Named ‘IMPACT’ (intein mediated purification with an affinity chitin binding tag), the procedure involves the expression of the target protein as a fusion with a chitin affinity tag, joined *via* a mutated intein.^[61] The mutant of yeast intein VMA is unable to complete the excision sequence, resulting in the formation of a thioether and subsequently the release of the target protein (Scheme 1.16). The reaction is triggered by thiol reagents, like DTT **5**, 2-thioethanol or cysteine, and proceeds cleanly at low temperatures, but progresses slowly. The reaction can be carried out while on the chitin column, allowing for elution of clean cleaved protein and leaving the tag still bound to the column.



Scheme 1.16. The mechanism of intein-mediated protein cleavage.^[61]

Cysteine also displays a capacity to spontaneously undergo an N→S acyl shift (or N→Se shift in case of selenocysteine),^[62] particularly when positioned at the C-terminal side of glycine, histidine or another cysteine.^[63] The reaction is catalysed by aqueous acid at elevated temperatures (60°C), leading to the formation and, subsequently, hydrolysis of the thioester bond, resulting in protein fragmentation (Scheme 1.17).

The cleavage reaction can be carried out under milder conditions (neutral pH and lower temperatures) when catalysed by small thiol reagents like thioacetic acid **6** or 3-thiopropionic acid.^[63] The thiol group can intercept the rearranged peptides, but due to the participation of the neighbouring carboxylate the peptide thioester spontaneously hydrolyses in water, leading to the production of a free peptide.



Scheme 1.17. The mechanism of thioether-mediated proteolysis.^[63]

The multitude of methods for partial protein hydrolysis (and the continuing development) demonstrates that protein digestion plays an important role in protein research. Site-specific protein cleavage is essential for protein preparation, but is also important for protein analysis. For all the diversity among the above described methods, one commonality is the long reaction times needed to achieve completion. A more rapid method to selectively digest proteins would be beneficial.

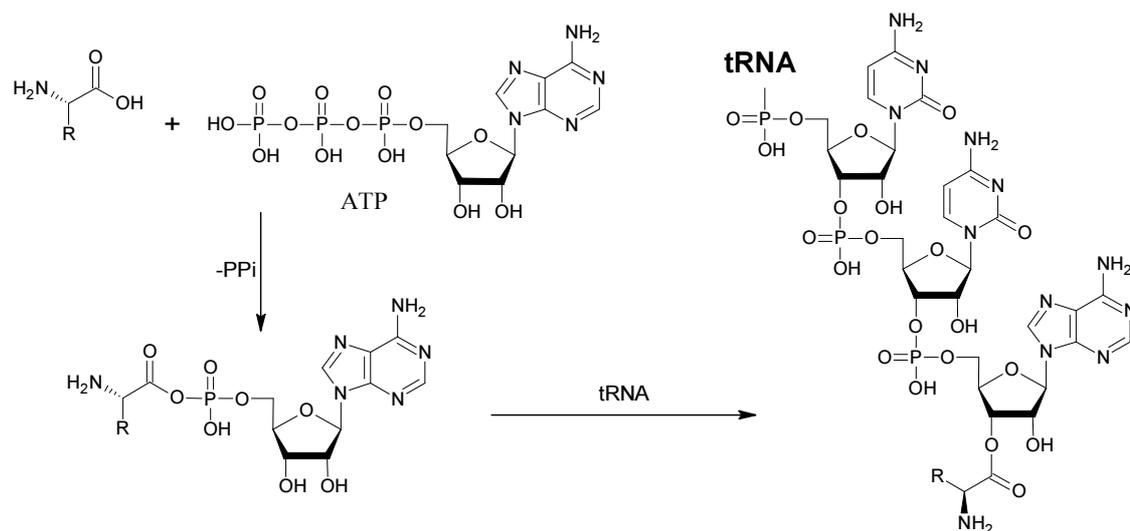
1.3. Unnatural amino acids in proteins

Proteins are made up primarily from 20 L- α -amino acids, known as canonical or proteogenic, encoded by the 64 codons of the genetic code. The sequence of amino acids in a protein is determined by the gene sequences during protein translation as the ribosome moves along messenger RNA (mRNA), transcribed from DNA.^[64] Transfer RNAs (tRNAs) with anticodons complementary to the mRNA codons deliver their respective amino acids to the ribosome as the nascent peptide chain grows.

The fidelity of protein translation relies on correct codon-anticodon match between the mRNA and tRNA in the ribosome as well as the tRNA being charged with the correct amino acid. Every proteogenic amino acid has its respective amino acyl-tRNA synthetase (AARS) that has evolved to discriminate the native, or cognate amino acid and the corresponding tRNAs.^[65, 66] The charging of a tRNA is a two stage process, involving first the activation of an amino acid by reacting it with ATP, followed by the reaction with the 3' stem of the tRNA (Scheme 1.18); throughout the whole process the amino acid remains bound to the enzyme.^[67]

The macromolecular interaction between the AARSs and the corresponding tRNAs enables correct substrate recognition.^[65] Discrimination of small amino acids, on the other hand, is significantly more difficult.^[67-69] Half of the synthetases are able to distinguish their substrate with complete selectivity by virtue of a specific interactions.^[69] Tyrosyl-tRNA synthetase (YRS) ensures proper recognition of the substrate by forming highly specific hydrogen bonds with the phenolic hydroxy group,^[70] while chemical properties of thiol groups enable cysteinyl-tRNA synthetase

(CRS) to selectively bind and activate cysteine even in the presence of isosteric serine.^[71]



Scheme 1.18. The charging of an amino acid.

The other half of AARSs requires stringent proof-reading ability to discard incorrectly recognised amino acids. Many possess two active sites, one for activation of potential substrates and one for proof-reading.^[67-69] This is referred to as the ‘double sieve’ mechanism:^[72] due to the difference of a single carbon unit, both isoleucine **7** and valine **8a** are activated by the activating domain of isoleucyl-tRNA synthetase (IRS), as valine **8a** is small enough to fit the active site, yet big enough to bind with sufficient affinity.^[73,74] The active site in the editing domain, the ‘second sieve’, however, is small enough to only accommodate and hydrolyse valine **8a**. (Later studies have also found proof-reading ability in the activating domain.) As a result of the rigorous screening valine **8a** replaces isoleucine **7** in proteins only 3 times out of 10000,^[75] contrary to the

early estimations by Linus Pauling in 1957.^[76] Similarly, the editing domain of valyl-tRNA synthetase (VRS) only accepts polar threonine that is activated along with the isosteric valine **8a** by the activating domain.^[77]

When a structurally related amino acid that the system had no exposure to throughout evolution is introduced, however, selection mechanisms may fail and the analogue becomes a potential substrate for the enzyme, albeit with lesser affinity. Dozens of non-canonical amino acids that are erroneously translationally incorporated into proteins due to structural similarity have been identified and utilised for research applications.^[78-81] (*S*)-Canavanine **9** (Figure 1.3), discovered in the 1960s, is an analogue of arginine **10** that leguminous plants produce to fend off insects; upon ingestion of plant material containing cavananine, the insects start producing non-functional protein with lethal consequences.^[82] Furanomycin **11** is an antibacterial agent isolated from a species of *Streptomyces*.^[83] While at first glance it bears little resemblance to any of the canonical amino acids, in 3D it is conformationally very similar to isoleucine **7** and was shown as a substitute in *E. coli* systems during protein synthesis.^[84] Azetidine-2-carboxylic acid **12** is an analogue of proline **13** found in sugar beets that was shown to be toxic to a wide range of species.^[85]

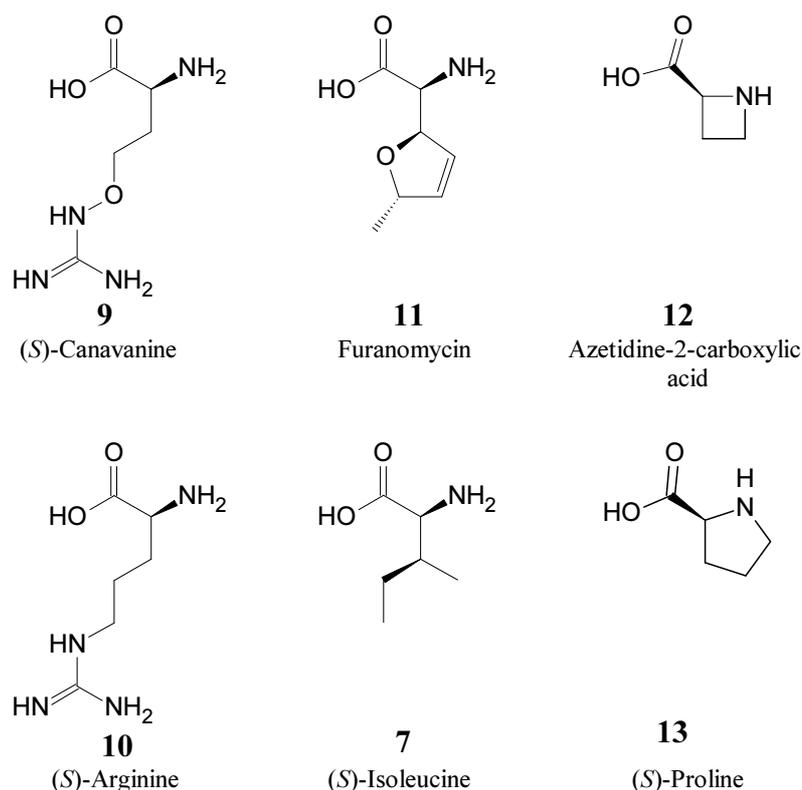


Figure 1.3. Some of the naturally occurring non-canonical amino acids (above) that are translationally incorporated into proteins and their proteogenic isosteres (below).

This promiscuity has been taken advantage of in research and the scope of mis-recognised unnatural amino acids has been further widened. (S)-Canavanine **9**, due to the non-discriminate disruption of protein function, was one of the most widely used analogues in protein function studies.^[78] Other amino acids were applied in more specific circumstances. Azetidine-2-carboxylic acid **12** was useful in collagen studies,^[78] the threonine analogue β -hydroxynorvaline **14** was found to interfere with protein glycosylation,^[78] 2-amino-3-chlorobutyric acid **15a** was used in protein degradation experiments in the 1960s through the 80s,^[86] and aza-tryptophan **16**, when substituted for tryptophan, produces fluorescent proteins that can be tracked without

further labelling^[87] (Figure 1.4). Even without detailed knowledge about the active site of the enzyme some predictions can be made about possible allowed modifications of the substrate, and with screening and improving knowledge of the AARSs the number of potential amino acid analogues has increased tremendously.^[79,80] Carefully selected analogues have been shown to produce functional proteins without apparent aberrations.

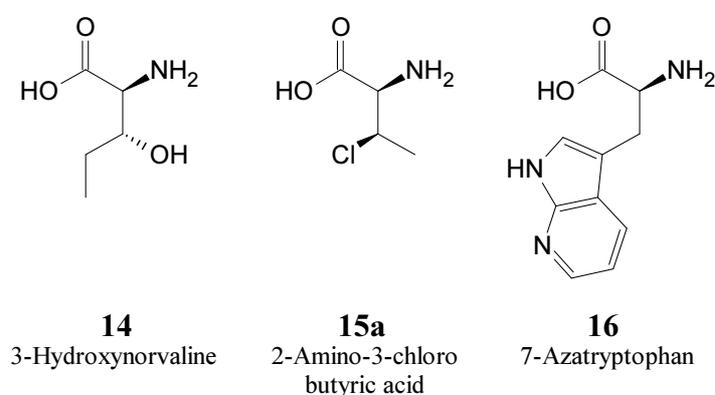


Figure 1.4. Some of the non-canonical amino acids used in protein studies.

Even though foreign amino acids are recognised by native AARSs, the cognate amino acids still bind with higher affinity and out-compete the analogues.^[69] Incorporation of toxic amino acids like canavanine **9**^[82] or furanomycin **11**^[84] that produce aberrant proteins does not have to be extensive for the toxin to be effective as small amounts of non-functional protein can disrupt cellular function, although a surplus of the cognate amino acid can alleviate the toxic effects. For practical applications, however, high degree of substitution is required; therefore, the natural amino acid must be excluded from the system. *In vivo* studies benefit from the innate inability of many tissues to synthesise essential amino acids that are often the targets for replacement. When the

work is carried out in competent organisms, however, this auxotrophy has to be introduced by a deliberate deletion of a gene involved in the biosynthesis of the substituted amino acid.^[88] Auxotrophic bacterial strains were successfully used to prepare proteins with high degree substitution with unnatural amino acids analogues, but a separate strain is required for each tested amino acid.^[80,89-91] Cell-free protein expression systems are more favourable than whole-cell approaches when unnatural amino acids are involved as the natural counterparts can be completely excluded during the system reconstitution.^[14,16,17] Therefore, cell-free systems tend to give higher degree of amino acid substitution due to lower background levels. Recent advances enabled incorporation of multiple (as many as five) unnatural amino acids in a single protein,^[81] whereas multiply auxotrophic cells are difficult to generate. Additionally, cell-free systems are not poisoned by toxic amino acids (regardless of mechanism of toxicity).

More recently, the scope of possible analogues was further widened by genetic manipulation of the AARSs. In 1991 phenylalanyl-tRNA synthetase (FRS) was mutated to relax the size restriction of the substrate.^[92] While the wild type synthetase could mis-incorporate *p*-fluorophenylalanine **17a** in place of phenylalanine **17**, the mutant with the relaxed active site was able to recognise *p*-chlorophenylalanine **17b** and *p*-bromophenylalanine **17c** as well. Directed evolution coupled with structural knowledge of AARSs allowed for the incorporation of nearly a hundred unnatural amino acids with novel functionalities that facilitated protein chemical labelling, crosslinking or spectroscopic monitoring capabilities.^[93,94] Through genetic modification of the tRNAs to recognise one of the 'stop' codons or an extended 4-base codon, the foreign amino acid recognised by the engineered AARS becomes the 21st

amino acid, incorporated site-specifically into an engineered protein; this is generally referred to as the ‘expanded genetic code’.

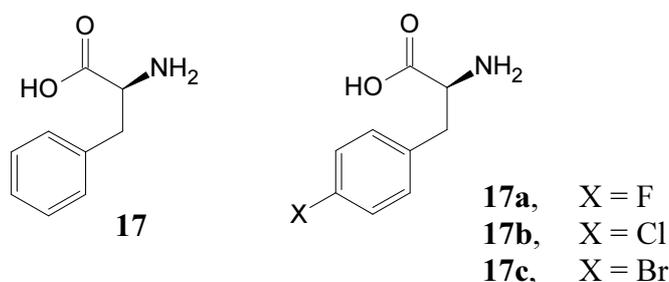
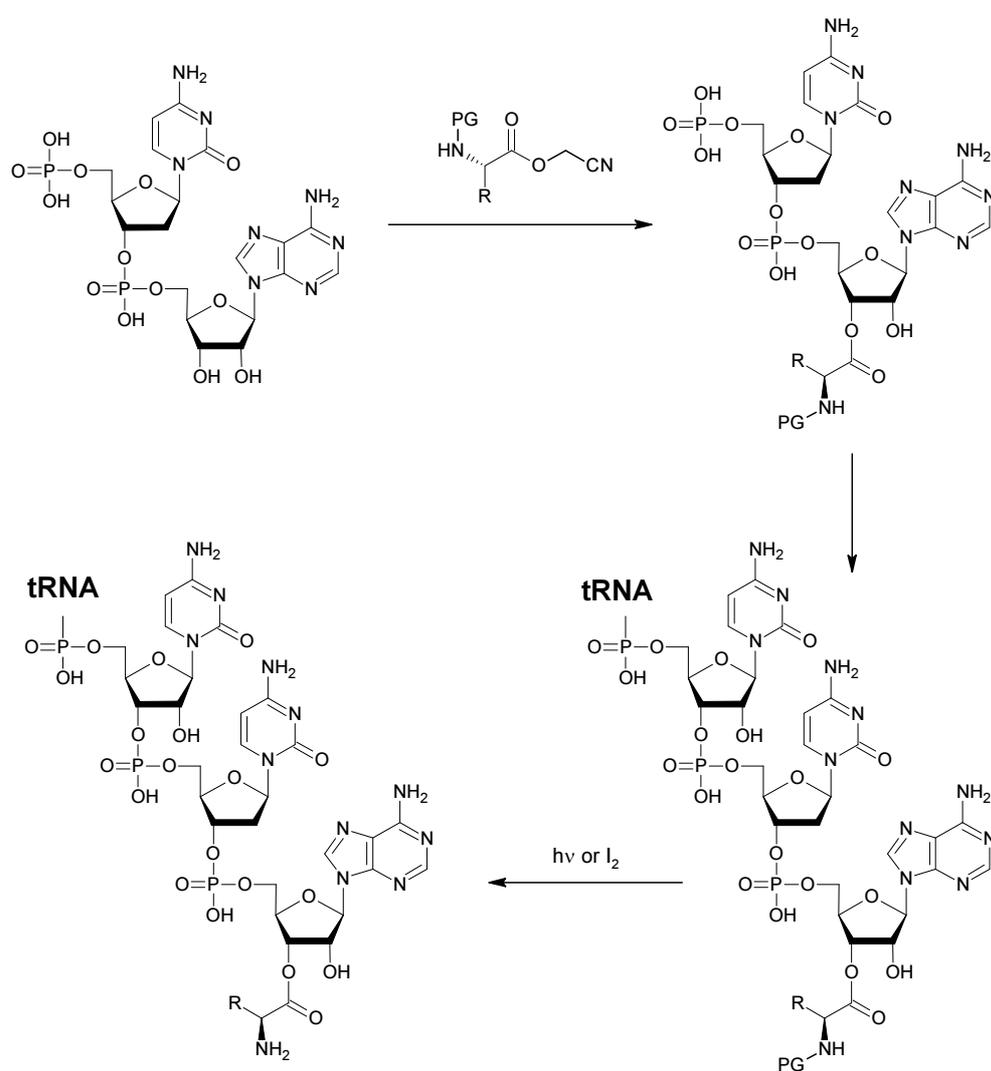


Figure 1.5. *Para*-halogenated analogues of (*S*)-phenylalanine **17**.

Chemical protein synthesis allows even greater diversification of the protein structure, alleviating restriction of the polymer assembly to (*2S*)- α -amino acids,^[25,27] and native chemical ligation strategies enable the fusion of such modified peptides with longer proteins.^[34] The chemical synthesis approach allows almost unlimited product design, but more elaborate functionality requires customised synthetic procedures, limiting the use of automation and thus making the procedures prohibitively long and complicated. Hence, unnatural compounds most commonly introduced into peptides through chemical peptide synthesis are still highly reminiscent of the natural counterparts,^[3] with inverted stereocentres (D-amino acids), *N*-methylation (*e.g.* sarcosine), modified chain length, expanded aromatic systems, hydrogen-to-fluorine substitutions, β -amino acids and des-amino appendages at the N-termini.

A hybrid method between chemistry and biology for the incorporation of modified amino acids was also developed, whereby the tRNA is chemically charged with an amino acid that is then incorporated into a growing protein chain *via* ribosomal

translation.^[52,53,60,95] A dinucleotide from the end of tRNA is acylated with an *N*-protected amino acid and subsequently enzymatically ligated onto a truncated tRNA molecule; the photo-labile *N*-6-nitroverastroxycarbonyl (NVOC) or the iodine-labile *N*-4-pentenoyl protecting group is then removed (Scheme 1.19). The charged tRNA is isolated and supplied to the expression mixture.



Scheme 1.19. Chemical acylation of tRNA.

The benefits of the expanded range of amino acids incorporated into proteins are undeniable. However, considerable efforts have to be invested into the engineering of artificial systems that reliably produce modified proteins. The natural promiscuity of amino acyl-tRNA synthetases, on the other hand, can be easily exploited for the incorporation of unnatural amino acids and still holds a lot of unexplored potential. The innate promiscuity of the protein synthetic machinery was utilised in the work described in this Thesis.

1.4. Analogues of the branched-chain amino acids

The three amino acids that comprise the branched-chain amino acid (BCAA) family, (*S*)-valine **8a**, (*S*)-leucine **18** and (*2S,3S*)-isoleucine **7**, all have simple and relatively similar branched hydrocarbon chains as side-chain. Thus, to distinguish between the them, the corresponding synthetases all possess editing sites, making discrimination based on hydrophobic size and shape possible.^[69,74,77,96] Structurally related compounds, however, are able to compete with the canonical substrates and evade proof-reading, as exemplified by the analogues **19-21** (Figure 1.6).^[78]

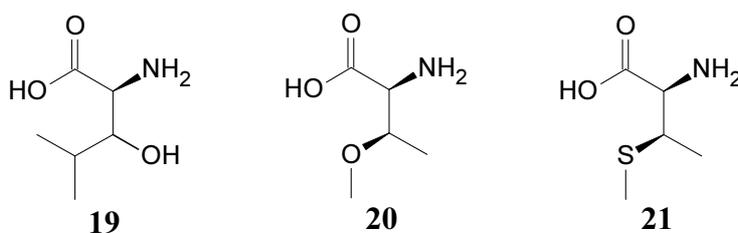
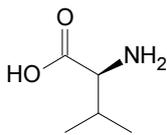
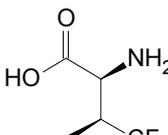
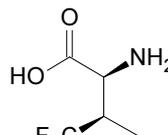
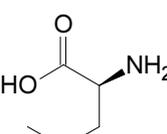
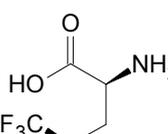
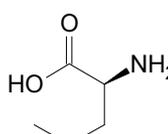
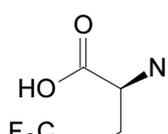
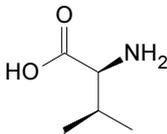
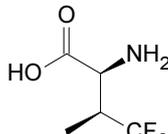
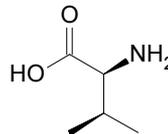
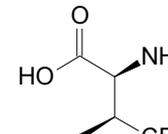
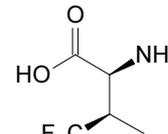


Figure 1.6. Analogues of (*S*)-leucine **18** and (*2S,3S*)-isoleucine **7** used to study protein processing.

Fluorine is a typical isostere for hydrogen.^[97,98] The recognition of fluorinated BCAA isosteres by the protein synthetic machinery was tested through a number of studies and in most cases the hydrogen-to-fluorine substitution did not preclude the incorporation of the tested analogues (Table 1.1). For some analogues the synthetase had to be over-expressed to make the incorporation noticeable, but the active site was not altered.

Table 1.1. Tested fluorinated analogues of the branched chain amino acids.

Natural substrate	Unnatural analogues				
 <p>8a</p>	 <p>22a^[99] ✘</p>	 <p>22b^{*,[99]} ✔</p>			
 <p>18</p>	 <p>23a^[100] ✔</p>	 <p>23b^[100] ✔</p>	 <p>23c^{*,[101]} ✔</p>		
 <p>7</p>	 <p>24a^[91] ✘</p>	 <p>24b^[91] ✔</p>	 <p>22a^[99] ✘</p>	 <p>22b^{*,[99]} ✔</p>	

✔-amino acid was incorporated into protein

*-the synthetase was over-expressed

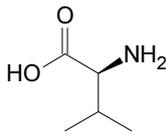
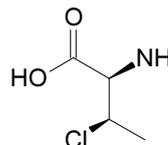
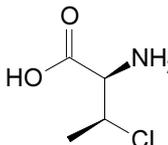
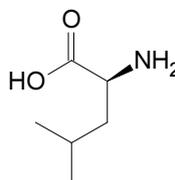
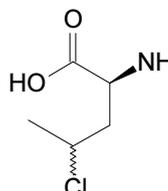
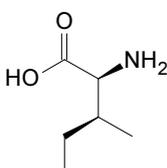
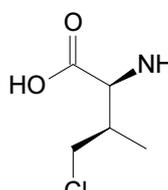
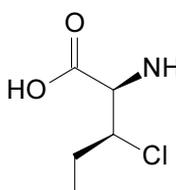
✘-amino acid was tested for incorporation, but no incorporation detected

Hydrogen and fluorine are similar in size, but, when grouped together, the trifluoromethyl group becomes substantially bigger than a methyl group.^[98] One diastereomer of trifluorinated valine **22a** and one regiomer of modified isoleucine **24a** were rejected by the corresponding synthetases; no enzyme activity was detected in the presence of the two analogues, suggesting they are rejected by the activating domain (the ‘first sieve’),^[91,99] while the analogue of valine **22b** was recognised by IRS.^[99]

A recent study re-evaluated the ability of chlorinated isosteres to emulate BCAAs (Table 1.2).^[102] The analogue **15a** was shown to be so potent that it could even effectively compete with (*S*)-valine **8a**, the natural substrate of VRS; the diastereomer **15b** was less efficient. Leucyl-tRNA synthetase (LRS) was also unable to discriminate against the chloride **25**, although IRS appeared to reject the chlorinated analogue **26a**, an isostere of (*2S,3S*)-isoleucine **7** (the regiomer **27** was not tested in the study). ATP turn-over by IRS in the presence of **26a** indicated proof-reading activity. In light of these data it was proposed that the difference in size between the van der Waals radii of a chloro and a methyl groups (1.75 Å vs 2.0 Å) is more apparent on a primary chloride and IRS is able to register the difference.

Based on these studies IRS appeared to have the capacity to register very tiny changes in the size of potential substrates. The seemingly opposite substitution pattern of the rejected isoleucine isosteres **26a** and **24a** prompted an investigation into the recognition mode of IRS, which is described in Chapter 2 of this Thesis. The chemical properties of the amino acid **26a** uncovered in the study led to the development of a new method for the preparation of small peptides; this application is described in Chapter 3.

Table 1.2. Chlorinated analogues of the branched chain amino acids with a single methyl group-to-chlorine substitution.

Natural substrate	Unnatural analogues	
 <p>8a</p>	 <p>15a^[102] ✓</p>	 <p>15b^[102] ✓</p>
 <p>18</p>	 <p>25^[102] ✓</p>	
 <p>7</p>	 <p>26a^[102] ✗</p>	 <p>27*</p>

✓-amino acid was incorporated into protein

✗-amino acid was tested for incorporation, but no incorporation detected

*-not tested

The chloride **15a** was known to interfere with protein function and had been utilised in studies of cellular protein expression.^[86] Antimicrobial effect against select species that was mitigated by leucine was also observed for the chloride **25**,^[103] with the

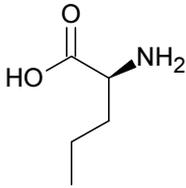
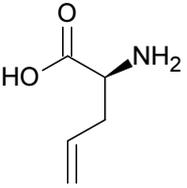
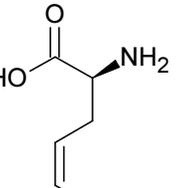
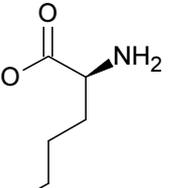
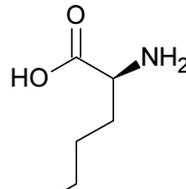
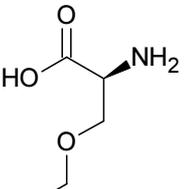
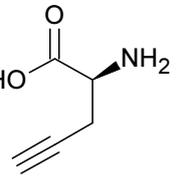
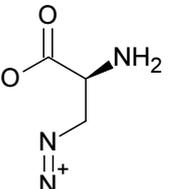
incorporation studies^[102] pointing to a potential mode of action. The underlying effects on protein, however, had not been explored.

The effects of (*S*)-canavanine **9** are well understood: the additional oxygen atom reduces the basicity of the guanidino group in turn reducing the ability to form salt bridges and other polar interactions that proteins rely on for function.^[82] The mode of action of many other analogues found in nature is not so clear cut, and while hypotheses have been proposed, the true mechanism is not yet certain and warrants further investigation. To investigate the potential effects of halogenated amino acids on proteins, bovine pancreatic trypsin inhibitor (BPTI), also known as aprotinin, was expressed with valine analogues **15a** and **15b** and crystal structures of the modified protein were obtained. This work is described in Chapter 4.

1.5. Substitution of methionine

Methionyl-tRNA synthetase (MRS) is particularly notorious for substrate promiscuity.^[104] Table 1.3 lists the structural analogues of (*S*)-methionine **28** that are accepted by MRS from *E. coli*, the largest array of substrate analogues of any AARS, as well as structurally related rejected compounds. Some of the amino acids were only incorporated into proteins when the enzyme was over-expressed and/or the amino acid was supplied in unnaturally high concentrations, but the enzyme was not modified. Taking advantage of selenomethionine **29** that revolutionised protein crystallography^[105] and effective protein tagging with the introduction of azidohomoalanine **34**^[106] and the unsaturated analogues **35-36**,^[107,108] MRS is becoming the most frequently exploited synthetase. Similar substrate promiscuity is observed with orthologues from yeast^[109] and multicellular organisms, with even more diverse diazo-compounds incorporated in place of (*S*)-methionine **28**.^[110]

Table 1.3. continued.

MRS overexpressed; analogue supplied in large concentration				
	42 ^[107]	43 ^[107]	44 ^[107]	45 ^[107]
Not incorporated				
	46 ^[107]	47 ^[107]	48 ^[107]	49 ^[106]

Methionyl-tRNA synthetase (MRS) only possesses one active site, used for both activation and proof-reading.^[117] The hydrophobic pocket is created upon substrate binding, resulting in extensive rearrangement of the protein, and a bonding interaction is established between the thioether of methionine side-chain and ζ O of tyrosine Y260 in the active site (Figure 1.7) to ensure correct substrate recognition.

The proof-reading ability of the active site relies on the chemistry of bound substrate. The primary competitor for methionine **28**, present in cells under physiological conditions, is the precursor homocysteine **50**.^[75] Effective rejection of homocysteine **50**

by MRS is critical as erroneous production of proteins with the thiol **50** (that can occur *via* a different pathway) has lethal consequences for the organism.^[118] As a free thiol, homocysteine **50** can undergo thio-lactonisation to form a 5-membered ring **51**, whereas methionine **28**, as a thioether, is unable to follow that pathway. Following mis-activation of homocysteine **50**, the synthetase then facilitates cyclisation of the adenylylated amino acid and release of the degraded thiolactone **51** (Scheme 1.20),^[75] whereas methionine **28** and other non-hydrolysable analogues are transferred onto the ^{Met}tRNA and incorporated into proteins.

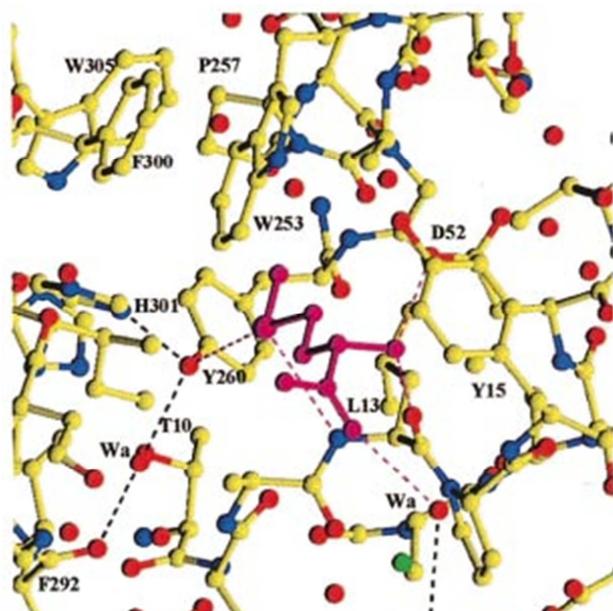
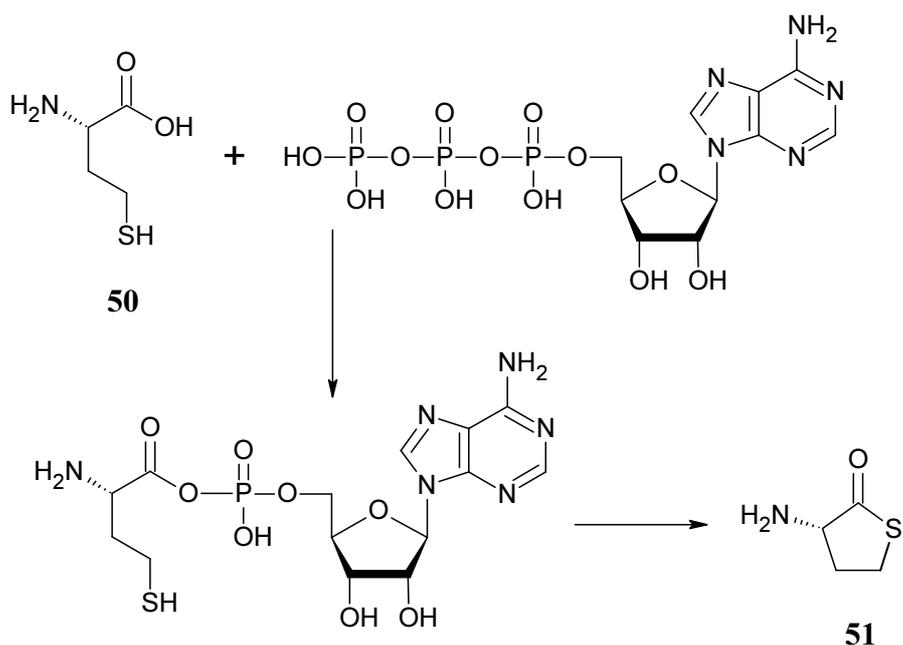


Figure 1.7. The active site of methionyl-tRNA synthetase with methionine **28** (purple) bound. A bond between the sulfur of methionine **28** and ζ O of Tyr260 is important for substrate recognition. Reproduced with permission from Elsevier.^[117]



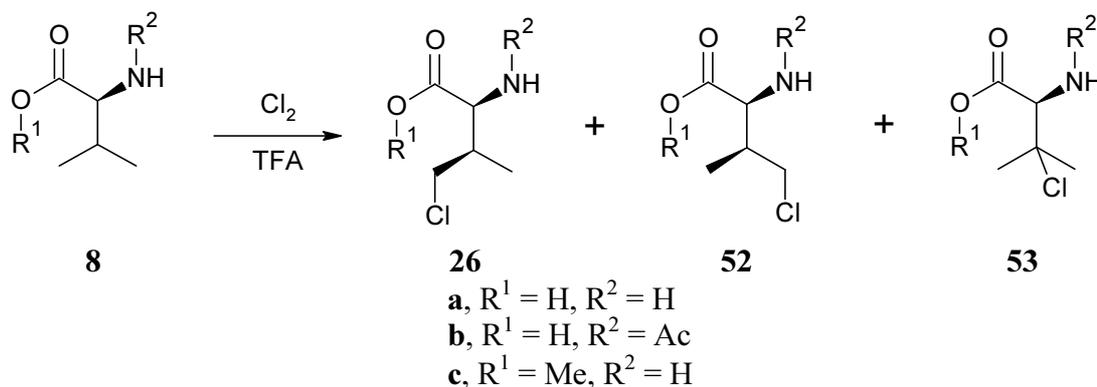
Scheme 1.20. The proofreading mechanism of mis-activated homocysteine **50** by MRS.

The highly specialised editing mechanism of MRS directed against a single target makes mis-incorporation of a vast array of analogues possible. The promiscuity of MRS was exploited in the introduction of peptide chain-cleaving analogues of (*S*)-methionine **28**; this work is described in Chapter 5.

Chapter 2. Substrate Selectivity of Isoleucyl-tRNA Synthetase



As indicated in the Introduction, the chlorinated isoleucine analogue **26a** was reported to be rejected by IRS during protein expression, standing out from the tested chlorinated analogues of the BCAA family (Table 1.2). γ -Chlorovaline **26a** was produced by direct treatment of (*S*)-valine **8a** with chlorine in trifluoroacetic acid (Scheme 2.1). The mixture of the isomers **26a**, **52a** and **53a** and the unreacted starting material **8a** was used for protein synthesis in the original study as only the acid **26a** was a potential substitute for isoleucine 7.^[102] For more detailed enzyme binding studies, however, pure material was required.



Scheme 2.1. Preparation of γ -chlorovaline **26a** and its derivatives by direct chlorination.

It was anticipated that the products of chlorination with protected N-termini would be easier to fractionate on reverse phase HPLC as reduced polarity of the compounds

would increase interactions with the stationary phase. Free and *N*-acetylated amino acids lead to similar patterns of side-chain modification during the chlorination reaction,^[119] hence (*S*)-*N*-acetylvaline **8b** was used for chlorination (Scheme 2.1; $R^1 = H, R^2 = Ac$). The obtained mixture was analysed by reverse phase HPLC and good resolution was achieved on an Alltima C18 column (22×250 mm, 5 μ) eluting with 20% methanol in 0.1% TFA (Figure 2.1). Three peaks (*) matching the mass of the isomers **26b**, **52b** and **53b** were identified by ESI-MS.

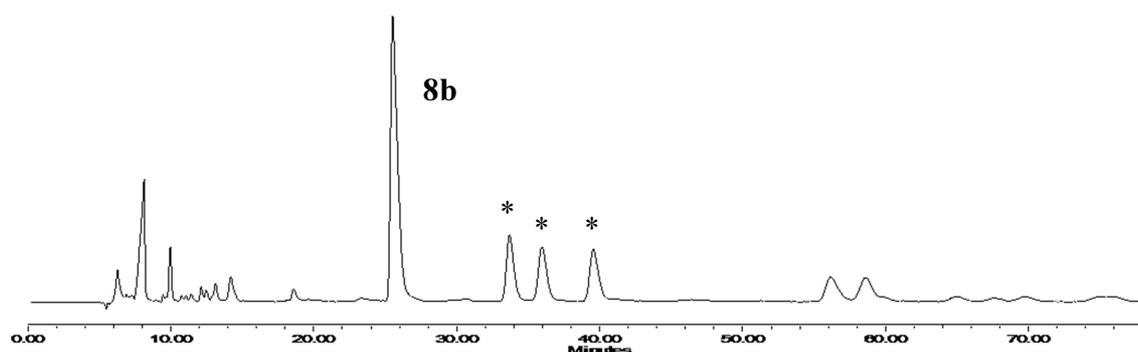


Figure 2.1. HPLC chromatogram showing the separation of the isomers **26b**, **52b** and **53b** on an Alltima C18 column (22×250 mm, 5 μ), eluting with 20% methanol in 0.1% TFA, 10 mL/min.

H_β signals in the 2 ppm to 3 ppm region in the proton NMR spectrum are the most diagnostic in the identification of the chlorination products of (*S*)-*N*-acetylvaline **8b** due to great separation (Figure 2.2). The H_β multiplet of the unreacted starting material **8b** resonates at δ 2.16 ppm and the corresponding protons of the γ -chlorinated products **26b** and **52b** appear at δ 2.28 ppm and δ 2.46 ppm (stereochemistry not assigned in the original study).^[119] The β -chlorinated product **53b** is not visible in the chosen region as

it lacks H β protons and is identified by the signals of the methyl groups at δ 1.73 ppm and δ 1.64 ppm.

Material eluting at 39.6 min had the signals attributable to the β -chloride **53b** and was thus no longer pursued. Signals at δ 2.46 ppm and δ 2.28 ppm (weak) corresponding to the sought diastereomers **26b** and **52b** were detected in the 33.7 min and 36.0 min fractions, respectively, but the same fractions also contained very prominent signals at δ 2.61 ppm and δ 2.77 ppm that were very weak or not observed at all in different batches of the original mixture of chlorination products.

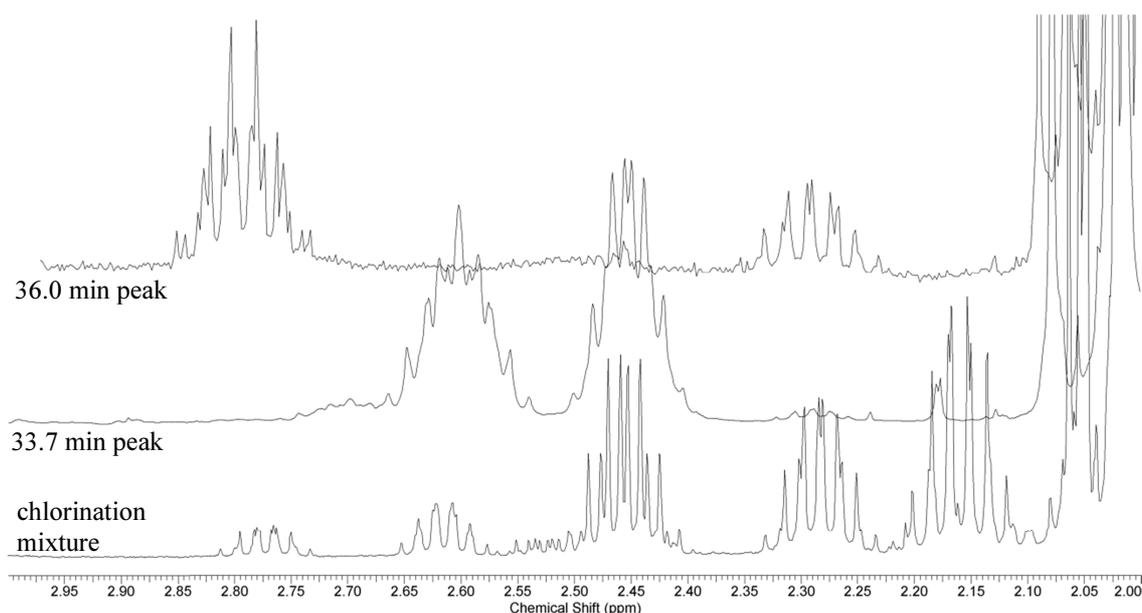


Figure 2.2. The diagnostic region of ¹H NMR of the chlorination mixture of (*S*)-*N*-acetylvaline **8b** and two purified HPLC fractions.

Sample contamination was ruled out after several purification attempts as the ‘contamination’ pattern was reproducible and unique to individual fractions. It thus

became apparent the unidentified peaks at δ 2.61 ppm and δ 2.77 ppm were arising due to degradation of the chlorinated compounds **26b** and **52b**. Lactone formation was considered as the most likely route to decomposition as 5-membered rings are highly favourable thermodynamically. The degradation products were isolated by HPLC and analysed by ^1H NMR (Figure 2.3).

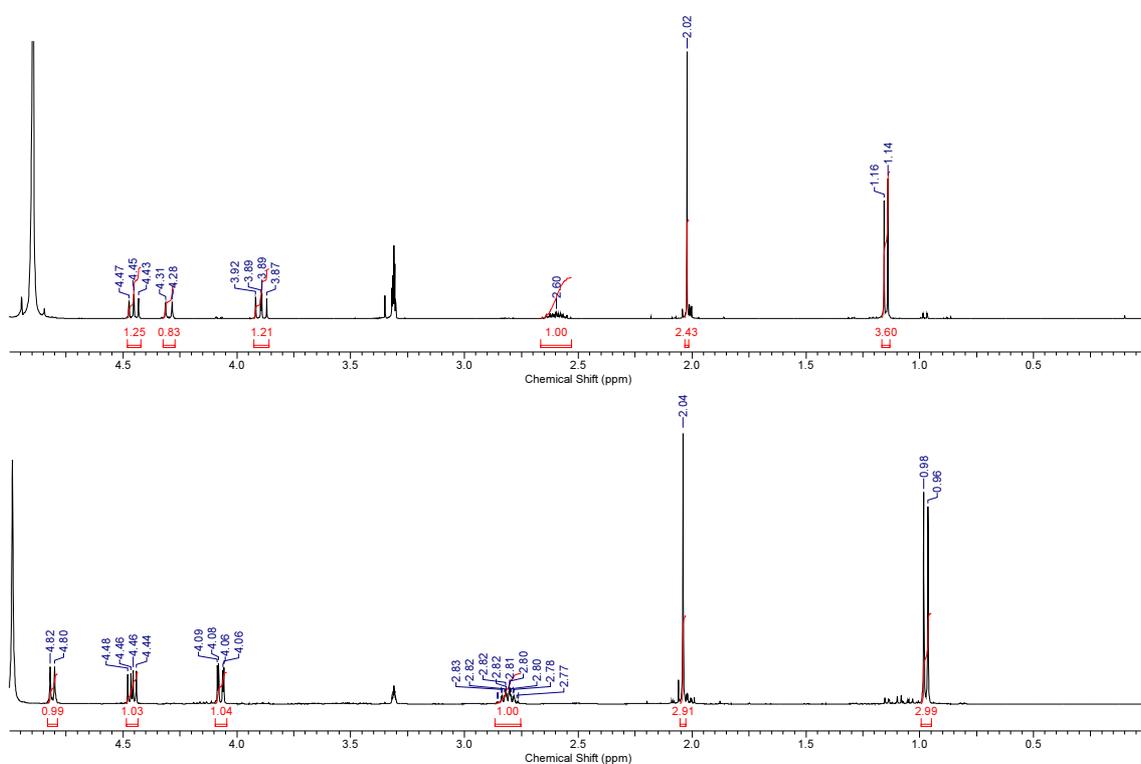


Figure 2.3. ^1H NMR spectra of the decomposition products of the chlorinated products eluting at 33.7 min (above) and 36.0 min (below).

The downfield shift of the H_γ proton signals from δ 3.5 ppm for both diastereomers **26b** and **52b**^[119] indicated the chlorides were substituted by the more de-shielding carboxylate groups and together with the geminal splitting of the H_γ protons

(δ 4.45 ppm and δ 3.89 ppm for the 33.7 min peak and δ 4.46 ppm and δ 4.07 ppm for the 36.0 min peak) confirmed the presence of the rings. Thus, the NMR analysis confirmed the decomposition products were the 5-membered lactones.

Stereochemistry of the lactones was then investigated. Configuration at the α -centre was known as (2*S*)-amino acids were used for the synthesis. Conformational analysis was employed to determine the relative stereochemistry between the two stereo centres. Avogadro modelling software was used to model the geometry of the lactone products. For the *trans*-lactone the dihedral angle between the H_α and H_β hydrogens was calculated to be 164° and the angles between H_β and each of the H_γ protons were 39° and 162° ; for the *cis*-lactone the dihedral angle between the H_α and H_β was 39° and the angles between H_β and each of the H_γ were 37° and 86° (Figure 2.4).

For related systems, a dihedral angle of 164° between two protons is expected to result in substantially stronger coupling between the two protons than 39° , hence the lactone with the coupling constant $^3J_{\alpha\beta} = 11.3$ Hz at the H_α position and was identified as the *trans*-lactone **54b** and the decomposition product with the coupling constant $^3J_{\alpha\beta} = 7.8$ Hz and was identified as the *cis*-compound **54a**. The coupling constants $^3J_{\beta\gamma} = 10.5$ Hz and 8.3 Hz were consistent with the predicted 162° and 39° angles for the *trans*-lactone and $^3J_{\beta\gamma} = 5.9$ Hz and 2.1 Hz corresponded well with 37° and 86° angles for the *cis*-ring. As the stereocentres of the chlorinated products are not affected during lactonisation, the established stereochemistry on the rings **54a** and **54b** could be translated to the starting materials **26b** and **52b** (Scheme 2.2).

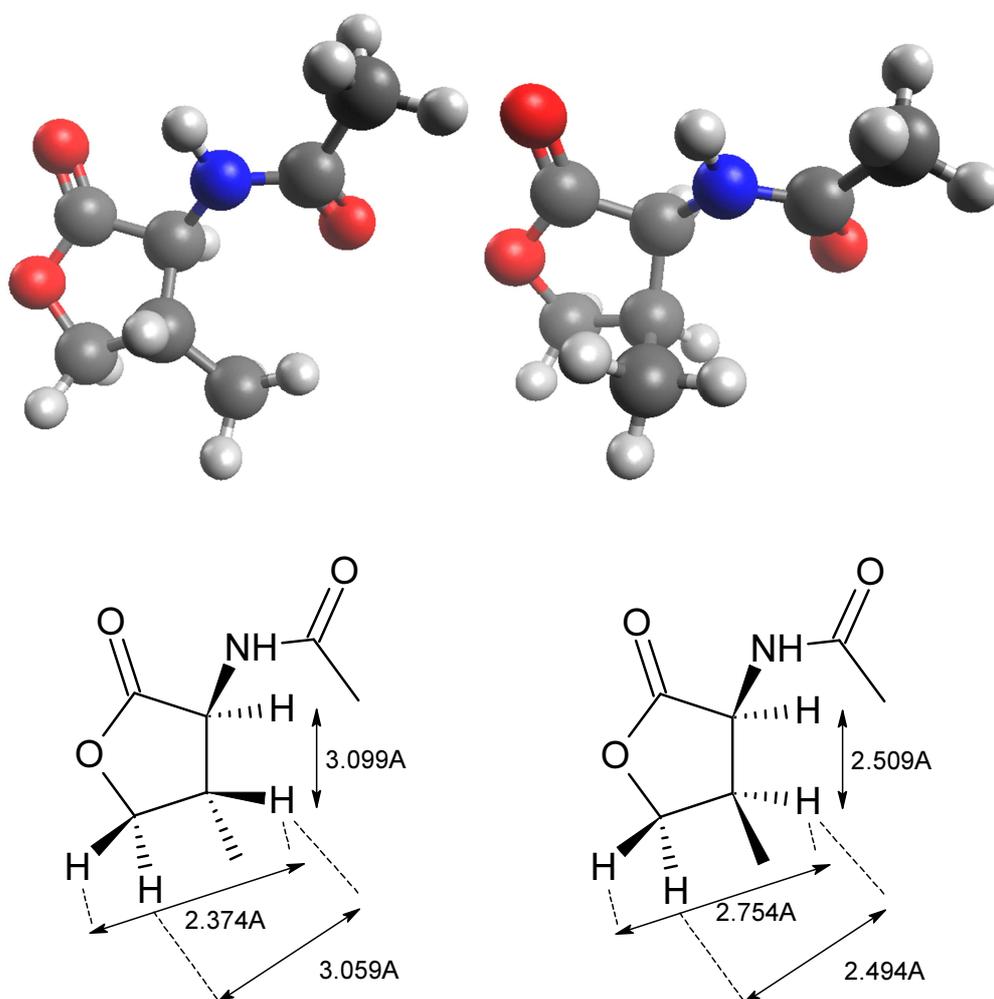
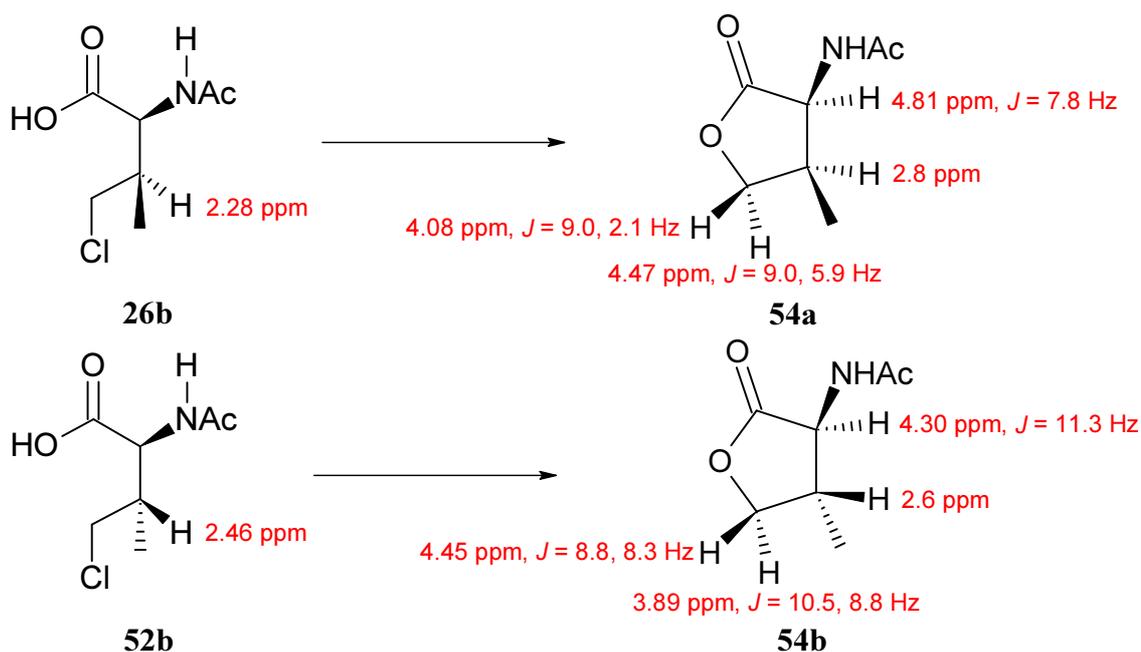


Figure 2.4. Top: ball and stick models of the *trans* (left) and *cis* (right) lactones. The geometries were optimised with Avogadro software. Bottom: distances between the spatially fixed protons of the two lactones, measured from the optimised structures.



Scheme 2.2. Decomposition of the diastereomers **26b** and **52b**.

Nuclear Overhauser Effect (NOE) was employed to verify the coupling constant analysis. Selective irradiation of a single proton cross-excites the other protons in the molecule, and the magnitude of excitation is proportional to proximity through space, hence allowing relative distance measurements between the interacting protons. As the experiment is repeated with longer mixing times, *i.e.* the protons are allowed to interact for longer, a linear relationship between time and intensities of the signals of the cross-excited protons is expected, and a larger gradient of the correlation trend indicates the protons are closer in space.

Based on the modelled structures of the lactones **54a** and **54b**, the spatial distance between H_{α} and H_{β} is comparable to the distance between H_{β} and the H_{γ} proton on the same side of the ring (Figure 2.4; bottom). Hence, measurement of relative distances between the concerned protons could be used to confirm relative stereochemistry.

For the *cis*-lactone **54a** the resonance at δ 2.8 ppm was irradiated (Figure 2.5). As expected, the H_α peak at δ 4.81 ppm and the two H_γ protons at δ 4.47 ppm and δ 4.08 ppm were cross-excited. A weaker signal at δ 1 ppm for the methyl group was also visible, but that was less useful for structural analysis due to the free rotation of the C-C bond.

As the mixing time τ was increased the NOE became stronger (linearly with respect to the mixing time initially, then started to plateau at $\tau > 400$ ms), but the relative intensities between the signals remained constant. The NOE of the δ 4.47 ppm signal was significantly stronger than the δ 4.08 ppm signal, suggesting the δ 4.47 ppm signal corresponds to the H_γ on the same face of the ring as the H_β and the δ 4.08 ppm signal corresponds to the H_γ on the opposite face. The δ 4.81 ppm NOE was marginally stronger than the δ 4.47 ppm signal, indicating the H_α and the H_β protons are on the same side of the ring, thus confirming the identity of the lactone **54a** as *cis*, consistent with the coupling constant analysis.

A similar analysis was attempted on the lactone **54b**, but was not successful as the NOE resulted in partially inverted signal.

Finally, the measurements were compared to previous reports.^[120] The *cis*-lactone **54a** had been isolated from a marine source. For consistency, all analysis of the compounds in the current work was carried out in methanol. When recorded in deuterated chloroform, as reported,^[120] the spectrum of the lactone **54a** was a perfect match to the reported values, thus unambiguously confirming the determined stereochemistry.

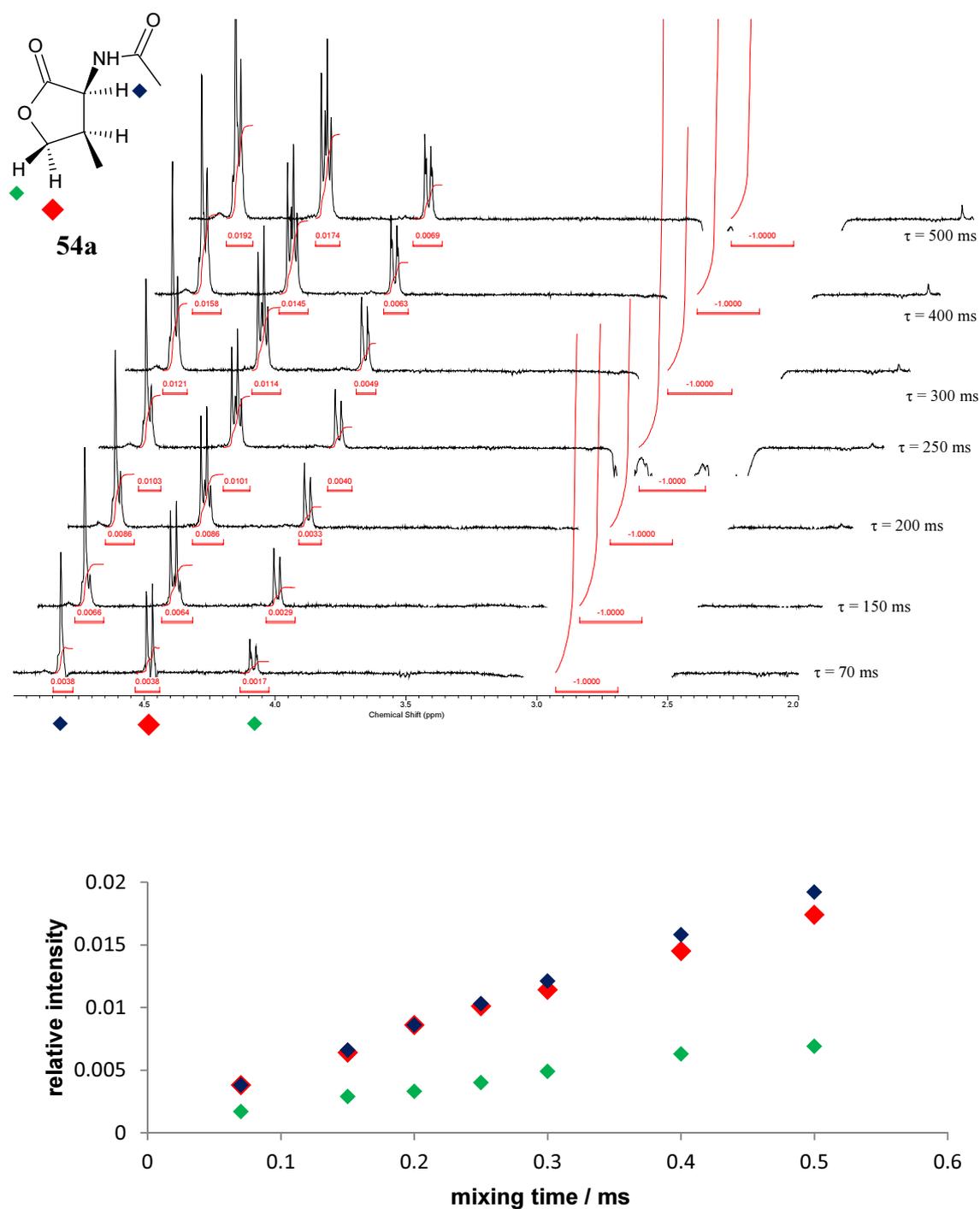


Figure 2.5. Top: NOE measurements of the lactone **54a**. The H_β at δ 2.8 ppm was selectively irradiated and cross-excitation was measured after 70-500 ms. Bottom: the plot of relative intensities of the cross-excited δ 4.81 ppm (◆), δ 4.47 ppm (◆) and δ 4.08 ppm (◆) peaks against the mixing time.

The mixture of the chlorinated compounds **26b** and **52b** could be stored indefinitely without decomposition, likely due to the residual TFA suppressing the nucleophilicity of the carboxylate, but under aqueous conditions lactonisation of the labile species made purification impossible. As the acetamido group does not contribute to the process of lactonisation, similar behaviour would be expected from the free amino acids **26a** and **52a**. Interestingly, no decomposition had been previously observed.^[102] Decomposition of the substrate **26a** under physiological conditions, depending on the rate, could account for the apparent rejection of the chloride **26a** by IRS, thus stability of the free amino acids was re-examined.

The key regions of ¹H NMR spectra were nearly identical for the chlorination mixtures of (*S*)-valine **8a** and (*S*)-*N*-acetylvaline **8b**, with the H_β resonances of the acetylated compounds shifted up-field correspondingly (Figure 2.6). Thus, the structural analysis conducted on the acetylated compounds **26b** and **52b** was easily correlated to the free amino acids **26a** and **52a**. Based on this, the H_β signal of the biologically relevant (*2R,3S*)-3-chlorovaline **26a** was at δ 2.40 ppm and the δ 2.62 ppm signal corresponded to the (*2S,3S*)-diastereomer **52a**, with the signal of the starting material (*S*)-valine **8a** at δ 2.31 ppm.

To assess the stability of γ-chlorinated amino acids **26a** and **52a** under physiological conditions, the mixture of chlorinated isomers was dissolved in D₂O and pD was adjusted to *ca.* 7. ¹H NMR spectra were recorded every 10 min for 2 h at 37°C and H_β peak intensities of the amino acids **26a** and **52a** were measured relative to that of valine **8a**. Even during the first scan (*t* = 0), the γ-chlorovaline **26a** resonance at δ 2.40 ppm was significantly smaller than the diastereomer **52a** signal at δ 2.62 ppm (Figure 2.7),

having partially decomposed during the set-up of the experiment, signifying the biologically relevant diastereomer **26a** is significantly less stable of the two; they are produced in roughly equal amounts during the chlorination reaction. The amino acid **26a** was essentially fully consumed within the first hour of the experiment and could no longer be reliably measured against the background, whereas decomposition of the diastereomer **52a** was tracked throughout the experiment. Logarithmic plot showed 1st order kinetics and the compound **26a** was found to decompose with a half-life of *ca.* 15 min, whereas the by-product **52a** was substantially more stable with a half-life of just under 4 h. Formation of the lactonisation products was also visible during the experiment (δ 2.8 ppm), but the compounds could not be used to monitor the decomposition of the chlorides as they were themselves unstable.

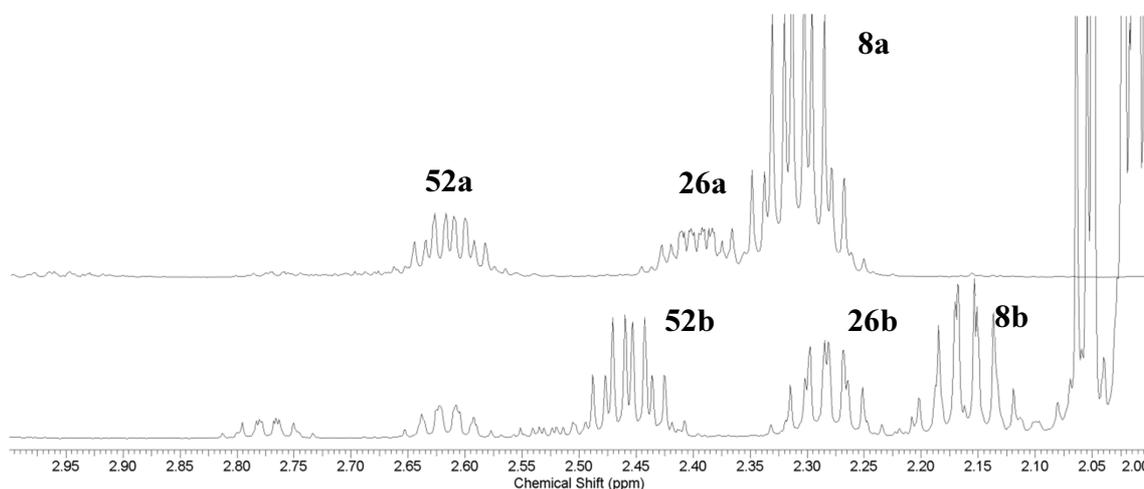


Figure 2.6. Comparison of ¹H NMR spectra of free (above) and acetylated (below) amino acids in methanol.

Chlorinated Amino Acids in Peptide Production

Chapter 2. Substrate Selectivity of Isoleucyl-tRNA Synthetase

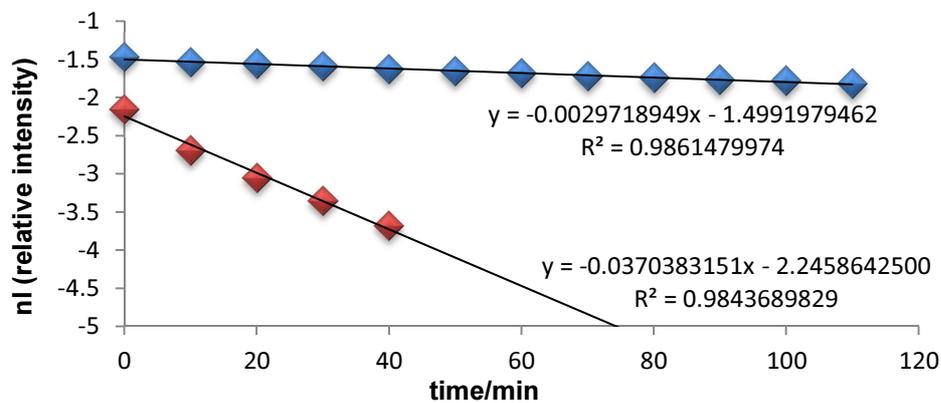
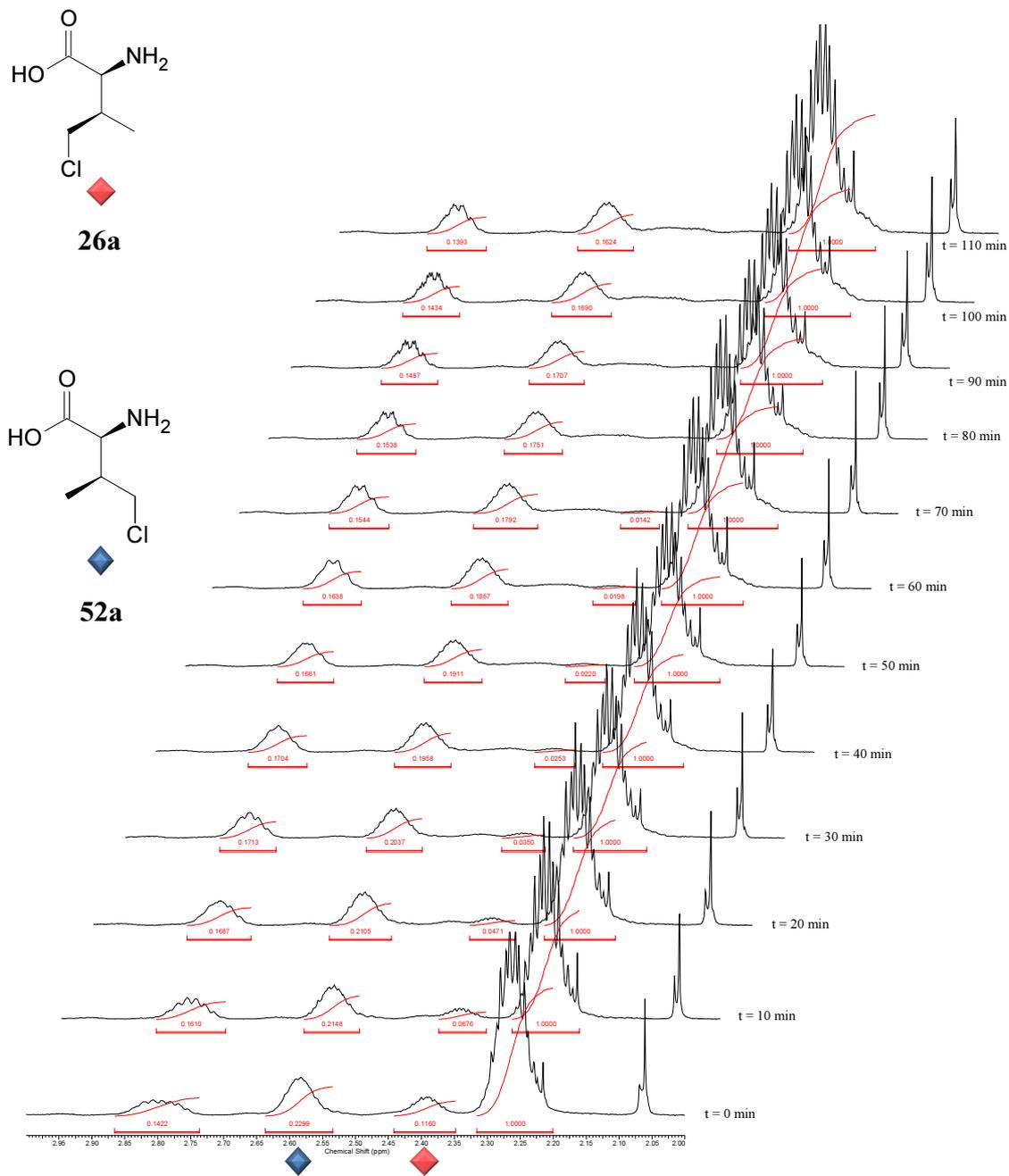


Figure 2.7. Decomposition of the amino acids **26a** (◆) and **52a** (◆) under physiological conditions. Top: intensities of H_β signals relative to H_β signal of valine **8a**. Bottom: logarithmic plots of the intensities; points for the chloride **26a** were ignored past 40 min.

The findings of the kinetic test are in direct contradiction to the previous report.^[102] The propensity of the isoleucine analogue **26a** to lactonise under physiological conditions might lead to the complete consumption of the IRS substrate before the cell-free reaction was even set up, thus appearing as though the amino acid **26a** was rejected. However, the chlorinated analogue of leucine **25** (whose two diastereomers were reported to lactonise with half-lives of 2.0 and 2.6 h)^[102] was successfully incorporated into proteins in the same study. To test whether the instability of the isoleucine analogue **26a** under physiological conditions was precluding protein synthesis, the reaction protocol was altered to ensure that the chloride **26a** was present in the mixture at the start of the reaction.

When amino acids produced by chlorination are used for protein expression, the sample is first neutralised as even drying under high vacuum fails to completely remove TFA. (The residual TFA inhibits lactonisation, thus crude samples can be stored on shelf indefinitely.) To ensure the desired amino acid was present in the reaction mixture at least in the early stages of the experiment, mixtures containing γ -chlorovaline **26a** and *N*-acetyl- γ -chlorovaline **26b** were neutralised in the final minutes before the assembly of the reaction chamber. Under these conditions small amounts of the test protein peptidyl-prolyl *cis-trans* isomerase B (His₆PpiB) were produced (Figure 2.8).

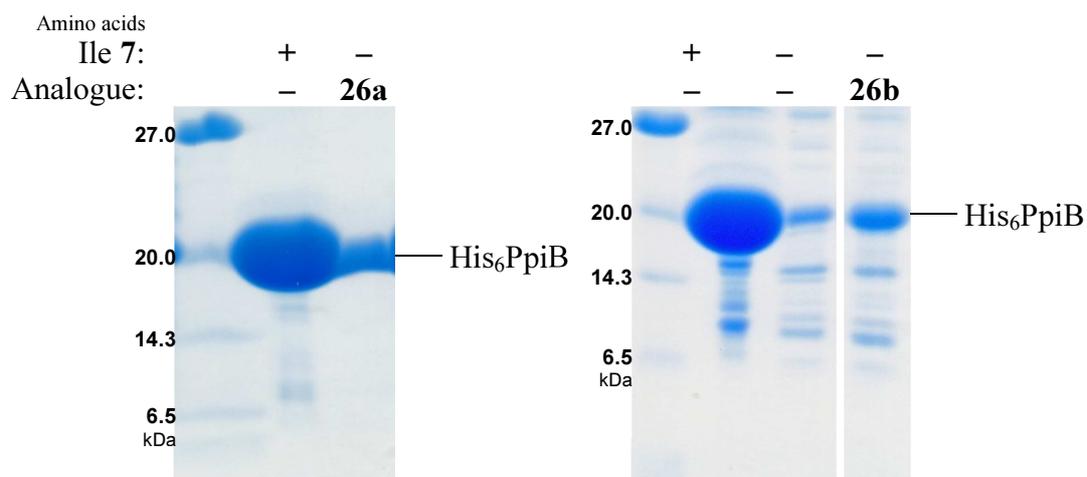


Figure 2.8. SDS-PAGE analysis of His₆PpiB produced *via* cell-free protein expression with mixtures containing the chlorides **26a** (left) and **26b** (right) instead of (2*S*,3*S*)-isoleucine 7.

In the case of the free amino acid **26a**, discrete peaks of partially substituted His₆PpiB were observed in the MS spectrum (Figure 2.9). The strongest peaks at 19413.3 Da and 19433.4 Da correspond to 8/10 and 9/10 isoleucine residues replaced by the chloride **26a**, respectively. Some fully substituted protein (19453.9 Da) was also seen, while the species with fewer substitutions (7/10 to 1/10) formed an array of decreasing peaks. This indicated that in the initial stages of the experiment the chloride **26a** was used for protein expression until the concentration rapidly decreased and expression ceased. Small, but rising amounts of background valine **8a** became available as the reaction progressed, leading to some production of all-natural protein (19249.5 Da).

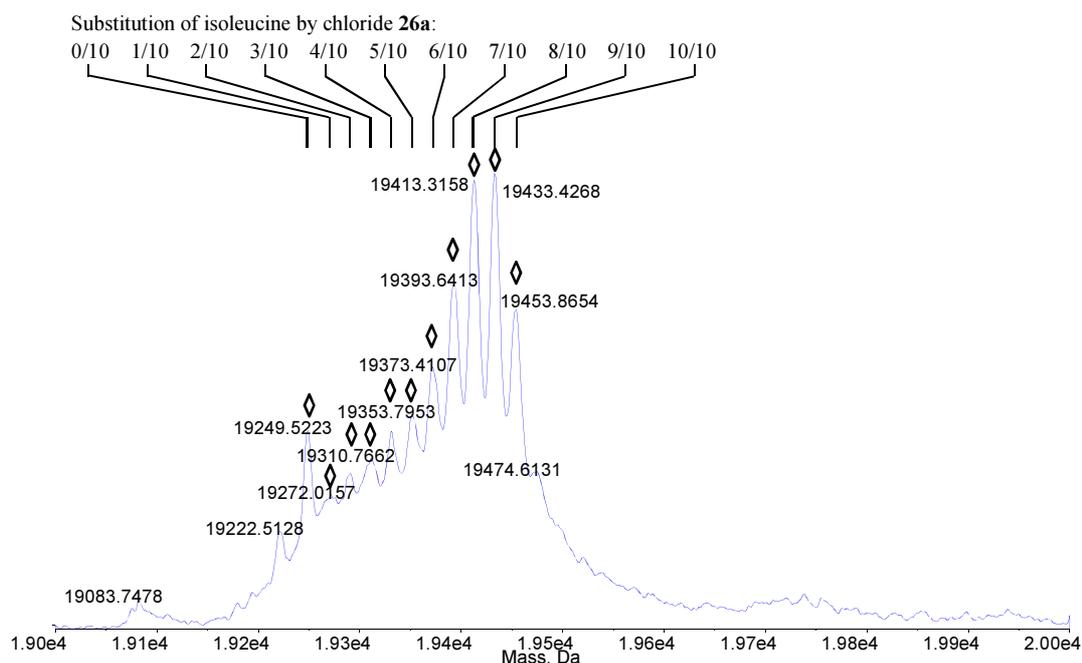


Figure 2.9. Deconvoluted ESI-MS spectrum of His₆PpiB produced in the presence of unpurified chloride **26a**.

With the acetylated analogue **26b**, MS indicated incorporation of more than 10 chlorine atoms into the protein (Figure 2.10). This could be explained by the presence of small amounts of over-chlorinated species bearing several chlorine atoms in the crude reaction mixture that is also accepted by IRS. However, due to time constraints and the complexity of the chlorination mixture the identity of the suggested species was not pursued.

The pattern of incorporation of the chloride **26a** is highly indicative that the instability of the chlorinated amino acid is responsible for lack of protein production in the previous study.^[102] However, in order to completely rule out the discrimination of the analogue by the synthetase, a way of sustaining a steady concentration of the potential substrate **26a** in solution was required.

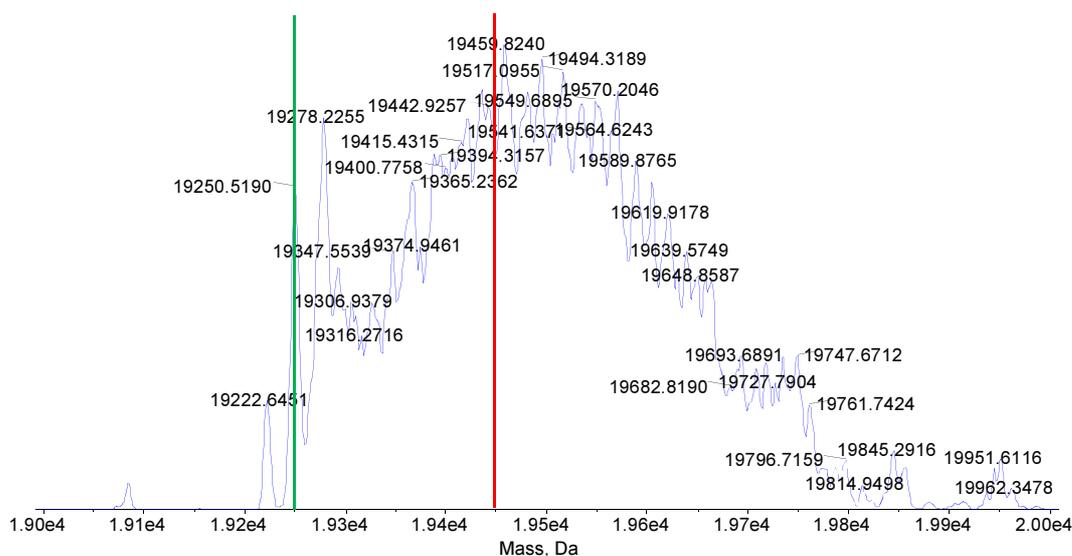


Figure 2.10. Deconvoluted ESI-MS spectrum of His₆PpiB produced in the presence of unpurified chloride **26b**. The red line indicates the expected mass of His₆PpiB upon full substitution of the 10 isoleucine residues with the analogue **26a** (19454 Da), the green line marks the all-native unsubstituted protein (19250 Da).

As the instability of the chlorinated species **26a** and **52a** was investigated, parallel work in the group had just shown that a range of common protecting groups, like methyl esters, were removed from (2*S*)-amino acids by the S30 (*vide infra*). Several synthetic amino acids were deprotected *in situ* during protein synthesis, thus reducing the steps of chemical synthesis. Hence, it was considered to supply the chlorinated amino acid **26a** with the carboxylate protected during protein expression: lactonisation would not be possible until the protecting group was removed.

In order to assess the viability of using protected amino acids for cell-free expression, the rate of amino acid deprotection was compared to the rate of protein synthesis. Protein production was found to be continuous throughout the 6 hour experiment (*vide infra*). The rate of deprotection of two potential protecting groups, methyl ester and

tert-butyl ester, that were shown to be fully and partially deprotected in the time frame of protein expression, respectively, was measured, by monitoring the deprotection of (*S*)-leucine methyl ester **55a** and (*S*)-leucine *tert*-butyl ester **55b**.

The protected amino acids **55a** and **55b** were incubated at 37°C with S30 for 6 h, and aliquots of the mixtures were taken throughout the experiment. Consumption of the (*S*)-leucine derivatives **55a** and **55b** and production of (*S*)-leucine **18** was monitored using the Waters AccQ.Tag method (*vide infra*).

The quinolone-based AccQ.Tag derivatisation reagent and most derivatised amines have an absorption maximum at 250 nm. Interestingly, the absorption maximum of derivatised (*S*)-leucine methyl ester **55a** shifted to 230 nm, making the peak difficult to track at 250 nm, thus the reaction was monitored at both frequencies (Figure 2.11, Figure 2.12).

Consumption of leucine methyl ester **55a** and the concomitant production of leucine **18** were mostly linear for the first 2.5-3 h of the experiment, slowing down in the final hours of the reaction after the amount of the ester **55a** had decreased substantially.

The deprotection reaction of (*S*)-leucine *tert*-butyl ester **55b** also progressed linearly, albeit significantly slower, with just under half of the material consumed in the 6 h window (Figure 2.13).

A similar zero order reaction profile can be expected for all deprotection reactions executed by the enzymes in the S30. The exact rate is likely determined by the amounts of the involved enzymes present in the extract as well as the activities of the enzymes towards the particular substrates.

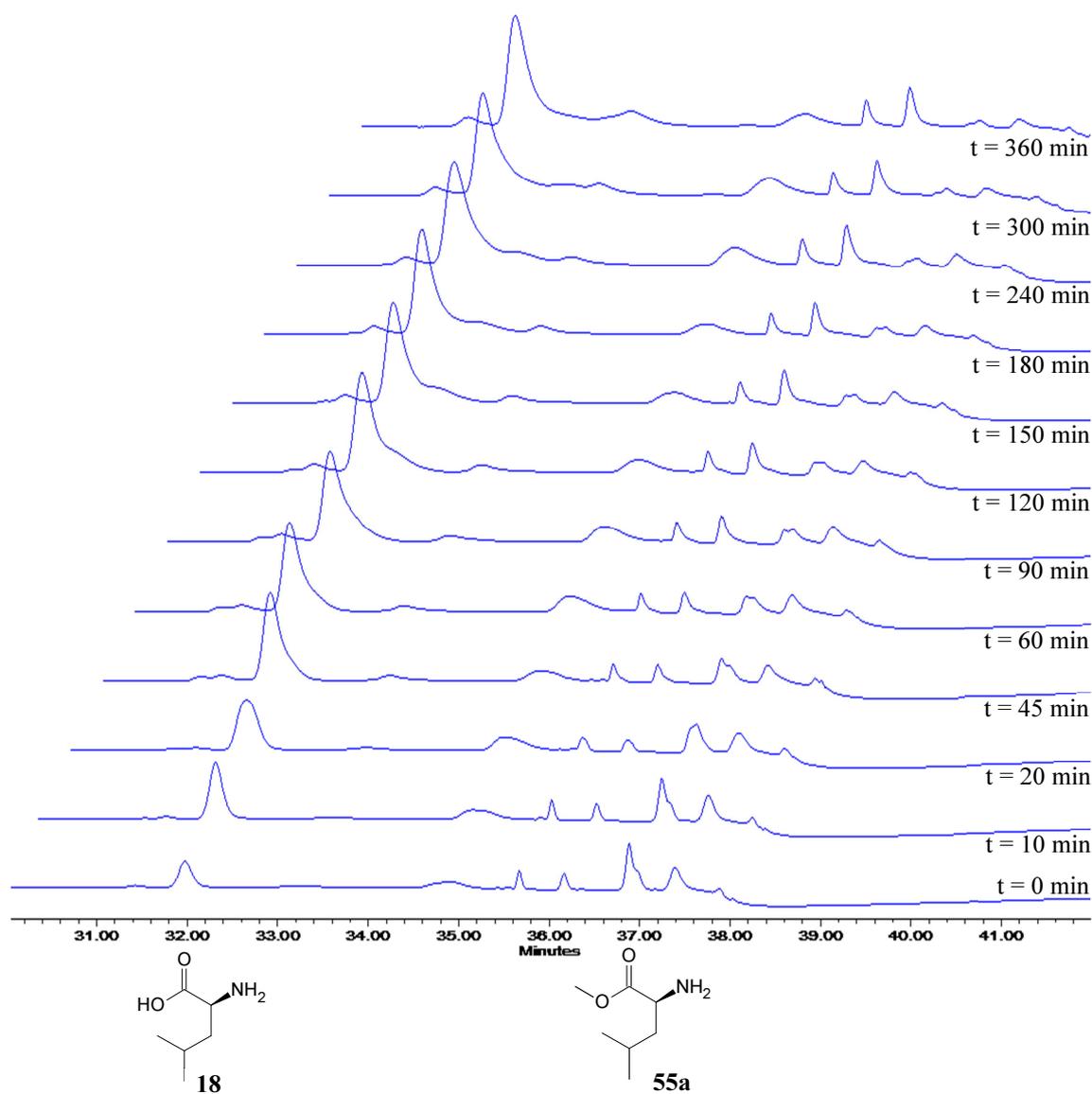


Figure 2.11. HPLC chromatograms of AccQ.TagTM-derivatised reaction mixture showing the consumption of (*S*)-leucine methyl ester **55a** incubated with the S30 extract at 37°C for 6 h, and the production of (*S*)-leucine **18**, detected by UV at 250 nm.

The rate of deprotection of leucine methyl ester **55a** matched the rate of protein expression (*vide infra*), whereas the tert-butyl ester **55b** reacted significantly slower. A

methyl ester was therefore considered the best protecting group for the unstable isoleucine analogue **26a**.

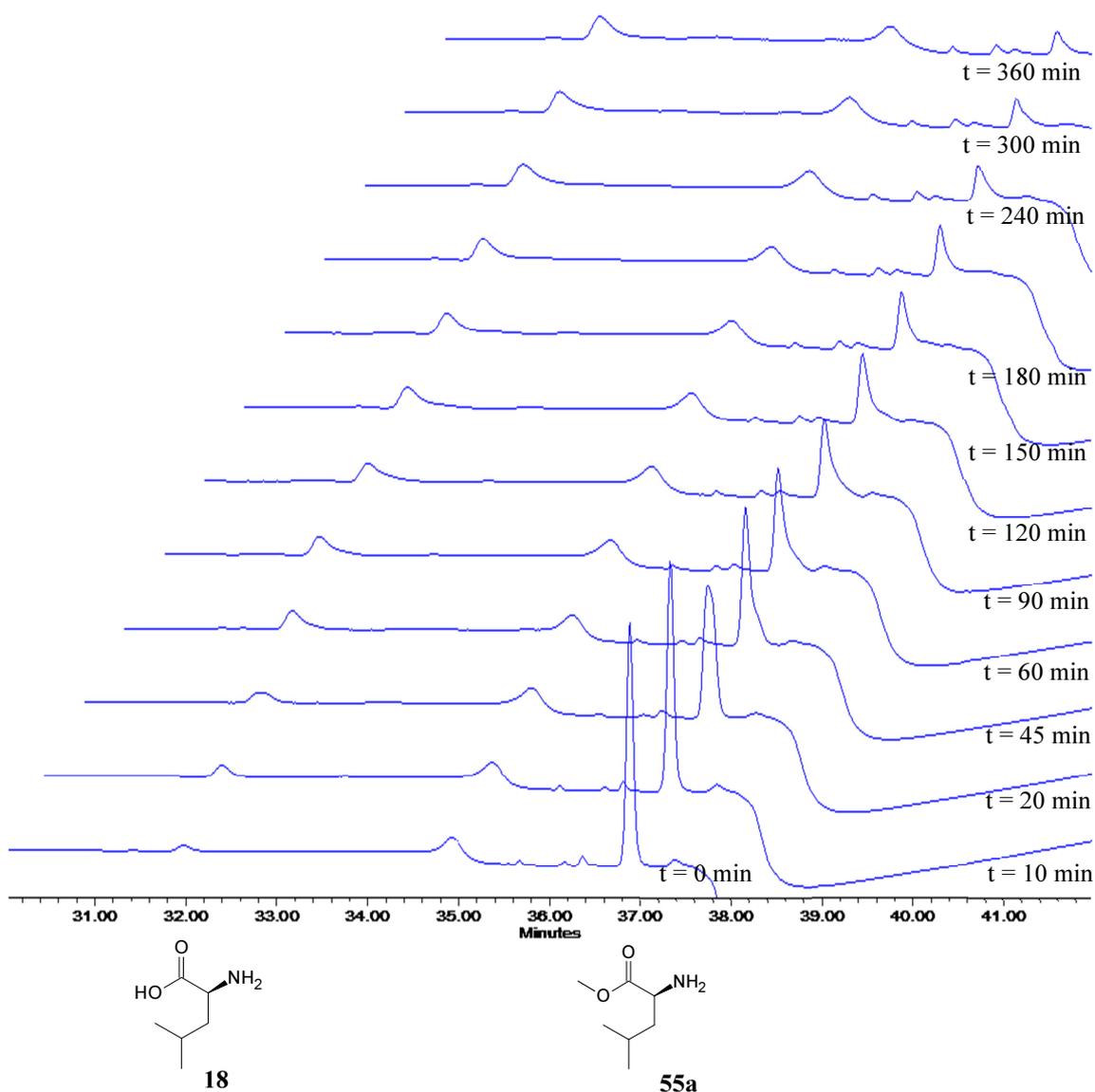


Figure 2.12. HPLC chromatograms of AccQ.TagTM-derivatised reaction mixture showing the consumption of (S)-leucine methyl ester **55a** incubated with the S30 extract at 37°C for 6 h, and the production of (S)-leucine **18**, detected by UV at 230 nm.

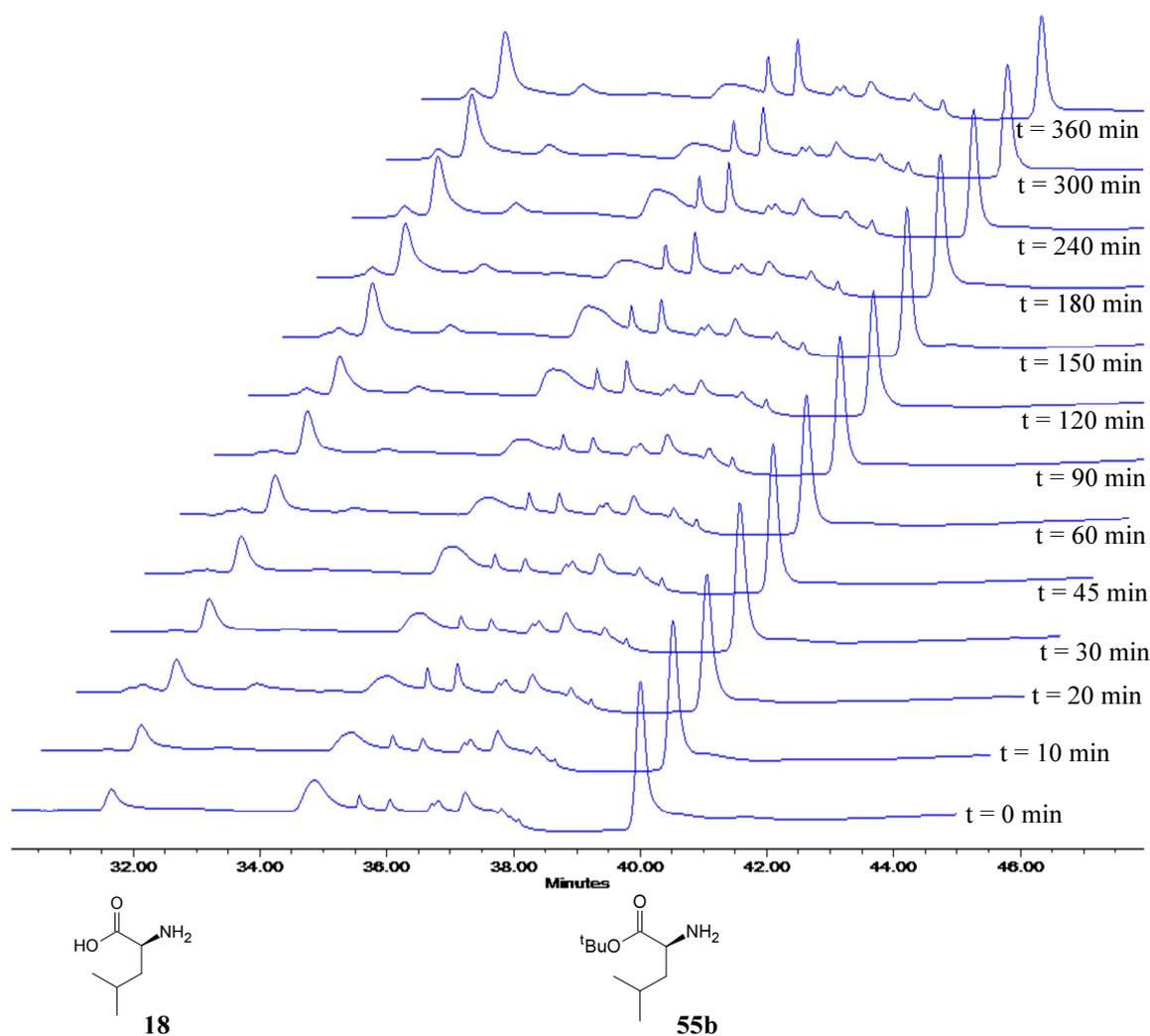


Figure 2.13. HPLC chromatograms of AccQ.TagTM-derivatised reaction mixture showing the consumption of (*S*)-leucine *tert*-butyl ester **55b** incubated with the S30 extract at 37°C for 6 h, and the production of (*S*)-leucine **18**, detected by UV at 250 nm.

As amino acid side-chain reactivity was not perturbed by functionalisation of the amino groups, it was anticipated simple derivatisation of the carboxy group would not influence the reaction either. Hence, (*S*)-valine methyl ester **8c** was chlorinated (Scheme 2.1; R¹ = Me, R² = H). Formation of the predicted products **26c**, **52c** and **53c** was confirmed by ¹H NMR by comparison with the spectra of the free acids; the two classes

of compounds had nearly identical spectra (apart for the ester signals). The similarity of the spectra also enabled the assignment of stereochemistry to the esters, with the δ 2.44 ppm signal attributable to the (2*R*,3*S*)-ester **26a** and the δ 2.60 ppm signal assigned to the (2*S*,3*S*)-diastereomer **52a** (Figure 2.14). Both esters were stable in aqueous solutions at physiological conditions.

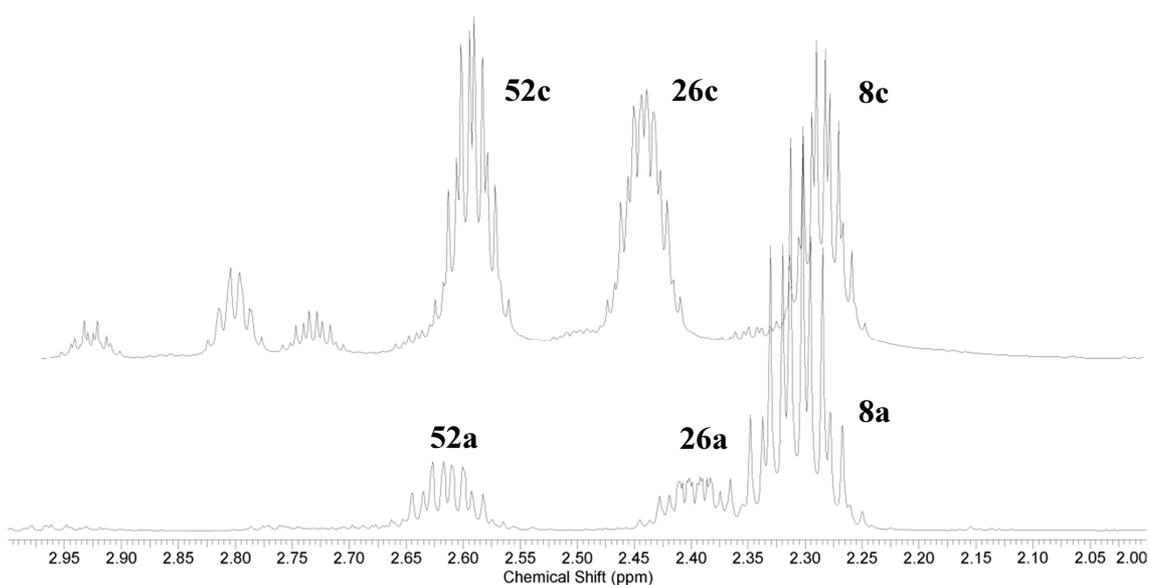


Figure 2.14. Comparison of ^1H NMR spectra of the chlorinated amino acids (below) and their methyl esters (above).

The mixture containing the diastereomers **26c** and **52c** was used for the initial protein expression test. Successful production of the test protein His₆PpiB confirmed the ester **26c** supports protein synthesis (Figure 2.15); ESI-MS analysis of the product showed that 10/10 isoleucine residues in the protein were substituted by the chlorinated amino acid **26a**. This confirmed that the analogue **26a** is prone to being mistaken for isoleucine

7 by IRS and the reported rejection was an artefact of the misrecognised instability of the substrate **26a** under physiological conditions.

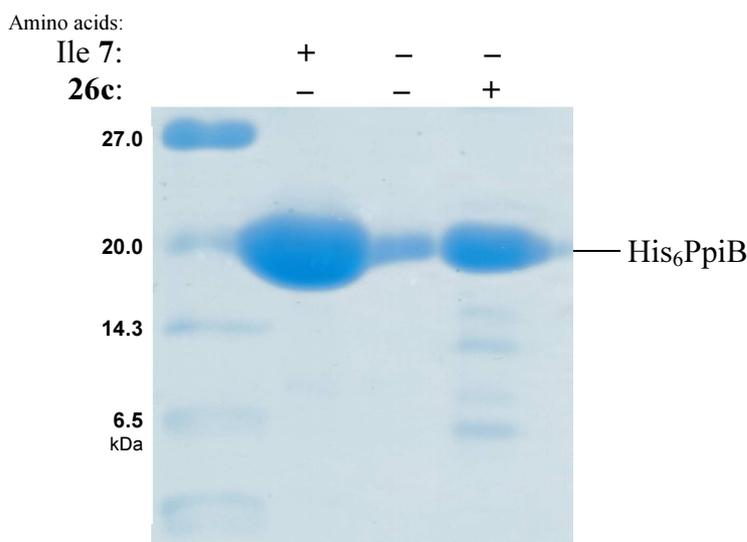


Figure 2.15. SDS-PAGE analysis of His₆PpiB produced *via* cell-free protein expression with unpurified analogue **26c** instead of isoleucine **7**.

As the methyl ester **26c** is chemically stable, its isolation from the reaction mixture was attempted. Reverse phase HPLC was used, but the free amines were causing severe tailing, likely due to strong interactions with the silanols of the stationary phase, making the separation more complicated. Figure 2.16 shows a typical trace obtained from a C18 HPLC column, even though the column was specifically designed for separation of highly polar compounds. The starting material **8c** was separable from the chlorinated species, but the three major chlorinated products **26c**, **52c** and **53c** tended to co-elute.

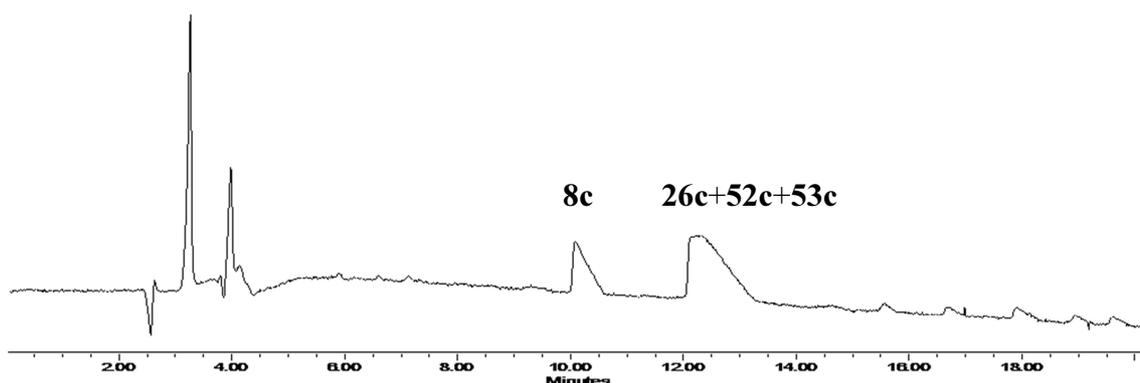


Figure 2.16. HPLC chromatogram of the chlorinated esters **26c**, **52c** and **53c** on an Atlantis dC18 column (4.6×250 mm, 5 μ), eluting with acetonitrile gradient from 5% to 20% in 0.1% TFA over 20 min, 1 mL/min.

A variety of reverse phase columns (Atlantis dC18, Fortis C18, Alltima C18), buffers and solvent gradients gave similar results. Hydrophilic interaction chromatography (HILIC), a version of normal phase chromatography increasingly used for the separation of polar compounds,^[121] also failed to separate the mixture. More promising results were obtained with a mixed mode column (Primesep 100) that separates the analytes both by cation exchange and hydrophobic interaction. For the first time some resolution of the chlorinated species **26c**, **52c** and **53c** was achieved (Figure 2.17). This suggested the cation exchange component is necessary for the separation of the ester **26c**, but no separation was achieved on ion exchange columns. Ion pairing chromatography was then attempted. While TFA is routinely used for HPLC separation of polar molecules, the ion pairing capability of trifluoroacetate counterion is limited and the interaction with the stationary phase is weak. Sodium dodecylsulfate (SDS) as an amphiphilic moiety with the longest commonly available hydrophobic chain was used instead of TFA to maximise the effects of ion pairing and minimise tailing. After optimisation the

components of the reaction mixture eluted as three well defined peaks, with β -chloride **53c** in a separate fraction from the γ -chlorides **26c** and **52c**, but the two diastereomers of γ -chlorovaline methyl ester **26c** and **52c** were inseparable by this approach (Figure 2.18).

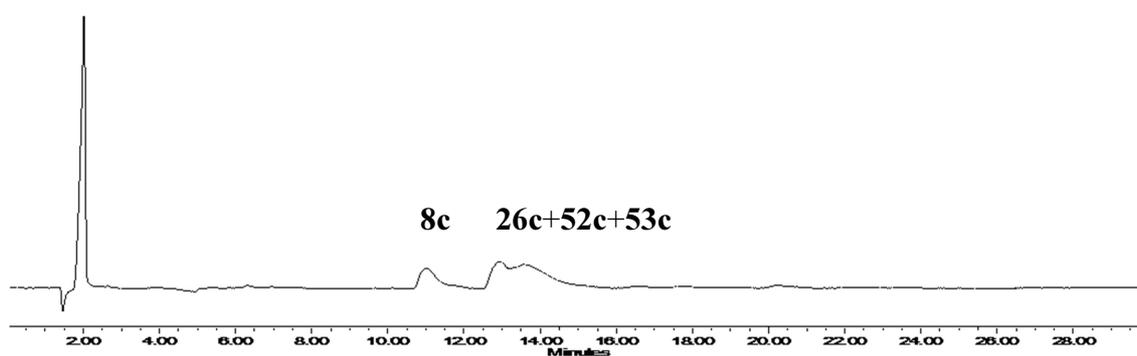


Figure 2.17. HPLC chromatogram of the chlorinated esters **26c**, **52c** and **53c** on a Primesep 100 column (4.6×150 mm, 5 μ), eluting with 20% acetonitrile in 0.1%TFA, 1 mL/min.

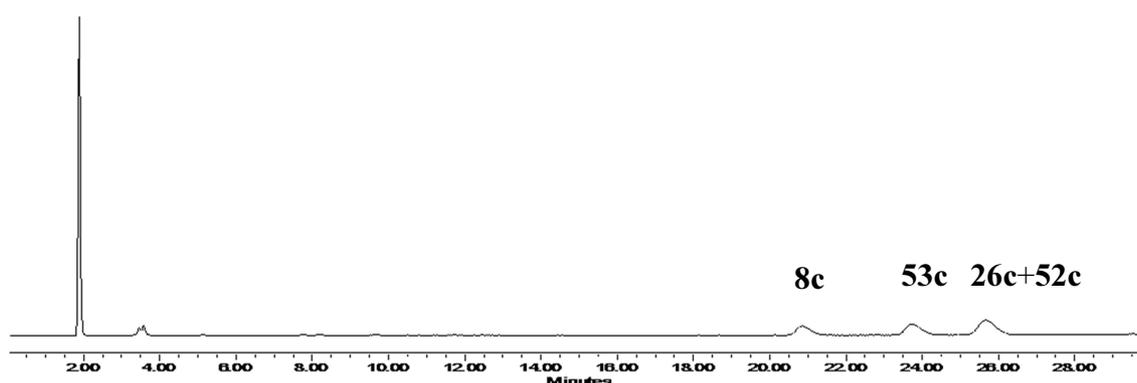


Figure 2.18. HPLC chromatogram of the chlorinated esters **26c**, **52c** and **53c** on an Alltima C18 column (4.6×250 mm, 5 μ), eluting with acetonitrile gradient from 30% to 40% in 20 mM SDS over 30 min, 1 mL/min.

Purification with sodium perchlorate in the mobile phase was attempted as a last resort as perchlorates were shown to improve the behaviour of amines on reverse phase.^[122] Initial attempts were promising as separate peaks for all four analytes **8c**, **26c**, **52c** and **53c** were observed for the first time on the Atlantis dC18 (4.6×250 mm, 5 μ) column (Figure 2.19). Optimisation of solvent gradient failed to improve resolution, so a second Atlantis dC18 column (4.6×150 mm, 3 μ) was connected in tandem to increase the length of stationary phase. Solvent flow rate had to be reduced to 0.8 mL/min to avoid overpressure, but the resolution improved to the extent that individual peaks of the analytes **26c**, **52c** and **53c** could be collected and analysed by NMR. Furthermore, the sample could be injected twice per run (2 min apart), doubling the amount of purified material (Figure 2.20). Residual perchlorate was removed by counter-ion exchange HPLC on a Primesep 100 (4.6×150 mm, 5 μ) column eluting with 0.4% TFA in 20% acetonitrile.

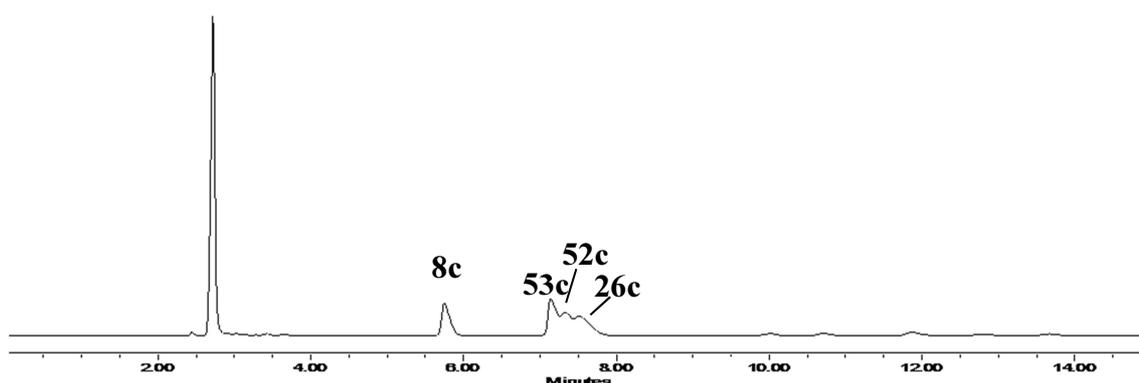


Figure 2.19. HPLC chromatogram of the chlorinated esters **26c**, **52c** and **53c** on an Atlantis dC18 column (4.6×250 mm, 5 μ), eluting with 20% acetonitrile in 100 mM NaClO₄, 1 mL/min.

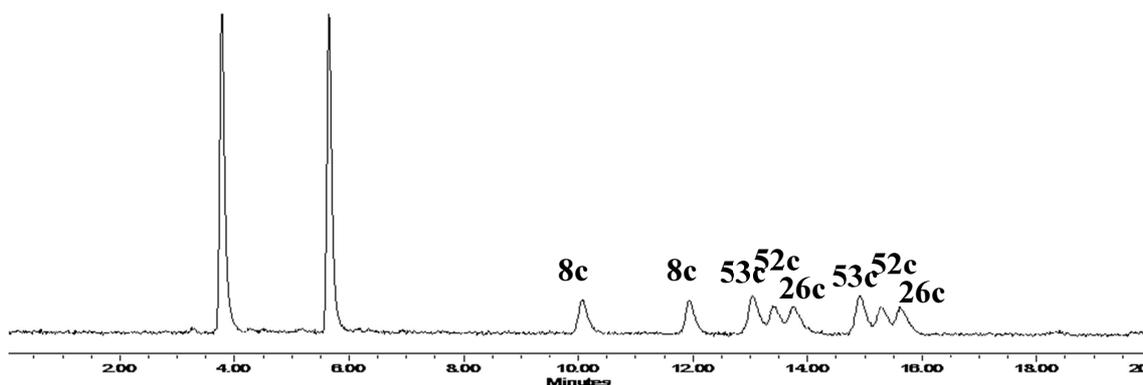


Figure 2.20. HPLC chromatogram of the chlorinated esters **26c**, **52c** and **53c** on an Atlantis dC18 column (4.6×150 mm, 3 μ) in tandem with an Atlantis dC18 column (4.6×250 mm, 5 μ), eluting with 20% acetonitrile in 100 mM NaClO₄, 0.8 mL/min. Two injections per run were achieved by stopping the run immediately after the first injection and injecting again after 2 min.

Only 1.3 mg of the target compound **26c** was obtained as large scale purification on analytical columns is not feasible, but that was sufficient for an incorporation study. The test protein His₆PpiB was successfully produced with the purified isoleucine analogue **26c** (Figure 2.21), thus validating the assignment of stereochemistry. Incorporation of the chlorinated analogue of isoleucine **26a** was confirmed by MS.

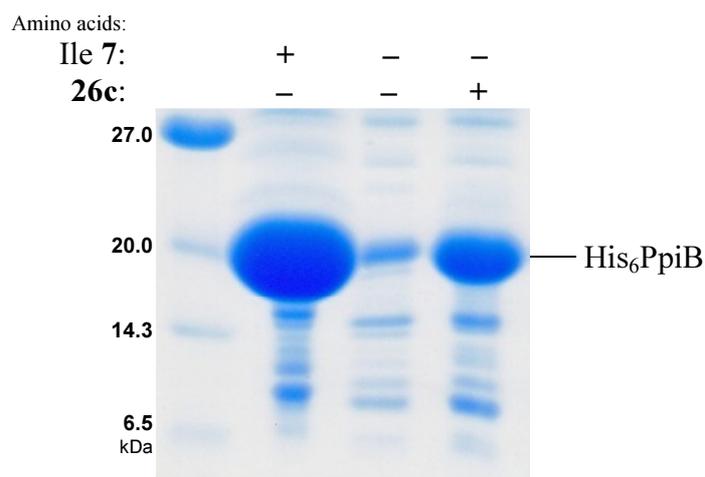
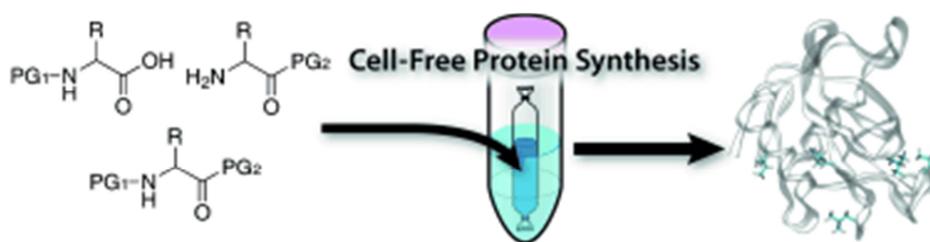


Figure 2.21. SDS-PAGE analysis of His₆PpiB produced *via* cell-free protein expression with the purified ester **26c** instead of (2*S*,3*S*)-isoleucine **7**.

The successful utilisation of the methyl ester **26c** in protein expression confirmed the previous attempts of protein production with the free amino acid **26a** were unsuccessful solely due to the instability of the substrate, and the biological selection mechanism employed by IRS does not have the capacity to discriminate between primary chloro and methyl groups. This work has been published^[123] and copied on the following pages:

In Situ Deprotection and Incorporation of Unnatural Amino Acids during Cell-Free Protein Synthesis

Dr. Isaac N. Arthur, Dr. James E. Hennessy, Dr. Dharshana Padmakshan, Dr. Dannon J. Stigers, Stéphanie Lesturgez, Samuel A. Fraser, Mantas Liutkus, Prof. Gottfried Otting, Dr. John G. Oakeshott and Prof. C. J. Easton



DOI: 10.1002/chem.201203923

In Situ Deprotection and Incorporation of Unnatural Amino Acids during Cell-Free Protein Synthesis

Isaac N. Arthur,^[a, b] James E. Hennessy,^[a, b] Dharshana Padmakshan,^[a, b]
Dannon J. Stigers,^[a, b] Stéphanie Lesturgez,^[a, b] Samuel A. Fraser,^[a, b] Mantas Liutkus,^[a, b]
Gottfried Otting,^[a] John G. Oakeshott,^[c] and Christopher J. Easton^{*, [a, b]}

Abstract: The S30 extract from *E. coli* BL21 Star (DE3) used for cell-free protein synthesis removes a wide range of α -amino acid protecting groups by cleaving α -carboxyl hydrazides; methyl, benzyl, *tert*-butyl, and adamantyl esters; *tert*-butyl and adamantyl carboxamides; α -amino form-, acet-, trifluoroacet-, and benzamides; and side-chain hydrazides and esters. The free amino acids are produced and incorporated into a protein under standard

conditions. This approach allows the deprotection of amino acids to be carried out in situ to avoid separate processing steps. The advantages of this approach are demonstrated by the efficient incorporation of the chemically intractable (*S*)-4-fluoroleucine, (*S*)-4,5-

dehydroleucine, and (2*S*,3*R*)-4-chlorovaline into a protein through the direct use of their respective precursors, namely, (*S*)-4-fluoroleucine hydrazide, (*S*)-4,5-dehydroleucine hydrazide, and (2*S*,3*R*)-4-chlorovaline methyl ester. These results also show that the fluoro- and dehydroleucine and the chlorovaline are incorporated into a protein by the normal biosynthetic machinery as substitutes for leucine and isoleucine, respectively.

Keywords: amino acids • cell-free synthesis • enzyme catalysis • protecting groups • protein expression

Introduction

Cell-free synthesis^[1–9] has emerged as a useful tool for the controlled incorporation of unnatural amino acids into proteins.^[6–14] Relative to expression systems in vivo,^[15–18] cell-free synthesis involves the prior extraction and partial purification of cellular material. The extract is then supplemented with the necessary DNA, amino acids, and other ingredients to enable protein synthesis in vitro under fully controlled conditions. Cell-free synthesis allows the substitution of a natural amino acid with an unnatural one. Incorporation of the unnatural amino acid then depends upon it being structurally similar to the one that is normally encoded and competing for loading onto the corresponding tRNA by the cog-

nate aminoacyl tRNA synthetase. The cell-free system provides a method to bias this competition in favour of the unnatural amino acid by supplementing it in place of the normal one. By using this approach, proteins containing high levels of isotopically labelled amino acids^[2,8] selenomethionine,^[6,7] 3,4-dihydroxyphenylalanine,^[10] and chlorinated amino acids^[11] have been synthesized. Alternatively, the cell-free system allows for the incorporation of an unnatural amino acid in addition to the normal ones through supplementation with an appropriate mutant aminoacyl tRNA synthetase and cognate suppressor tRNA.^[12–14]

The unnatural amino acids are prepared through synthesis, often by using routes in which the final steps involve removal of amino and carboxyl protecting groups.^[19] Our serendipitous observation that (*S*)-4-fluoroleucine hydrazide (**23**) was cleaved to (*S*)-4-fluoroleucine (**26**), which was incorporated into a protein in place of (*S*)-leucine during cell-free synthesis with the S30 extract of *E. coli* BL21 Star (DE3)^[20] prompted us to explore the scope of this in situ deprotection and incorporation of amino acids. The obvious advantage of this approach of deprotecting the amino acids in situ is that it avoids separate processing steps that are sometimes challenging, particularly with free amino acids that are chemically labile.

Herein, we report our studies of a variety of backbone- and side-chain-protected amino acids and related compounds. Esters and amides were used in addition to hydrazides to investigate carboxyl protecting groups, while a range of amides and carbamates were employed as protected amines. This initial screen was carried out with protected forms of those amino acids typically found in proteins and

[a] Dr. I. N. Arthur, Dr. J. E. Hennessy, Dr. D. Padmakshan, Dr. D. J. Stigers, S. Lesturgez, S. A. Fraser, M. Liutkus, Prof. G. Otting, Prof. C. J. Easton
Research School of Chemistry
Australian National University
Canberra, ACT 0200 (Australia)
Fax: (+61)2-6125-8114
E-mail: easton@rsc.anu.edu.au

[b] Dr. I. N. Arthur, Dr. J. E. Hennessy, Dr. D. Padmakshan, Dr. D. J. Stigers, S. Lesturgez, S. A. Fraser, M. Liutkus, Prof. C. J. Easton
ARC Centre of Excellence for Free Radical Chemistry and Biotechnology
Australian National University, Canberra, ACT 0200 (Australia)

[c] Dr. J. G. Oakeshott
CSIRO Ecosystem Sciences, Canberra, ACT 2601 (Australia)

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/chem.201203923>.

shows that the S30 extract of *E. coli* BL21 Star (DE3) has the capacity to deprotect a wide range of amino acid derivatives with various α -amino, α -carboxyl and side-chain protecting groups. In many cases, the deprotection process is sufficiently facile that the rate of production of the free amino acid does not limit the quantity of protein formed through cell-free synthesis under standard conditions. In addition to the initial assessment of the reactivity of various protecting groups, we demonstrate the use of this methodology for the deprotection of chemically intractable unnatural amino acids in situ for their direct incorporation into a protein.

Results and Discussion

The compounds used in the initial screening phase of this study are illustrated in Figures 1–3. Compounds **1a–f**, **2**, **8**, **9a–c**, **15a,b**, and **16** were selected to investigate the scope of cleavage of amino acid hydrazides. Amino acid esters **3a–g**, **10a,c**, and **13b,c**, and amides **4** and **10b**, were used to examine the removal of other carboxylic acid protecting groups. Ester **13a** was chosen as a protected alcohol. The ability of S30 to unmask amino acid amino groups was also studied by using acetamides **5a–c** and **11a–d**, trifluoroacetamides **6a,b**, formamides **7a,b**, benzamide **12**, and other *N*-substituted amino acid derivatives **17a–c**. Compounds **14** and **18** were included in the study as amino acids with both their amino and carboxyl groups protected. Each of these compounds was incubated with the *E. coli* S30 extract for six hours at 37°C at a concentration of 2 mM under the standard conditions used for cell-free protein synthesis,^[8,10,21,22] except in the absence of the DNA required to produce a protein. The

mixtures were monitored by using HPLC for the consumption of the starting materials and formation of the corresponding deprotected products. Under these conditions, compounds **1a–f**, **2**, **3a–g**, **4**, **5a–c**, **6a,b**, **7a,b**, and **8** illustrated in Figure 1 were all completely deprotected (>90%) in this assay. Compounds **9a–c**, **10a–c**, **11a–d**, **12**, **13a–c**, and **14** shown in Figure 2 were partially deprotected (>20–<80%), but underwent complete reaction when the incubation time was extended to 24 hours or when proportionately more of the S30 extract was added to the mixture, whereas compounds **15a,b**, **16**, **17a–c**, and **18** (Figure 3) were unaffected (<10% reaction).

The deprotection of **1a–f** and **9a–c** establishes the broad substrate specificity of the S30 extract towards cleavage of amino acid hydrazides, while, to the extent that they are representative, the inertness of the *R* enantiomers **15a,b** of the (*S*)-leucine and (*S*)-valine derivatives **1a** and **1c** shows the reactions are stereoselective and therefore likely to be enzyme catalysed. The *N*-substituted derivative **2** of hydrazide **1a** is also processed, but the des-amino analogue **16** of the phenylalanine derivative **1b** is not. This reaction is not limited to hydrazides. Amino acid esters and amides are also converted, as shown with methyl esters **3a–c**, benzyl esters **3d–f** and **13b,c**, *tert*-butyl esters **3g** and **10a** and amide **10b**, and the adamantyl amide **4** and ester **10c**. The reactions of esters **3d–g**, **10a**, and **13b,c** are particularly noteworthy given the common usage of benzyl and *tert*-butyl ester protecting groups in amino acid and peptide chemistry.^[19]

The turnover of acetamides **5a–c** and **11a–d**, trifluoroacetamides **6a,b**, formamides **7a,b**, and benzamide **12** demonstrates the activity of the S30 extract with some of the commonly used amino protecting groups.^[19,23–27] Al-

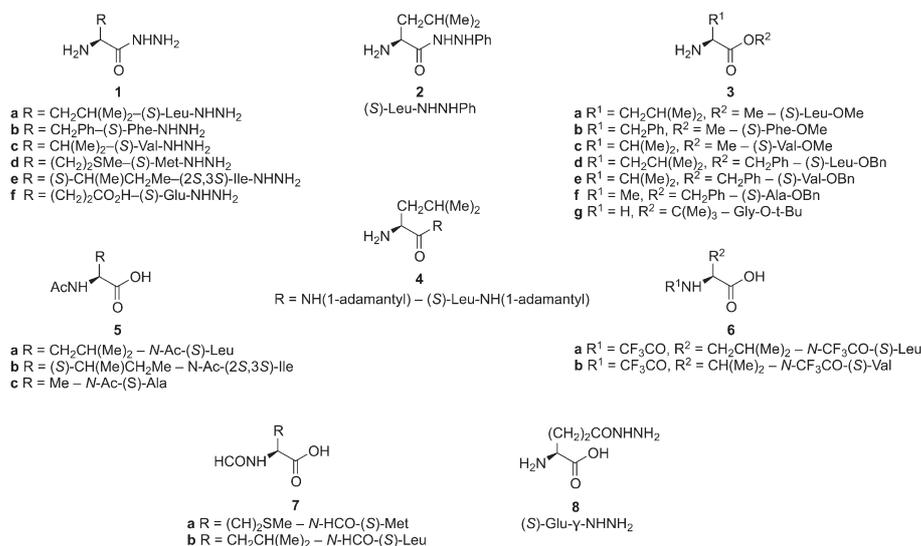


Figure 1. Amino acid derivatives fully deprotected (>90%) by *E. coli* S30 extract under standard conditions for cell-free protein synthesis.

Chlorinated Amino Acids in Peptide Production

Chapter 2. Substrate Selectivity of Isoleucyl-tRNA Synthetase

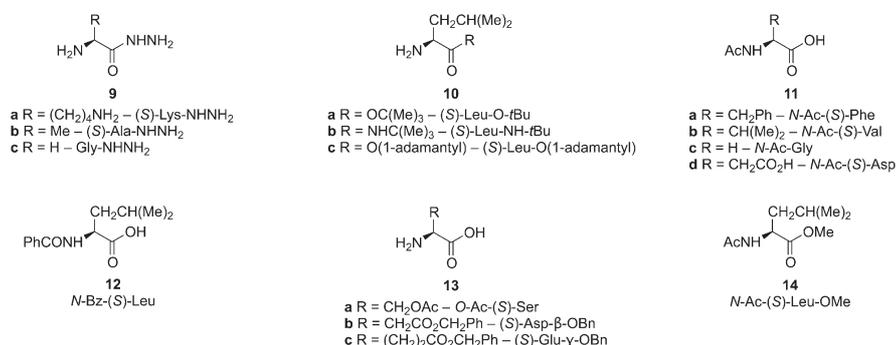


Figure 2. Amino acid derivatives partially deprotected (>20–<80%) by *E. coli* S30 extract under standard conditions for cell-free protein synthesis.

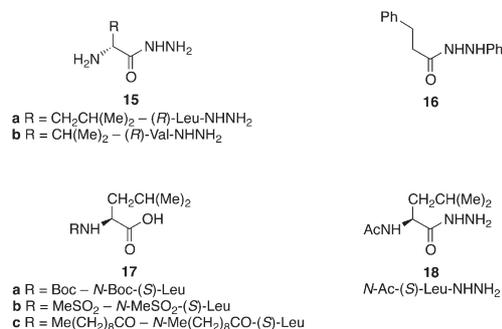


Figure 3. Amino acid derivatives unaffected (<10% reaction) by the *E. coli* S30 extract under standard conditions for cell-free protein synthesis. Boc = *tert*-butoxycarbonyl.

though acetamides are more common, a trifluoroacetamide functionality offers some specific advantages in synthesis because incorporation of the trifluoroacetyl group deactivates the amino acid α -position towards radical reactions.^[26,27] It is apparent that there are limitations to the removal of amino protecting groups by the S30 extract, however, as illustrated by the lack of reaction of carbamate **17a**, sulfonamide **17b**, and *N*-decanoylleucine **17c**.

Most of the compounds examined had either the α -amino or α -carboxyl group protected, but the reactions of glutamyl hydrazide **8**, serine acetate **13a**, and aspartate and glutamate benzyl esters **13b,c**, respectively, demonstrate that side-chain carboxylic acid derivatives are also cleaved. As for doubly protected amino acids, both protecting groups are removed from the *N*-acetylated leucine methyl ester **14**, but the corresponding hydrazide **18** is inert.

It does not affect the utility of the results, but it is nonetheless interesting to consider the enzymes likely to be in the S30 extract and responsible for the reactions observed. Leucine aminopeptidases (LAP, E.C. 3.4.11.1) are known to hydrolyse hydrazide **1a**^[28] and the presence of this enzyme in *E. coli* probably contributes to its cleavage by S30. This enzyme is also known to catalyse reactions of other protected amino acids, such as phenylalanine hydrazide **1b** and

ester **3b**,^[28] but in separate studies with this enzyme we found it does not show activity for other protected amino acids, such as valine hydrazide **1c** and ester **3c**. Therefore other enzymes, such as aminopeptidase N (PepN, E.C. 3.4.11.2), which is known to have broad substrate specificity,^[29–31] are also likely to be involved, as well as those enzymes with amino acid selectivity, such as methionine aminopeptidase Type 1 (MAP 1, E.C. 3.4.11.18),^[32] aminopeptidase B (PepB, E.C. 3.4.11.23), which is selective for acidic amino acids,^[33] and oligopeptidase B (Protease II, E.C. 3.4.21.83), which shows selectivity for basic amino acids.^[34,35]

Hydrazide cleavage appears to be dependent on enzymes that require their substrates to have a free amino group, such as LAP,^[28] because **1a** and **2** react, whereas the des-amino hydrazide **16** and *N*-acetylated leucine hydrazide **18** do not react. Ester cleavage is not limited in this way, however, as shown by the reaction of the acetylated leucine methyl ester **14**. β -Aspartyl peptidase (E.C. 3.4.19.5),^[36,37] γ -glutamyl transpeptidase (E.C. 2.3.2.2),^[38,39] and glutaminase A (E.C. 3.5.1.2)^[40] are likely to be involved with the removal of side-chain protecting groups, whereas carboxypeptidases probably account for the amino acid deacylation reactions. In any event, given the large number of enzymes expected to be in the S30 extract, there is likely to be redundancy.

Having determined the scope of the deprotection of amino acid derivatives by the S30 extract, examples of those shown in Figures 1 and 2 were tested for deprotection and incorporation into peptidyl-Pro *cis*–*trans* isomerase B (PpiB) in situ through cell-free protein synthesis.^[8,10,21,22] Figure 4 shows a representative SDS-PAGE analysis of the purified proteins: lane B shows the PpiB produced when each of the twenty normal amino acids (1 mM) was added, lane C shows a decreased amount of that protein when (*S*)-leucine was not added and only the background concentration of this amino acid was present (shown through separate HPLC analyses to increase from negligible to around 0.01 mM during the course of the experiment), and lanes D and E show the PpiB formed when hydrazide **1a** (2 mM) and methyl ester **3a** (2 mM) were added, respectively, instead of (*S*)-leucine. Mass-spectrometric analysis confirmed the native sequence of the PpiB that formed. These results

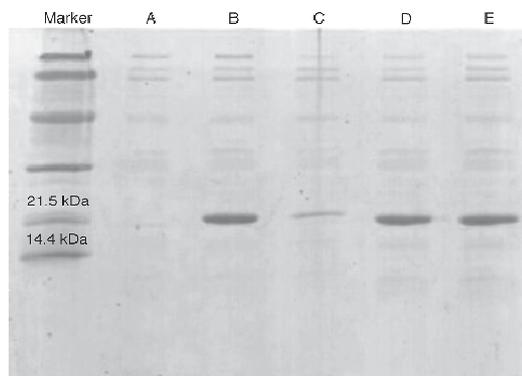
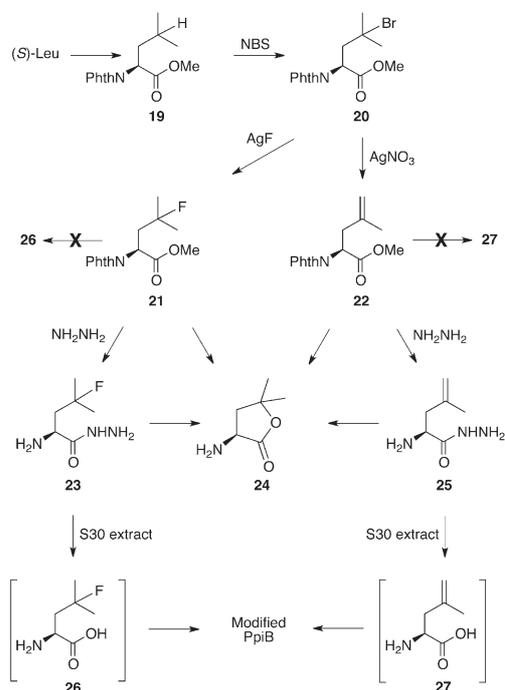


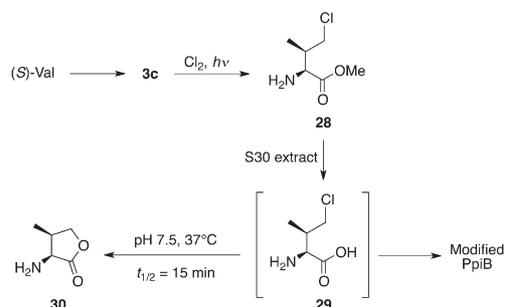
Figure 4. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) analysis of synthesized His₆-PpiB with A) no DNA, B) DNA, C) no (*S*)-leucine, D) no (*S*)-leucine but with hydrazide **1a**, and E) no (*S*)-leucine but with methyl ester **3a**.

illustrate that PpiB is efficiently produced by using either (*S*)-leucine, hydrazide **1a**, or ester **3a** and that conversion of hydrazide **1a** and ester **3a** into the free amino acid occurs sufficiently easily that it does not limit cell-free protein synthesis under the standard conditions. In identical fashion, it was demonstrated that ester **3d**, amide **4**, acetamide **5a**, trifluoroacetamide **6a**, and the doubly protected *N*-acetyl-(*S*)-leucine methyl ester **14** resulted in unrestricted synthesis of PpiB. Less protein formed with hydrazides **1b** and **9a**, ester **10a**, and amide **10b**, but even in those cases unrestricted PpiB production was achieved by adding more of the S30 extract, by preincubating the mixture before finally adding the PpiB DNA, or by increasing the assay time to 24 hours. In this context it is relevant to note that cell-free protein synthesis occurs continuously throughout the experiment (6 h) performed under standard conditions (as evident from the SDS-PAGE analysis provided in the Supporting Information), whereas 50 and 75 % of methyl ester **3a** had hydrolysed after 0.5 and 1 h, respectively, whereas only 50 % of *tert*-butyl ester **10a** was converted into free (*S*)-Leu after 6 h. Irrespective of this, the reactions carried out in situ were shown to be compatible with the cell-free protein synthesis for a variety of carboxylic acid derivatives including hydrazides, methyl esters, benzyl esters, *tert*-butyl esters and amides, and amino protecting acetamides and trifluoroacetamides.

The utility of these observations is demonstrated by the incorporation of the unnatural amino acids (*S*)-4-fluoro-leucine (**26**), (*S*)-4,5-dehydroleucine (**27**), and (2*S*,3*R*)-4-chloro-valine (**29**) into PpiB through the direct use of their respective protected forms **23**, **25**, and **28** in the cell-free system. The latter compounds were prepared through side-chain elaboration of (*S*)-leucine and (*S*)-valine^[20,41–43] (Schemes 1 and 2, respectively). All attempts to prepare the free amino acids **26** and **27** by deprotection of the corresponding leucine derivatives **21** and **22** were frustrated by formation of lactone **24**. Eventually, removal of the phthaloyl group from



Scheme 1. Side-chain elaboration of (*S*)-leucine. Leu = leucine, NBS = *N*-bromosuccinimide.



Scheme 2. Side-chain elaboration of (*S*)-valine.

fluoride **21** was achieved by treatment with hydrazine to give hydrazide **23**, which underwent oxidative cleavage with NBS to give amino acid **26** in addition to lactone **24**. Even by using this approach, the protected alkene **22** afforded hydrazide **25**, but treatment of it with NBS gave only lactone **24**. Nevertheless, as shown by the SDS-PAGE analysis illustrated in Figure 5, PpiB was efficiently produced through cell-free synthesis by using hydrazides **23** and **25** (lanes G and H, respectively), as substitutes for (*S*)-leucine (lane E). The most abundant peak at 19340 Da in the mass spectrum

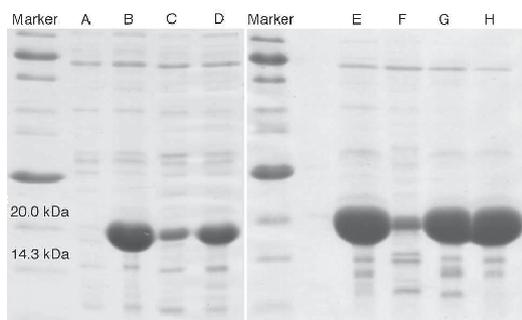


Figure 5. SDS-PAGE analysis of synthesized His₇-PpiB with A) no DNA, B) DNA, C) no (2*S*,3*S*)-isoleucine, D) no (2*S*,3*S*)-isoleucine but with chloride **28**, E) DNA, F) no (*S*)-leucine, G) no (*S*)-leucine but with hydrazide **23**, and H) no (*S*)-leucine but with hydrazide **25**.

of the protein formed from hydrazide **23** (Figure 6b) corresponds to substitution of fluoride **26** for all five of the (*S*)-leucine residues found in native PpiB (Figure 6a), while the next most abundant peak at 19321 Da in Figure 6b corresponds to substitution of four of the five (*S*)-leucine residues. (Each substitution increases the mass by 18 Da. However, due to the natural ¹³C isotope abundance, PpiB gives four dominant peaks 1 Da apart, which differ in intensity by less than 15%. These peaks were not resolved and any one may be labelled by the spectrometer, so the mass difference indicated for one substitution is 18 ± 4 Da.) Based on the

mass-spectrometric data, the average degree of incorporation of fluoride **26** is approximately 90%. In the case of hydrazide **25**, the mass spectrum (Figure 6c) shows that four or five of the (*S*)-leucine residues are replaced with alkene **27** (in this case, each substitution decreases the mass by 2 Da). This extent of incorporation was confirmed through amino acid analysis of the PpiB. There has been an earlier report of the incorporation of alkene **27** into leucine zipper peptides,^[44] although no details of that study have been published.

Previously, we had found that chloride **29** is not incorporated into PpiB in place of (2*S*,3*S*)-isoleucine, despite several other chlorides proving to be suitable substitutes for aliphatic amino acids.^[11] These studies had been based on the use of a crude mixture of chloride **29**, its 2*S*,3*S* diastereomer, (*S*)-3-chlorovaline, and recovered starting material, obtained by chlorination of (*S*)-valine,^[43] because it had not been practical to separate and test the pure material. Close analysis of the mixture showed that chloride **29** is unstable and converts into lactone **30** under the conditions of cell-free protein expression (i.e., pH 7.5, 37 °C), with a half-life of approximately 15 minutes. By comparison, when the valine methyl ester **3c** was chlorinated, the chloride **28** that formed was sufficiently stable for it to be separated through HPLC without lactonization. With the cell-free system, this material underwent deprotection in situ and product **29** was incorporated into PpiB (Figure 5, lane D) as a substitute for (2*S*,3*S*)-isoleucine (Figure 5, lane C). With chloride **28**, the most abundant peak in the mass spectrum of the PpiB at

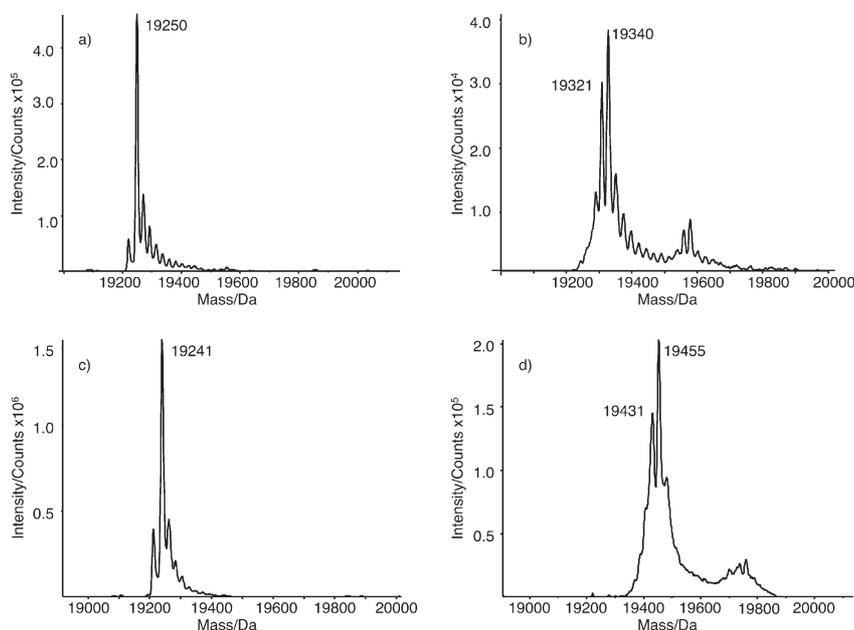


Figure 6. Mass spectra of a) native PpiB, b) PpiB produced using hydrazide **23** in place of (*S*)-leucine, c) PpiB produced by using hydrazide **25** in place of (*S*)-leucine, and d) PpiB produced by using chloride **28** in place of (2*S*,3*S*)-isoleucine.

19455 Da (Figure 6d) corresponds to substitution of the chlorinated free amino acid **29** for all ten of the (2*S*,3*S*)-isoleucine residues found in native PpiB (Figure 6a), while the next most abundant peak at 19431 Da results from nine replacements (each substitution that corresponds to replacing a methyl group with chlorine increases the mass by an average of 20.5 Da, although the natural carbon and chlorine isotope abundance in each of these chlorinated PpiBs results in five dominant ions 1 Da apart, differing in intensity by 15% or less, so the difference indicated by the spectrometer is (20 ± 5) Da). With chloride **29**, the mass spectrum shows that the average degree of incorporation is above 90%. For hydrazides **23** and **25**, the S30 extract simply provides a way to avoid having to carry out problematic deprotection as a separate step; however, for chloride **28** the S30 extract also continuously replenishes the free amino acid **29** as it is consumed through lactonization. As a result, incorporation is observed by using ester **28**, but not with the same initial concentration of the free amino acid **29**.

Conclusion

This study has demonstrated the ability of the S30 extract to remove a range of protecting groups for direct incorporation of the resulting amino acids into a protein. This approach is more efficient because it not only decreases the number of synthetic steps required in the preparation of unnatural amino acids, but also provides a versatile method to circumvent problems associated with chemical instability of amino acids during both their deprotection and protein synthesis. The method has been demonstrated to be suitable for the incorporation of the fluoro- and dehydroleucines **26** and **27** and chloroalanine **29** as substitutes for leucine and isoleucine, respectively, at levels of 90% or above. These high levels are more than adequate for applications such as isotopic labelling or the fluorination of proteins, for use in spectroscopic studies for example, in which the unmodified protein is not detectable. These levels are also suitable to investigate general rather than specific effects of amino acid modifications on protein structure and function. Further, the cell-free system with the S30 extract allows for a complete site-specific incorporation of an unnatural amino acid through the addition of a mutant aminoacyl tRNA synthetase and cognate suppressor tRNA, under which conditions the incorporation levels would be expected to be quantitative.

Experimental Section

Amino acid derivatives: With the exception of hydrazides **2** and **9a** and ester **10c**, all the amino acids illustrated in Figures 1–3 are available from Sigma–Aldrich, Merck Pty. Ltd., Aussep, TCI Chemicals, or Aurora Fine Chemicals LLC, although for the purposes of this investigation many were prepared from the corresponding free amino acids (see the Supporting Information). Phenylhydrazide **2** was synthesized from methyl ester **3a** by using phenylhydrazine.^[45] Hydrazide **9a** was prepared from (S)-lysine by esterification with thionyl chloride in methanol, followed by

treatment of the corresponding ester with hydrazine hydrate.^[46,47] Ester **10c** was prepared by the treatment of (S)-leucine with 1-adamantanol, dimethyl sulfoxide, and *para*-toluenesulfonic acid.^[48]

(S)-4-Fluoroisoleucine hydrazide (**23**) was prepared by the protection of (S)-leucine, bromination of phthalimide **19**, treatment of bromide **20** with silver fluoride, and reaction of fluoride **21** with hydrazine hydrate.^[20] (S)-4,5-Dehydroisoleucine hydrazide (**25**) was prepared from *N*-phthaloyl-4,5-dehydroisoleucine methyl ester (**22**)^[49] by reaction with hydrazine hydrate.^[20] M.p. 130–132 °C; ¹H NMR (300 MHz, D₂O): δ = 5.06 (m, 1H), 4.93 (m, 1H), 4.15 (dd, *J* = 9, 6 Hz, 1H), 2.62 (dd, *J* = 14, 6 Hz, 1H), 2.54 (dd, *J* = 14, 9 Hz, 1H), 1.78 ppm (s, 3H); ¹³C NMR (100 MHz, D₂O): δ = 168.9, 138.5, 116.9, 50.8, 39.8, 21.0 ppm; HRMS (ESI, +ve) *m/z*: calcd for C₈H₁₄N₃O: 144.1137; found: 144.1140 [M+H]⁺.

(2*S*,3*R*)-4-Chloroalanine methyl ester (**28**) was isolated through HPLC from the mixture obtained by chlorination of ester **3c**.^[43] ¹H NMR (400 MHz, CD₃OD): δ = 4.27 (d, *J* = 4 Hz, 1H), 3.87 (s, 3H), 3.74 (dd, *J* = 12, 8 Hz, 1H), 3.66 (dd, *J* = 12, 6 Hz, 1H), 2.35–2.45 (m, 1H), 1.10 ppm (d, *J* = 8 Hz, 3H); ¹³C NMR (100 MHz, CD₃OD): δ = 170.2, 55.7, 53.8, 47.0, 39.5, 13.5 ppm; HRMS (ESI, +ve) *m/z*: calcd for C₆H₁₃NO₂Cl: 166.0635 and 168.0605; found: 166.0633 and 168.0610 [M+H]⁺; *m/z*: calcd for C₆H₁₂NO₂ClNa: 188.0454 and 190.0425; found: 188.0452 and 190.0428 [M+Na]⁺; further details are provided in the Supporting Information.

Treatment of amino acid derivatives with S30 extract from *E. coli* BL21 Star (DE3): S30 extract from *E. coli* BL21 Star (DE3) was prepared as previously reported.^[22] Stock solutions of the amino acid derivatives **1a–f**, **2**, **3a–g**, **4**, **5a–c**, **6a,b**, **7a,b**, **8**, **9a–c**, **10a–c**, **11a–d**, **12**, **13a–c**, **14**, **15a,b**, **16**, **17a–c**, and **18** were prepared in water, ethanol, or dimethyl sulfoxide (DMSO) according to solubility. An aliquot (4 μL) of each stock solution was diluted to a final concentration of 2 mM with 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (116 μL, 50 mM, pH 7.5) and the S30 extract (80 μL). The mixtures were incubated at 37 °C for 6 h and centrifuged at 12000 rpm for 10 min. Each supernatant was passed through an Amicon Ultra-4 (YM-10) centrifugal filter device, and the filtrates were analysed with HPLC by using the Waters AccQ.Tag method, with reference to amino acid standard solutions and the background amino acid concentration of the S30 extract. Accordingly, a sample of each filtrate (20 μL) was treated with AccQ.Flour borate buffer (80 μL) and reconstituted AccQ.Flour reagent (20 μL). The mixtures were analysed by using an AccQ.Tag column (C18, 4 μm, 150 × 3.9 mm), eluting with a gradient of acetonitrile in AccQ.Tag eluent. Representative HPLC traces are provided in the Supporting Information.

Cell-free protein synthesis: Plasmid DNA encoding for His₆-PpiB with expression under control of the phage T7 promoter (pND1098) was carried out according to a previous report.^[10] Plasmid DNA was prepared from *E. coli* DH5α/pND1098 with the Qiagen Plasmid Maxi kit. T7 RNA polymerase (50000 U mL⁻¹) was obtained from New England Biolabs Inc. (MA, USA). Cell-free protein synthesis was carried out by using a reported procedure^[8,10,21,22] with the following few modifications. (S)-Alanine and RNasin were not added to the inner mixture (500 μL). T7 RNA polymerase (2 μL) was added to each reaction mixture instead of the plasmid encoding for this enzyme. An aliquot of tRNA solution of 5 μL was added instead of 10 μL. The final concentration of the solvent (ethanol or DMSO) used to dissolve some of the amino acid derivatives **1a–f**, **2**, **3a–g**, **4**, **5a–c**, **6a,b**, **7a,b**, **8**, **9a–c**, **10a–c**, **11a–d**, **12**, **13a–c**, **14**, **15a,b**, **16**, **17a–c**, and **18** was ≤ 2%, which was established through control experiments to have no effect on protein synthesis. The His₆-PpiB sequence (with an additional C-terminal asparagine residue,^[10] mass = 19221 Da, *N*-formyl-His₆-PpiB = 19250 Da) is MHHHHHHMVT FHTNHGDIVI KTFDDKAPET VKNFLDYCRE GFYNNTIFHR VINGFMIQGG GFEPGMKOKA TKEPIKNEAN NGLKNTRGTL AMARTQAPHS ATAQFFINVV DNDFLNFSGE SLOGWGVCVF AEVVDGMDVV DKIKGVATGR SGMHQDVPKE DVIIESVTVS EN. The in vitro cell-free reaction mixture containing expressed His₆-PpiB was centrifuged at 12000 rpm for 2 min and the supernatant was applied to a Ni-ion affinity column equilibrated with 20 mM sodium phosphate, 0.5 M NaCl, and 20 mM imidazole at pH 7.5 and 4 °C. Bound proteins were eluted by application of 20 mM sodium phosphate, 0.5 M NaCl,

Chlorinated Amino Acids in Peptide Production

Chapter 2. Substrate Selectivity of Isoleucyl-tRNA Synthetase

and 500 mM imidazole at pH 7.5 and 4°C. The eluted protein fractions were concentrated using an Amicon Ultra-4 (YM-10) centrifugal filter device and analysed by 20% SDS-PAGE and mass spectrometry (see the Supporting Information for further details).

Acknowledgements

We gratefully acknowledge the support provided for this work through the CSIRO Emerging Science Initiative for Synthetic Enzymes and the ARC Centre of Excellence for Free Radical Chemistry and Biotechnology.

- [1] A. S. Spirin, V. I. Baranov, L. A. Ryabova, S. Y. Ovodov, Y. B. Alakhov, *Science* **1988**, *242*, 1162–1164.
- [2] T. Kigawa, T. Yabuki, Y. Yoshida, M. Tsutsui, Y. Ito, T. Shibata, S. Yokoyama, *FEBS Lett.* **1999**, *442*, 15–19.
- [3] J. R. Swartz, M. C. Jewett, K. A. Woodrow in *Methods in Molecular Biology, Vol. 267: Recombinant Gene Expression*, 2nd ed. (Eds.: P. Balbás, A. Lorence), Humana Press Inc., Totowa, NJ, **2004**, pp. 169–182.
- [4] M. Ohuchi, H. Murakami, H. Suga, *Curr. Opin. Chem. Biol.* **2007**, *11*, 537–542.
- [5] Y. Shimizu, T. Ueda in *Methods in Molecular Biology, Vol. 607: Cell-Free Protein Production*, 1st ed. (Eds.: Y. Endo, K. Takai, T. Ueda), Humana Press Inc., New York, NY, **2010**, pp. 11–21.
- [6] T. Kigawa, E. Yamaguchi-Nunokawa, K. Kodama, T. Matsuda, T. Yabuki, N. Matsuda, R. Ishitani, O. Nureki, S. Yokoyama, *J. Struct. Funct. Genomics* **2002**, *2*, 29–35.
- [7] V. Bergo, S. Mamaev, J. Olejnik, K. J. Rothschild, *Biophys. J.* **2003**, *84*, 960–966.
- [8] K. Ozawa, M. J. Headlam, P. M. Schaeffer, B. R. Henderson, N. E. Dixon, G. Otting, *Eur. J. Biochem.* **2004**, *271*, 4084–4093.
- [9] F. Katzen, G. Chang, W. Kudlicki, *Trends Biotechnol.* **2005**, *23*, 150–156.
- [10] K. Ozawa, M. J. Headlam, D. Mouradov, S. J. Watt, J. L. Beck, K. J. Rodgers, R. T. Dean, T. Huber, G. Otting, N. E. Dixon, *FEBS J.* **2005**, *272*, 3162–3171.
- [11] D. J. Stigers, Z. I. Watts, J. E. Hennessy, H.-K. Kim, R. Martini, M. C. Taylor, K. Ozawa, J. W. Keillor, N. E. Dixon, C. J. Easton, *Chem. Commun.* **2011**, *47*, 1839–1841.
- [12] A. R. Goerke, J. R. Swartz, *Biotechnol. Bioeng.* **2009**, *102*, 400–416.
- [13] I. N. Ugwumba, K. Ozawa, Z.-Q. Xu, F. Ely, J.-L. Foo, A. J. Herlt, C. Coppin, S. Brown, M. C. Taylor, D. L. Ollis, L. N. Mander, G. Schenk, N. E. Dixon, G. Otting, J. G. Oakeshott, C. J. Jackson, *J. Am. Chem. Soc.* **2011**, *133*, 326–333.
- [14] K. V. Loscha, A. J. Herlt, R. Qi, T. Huber, K. Ozawa, G. Otting, *Angew. Chem.* **2012**, *124*, 2286–2289; *Angew. Chem. Int. Ed.* **2012**, *51*, 2243–2246.
- [15] T. L. Hendrickson, V. de Crécy-Lagard, P. Schimmel, *Annu. Rev. Biochem.* **2004**, *73*, 147–176.
- [16] B. Holzberger, M. Rubini, H. M. Möller, A. Marx, *Angew. Chem.* **2010**, *122*, 1346–1349; *Angew. Chem. Int. Ed.* **2010**, *49*, 1324–1327.
- [17] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444.
- [18] M. G. Hoesl, N. Budisa, *Angew. Chem.* **2011**, *123*, 2948–2955; *Angew. Chem. Int. Ed.* **2011**, *50*, 2896–2902.
- [19] A. Isidro-Llobet, M. Álvarez, F. Albericio, *Chem. Rev.* **2009**, *109*, 2455–2504.
- [20] D. Padmakshan, S. A. Bennett, G. Otting, C. J. Easton, *Synlett* **2007**, 1083–1084.
- [21] L. Guignard, K. Ozawa, S. E. Pursglove, G. Otting, N. E. Dixon, *FEBS Lett.* **2002**, *524*, 159–162.
- [22] M. A. Apponyi, K. Ozawa, N. E. Dixon, G. Otting in *Methods in Molecular Biology, Vol. 426: Structural Proteomics: High Throughput Methods*, (Eds.: B. Kobe, M. Guss, T. Huber), Humana Press Inc., Totowa, NJ, **2008**, pp. 257–268.
- [23] M. Schelhaas, H. Waldmann, *Angew. Chem.* **1996**, *108*, 2192–2219; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 2056–2083.
- [24] J. F. Carson, *Synthesis* **1980**, 730–733.
- [25] C. J. Easton, *Chem. Rev.* **1997**, *97*, 53–82.
- [26] A. K. Croft, C. J. Easton, L. Radom, *J. Am. Chem. Soc.* **2003**, *125*, 4119–4124.
- [27] A. K. Croft, C. J. Easton, K. Kociuba, L. Radom, *Tetrahedron: Asymmetry* **2003**, *14*, 2919–2926.
- [28] S. Fittkau, U. Förster, C. Pascual, W.-H. Schunck, *Eur. J. Biochem.* **1974**, *44*, 523–528.
- [29] C. Lazdunski, J. Busuttill, A. Lazdunski, *Eur. J. Biochem.* **1975**, *60*, 363–369.
- [30] F. C. Golich, M. Han, M. W. Crowder, *Protein Expression Purif.* **2006**, *47*, 634–639.
- [31] D. Chandu, D. Nandji, *Microbiology* **2003**, *149*, 3437–3447.
- [32] A. Ben-Bassat, K. Bauer, S. Y. Chang, K. Myambo, A. Boosman, S. Chang, *J. Bacteriol.* **1987**, *169*, 751–757.
- [33] Z. Mathew, T. M. Knox, C. G. Miller, *J. Bacteriol.* **2000**, *182*, 3383–3393.
- [34] T. Juhász, Z. Szeltner, V. Renner, L. Polgár, *Biochemistry* **2002**, *41*, 4096–4106.
- [35] T. H. T. Coetzer, J. P. D. Goldring, L. E. J. Huson, *Biochimie* **2008**, *90*, 336–344.
- [36] E. E. Haley, *J. Biol. Chem.* **1968**, *243*, 5748–5752.
- [37] J. D. Gary, S. Clarke, *J. Biol. Chem.* **1995**, *270*, 4076–4087.
- [38] H. Suzuki, S. Izuka, H. Minami, N. Miyakawa, S. Ishihara, H. Kumagai, *Appl. Environ. Microbiol.* **2003**, *69*, 6399–6404.
- [39] H. Suzuki, H. Kumagai, T. Tochikura, *J. Bacteriol.* **1986**, *168*, 1325–1331.
- [40] S. Prusiner, R. E. Miller, R. C. Valentine, *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 2922–2926.
- [41] C. J. Easton, C. A. Hutton, G. Rositano, E. W. Tan, *J. Org. Chem.* **1991**, *56*, 5614–5618.
- [42] C. J. Easton, C. A. Hutton, M. C. Merrett, E. R. T. Tiekink, *Tetrahedron* **1996**, *52*, 7025–7036.
- [43] Z. I. Watts, C. J. Easton, *J. Am. Chem. Soc.* **2009**, *131*, 11323–11325.
- [44] J. C. M. van Hest, D. A. Tirrell, *Chem. Commun.* **2001**, 1897–1904.
- [45] G. Verardo, N. Toniutti, A. Gorassini, A. G. Giumanini, *Eur. J. Org. Chem.* **1999**, 2943–2948.
- [46] R. E. Dawson, A. Hennig, D. P. Weimann, D. Emery, V. Ravikumar, J. Montenegro, T. Takeuchi, S. Gabutti, M. Mayor, J. Mareda, C. A. Schalley, S. Matile, *Nat. Chem.* **2010**, *2*, 533–538.
- [47] A. K. H. Hirsch, E. Buhler, J.-M. Lehn, *J. Am. Chem. Soc.* **2012**, *134*, 4177–4183.
- [48] S. M. Iossifidou, C. C. Froussios, *Synthesis* **1996**, 1355–1358.
- [49] C. J. Easton, A. J. Edwards, S. B. McNabb, M. C. Merrett, J. L. O’Connell, G. W. Simpson, J. S. Simpson, A. C. Willis, *Org. Biomol. Chem.* **2003**, *1*, 2492–2498.

Received: November 2, 2012

Revised: February 15, 2013

Published online: March 27, 2013

CHEMISTRY
A EUROPEAN JOURNAL

Supporting Information

© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2013

**In Situ Deprotection and Incorporation of Unnatural Amino Acids during
Cell-Free Protein Synthesis**

**Isaac N. Arthur,^[a, b] James E. Hennessy,^[a, b] Dharshana Padmakshan,^[a, b]
Dannon J. Stigers,^[a, b] Stéphanie Lesturgez,^[a, b] Samuel A. Fraser,^[a, b] Mantas Liutkus,^[a, b]
Gottfried Otting,^[a] John G. Oakeshott,^[c] and Christopher J. Easton^{*, [a, b]}**

chem_201203923_sm_miscellaneous_information.pdf

CONTENTS	PAGE
1. General experimental	S3
2. Syntheses	S3
3. Compound characterisations	S7
(<i>S</i>)-4-Fluoroleucine hydrazide 23	S7
(<i>S</i>)-4,5-Dehydroleucine hydrazide 25	S10
(2 <i>S</i> ,3 <i>R</i>)-4-Chlorovaline methyl ester 28	S14
4. Representative HPLC analyses of the deprotection of amino acid derivatives	S21
<i>Figure S1</i> - S30 Control	S21
<i>Figure S2</i> - (<i>S</i>)-Phenylalanine	S21
<i>Figure S3</i> - (<i>S</i>)-Phenylalanine hydrazide 1b	S22
<i>Figure S4</i> - (<i>S</i>)-Phenylalanine hydrazide 1b + S30 extract	S22
<i>Figure S5</i> - (<i>S</i>)-Valine	S23
<i>Figure S6</i> - (<i>S</i>)-Valine hydrazide 1c	S23
<i>Figure S7</i> - (<i>S</i>)-Valine hydrazide 1c + S30 extract	S24
5. SDS-PAGE analyses	S24
<i>Figure S8</i> - PpiB synthesised using the hydrazide 1a , the ester 3a and the ester 10a in place of (<i>S</i>)-leucine.	S24
<i>Figure S9</i> - PpiB synthesised using the amide 10b , the amide 4 and the ester 3d in place of (<i>S</i>)-leucine.	S25
<i>Figure S10</i> - PpiB synthesised using the acetamide 5a and the trifluoroacetamide amide 6a in place of (<i>S</i>)-leucine.	S25
<i>Figure S11</i> - PpiB synthesised using the ester 3a , the acetamide 5a and the diprotected leucine 14 in place of (<i>S</i>)-leucine.	S26
<i>Figure S12</i> - PpiB synthesised using the hydrazide 1b in place of (<i>S</i>)-phenylalanine and the hydrazide 9a in place of (<i>S</i>)-lysine.	S26
<i>Figure S13</i> - Time-dependent wild-type PpiB synthesis.	S27
<i>Figure S14</i> – PpiB synthesised using the chloride 28 in place of (2 <i>S</i> ,3 <i>S</i>)-isoleucine, before Ni ion affinity chromatography	S27
6. ESI Mass spectrometry of purified proteins	S28
7. References	S30

1. General Experimental

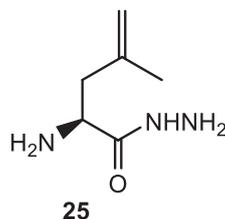
Nuclear Magnetic Resonance (NMR) spectra were recorded on Varian Mercury 300 and MR400 spectrometers operating at 300 MHz and 400 MHz respectively for ^1H , and 100 MHz for ^{13}C . The multiplicity of signals is abbreviated as follows: d = doublet, dd = doublet of doublets, m = multiplet, s = singlet. NMR solvents with purity of at least 99.8% were purchased from Cambridge Isotope Laboratories Inc. High resolution mass spectrometry was conducted using a Bruker Apex 4.7T FTICR spectrometer and a Waters LCT Premier XE spectrometer. Protein mass spectrometry was carried out by direct injection onto an Agilent 1100 series LC/MSD TOF instrument. HPLC was carried out using a Waters Alliance Separation Module 2695 with a Waters 2996 photodiode array detector. LC/MSD and HPLC solvents were purchased from Merck and ultrapure water (resistivity $>15\text{ M}\Omega\text{ cm}^{-1}$) was prepared using a Milli-Q[®] reagent system. Spectra/Por[®] dialysis membrane (#2, MWCO: 12-14,000) was purchased from Spectrum Laboratories Inc. SDS-PAGE gels were run on a Mini-PROTEIN[®] Tetra system and stained with Bio-Safe[®] Coomassie Blue stain from Bio-Rad with reference to molecular weight standards (low range) purchased from Bio-Rad. Proteins were purified with a His GraviTrap[®] Kit from GE Healthcare and concentrated with Amicon[®] Ultra-4 (YM-10) centrifugal devices purchased from Millipore.

2. Syntheses

With the exception of the hydrazides **2** and **9a**, and the ester **10c**, all the amino acids illustrated in Figures 1-3 are available from Sigma-Aldrich[®], Merck Pty. Ltd., Auspep, TCI Chemicals or Aurora Fine Chemicals LLC. For the purposes of this investigation, the hydrazides **1a-f**, **9b-c** and **15a,b** were prepared from the free amino acids by esterification with thionyl chloride in methanol,^[1] followed by treatment of the corresponding product

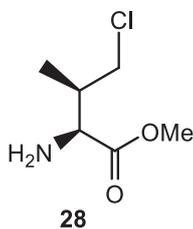
esters (including the leucine, phenylalanine and valine derivatives **3a-c**) with hydrazine hydrate.^[1b, 2] The phenylhydrazide **16** was synthesised from methyl 3-phenylpropanoate, using phenylhydrazine.^[3] *N*-Boc-(*S*)-leucine **17a** was prepared by treatment of (*S*)-leucine with di-*tert*-butyl dicarbonate.^[4] Reaction of *N*-Boc-(*S*)-leucine **17a** with benzyl bromide followed by treatment with trifluoroacetic acid gave the benzyl ester **3d**, with the same method being used to generate the benzyl esters **3e** and **3f**.^[4-5] Reaction of *N*-Boc-(*S*)-leucine **17a** with BOP and 1-aminoadamantane and *tert*-butylamine, respectively, followed by deprotection with trifluoroacetic acid, gave the amides **4** and **10b**.^[6] The acetamides **5a,b** and **11a,b** were obtained by acetylation of their respective free amino acids using acetic anhydride.^[7] Treatment of (*S*)-leucine and (*S*)-valine with ethyl trifluoroacetate yielded compounds **6a** and **6b**, respectively.^[8] The benzamide **12** and the sulfonamide **17b** were synthesised from the ester **3a** using benzoyl chloride^[9] and methanesulfonyl chloride,^[9-10] respectively, followed by ester cleavage using lithium hydroxide. BOP coupling of the ester **3a** with decanoic acid then treatment with base yielded compound **17c**.^[11] The hydrazide **8** was prepared by selective side chain esterification of (*S*)-glutamic acid using sulfuric acid in methanol, followed by reaction with hydrazine hydrate.^[12] Acetylation of the ester **3a** using acetic anhydride yielded the diprotected leucine **14**,^[13] which reacted with hydrazine hydrate to give the corresponding hydrazide **18**.^[14] The amides **5c**, **7a,b** and **11c,d**, the esters **3g** and **10a** and the side chain protected amino acids **13a-c** were purchased from Sigma-Aldrich[®], Merck Pty. Ltd., Auspep or TCI Chemicals.

(S)-4,5-Dehydroleucine hydrazide 25.



N-Phthaloyl-(*S*)- γ,δ -dehydroleucine methyl ester **22**^[15] (8 mg, 0.029 mmol) was dissolved in 2 mL ethanol, followed by addition of 80% hydrazine hydrate (17.8 μ L, 0.29 mmol). The resultant solution was heated at reflux for 1 h. After cooling and evaporation of the solvent, the colourless residue was subjected to reverse phase HPLC (YMC-Pack ODS-AQ, 250 x 4.6 mm, 5 μ m) running a mobile phase of 20% acetonitrile in 0.1% aqueous trifluoroacetic acid at a flow rate of 1 mL per minute. Upon evaporation of the solvent, white needles of the title compound as the TFA salt were obtained (1.9 mg, 25%). M.P. 130-132 °C. HPLC t_R = 2.8 minutes. $[\alpha]_D^{20}$ -19.3 (c = 0.8, H₂O). ¹H NMR (300 MHz, D₂O): δ 1.78 (s, 3H), 2.54 (dd, J = 14, 9 Hz, 1H), 2.62 (dd, J = 14, 6 Hz, 1H), 4.15 (dd, J = 9, 6 Hz, 1H), 4.93 (s, 1H), 5.06 (s, 1H). ¹³C NMR (100 MHz, D₂O) δ 21.0, 39.8, 50.8, 116.9, 138.5, 168.9. HRMS (ESI) calcd. for C₆H₁₄N₃O [M+H]⁺ m/z 144.1137; found 144.1140.

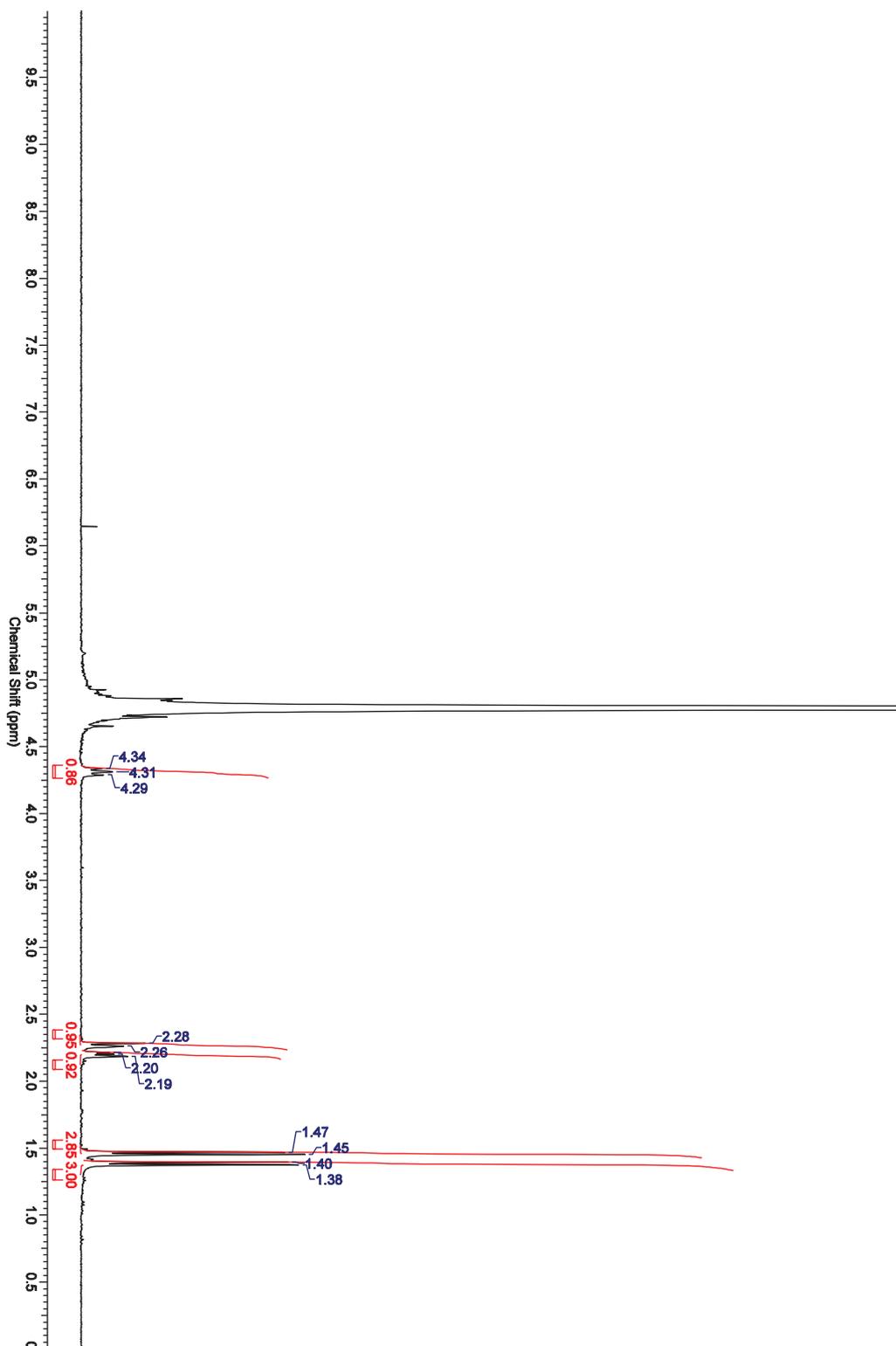
(2*S*,3*R*)-4-Chlorovaline methyl ester **28**



(*S*)-Valine methyl ester **3a**^[15] (2.0 g, 11.9 mmol (hydrochloride salt)) was dissolved in trifluoroacetic acid (20 mL) and chlorine gas was bubbled through the solution for 5 minutes whilst it was irradiated with a 300 W sunlamp. After evaporation of the solvent, a sample of this reaction mixture was subjected to reverse phase HPLC (Atlantis, 150 x 4.6 mm, 3 μ m, in tandem with Atlantis, 250 x 4.6 mm, 5 μ m) running a mobile phase of 20% acetonitrile in 100 mM NaClO₄, pH 2.1, at a flow rate of 0.8 mL per minute. Counter-ion exchange by HPLC (Primesep 100, 150 x 4.6 mm, 5 μ m) running a mobile phase of 20% methanol in 0.4% aqueous trifluoroacetic acid at a flow rate of 1 mL per minute, followed by evaporation of the solvent gave the TFA salt of the title compound as colourless powder. HPLC (Atlantis) t_R = 15.6 minutes. HPLC (Primesep) t_R = 5.4 minutes. ¹H NMR (400 MHz, CD₃OD) δ 1.10 (d, J = 8 Hz, 3H), 2.35-2.45 (m, 1H), 3.66 (dd, J = 12, 6 Hz, 1H), 3.74 (dd, J = 12, 8 Hz, 1H), 3.87 (s, 3H), 4.27 (d, J = 4 Hz, 1H). ¹³C NMR (100 MHz, CD₃OD) δ 13.5, 39.5, 47.0, 53.8, 55.7, 170.2. HRMS (ESI) calcd. for C₆H₁₃NO₂³⁵Cl [M+H]⁺ m/z 166.0635, 168.0605; found 166.0633, 168.0610; calcd. for C₆H₁₂NO₂³⁵ClNa [M+Na]⁺ m/z 188.0454, 190.0425; found 188.0452, 190.0428.

3. Compound Characterisations

(*S*)-4-Fluoroleucine hydrazide **23**



Elemental Composition Report

Page 1

Single Mass Analysis

Tolerance = 3.0 PPM / DBE: min = -1.5, max = 40.0

Element prediction: Off

Number of isotope peaks used for i-FIT = 3

Monoisotopic Mass, Even Electron Ions

40 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-60 H: 0-50 N: 0-3 O: 0-1 F: 0-1 23Na: 1-1

SF2-113FLUOROHYDRAZIDE/AJ

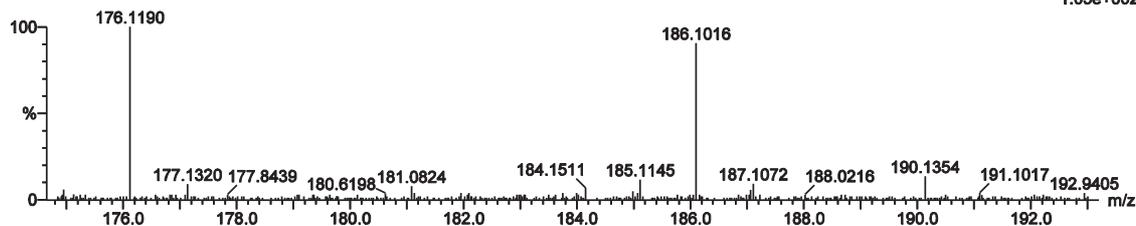
KE375

04-Jun-2012 14:04:18

17650

0712 124 (5.430)

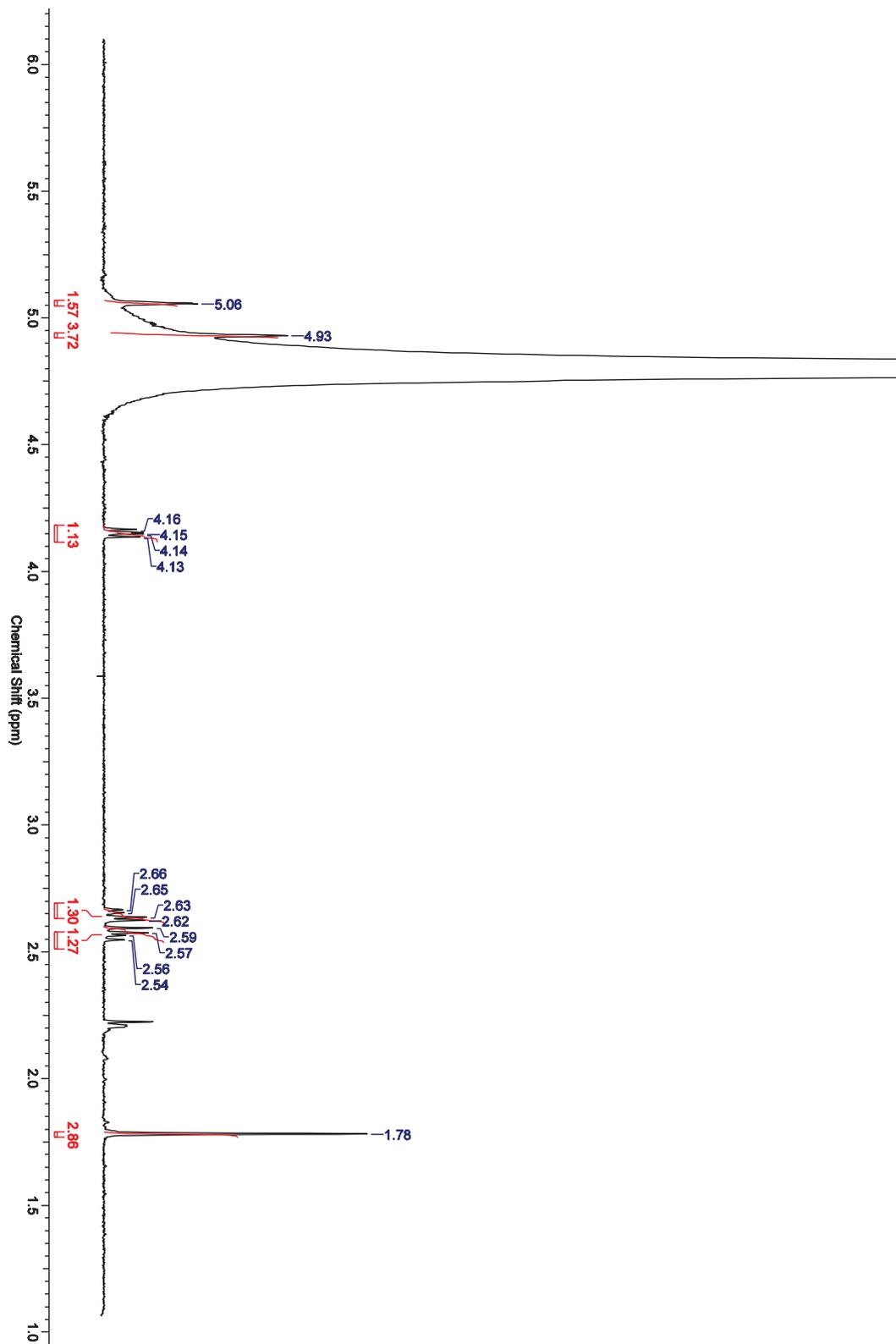
1: TOF MS ES+
1.05e+002

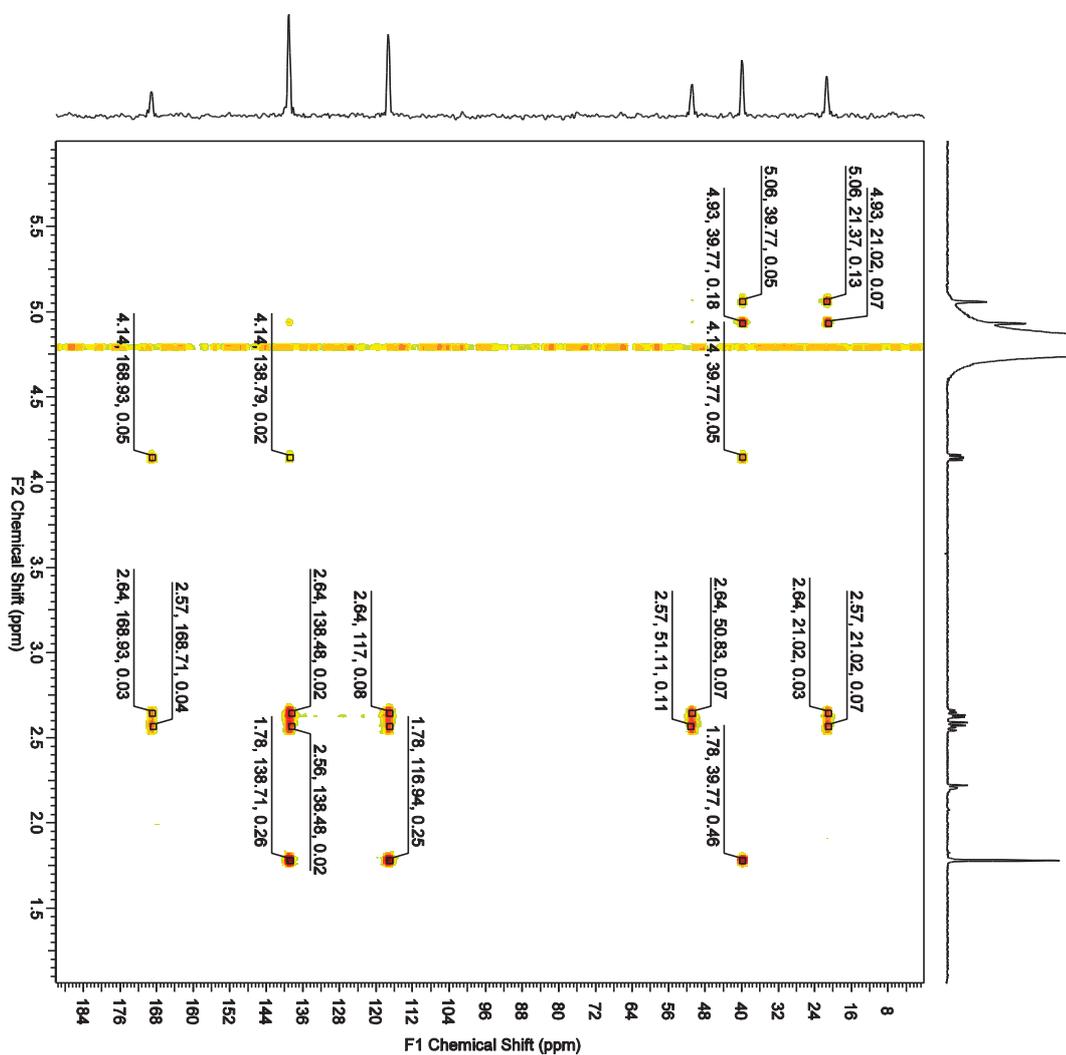


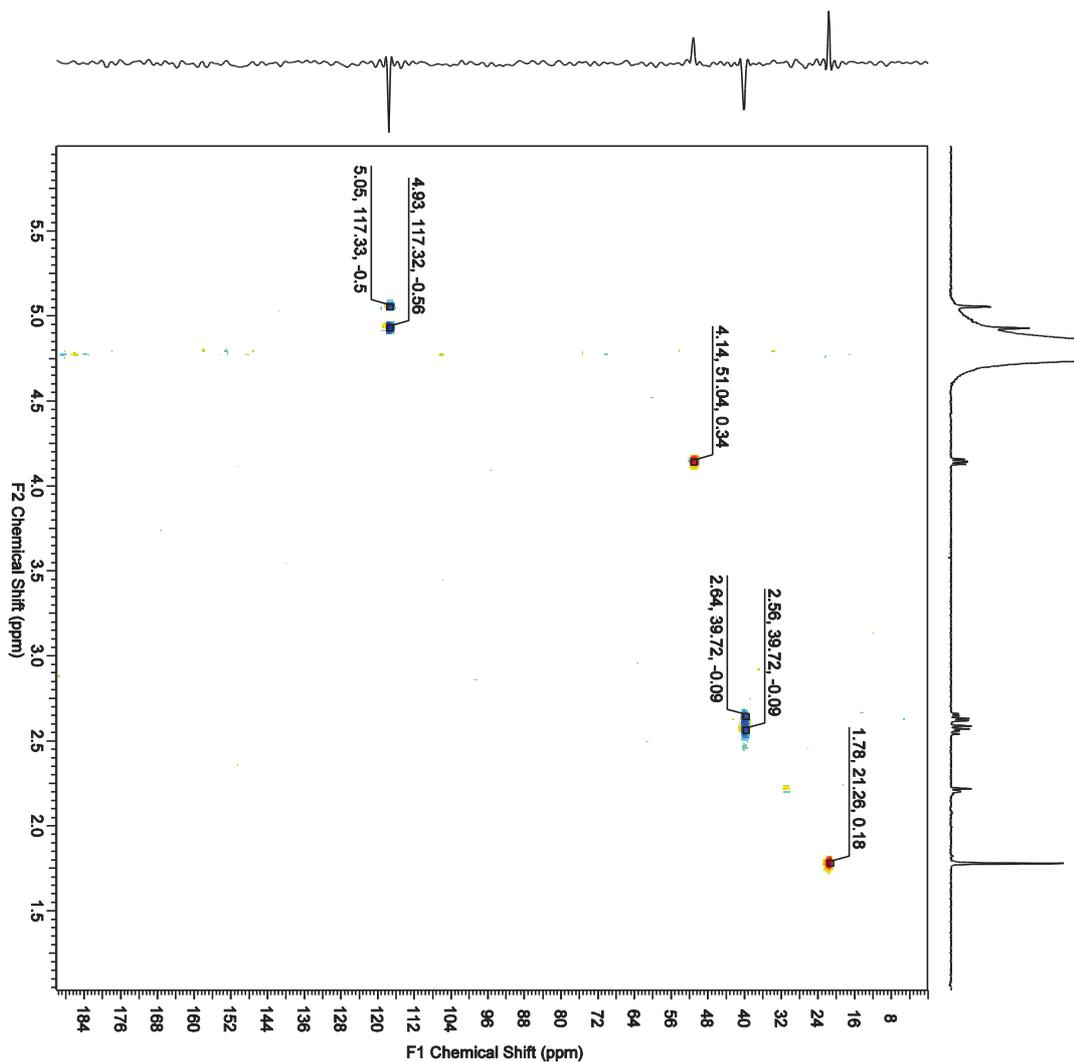
Minimum: -1.5
Maximum: 5.0 3.0 40.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
186.1016	186.1019	-0.3	-1.6	0.5	n/a	C6 H14 N3 O F 23Na

(S)-4,5-Dehydroleucine hydrazide **25**







Elemental Composition Report

Page 1

Single Mass Analysis

Tolerance = 3.0 PPM / DBE: min = -1.5, max = 20.0

Element prediction: Off

Number of isotope peaks used for i-FIT = 3

Monoisotopic Mass, Even Electron Ions

71 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-50 H: 0-50 N: 0-5 O: 0-5

SF2-80a/AJ

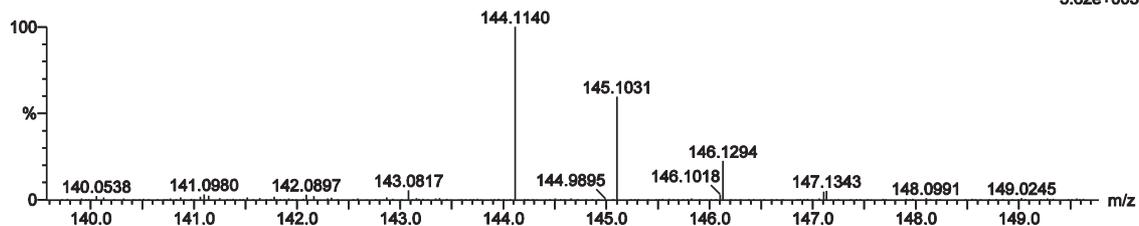
15292

0314 88 (3.849) Cm (86:88)

KE375

16-Mar-2012 09:46:37

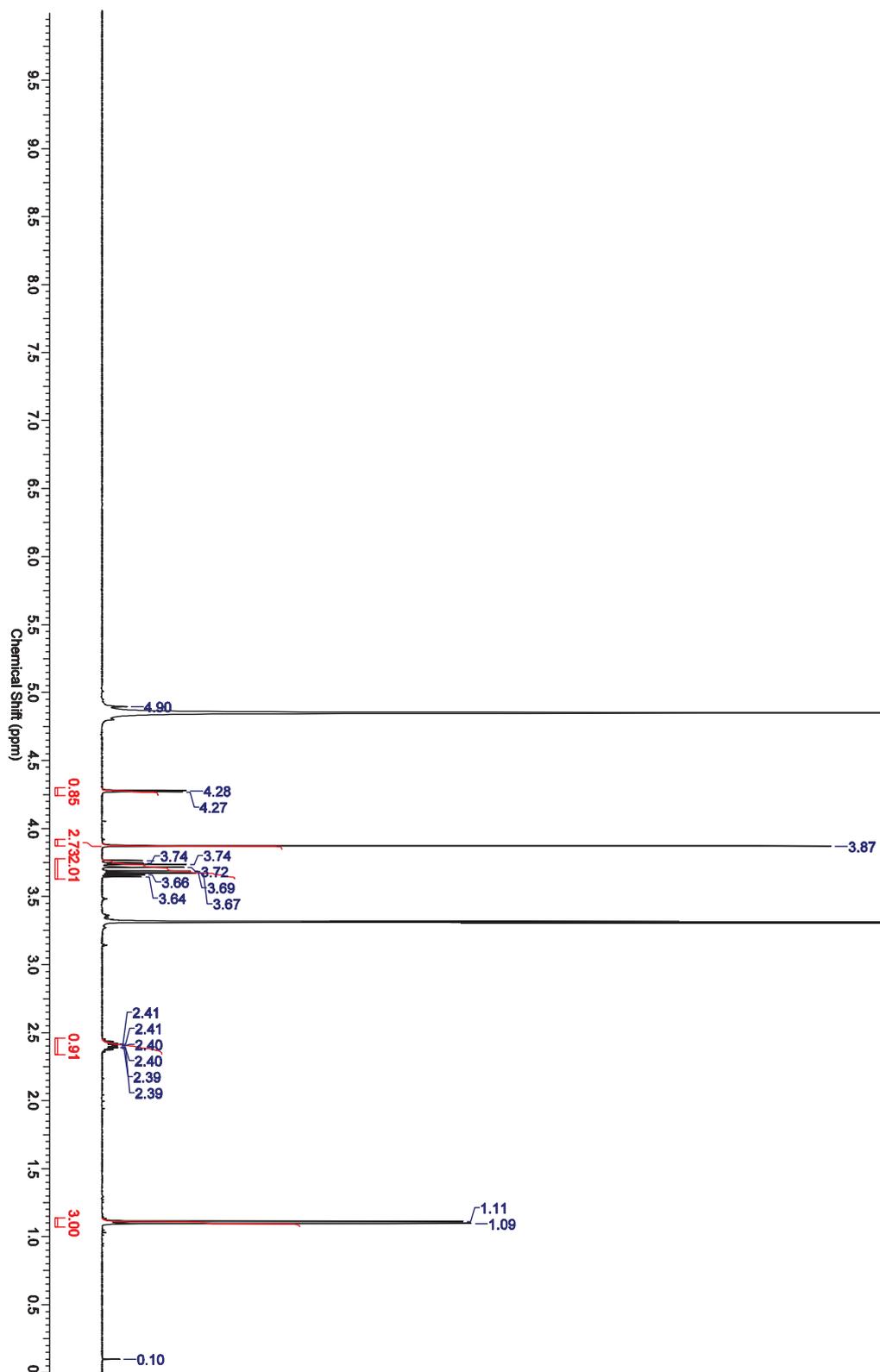
1: TOF MS ES+
3.62e+003

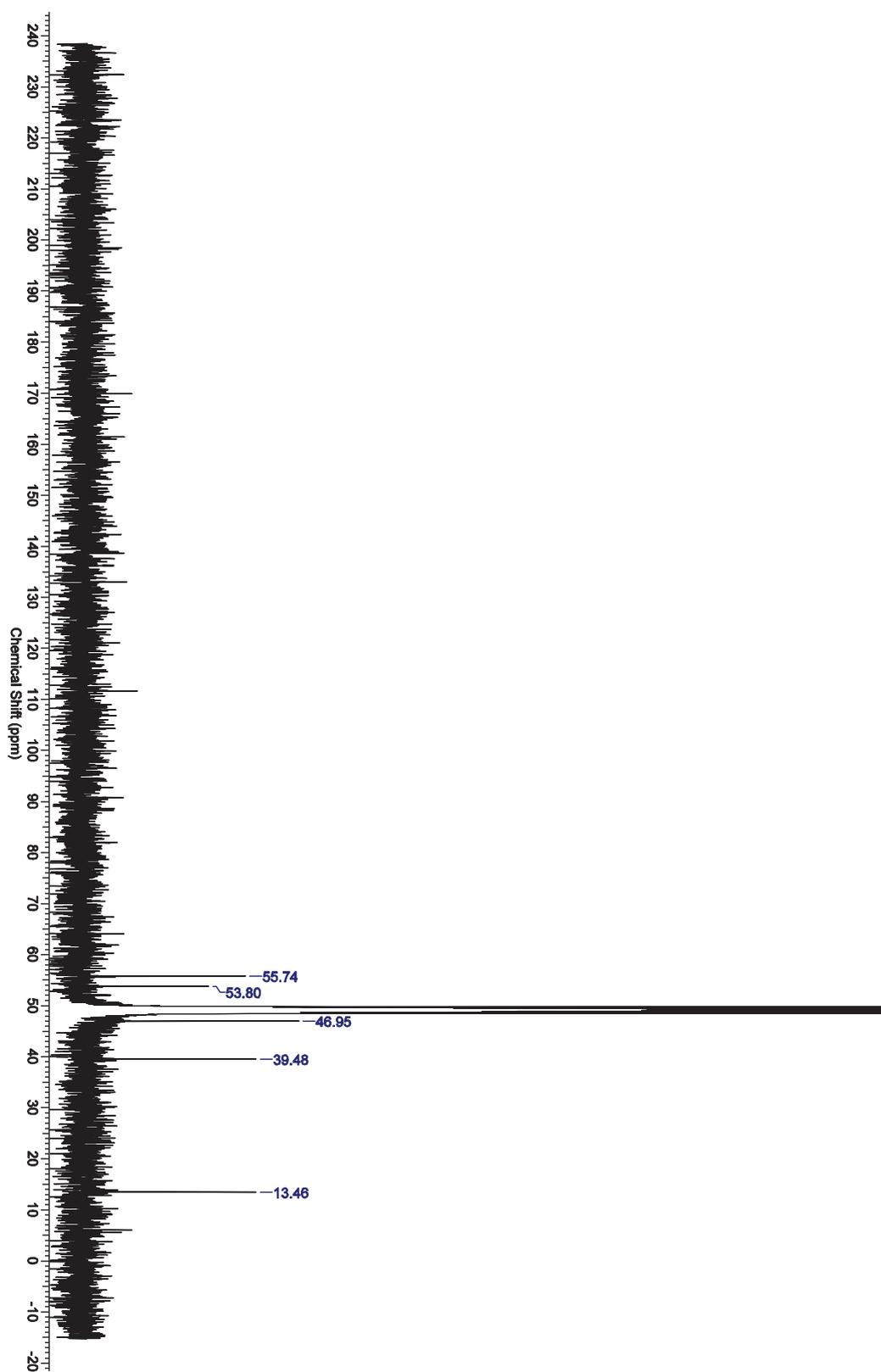


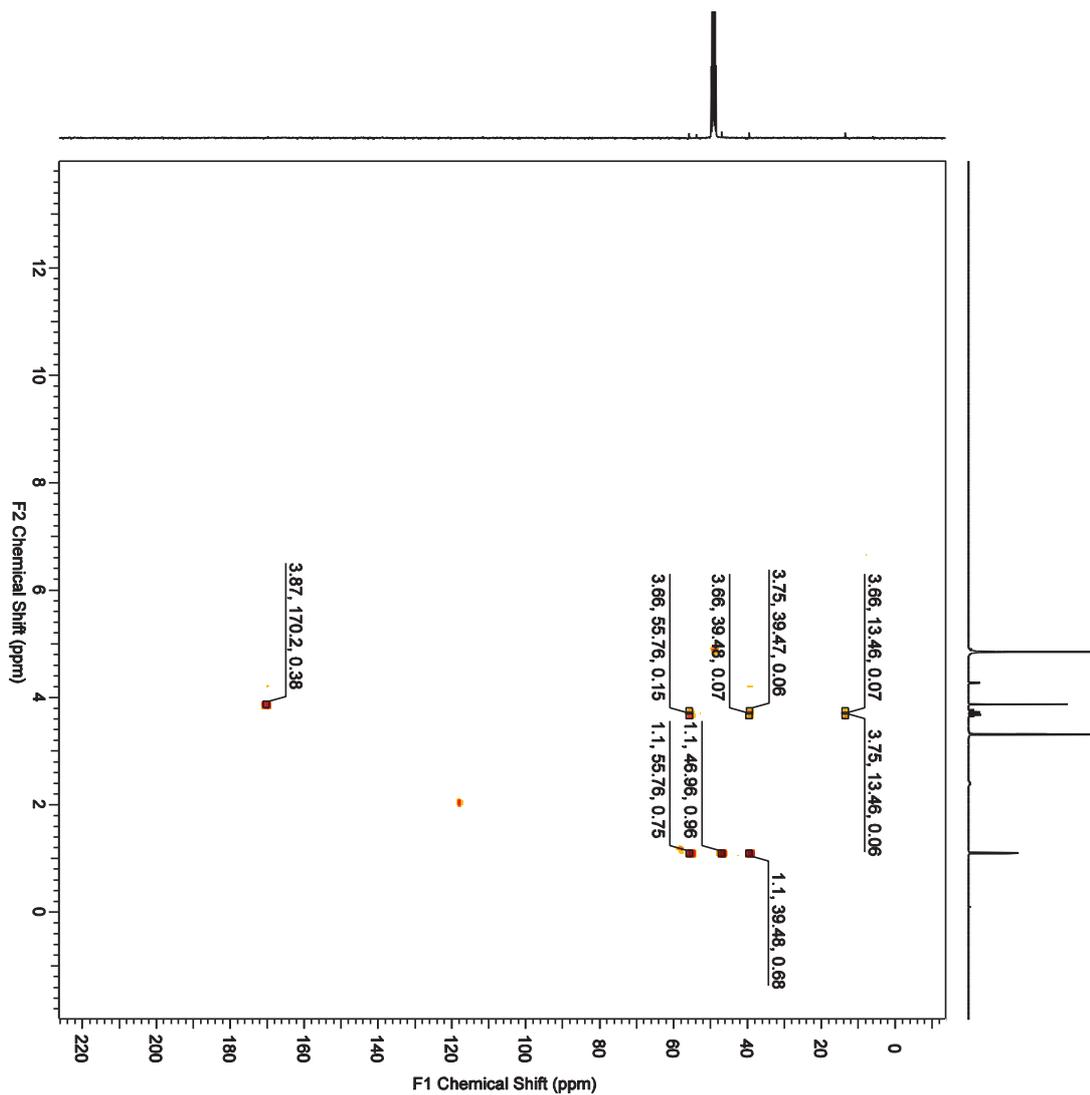
Minimum: -1.5
Maximum: 5.0 3.0 20.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
144.1140	144.1137	0.3	2.1	1.5	n/a	C6 H14 N3 O

(2*S*,3*R*)-4-Chlorovaline methyl ester **28**







Elemental Composition Report

Page 1

Single Mass Analysis

Tolerance = 3.0 PPM / DBE: min = -1.5, max = 20.0

Element prediction: Off

Number of isotope peaks used for i-FIT = 3

Monoisotopic Mass, Odd and Even Electron Ions

25 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-6 H: 0-13 N: 0-1 O: 0-2 35Cl: 0-1 37Cl: 0-1

ML121hC/AJ

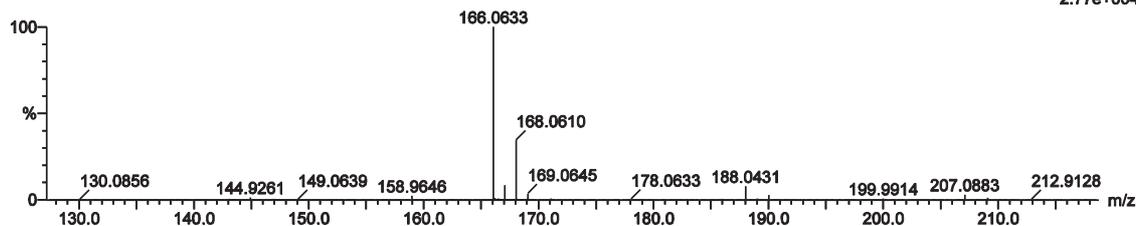
KE375

10-Jul-2012 13:55:11

18672

1: TOF MS ES+
2.77e+004

0853 11 (0.488)



Minimum: -1.5
Maximum: 20.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
166.0633	166.0635	-0.2	-1.2	0.5	4679.6	C6 H13 N O2 35Cl

Elemental Composition Report

Single Mass Analysis

Tolerance = 3.0 PPM / DBE: min = -1.5, max = 20.0
 Element prediction: Off
 Number of isotope peaks used for i-FIT = 3

Monoisotopic Mass, Even Electron Ions

23 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-6 H: 0-13 N: 0-1 O: 0-2 35Cl: 0-1 37Cl: 0-1

ML121hC/AJ

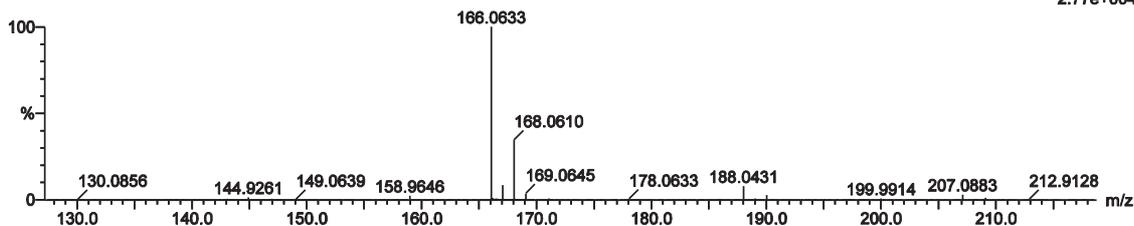
KE375

10-Jul-2012 13:55:11

18672

1: TOF MS ES+
 2.77e+004

0853 11 (0.488)



Minimum:

5.0 3.0 -1.5

Maximum:

5.0 3.0 20.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
168.0610	168.0605	0.5	3.0	0.5	40.3	C6 H13 N O2 37Cl

Elemental Composition Report

Page 1

Single Mass Analysis

Tolerance = 3.0 PPM / DBE: min = -1.5, max = 20.0

Element prediction: Off

Number of isotope peaks used for i-FIT = 3

Monoisotopic Mass, Even Electron Ions

49 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

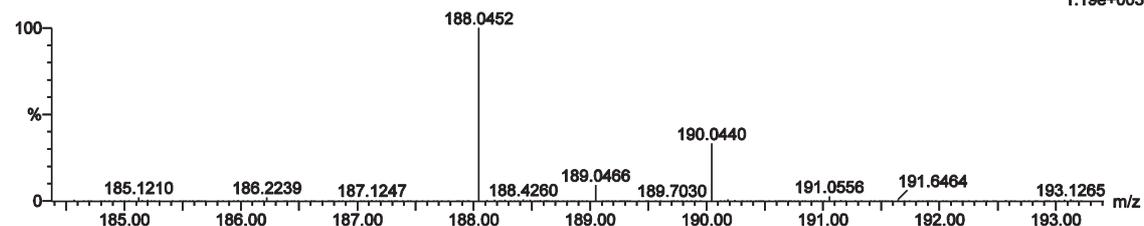
C: 0-6 H: 0-13 N: 0-1 O: 0-2 35Cl: 0-1 37Cl: 0-1 23Na: 0-1

ML121hC/AJ
18672
0853 3 (0.138)

KE375

10-Jul-2012 13:55:11

1: TOF MS ES+
1.19e+003



Minimum: -1.5
Maximum: 5.0 3.0 20.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
188.0452	188.0454	-0.2	-1.1	0.5	192.8	C6 H12 N O2 35Cl 23Na

Elemental Composition Report

Single Mass Analysis

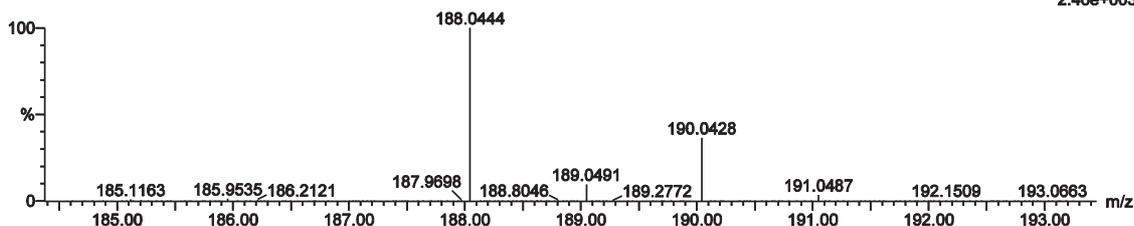
Tolerance = 3.0 PPM / DBE: min = -1.5, max = 20.0
 Element prediction: Off
 Number of isotope peaks used for i-FIT = 3

Monoisotopic Mass, Even Electron Ions
 47 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)
 Elements Used:
 C: 0-6 H: 0-13 N: 0-1 O: 0-2 35Cl: 0-1 37Cl: 0-1 23Na: 0-1

ML121hC/AJ
 18672
 0853 4 (0.172)

KE375

10-Jul-2012 13:55:11
 1: TOF MS ES+
 2.46e+003



Minimum: -1.5
 Maximum: 5.0 3.0 20.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
190.0428	190.0425	0.3	1.6	0.5	2.1	C6 H12 N 02 37Cl 23Na

4. Representative HPLC analyses of the deprotection of amino acid derivatives

Analysis of the deprotection of amino acid derivatives by S30 was carried out using the Waters AccQ.Tag™ method as described in the Experimental Section of the manuscript. Below are representative analyses used to determine the extents of deprotection of the hydrazides **1b** and **1c** by S30.

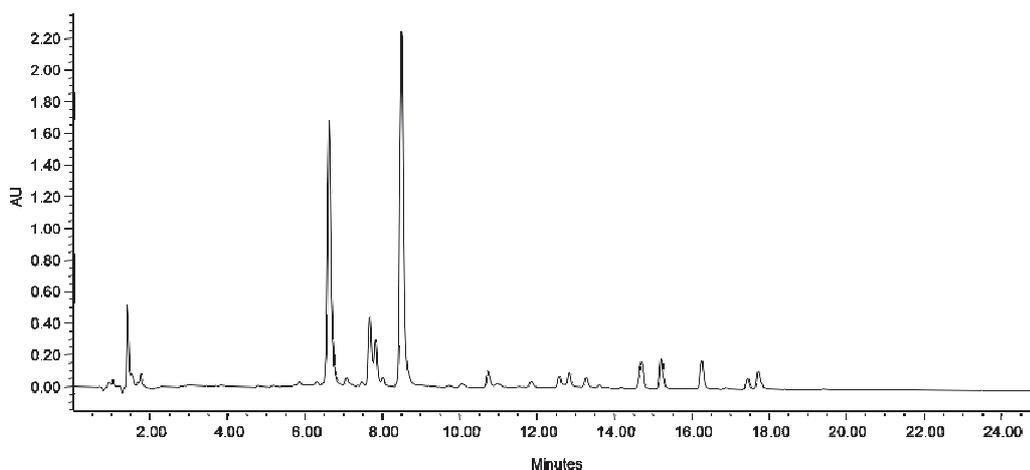


Figure S1. Chromatogram obtained through reverse phase HPLC of AccQ.Tag™ derivatized S30 extract incubated at 37 °C for 6 h.

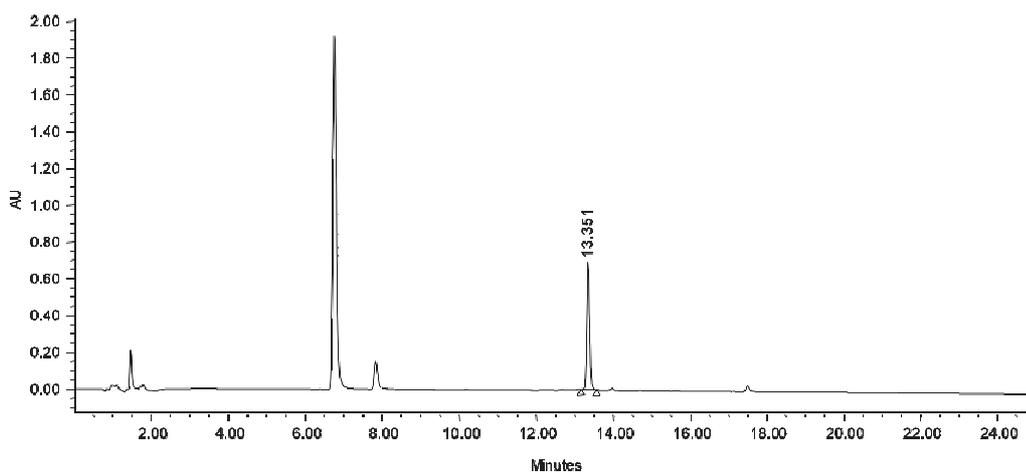


Figure S2. Chromatogram obtained through reverse phase HPLC of AccQ.Tag™ derivatized (*S*)-phenylalanine.

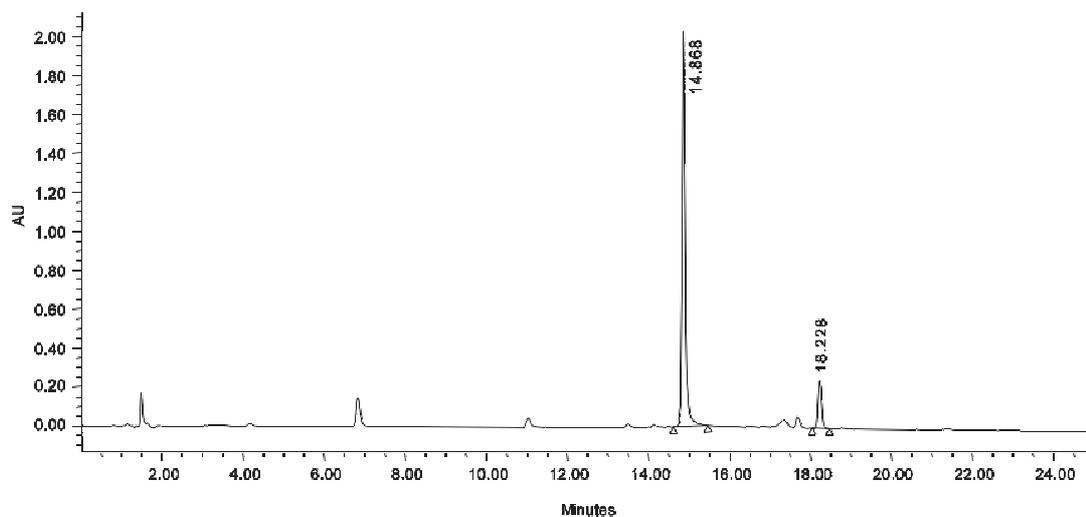


Figure S3. Chromatogram obtained through reverse phase HPLC of AccQ.Tag™ derivatized (*S*)-phenylalanine hydrazide **1b**.

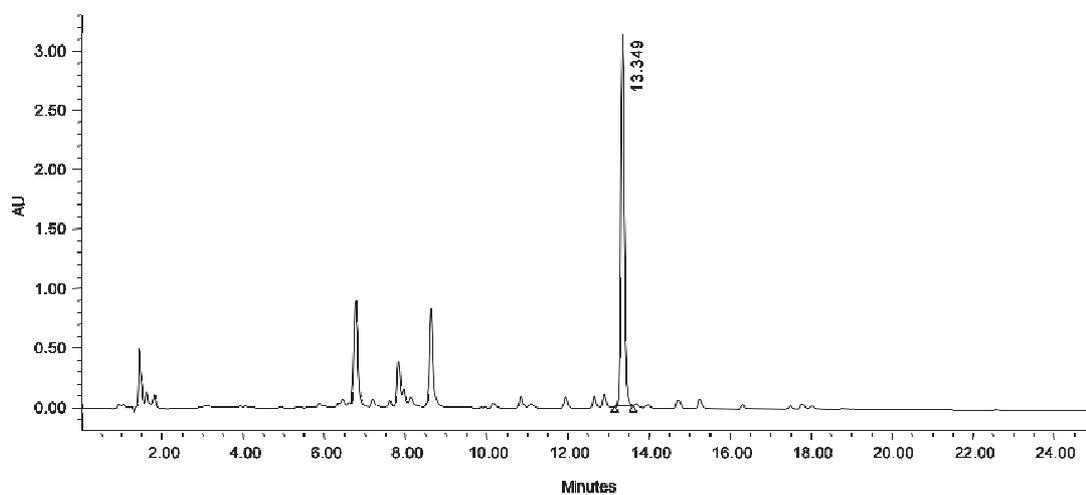


Figure S4. Chromatogram obtained through reverse phase HPLC of AccQ.Tag™ derivatized (*S*)-phenylalanine hydrazide **1b** incubated with S30 extract at 37 °C for 6 h.

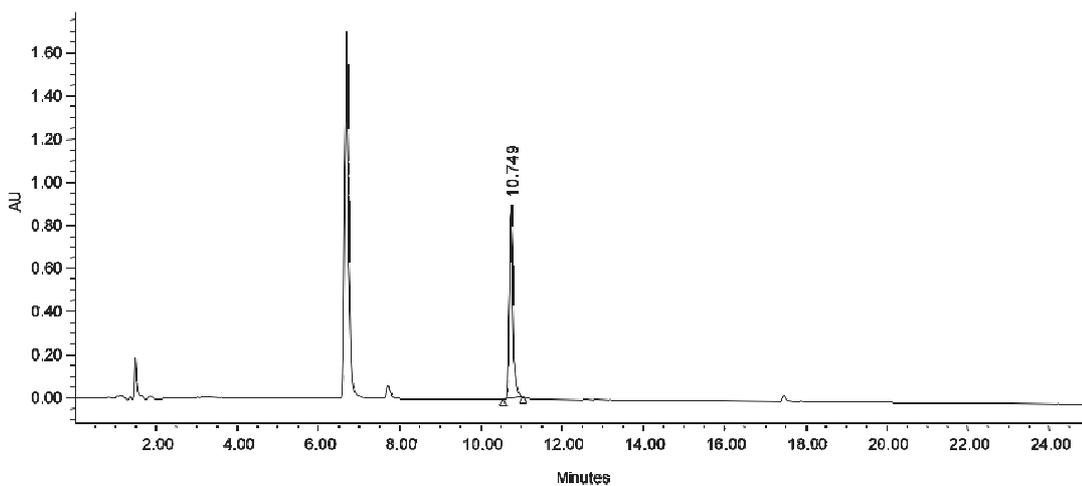


Figure S5. Chromatogram obtained through reverse phase HPLC of AccQ.Tag™ derivatized (*S*)-valine.

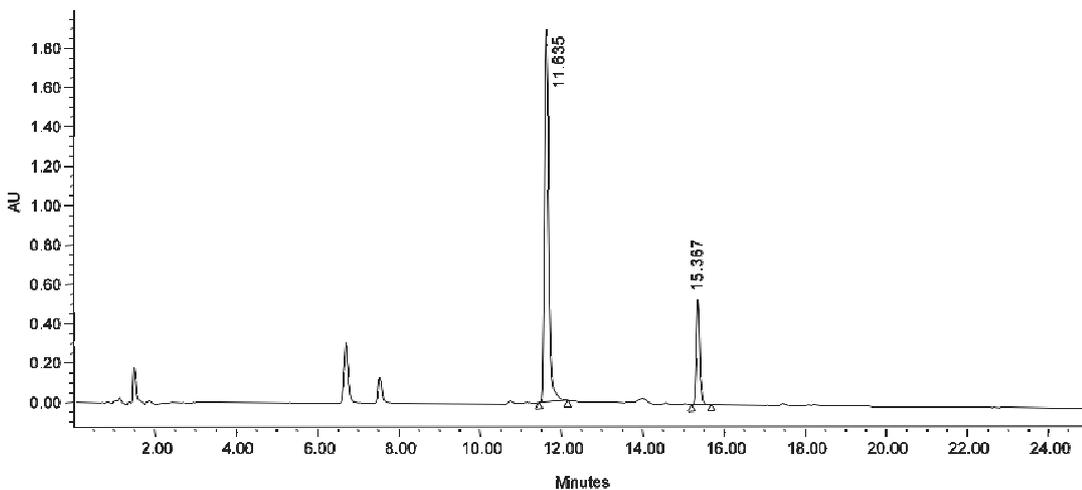


Figure S6. Chromatogram obtained through reverse phase HPLC of AccQ.Tag™ derivatized (*S*)-valine hydrazide **1c**.

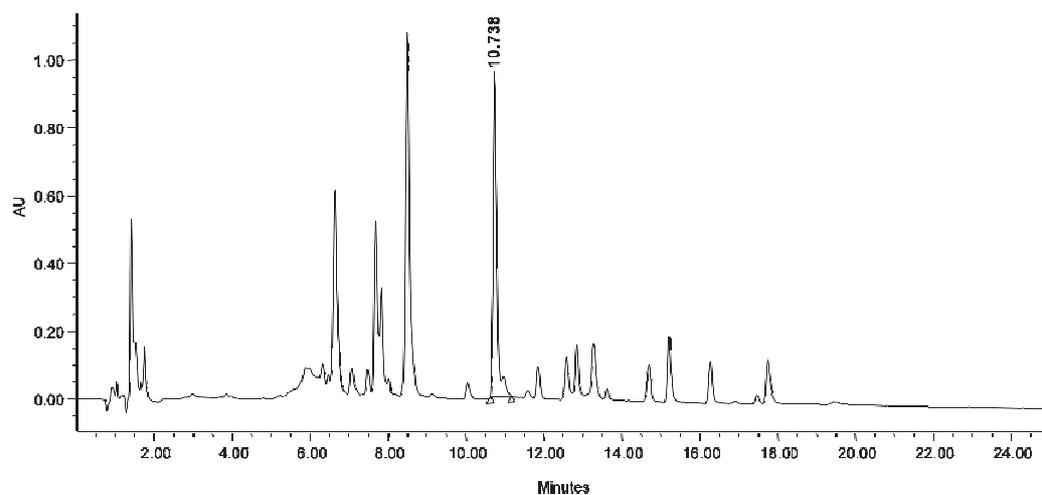


Figure S7. Chromatogram obtained through reverse phase HPLC of AccQ.Tag™ derivatized (*S*)-valine hydrazide **1c** incubated with S30 extract at 37 °C for 6 h.

5. SDS-PAGE analyses

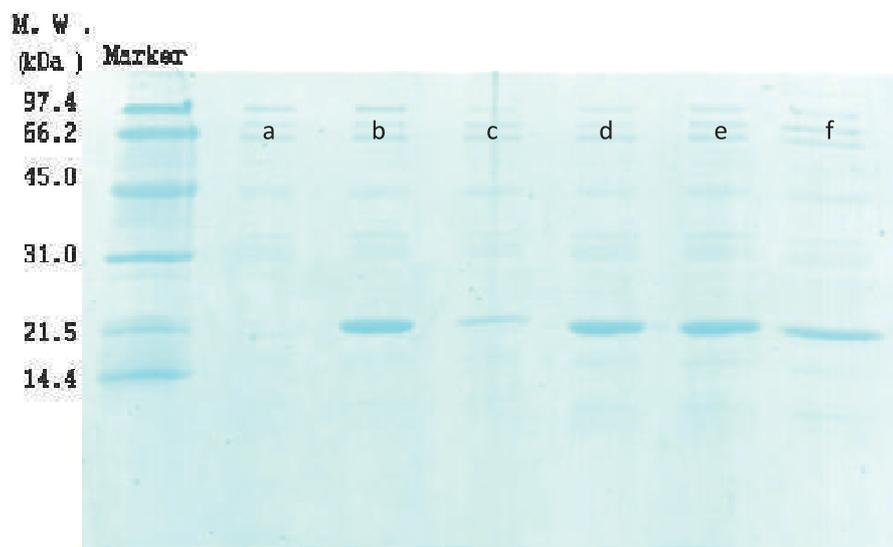


Figure S8. 20% SDS-PAGE of synthesised His₆-PpiB with: a) no DNA; b) DNA; c) no (*S*)-leucine; d) no (*S*)-leucine but with the hydrazide **1a**; e) no (*S*)-leucine but with the methyl ester **3a**; and f) no (*S*)-leucine but with the ester **10a**.

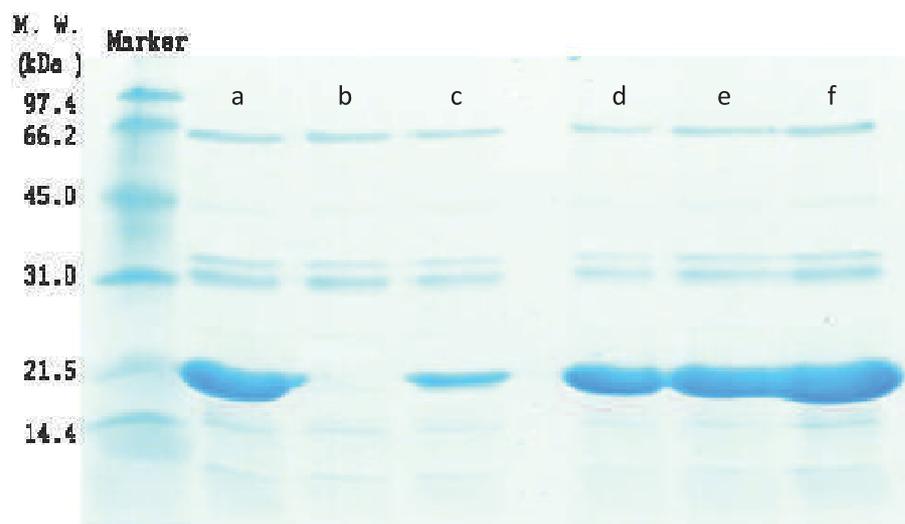


Figure S9. 20% SDS-PAGE of synthesised His₆-PpiB with: a) DNA; b) no DNA; c) no (*S*)-leucine; d) no (*S*)-leucine but with the amide **10b**; e) no (*S*)-leucine but with the amide **4**; and f) no (*S*)-leucine but with the ester **3d**.

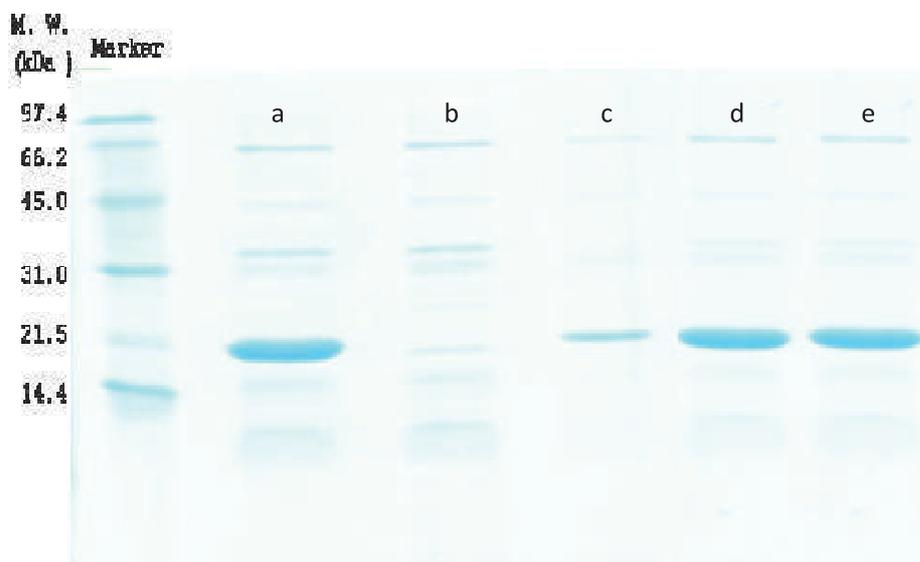


Figure S10. 20% SDS-PAGE of synthesised His₆-PpiB with: a) DNA; b) no DNA; c) no (*S*)-leucine; d) no (*S*)-leucine but with the acetamide **5a**; and e) no (*S*)-leucine but with the trifluoroacetamide **6a**.

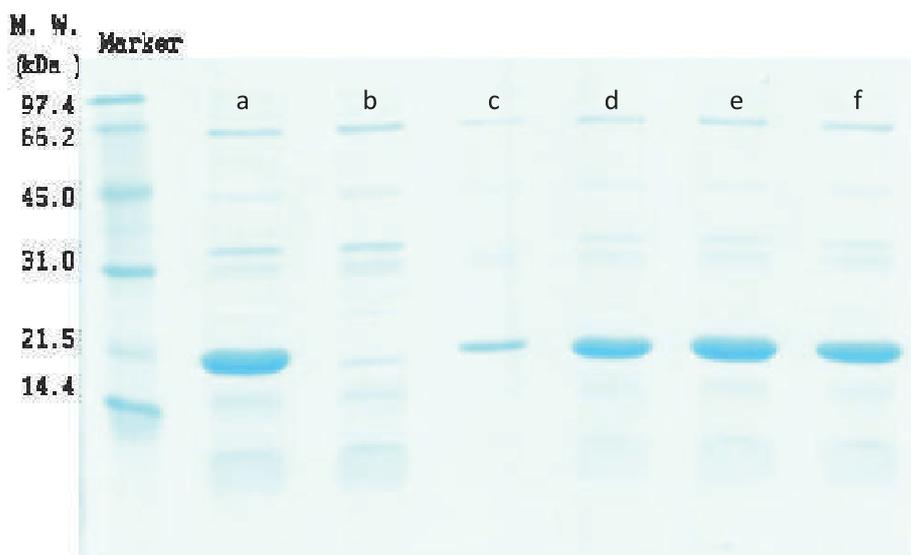


Figure S11. 20% SDS-PAGE of synthesised His₆-PpiB with: a) DNA; b) no DNA; c) no (S)-leucine; d) no (S)-leucine but with the ester **3a**; e) no (S)-leucine but with the acetamide **5a**; and f) no (S)-leucine but with the diprotected leucine **14**.

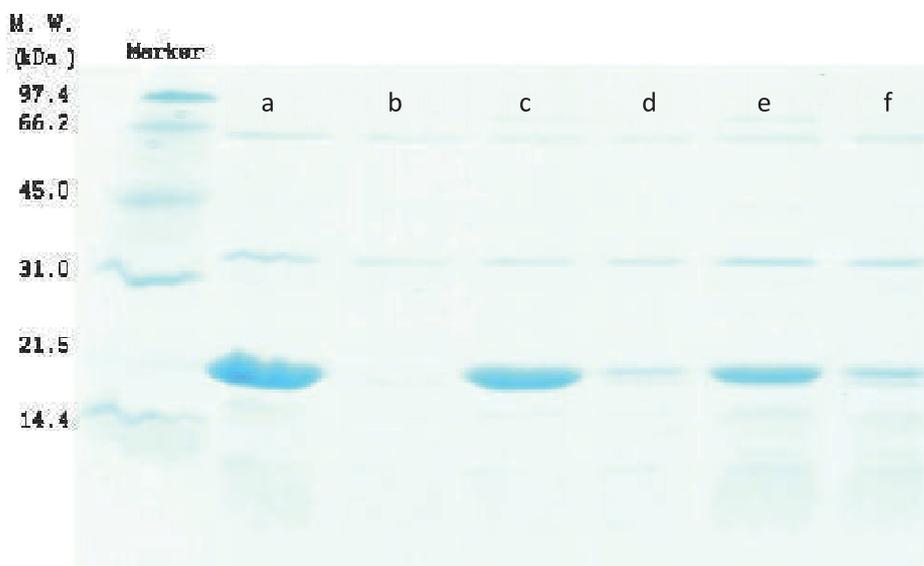


Figure S12. 20% SDS-PAGE of synthesised His₆-PpiB with: a) DNA; b) no DNA; c) no (S)-phenylalanine but with the hydrazide **1b**; d) no (S)-phenylalanine; e) no (S)-lysine but with the hydrazide **9a**; and f) no (S)-lysine.

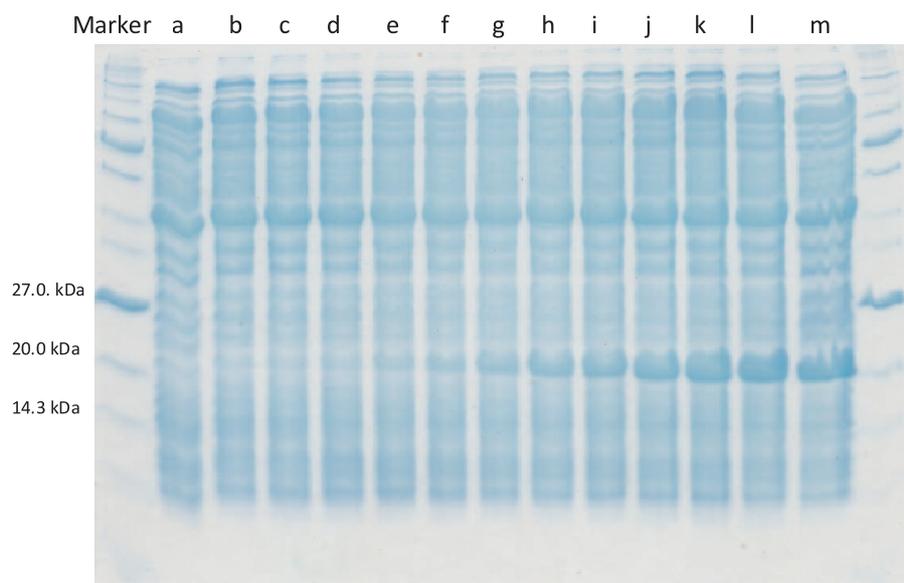


Figure S13. Time-dependent wild-type PpiB synthesis (19250 Da) illustrated by 20% SDS-PAGE of unpurified proteins after a) 10 mins; b) 20 mins; c) 30 mins; d) 40 mins; e) 50 mins; f) 60 mins; g) 90 mins; h) 120 mins; i) 150 mins; j) 180 mins; k) 240 mins; l) 300 mins; and m) 360 mins incubation at 37 °C.

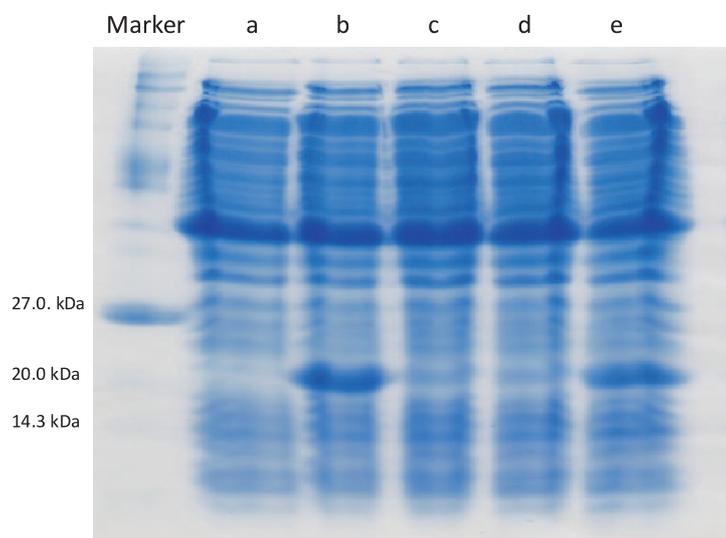


Figure S14. PpiB synthesised using the chloride **28** in place of (2*S*,3*S*)-isoleucine, before Ni ion affinity chromatography, illustrated by 20% SDS-PAGE showing His₆-PpiB (19250 Da) synthesised with: a) no DNA; b) DNA; c) no (*S*)-leucine; d) no (2*S*,3*S*)-isoleucine; and e) no (2*S*,3*S*)-isoleucine but with the chloride **28**.

6. ESI Mass spectrometry of purified proteins

Purified protein solutions (50 μ L) were analysed by ESI mass spectrometry by direct injection into the spectrometer running a mobile phase of a 50:50 (v/v) solution of 0.1% formic acid in acetonitrile:0.1% aqueous formic acid. Deconvoluted mass spectra of native PpiB and PpiB synthesised in the presence of the hydrazides **23** and **25**, and the chloride **28**, are shown below.

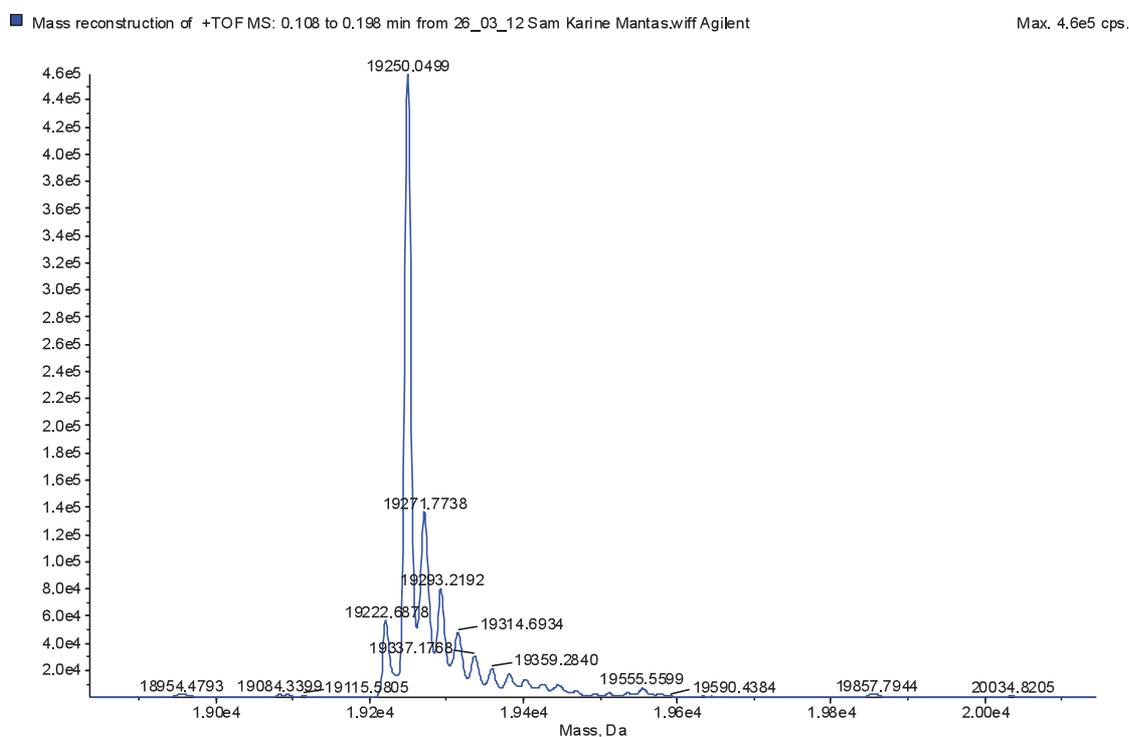


Figure S15. Deconvoluted mass spectrum of native PpiB

Chlorinated Amino Acids in Peptide Production Chapter 2. Substrate Selectivity of Isoleucyl-tRNA Synthetase

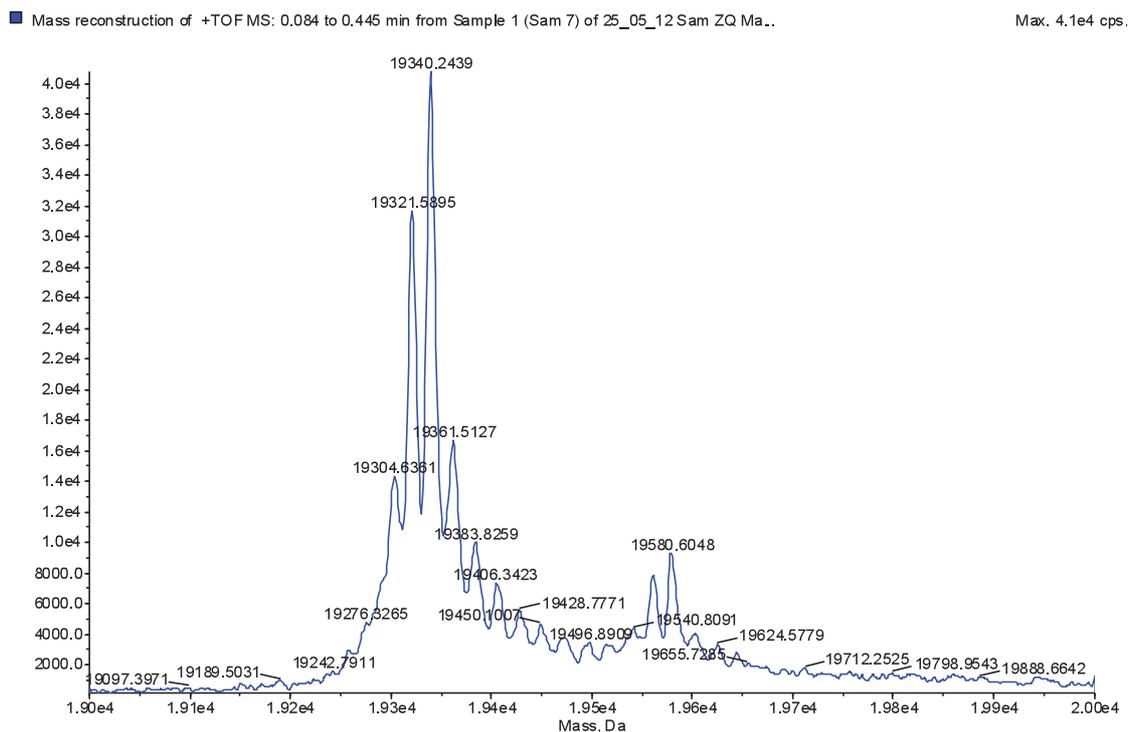


Figure S16. Deconvoluted mass spectrum of PpiB synthesised using the hydrazide **23**.

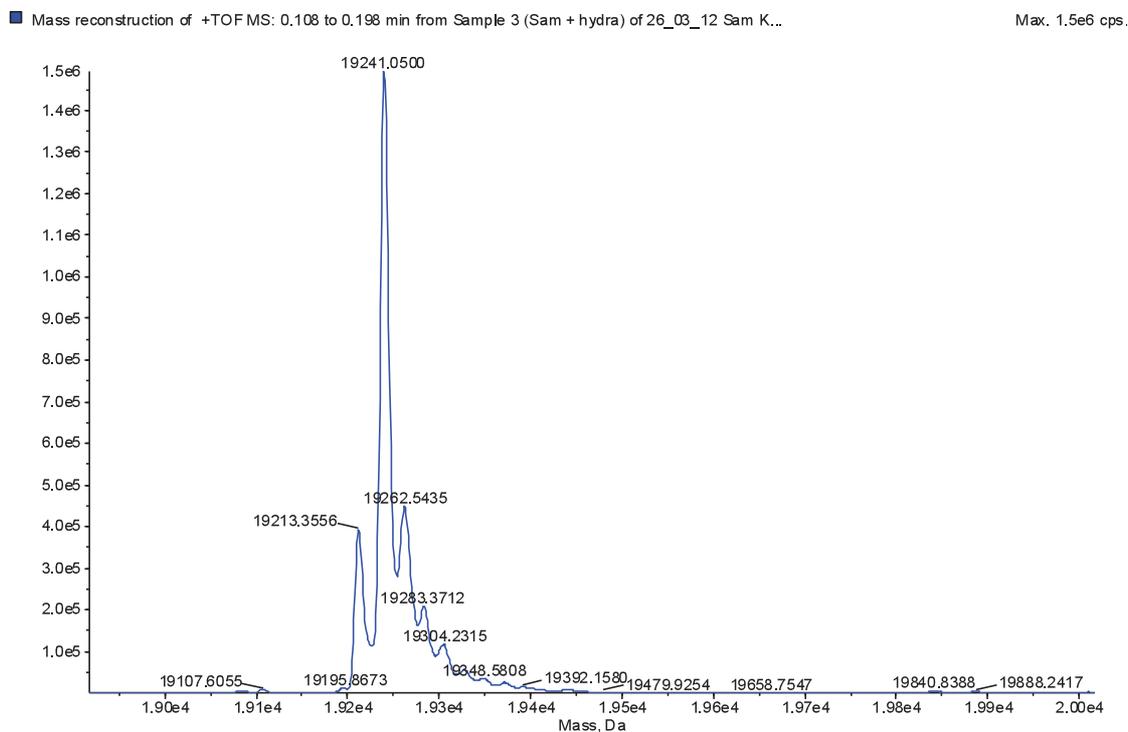


Figure S17. Deconvoluted mass spectrum of PpiB synthesised using the hydrazide **25**.

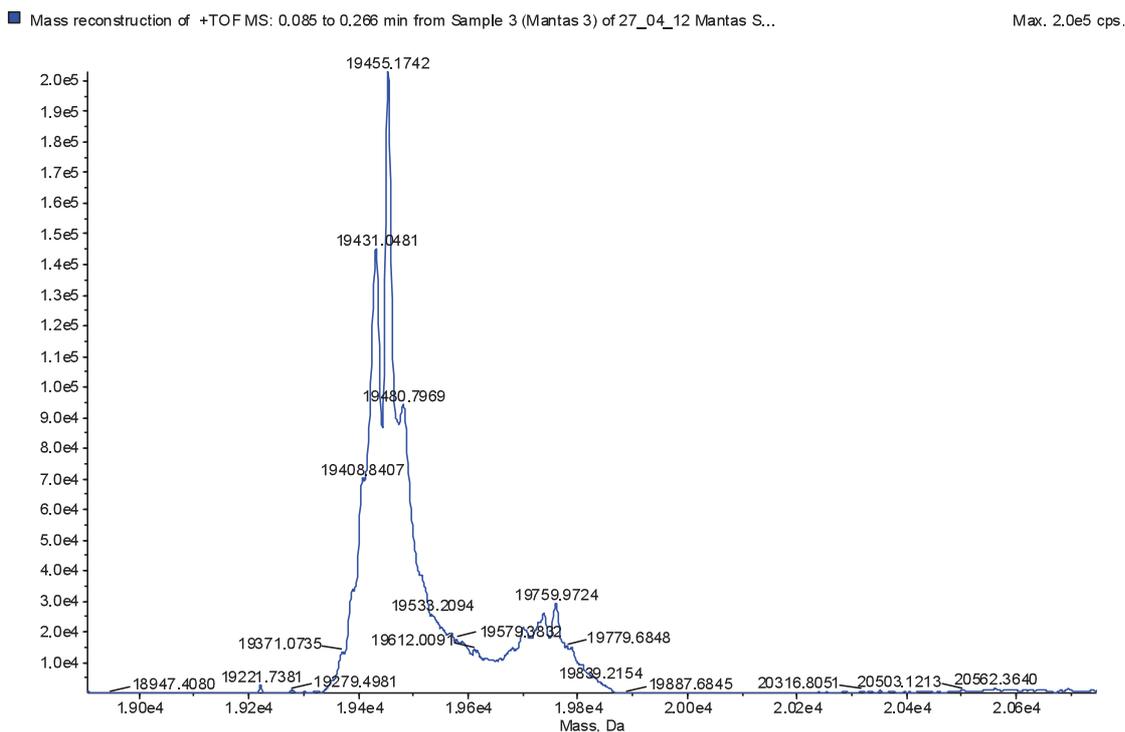


Figure S18. Deconvoluted mass spectrum of PpiB synthesised using the chloride **28**.

7. References

- [1] a) M. Brenner, W. Huber, *Helv. Chim. Acta* **1953**, *36*, 1109; b) Q. F. Liang, J. J. Liu, J. Chen, *Tetrahedron Lett.* **2011**, *52*, 3987; c) J. B. Li, Y. W. Sha, *Molecules* **2008**, *13*, 1111; d) Y. Zhou, M. Zhao, Y. Wu, C. Li, J. Wu, M. Zheng, L. Peng, S. Peng, *Bioorg. Med. Chem.* **2010**, *18*, 2165; e) L. Bi, Y. Zhang, M. Zhao, C. Wang, P. Chan, J. B. Tok, S. Peng, *Bioorg. Med. Chem.* **2005**, *13*, 5640.
- [2] a) R. E. Dawson, A. Hennig, D. P. Weimann, D. Emery, V. Ravikumar, J. Montenegro, T. Takeuchi, S. Gabutti, M. Mayor, J. Mareda, C. A. Schalley, S. Matile, *Nat. Chem.* **2010**, *2*, 533; b) A. K. H. Hirsch, E. Buhler, J. M. Lehn, *J. Am. Chem. Soc.* **2012**, *134*, 4177.
- [3] a) G. Verardo, N. Toniutti, A. Gorassini, A. G. Giumanini, *Eur. J. Org. Chem.* **1999**, 2943; b) T. Sakamoto, Y. Kikugawa, *Chem. Pharm. Bull.* **1988**, *36*, 800.
- [4] F. E. Dutton, B. H. Lee, S. S. Johnson, E. M. Coscarelli, P. H. Lee, *J. Med. Chem.* **2003**, *46*, 2057.
- [5] a) I. D'Acquarica, A. Cerreto, G. Delle Monache, F. Subrizi, A. Boffi, A. Tafi, S. Forli, B. Botta, *J. Org. Chem.* **2011**, *76*, 4396; b) K. Chandra, D. Dutta, A. K. Das, A. Basak, *Bioorg. Med. Chem.* **2010**, *18*, 8365.
- [6] Y. P. Lu, C. W. Zheng, Y. Q. Yang, G. Zhao, G. Zou, *Adv. Synth. Catal.* **2011**, *353*, 3129.
- [7] a) B. J. W. Barratt, C. J. Easton, D. J. Henry, I. H. W. Li, L. Radom, J. S. Simpson, *J. Am. Chem. Soc.* **2004**, *126*, 13306; b) T. Yajima, T. Horikawa, N. Takeda, E. Takemura, H. Hattori, Y. Shimazaki, T. Shiraiwa, *Tetrahedron: Asymmetry* **2008**, *19*, 1285; c) M. H. C. L. Dressen, B. H. P. V. van de Kruijs, J. Meuldijk, J. A. J. M. Vekemans, L. A. Hulshof, *Org. Process Res. Dev.* **2009**, *13*, 888; d) P. Dydio, C. Rubay, T. Gadzikwa, M. Lutz, J. N. H. Reek, *J. Am. Chem. Soc.* **2011**, *133*, 17176.

- [8] a) P. A. Jass, V. W. Rosso, S. Racha, N. Soundararajan, J. J. Venit, A. Rusowicz, S. Swaminathan, J. Livshitz, E. J. Delaney, *Tetrahedron* **2003**, *59*, 9019; b) J. Deblander, S. Van Aeken, J. Jacobs, N. De Kimpe, K. A. Tehrani, *Eur. J. Org. Chem.* **2009**, 4882.
- [9] G. Radau, J. Gebel, D. Rauh, *Archiv Der Pharmazie* **2003**, *336*, 372.
- [10] T. Polonski, *Tetrahedron* **1985**, *41*, 603.
- [11] E. Jungermann, J. F. Gerecht, I. J. Krems, *J. Am. Chem. Soc.* **1956**, *78*, 172.
- [12] A. Lepp, M. S. Dunn, *Biochem. Prep.* **1955**, *4*, 80.
- [13] L. Sun, C. P. Du, J. Qin, J. S. You, M. Yang, X. Q. Yu, *J. Mol. Catal. a-Chem.* **2005**, *234*, 29.
- [14] N. A. Smart, G. T. Young, M. W. Williams, *J. Chem. Soc.* **1960**, 3902.
- [15] C. J. Easton, A. J. Edwards, S. B. McNabb, M. C. Merrett, J. L. O'Connell, G. W. Simpson, J. S. Simpson, A. C. Willis, *Org. Biomol. Chem.* **2003**, *1*, 2492.

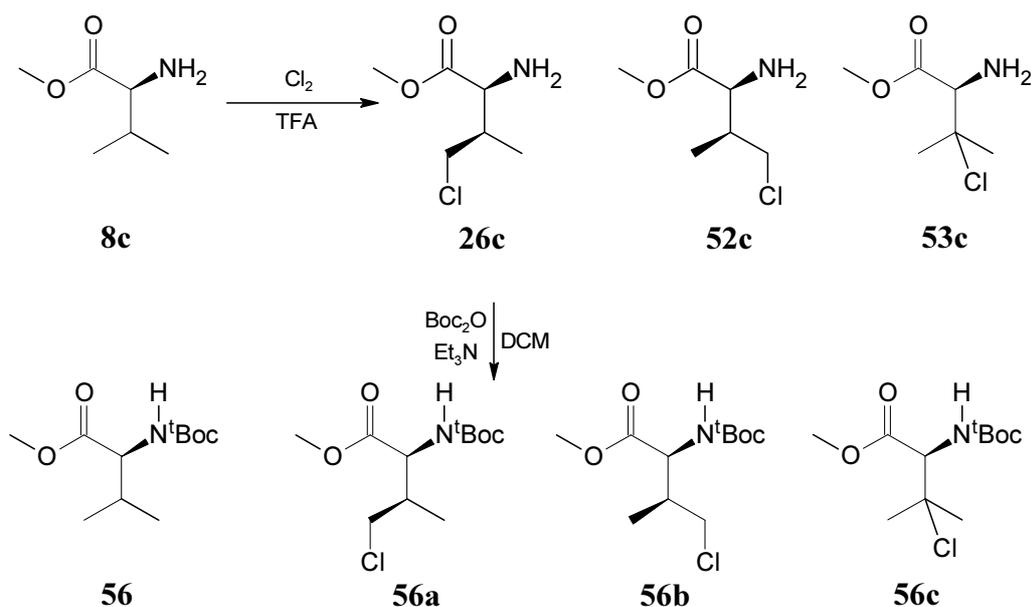
2.1. Alternate approaches to γ -chlorovaline methyl ester

The methyl ester **26c** is stable and can support protein expression and, when produced by chlorination of (*S*)-valine methyl ester **8c** without excessive chlorination, is the only product in the mixture that can replace (2*S*,3*S*)-isoleucine **7** (Scheme 2.1). As such, unfractionated mixture was successfully used to establish the incorporation of the isoleucine analogue **26a** (Figure 2.15). However, due to limited solubility of the esters in aqueous solutions, use of the mixture for protein production is problematic as the compounds interfere with the synthetic machinery when supplied in larger amounts. As the established purification method for the target **26c** (*vide supra*) is not suitable for preparative scale, a new synthetic route was sought.

Total synthesis of enantiomerically pure amino acid **26a** was reported^[124] and the method could be modified to produce the ester **26c**; however, the procedure requires 11 sequential synthetic steps. Direct chlorination of valine **8a** can lead to the target in a single step, thus development of a new purification method seemed more promising.

To increase the retention of the target compound on the reverse phase, the mixture of chlorination products obtained from chlorination of valine methyl ester **8c** was protected with the ^tBoc group (Scheme 2.3). The acid labile ^tBoc protecting group was chosen as the target compound **26c** is stable in strong acid.

Reverse phase HPLC was used for fractionation of the mixture. The γ -chlorinated diastereomers **56a** and **56b** separated from all other components in the mixture, but were not separable (Figure 2.22).



Scheme 2.3. Preparation of the protected analogues of isoleucine.

The diastereomers **56a,b** with flexible chains appeared inseparable, hence separation of conformationally restricted lactones **57a** and **57b** was considered, followed by reinstallation of the desired functionality afterwards (Scheme 2.4).

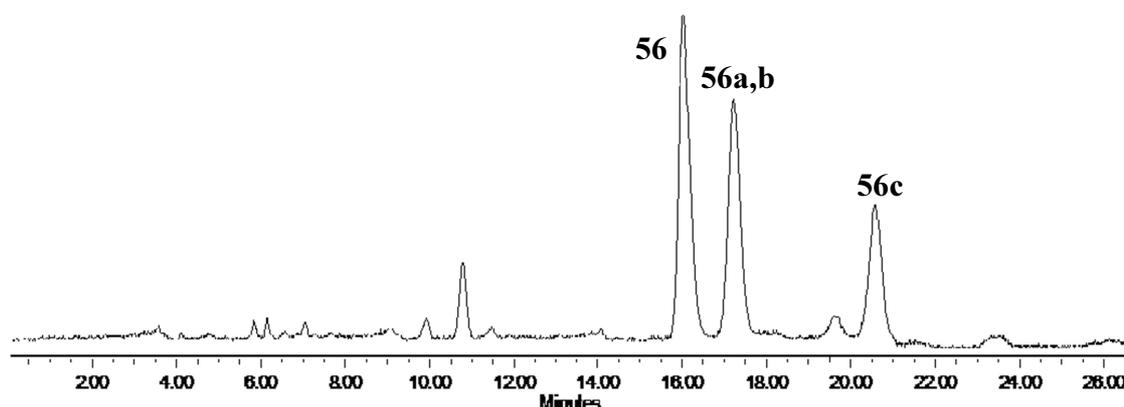
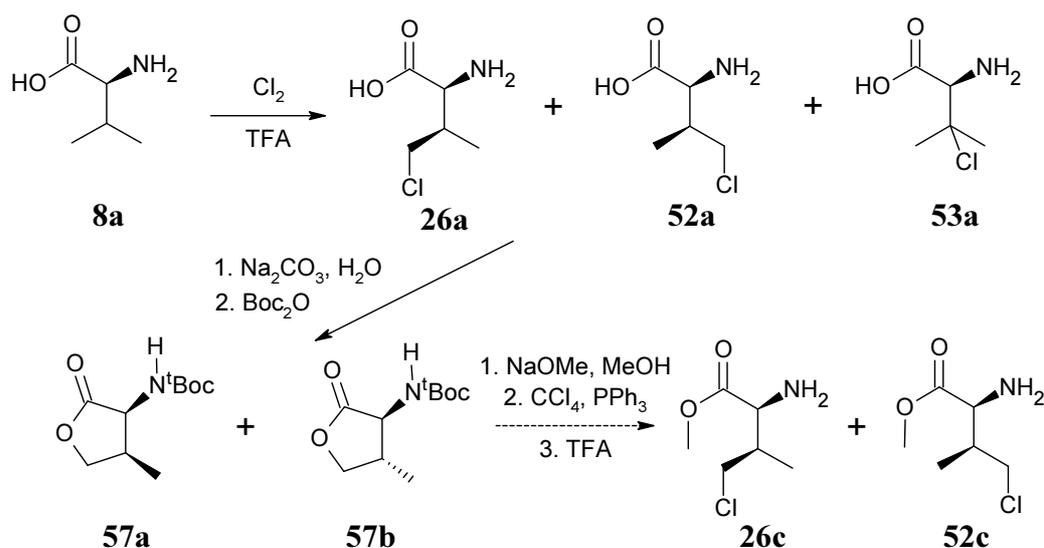


Figure 2.22. HPLC chromatogram of the separation of the chlorides **56a,b** on an Alltima C18 column (22×250 mm, 5 μ), eluting with 60% methanol, 10 mL/min.



Scheme 2.4. The proposed route for the large scale preparation of pure protected analogue **26c**.

After chlorination of valine **8a**, the mixture was dissolved in water and pH was neutralised to promote lactonisation of the chlorinated species **26a** and **52a**. The amino groups were subsequently protected and the mixture was analysed by HPLC. As expected, the diastereomeric lactones **57a** and **57b** eluted as separate fractions (Figure 2.23). Stereochemistry was established by ^1H NMR by comparison with the spectra of *N*-acetylated lactones **54a** and **54b** used in the previous study (*vide supra*) as well as the published spectra of the lactones **57a** and **57b**.^[125]

The lactones **57a** and **57b** are useful compounds for chemical synthesis^[125] and could be elaborated to the chlorides **26c** and **52c** on a large scale, as proposed in Scheme 2.4. However, this route was deemed too long for the preparation of analogues for protein expression and was not attempted.

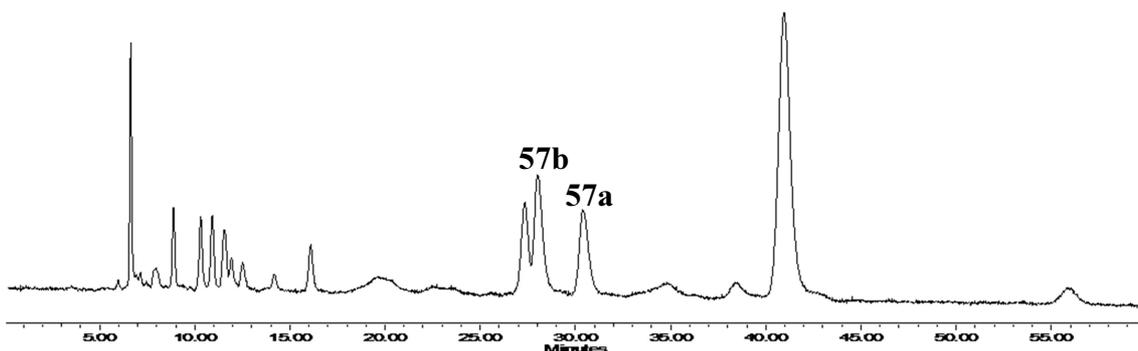


Figure 2.23. HPLC chromatogram showing the separation of the lactones **57a** and **57b** on an Alltima C18 column (22×250 mm, 5 μ), eluting with 40% methanol, 12 mL/min.

Instead, the mixture of the diastereomers **56a,b** isolated earlier (Figure 2.22) was deprotected with TFA and the resultant mixture of the chlorides **26c** and **52c** was used for all subsequent experiments. The two esters were completely soluble in aqueous mixtures at concentrations needed for effective protein expression, and the diastereomer **52c** was not recognised by the protein synthetic machinery.

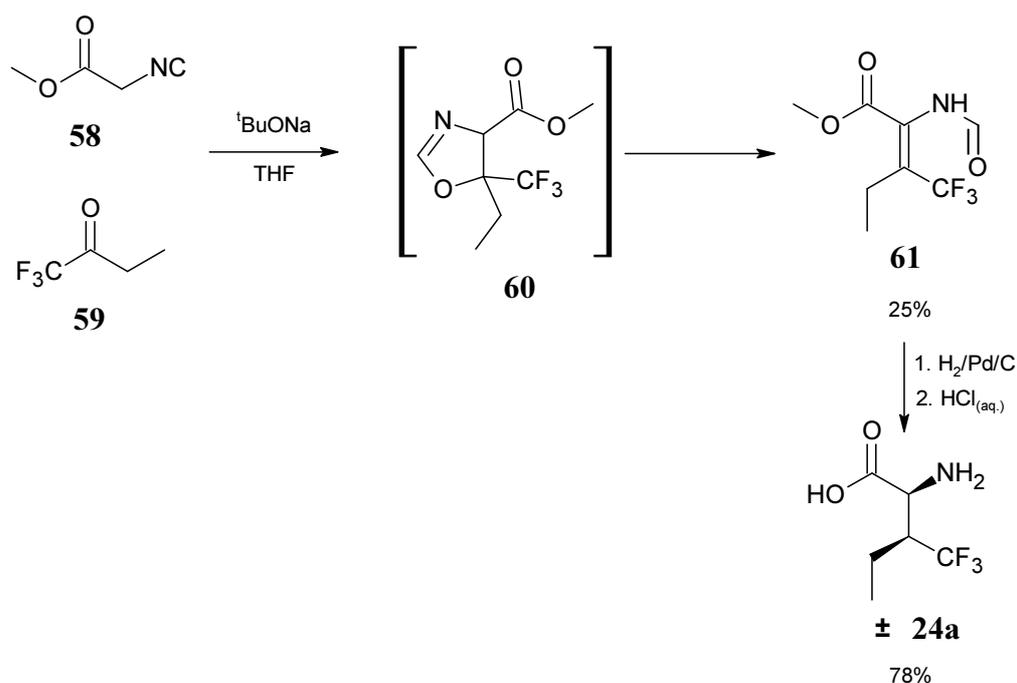
2.2. Incorporation of 3',3',3' – trifluoroisoleucine

As mentioned in the Introduction, 3',3',3'-trifluoroisoleucine **24a** was not a replacement for isoleucine **7** during protein expression and ATP consumption assays with purified IRS showed no activity, suggesting the analogue **24a** was rejected by the activating domain of the enzyme.^[91] The fluoride **24a** was reinvestigated alongside the chloride **26a**.

A literature procedure^[126] was employed for the preparation of the analogue **24a** (Scheme 2.5). Methyl isocyanoacetate **58** was coupled with 1,1,1-trifluorobutanone **59**

in the presence of sodium *tert*-butoxide; the strong base formed the alkene **61** via the oxazoline intermediate **60** with 25% yield in a single step. Technical grade methyl isocyanoacetate **58** was used for the synthesis, which accounts for the lower than expected yield.

The alkene intermediate **61** was then hydrogenated over Pd/C and deprotected with aqueous hydrochloric acid. As exclusively the *Z*-olefin was formed during the coupling, the amino acid **24a** was obtained as a single diastereomer.



Scheme 2.5. The synthesis of trifluoroisoleucine **24a**.

Only background levels of the test protein His₆PpiB were produced when the fluoride **24a** was supplied in place of isoleucine **7** during cell-free expression (Figure 2.24). MS

analysis of the produced protein showed trace incorporation of the fluoride **24a**, resulting in partially fluorinated His₆PpiB (Figure 2.25).

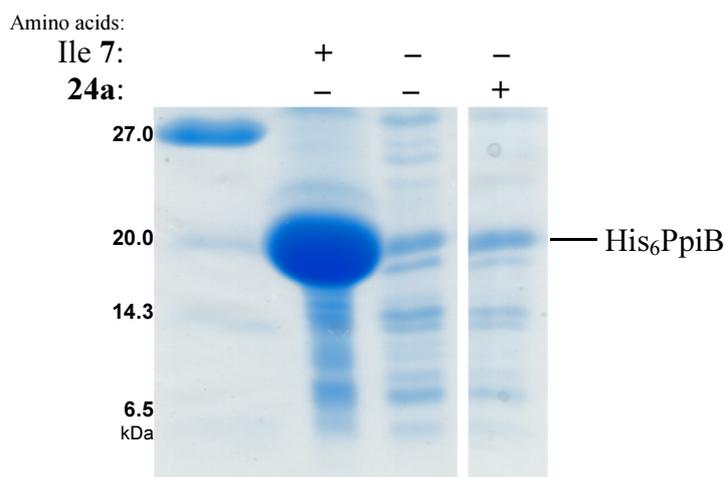


Figure 2.24. SDS-PAGE analysis of His₆PpiB produced *via* cell-free expression with the fluoride **24a** instead of (2*S*,3*S*)-isoleucine 7.

These findings confirm the analogue **24a** is for the most part rejected by IRS. While trace amounts of the amino acid were incorporated into protein, the error rates are far too low to support protein synthesis.

These observations are consistent with the proposed size exclusion mechanism. While a single hydrogen-to-fluorine substitution is considered inconsequential, the entire methyl group is replaced by a trifluoromethyl group in the analogue **24a**, increasing the steric bulk of the residue. The cumulative effect of the triple substitution on the shorter arm of isoleucine is likely registered by the synthetase as discrimination of size on the shorter branch of the substrate is critical for the discrimination against (2*S*,3*R*)-*allo*-isoleucine

62, a diastereomer of isoleucine that is not found in proteins, but none the less is present in the cell (Figure 2.26).^[127]

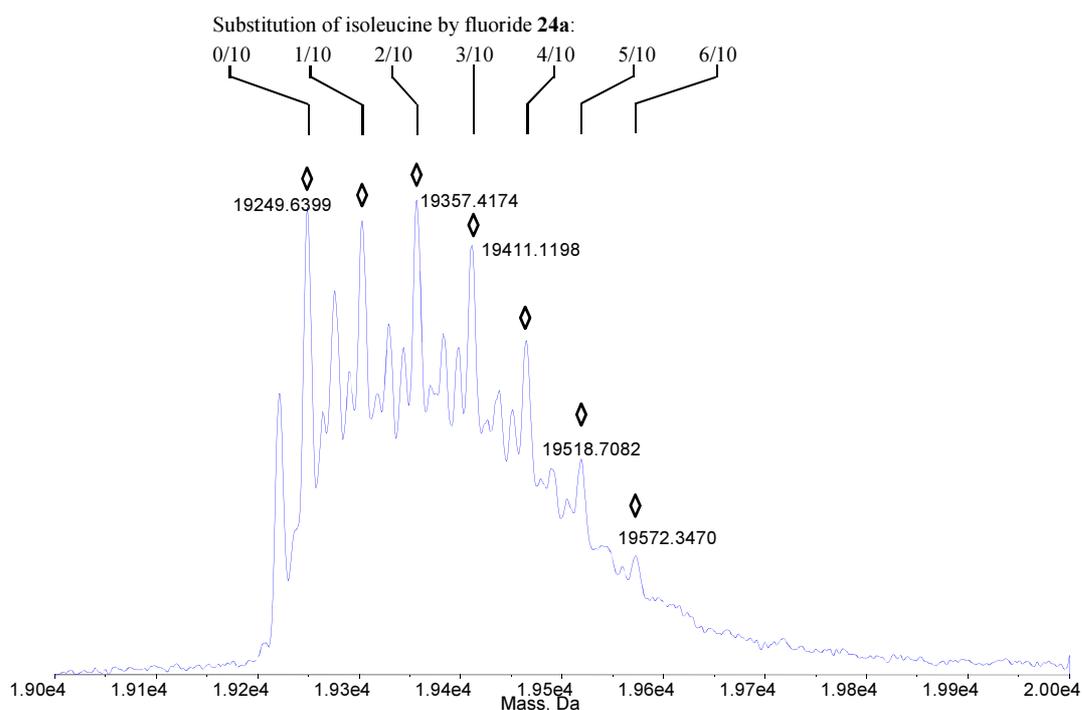


Figure 2.25. Deconvoluted ESI-MS spectrum of the trace amounts of His₆PpiB produced in the presence of the analogue **24a**. The mass increase of 54 Da between the protein species indicates a higher level of substitution with the fluoride **24a**.

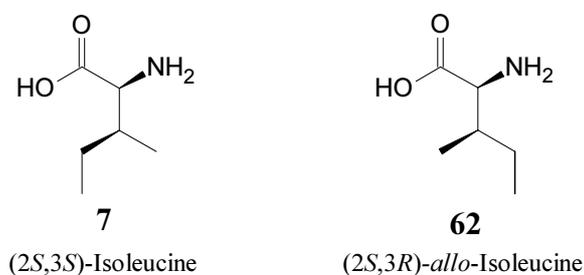
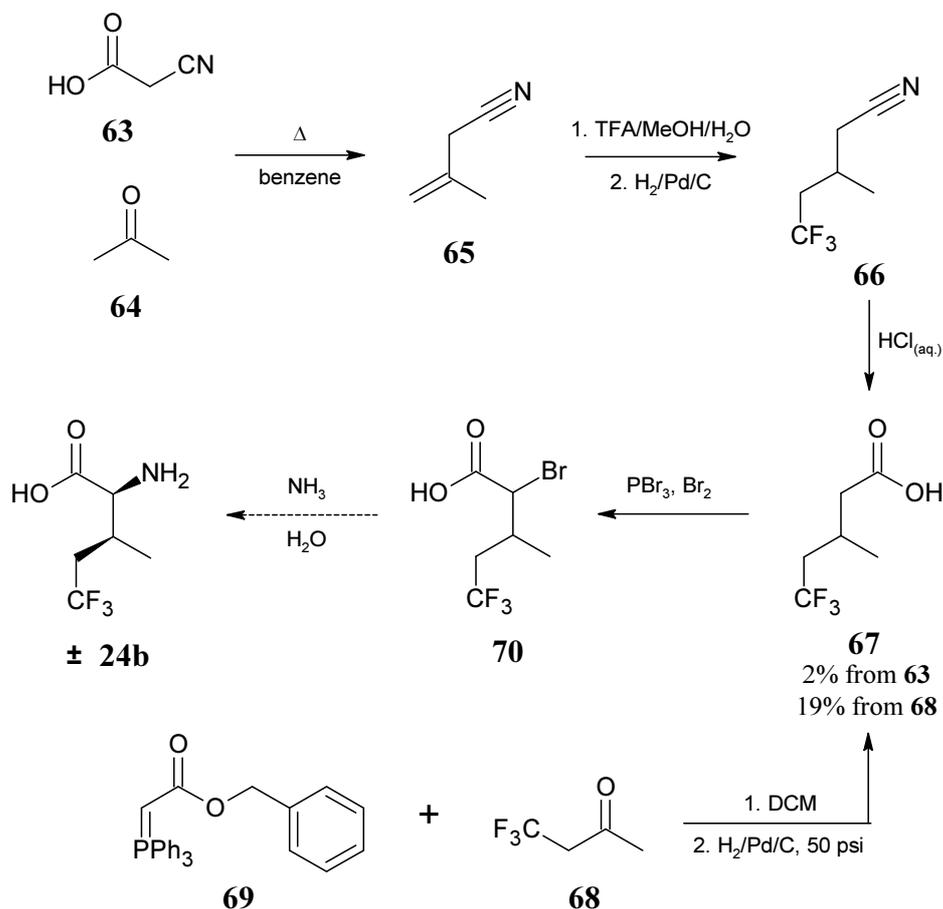


Figure 2.26. The diastereomers of (2*S*)-isoleucine.

2.3. Synthesis of 5,5,5 – trifluoroisoleucine



Scheme 2.6. The attempted synthesis of 5,5,5-trifluoroisoleucine **24b**.

For direct comparison of the trifluorinated analogues of isoleucine, the regioisomer 5,5,5-trifluoroisoleucine **24b** was also pursued (Scheme 2.6). Following a literature procedure,^[128,129] cyanoacetic acid **63** and acetone **64** were condensed under Dean-Stark conditions. The condensate was then distilled to produce 3-methylbut-3-enitrile **65**.

During the distillation, the olefin **65** was prone to isomerise to the more favourable 3-methylbut-2-enitrile, and the distillate contained a mixture of the two isomers, but was used without further purification.

The mixture was electrolysed in the presence of trifluoroacetic acid. Eight major triplets ($^3J_{\text{FH}} = 11 \text{ Hz}$) were seen in the ^{19}F NMR spectrum of the electrolysate, indicating the presence of (2,2,2-trifluoroethyl)- groups, thus confirming the alkene **65** was trifluoromethylated, but also suggesting the intermediate radical species may have partially undergone subsequent polymerisation, thus consuming part of the product.

The electrolysed mixture was distilled under vacuum to remove any polymeric species and then hydrogenated over Pd/C. This significantly reduced the complexity of the mixture (down to four major triplets in the ^{19}F NMR spectrum). Isolation of 5,5,5-trifluoro-3-methylpentannitrile **66** was attempted by distillation. Due to the nature of the mixture, however, complete purity of the nitrile **66** was unattainable; Figure 2.27 displays a ^1H NMR spectrum of the fluoride **66** obtained after repeated distillation, including through a Vigreux column. Nevertheless, analytical purity was not required at this stage of the synthesis, and the crude material was carried over to the next step.

While the intermediate **66** was not isolated as a single species, greater purity was obtained than had been in previous attempts,^[91] thus enabling unambiguous assignment of the signals in the ^1H NMR spectrum; previously it had been impossible to even confirm the δ 1.19 ppm peak belongs to the compound **66**, with the δ 1.02 ppm signal potentially assigned.

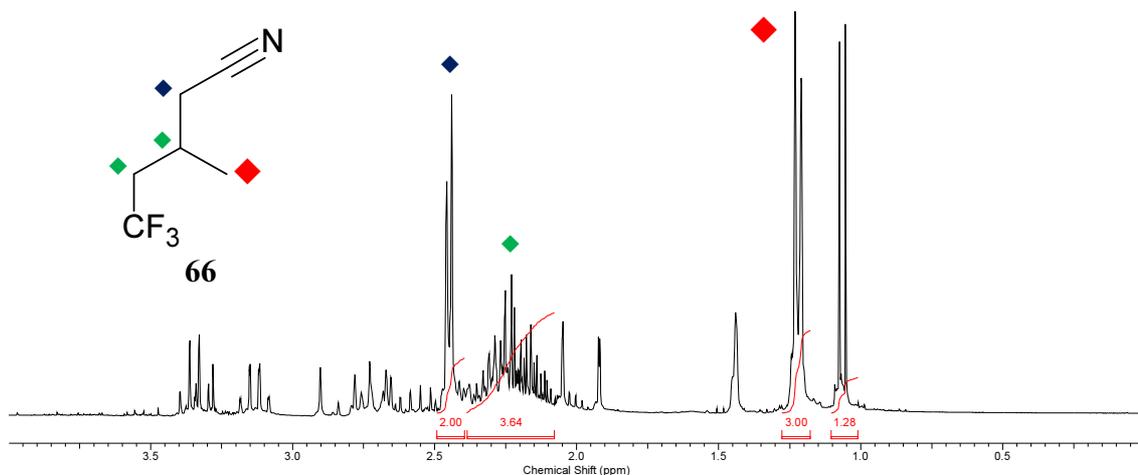


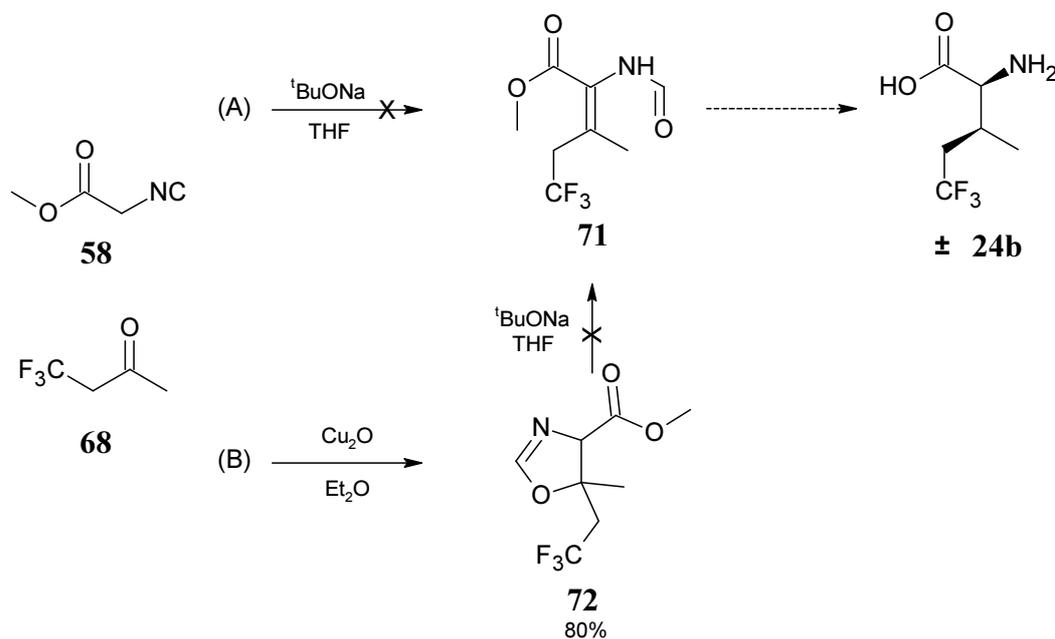
Figure 2.27. ¹H NMR spectrum of partly purified 5,5,5-trifluoro-3-methylpentannitrile **66**.

5,5,5-Trifluoro-3-methylpentannitrile **66** was treated with aqueous hydrochloric acid to hydrolyse the cyano group. Upon neutralisation of the hydrolysis solution the product 5,5,5-trifluoro-3-methylpentanoic acid **67** dissolved in the aqueous solution, whereas non-hydrolysable material (likely partly polymerised hydrocarbons) remained in a separate layer. After re-acidification of the aqueous layer, 5,5,5-trifluoro-3-methylpentanoic acid **67** separated into a separate organic layer and after redistillation was obtained as a colourless foul-smelling liquid.

5,5,5-trifluoro-3-methylpentanoic acid **67** was isolated in only 2% yield, based on cyanoacetic acid **63**. The low yield is primarily attributable to the observed polymerisation during the electrolysis, largely facilitated by the lack of proper apparatus and electrodes for efficient current control, as well as loss of material during the attempted purification of 5,5,5-trifluoro-3-methylpentannitrile **66**. A more conventional synthesis of the acid **67** was then considered. A Wittig reaction between 4,4,4-trifluorobutanone **68** and benzyl (triphenylphosphoranylidene)acetate **69** followed

by hydrogenation gave **67**. However, only 19% yield was achieved. Due to the limited availability of the materials scaling up of this reaction was problematic and the low yields made it difficult to attain large amounts of the product.

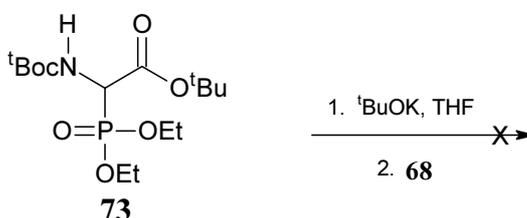
Hell-Volhard-Zelinsky reaction with the acid **67**, followed by treatment with aqueous ammonia of the resultant bromide **70** was used for the preparation of 5,5,5-trifluoroisoleucine **24b**. The amino acid was detected in the reaction mixture by MS, but the material was lost during purification attempts. Low yields throughout the synthesis partly complicated the purification efforts and resulted in loss of material. It was therefore decided to develop a new route to the analogue **24b**.



Scheme 2.7. The attempted synthesis of 5,5,5-trifluoroisoleucine **24b** via the Schöllkopf oxazoline **72**.

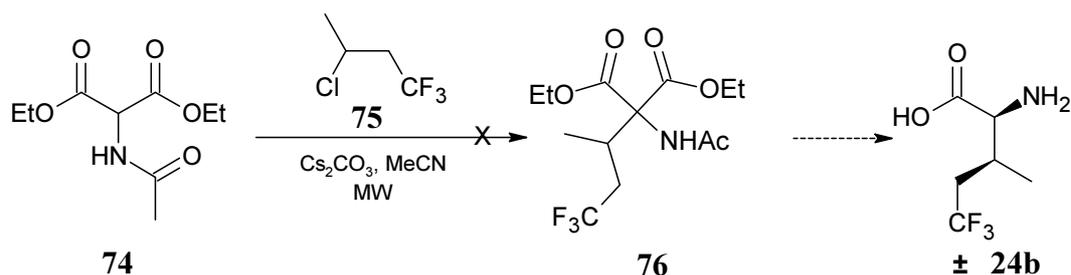
Based on the synthesis of the fluoride **24a** *via* the Schöllkopf oxazoline **60** (Scheme 2.5), an analogous route to the fluoride **24b** *via* the alkene **71** was considered (Scheme 2.7). A one-pot procedure to generate the intermediate **71** from methyl isocyanoacetate **58** and 4,4,4-trifluorobutanone **68** was attempted (Scheme 2.7, route A), as was done with the fluoride **61**, but the target was not detected. The intermediate oxazoline **72** was then synthesised using gentler conditions^[130] (Scheme 2.7, route B). Base treatment, however, led to complete decomposition of the substrate **72**.

The difficulty of handling 4,4,4-trifluorobutanone **68** is a consequence of the β -trifluoromethyl-carbonyl arrangement of the molecule, making it extremely susceptible to enolisation and/or HF-elimination. This susceptibility persists in derivatives of the ketone **68**, thus explaining the decomposition of the oxazolinone **72** under basic conditions. As the butanone **68** could participate in a Wittig reaction (Scheme 2.6), it was attempted to pre-install the amino acid functionality on the phosphonate prior to the coupling reaction. Unfortunately, the Wittig-Horner reaction with the phosphonate **73** was not successful (Scheme 2.8).



Scheme 2.8. Wittig-Horner reaction with 4,4,4-trifluorobutanone **68**.

Finally, the synthesis of 5,5,5-trifluoroisoleucine **24b** was attempted through microwave-assisted alkylation of diethyl acetamidomalonate **74** with 3-chloro-1,1,1-trifluorobutane **75** (Scheme 2.9).^[131] Unfortunately, only starting material **74** was recovered; the chloride **75** was likely consumed by elimination. Elimination is known to compete with substitution when saturated alkyl halides are used for alkylation, but at least small amounts of the alkylation product are generally achieved.^[131] The trifluoromethyl group, however, appears to overwhelmingly promote elimination by increasing the acidity of the neighbouring methylene group, resulting in rapid consumption of the substrate **75**.

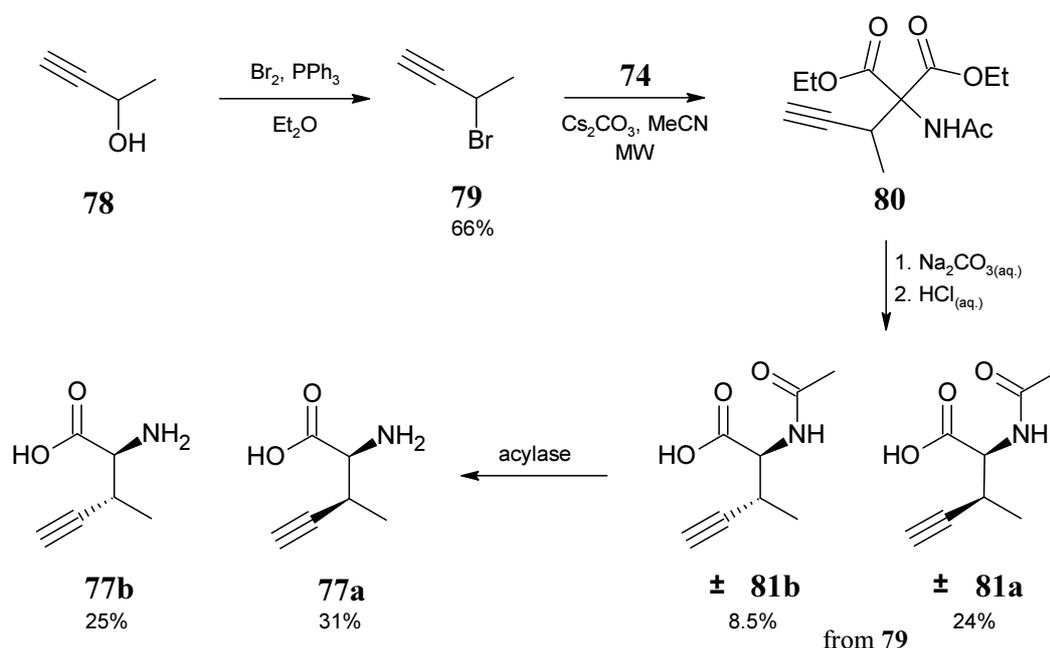


Scheme 2.9. Attempted alkylation of diethyl acetamidomalonate **74**.

After the unsuccessful attempts to synthesise the isoleucine analogue **24b** it was apparent that the pathway *via* the acid **67** was the most viable. At this stage, however, the discrepancy regarding the incorporation of unnatural analogues was resolved; hence the synthesis of fluoride **24b** was abandoned.

2.4. Incorporation of amino acids for ‘click’ chemistry

The permissiveness of AARSs can be utilised for incorporation of amino acids for residue-specific modification. In particular, the alkyne-bearing analogue of isoleucine **77a** could enable rapid labelling of protein *via* azide-alkyne cycloaddition. Protein with the alkyne **77a** had been prepared in the past, but live cell protein expression was used and the amino acid was supplied to the medium as a racemic mixture.^[89] In the current study the diastereomer precursors **81a** and **81b** were separated and enzymatic deacylation was used to obtain the target amino acid **77a** as a single enantiomer (Scheme 2.10).



Scheme 2.10. The synthesis of 2-amino-3-methylpent-4-ynoic acids **77a** and **77b**.

But-3-yn-2-ol **78** was brominated by triphenylphosphite dibromide generated from triphenylphosphite and bromine.^[132] The alcohol **78** in ether was added to the solid triphenylphosphite dibromide and the heterogeneous reaction was allowed to proceed for 3 h. Afterwards the reaction mixture was vacuum distilled and 3-bromobut-1-yne **79** was collected as a colourless fruity-smelling liquid in 66% yield.

During alkylation of diethyl acetamidomalonate **74**,^[133] 3-bromobutyne **79** underwent both direct S_N2 substitution to give the desired product **80**, and S_N2' substitution to produce some undesired by-product. Regardless, both of the products were decarboxylated, and the mixture was then purified by reverse phase HPLC.

Upon decarboxylation of the malonate **80** the desired (2*S*,3*S*)-(±)-enantiomeric pair **81a** formed preferentially. After deprotection with acylase the isoleucine analogue **77a** was used in the production of His₆PpiB (Figure 2.28).

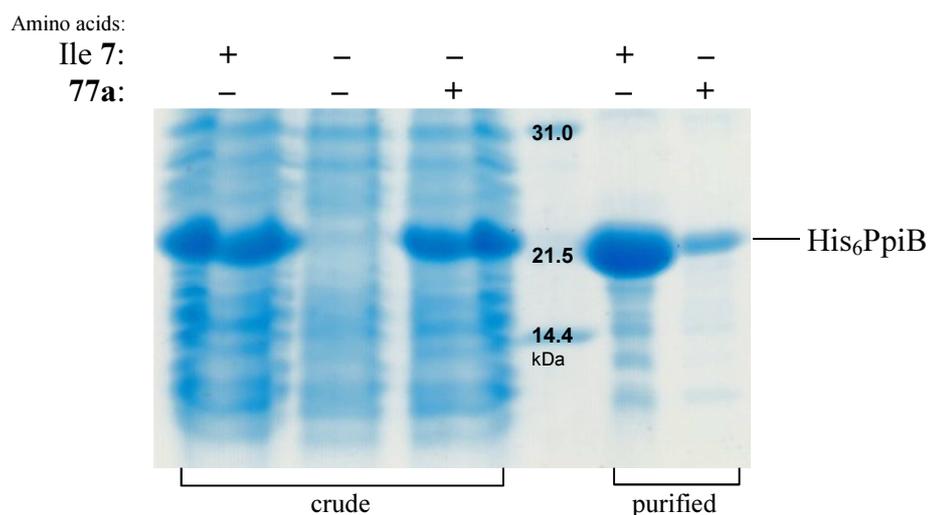


Figure 2.28. SDS-PAGE analysis of His₆PpiB produced *via* cell-free expression with the alkene **77a** instead of (2*S*,3*S*)-isoleucine **7**.

As expected, large amounts of protein were produced when (2*S*,3*S*)-isoleucine **7** was replaced with the alkyne **77a**, confirming the analogue is a suitable replacement for isoleucine.^[89] Surprisingly, though, nickel affinity purification of substituted His₆PpiB failed, as most of the protein did not bind the metal resin and was collected in the flow-through. Such failure of hexahistidine-tagged protein to bind Ni(II)-NTA resin has not been reported. It was hypothesised that the alkyne, due to the affinity to transition metals, could in some way interfere with nickel-histidine interaction. In the previous expression^[89] of dihydroxyfolate reductase (DHFR) with the analogue **77a** the product was successfully isolated by affinity to immobilised nickel, although only *ca.* 70% substitution of isoleucine was achieved. Nevertheless, the mechanism for reduction of binding affinity is difficult to propose, and addition of butynol **78** to wash buffers did not obstruct the purification of unsubstituted His₆PpiB.

As a means to counter the poor binding, the hexahistidine-tag on the protein was extended to a dodecamer. The stronger binding of the His₁₂PpiB enabled the use of more stringent conditions during purification, resulting in cleaner product. Even with this modification, however, the alkyne **77a**-bearing protein had poor affinity to the resin (Figure 2.29) and required several cycles of purification to isolate all of the product for analysis.

The incorporation levels of the analogue **77a** into His₁₂PpiB were analysed by 2D NMR. ¹³C-HSQC spectra were recorded in 6 M urea-D₆ in D₂O to ensure the protein is unfolded (Figure 2.30). Signals corresponding to the methyl groups of isoleucine **7** residue had reduced to background levels in the sample containing modified His₁₂PpiB and strong peaks due to the analogue **77a** were observed.

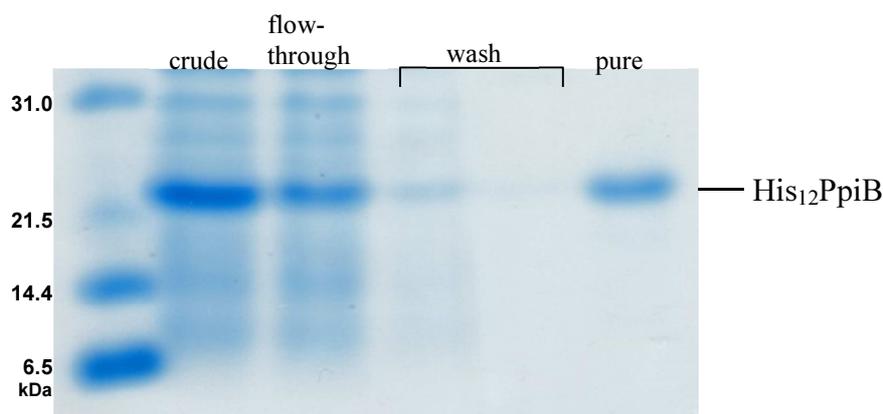


Figure 2.29. SDS-PAGE analysis showing the poor binding of His₁₂PpiB produced with the alkyne **77a** to Ni(II)-NTA resin during protein purification.

Quantification of the substitution of isoleucine **7** based on the NMR signal strength was attempted. Peak intensities in 2D NMR spectra depend heavily on the nature of the species, making quantification unreliable, but signals corresponding to chemically similar groups are expected to have similar intensities.^[134] For that reason signals corresponding to the methyl groups of leucine **18**, isoleucine **7** and valine **8a** were integrated and divided by the number of protons they represent (based on protein sequence). Similar intensities per proton were obtained for valine **8a** and leucine **18**, whereas signals for isoleucine **7** were smaller by about 20%, thus signifying a large error margin associated with the employed method (Figure 2.31). Relative to leucine **18**, the signals for isoleucine in protein expressed with the alkyne **77a** were reduced by more than 98%, indicating that only background levels of isoleucine **7** were incorporated into the modified protein. However, the intensity of the signal attributed to the methyl group of the analogue **77a** was weaker than the signals of the excluded isoleucine. Thus, while the NMR analysis showed the alkene **77a** was very effectively

incorporated into protein in place of isoleucine 7, due to the error associated with the method exact level of substitution cannot be established.

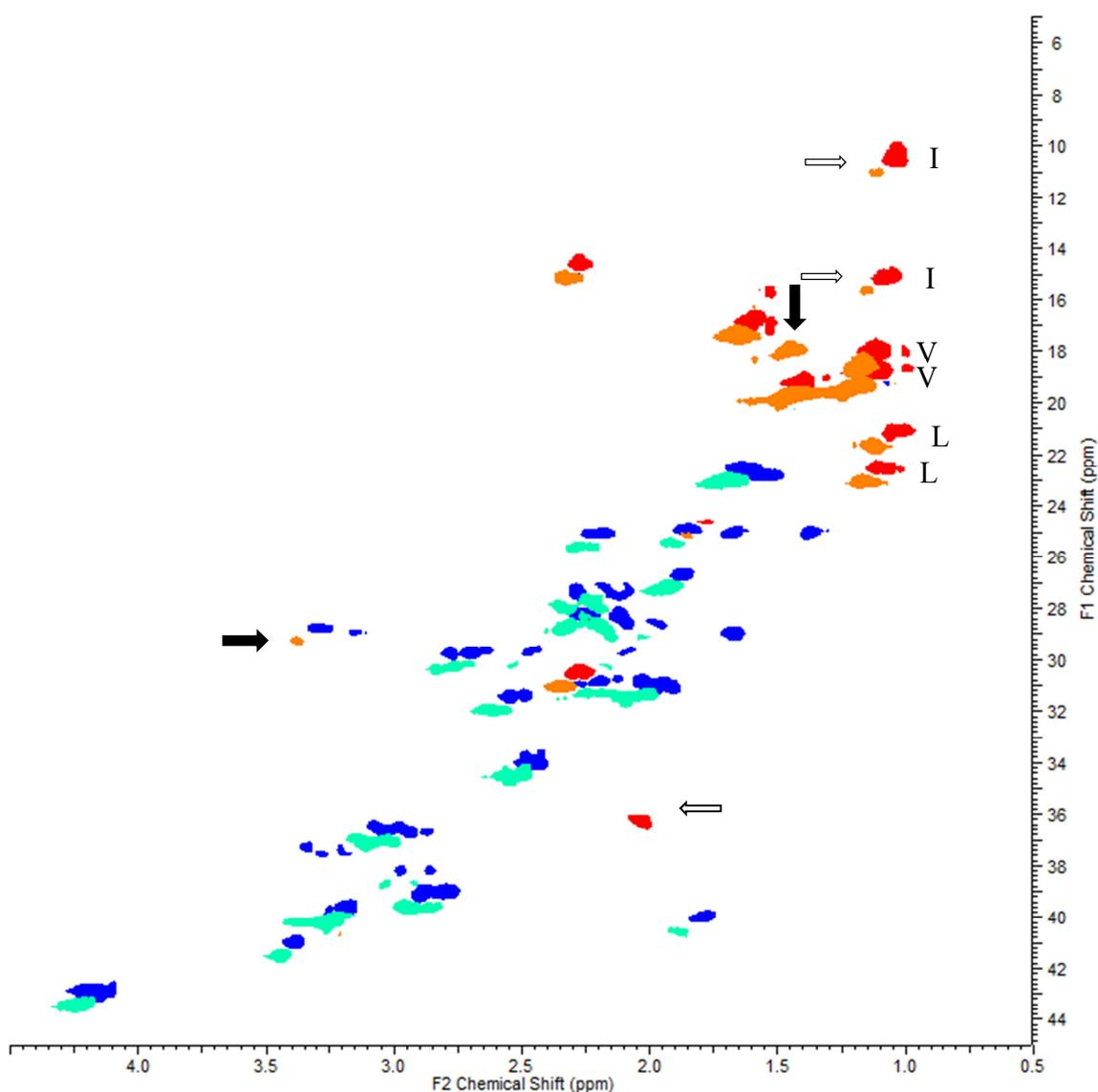


Figure 2.30. Detection of **77a** in His₁₂PpiB by ¹³C-HSQC. The spectrum of modified protein (orange and light blue) is off-set from the control (red and blue). Signals attributable to **77a** are clearly visible in the modified protein (solid arrows), whereas those of isoleucine 7 are significantly weaker (hollow arrows). The signals of the methyl groups of BCAA are assigned.

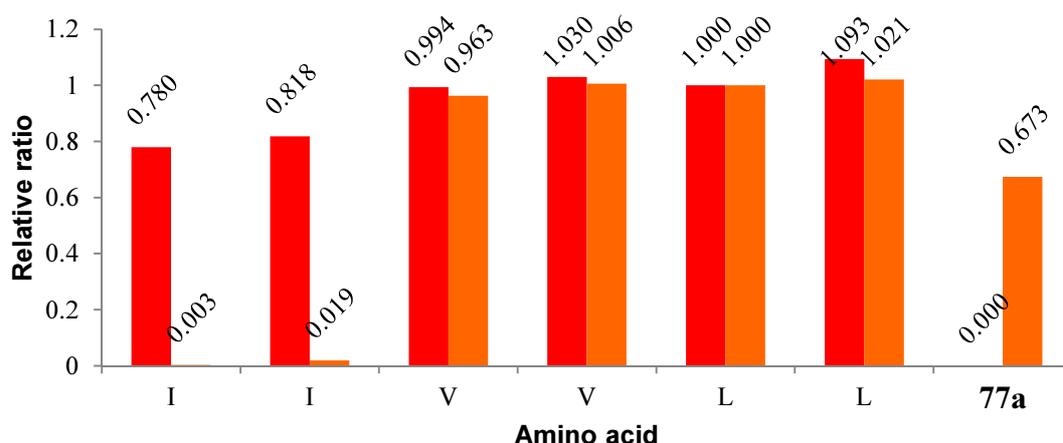


Figure 2.31. Normalised ^{13}C -HSQC signal strength (per proton) of the methyl groups of BCAA in spectra of natural His₁₂PpiB (■) and modified with 77a (■). The signals of isoleucine 7 are reduced by >95% in the spectrum of the modified protein.

2.5. Conclusions

In the present study the chlorinated analogue of leucine **26a** was shown to be unstable under physiological conditions and the instability was demonstrated to be the underlying reason for failure to support protein synthesis. The methyl ester **26c** that is deprotected *in situ* at a rate that matches the rate of protein expression enables the incorporation of the unstable chloride **26a**, thus confirming that a chlorine atom is a suitable replacement for a methyl group. A practical route for the preparation of the ester **26c** was established.

In contrast, the rejection of trifluorinated analogue 3',3',3'-trifluoroisoleucine **24a** during protein synthesis was confirmed. The miniscule levels of incorporation of the analogue are consistent with the proposed size exclusion mechanism, with the trifluoromethyl group being on the boundary of what is accepted into the active site of

IRS. It is theorised that the increase in size due to the triple hydrogen-to-fluorine replacement is the reason for rejection.

An alkyne analogue of isoleucine **77a** was prepared as a single enantiomer and was incorporated into protein. The incorporation was confirmed by 2D ^{13}C -HSQC.

Chapter 3. Cleavage of Peptide Bonds



As described in the previous chapter, the isostere of isoleucine γ -chlorovaline **26a** was prone to rapid decomposition to a 5-membered lactone under physiological conditions, making direct incorporation into proteins through cell-free expression impossible. The problem was circumvented with the use of the methyl ester **26c** that was deprotected *in situ*, thus providing a constant supply of the analogue **26a** and allowing the production of modified protein. Subsequently, it was observed that the propensity of the chloride **26a** to lactonise was retained after incorporation into protein, leading to the cleavage of the proximal peptide bond, following a mechanism reminiscent of cyanogen bromide-mediated methionine-specific proteolysis, but required elevated temperatures to proceed; chlorinated protein was stable at or below physiological temperature, but began to fragment at 60°C and decomposed within minutes at 100°C. Both diastereomers of the analogue of leucine 4-chloronorvaline **25** were found to impart the same properties onto modified proteins, signifying heat-induced mediation of peptide bond cleavage is a common property of γ -chlorinated amino acids, driven by the formation of highly favourable 5-membered rings. The mechanism of fragmentation was confirmed by ESI-MS analysis of peptide fragments generated after heat treatment of chlorinated test protein His₆PpiB. In contrast, β -chlorinated analogues of valine **15a** and **15b** and the previously uninvestigated

isostere of isoleucine (2*R*,3*S*)-(±)-3-chloronorvaline **27** led to proteins that were not fragmented in response to heat, consistently with the properties of free amino acids.

The latent leucine- and isoleucine-specific heat-controlled peptide bond cleavage was developed as an alternative method to the more traditional biological or chemical proteolysis for the production of small peptides, expressed as parts of larger fusion proteins. The validity of the method was demonstrated through the preparation of gastrin releasing peptide prohormone (GRPG), utilising the isostere of isoleucine **26c** to exact the bond cleavage, and cholecystokinin prohormone (CCKG), using the analogue of leucine **25** for the release of the peptide, as well as modified analogues of the two peptides bearing perdeuterated or fluorinated residues. Preparation of prolactin releasing peptide prohormone (PRPG) *via* this route was not successful, however, due to the N-terminal serine of PRPG interfering with the bond hydrolysis reaction.

Concurrently, dehydro isosteres of leucine and isoleucine were shown to induce peptide bond cleavage upon exposure to iodine at 0°C following an analogous mechanism, as demonstrated with allylglycine^[52,53] (Scheme 1.6). The utility of the iodination-sensitive amino acids was demonstrated through the preparation of GRPG and PRPG, following an analogous approach.

The two new methods of bond cleavage for peptide production were contrasted with traditional enzymatic protein digestion, exemplified by preparation of calcitonin prohormone CTG and six modified analogues through digestion of the corresponding fusion proteins with enterokinase. The latent chemical cleavage methods had distinct advantages over the enzymatic approach, like shorter reaction times, the ability to digest insoluble material and the lack of need for expensive reagents.

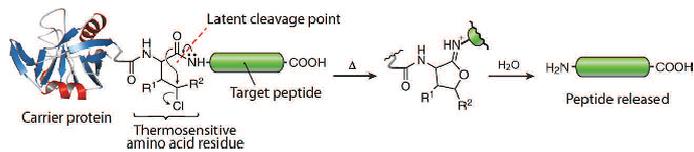
In the work described, an application of functionalised ‘activated’ amino acids introduced into proteins in place of aliphatic amino acids with non-functionalised side-chains was demonstrated. All possible singly-chlorinated isosteres of the BCAA family were incorporated into proteins through cell-free protein expression. The γ -chlorides were shown to be latent peptide bond cleavage triggers, and the resulting protein fragmentation was utilised in the production of small peptides, whereas the β -chlorides led to stable proteins, potentially enabling orthogonal chemistry to be carried out on the two classes of the chlorinated amino acids. This work has been published in *ChemBioChem* and the following pages contain the published manuscript.

COMMUNICATIONS

M. Liutkus, S. A. Fraser, K. Caron,
D. J. Stigers, C. J. Easton*



 **Peptide Synthesis through Cell-Free
Expression of Fusion Proteins
Incorporating Modified Amino Acids
as Latent Cleavage Sites for Peptide
Release**



Exploiting promiscuity: Native peptide biosynthetic machinery permits misincorporation of chlorinated and dehydro analogues of Leu and Ile, for use as latent peptide cleavage sites upon heat-

ing and treatment with iodine, respectively. The utility of the methods is illustrated by the synthesis of a range of natural and modified peptides, after expression as fusion proteins.

Peptide Synthesis through Cell-Free Expression of Fusion Proteins Incorporating Modified Amino Acids as Latent Cleavage Sites for Peptide Release

Mantas Liutkus, Samuel A. Fraser, Karine Caron, Dannon J. Stigers, and Christopher J. Easton*^[a]

Chlorinated analogues of Leu and Ile are incorporated during cell-free expression of peptides fused to protein, by exploiting the promiscuity of the natural biosynthetic machinery. They then act as sites for clean and efficient release of the peptides simply by brief heat treatment. Dehydro analogues of Leu and Ile are similarly incorporated as latent sites for peptide release through treatment with iodine under cold conditions. These protocols complement enzyme-catalyzed methods and have been used to prepare calcitonin, gastrin-releasing peptide, cholecystokinin-7, and prolactin-releasing peptide prohormones, as well as analogues substituted with unusual amino acids, thus illustrating their practical utility as alternatives to more traditional chemical peptide synthesis.

Table 1. Peptides produced in this study.

Peptide	Sequence	
1 a	CTG	CGNLSTCMLGTYTQDFNKFHTFPQTAIGVGAPG
1 b	5 ₂ -CTG	CGNLSTCMLGTYTQD5* ⁵ NK5* ⁵ HT5* ⁵ PQTAIGVGAPG
1 c	6 ₂ -CTG	CGNLSTCMLGTYTQD6* ⁶ NK6* ⁶ HT6* ⁶ PQTAIGVGAPG
1 d	7 ₂ -CTG	CGNLSTCMLGTYTQD7* ⁷ NK7* ⁷ HT7* ⁷ PQTAIGVGAPG
1 e	8 ₂ -CTG	CGNLSTCMLGTYTQD8* ⁸ NK8* ⁸ HT8* ⁸ PQTAIGVGAPG
1 f	9-CTG	CGNLSTCMLG19* ⁹ TQDFNKFHTFPQTAIGVGAPG
1 g	14-CTG	CGNLSTCMLGTYTQDFNKFHTFPQTAIG14* ¹⁴ GAPG
2 a	GRPG	VPLPAGGGTVLTKMYPRGNHWAVGHLMG
2 b	[D ₁₀]-12 ₂ -GRPG	VPI[D ₁₀]-12* ¹⁰ PAGGGTV[D ₁₀]-12* ¹⁰ TKMYPRGNH-WAVGH[D ₁₀]-12* ¹⁰ MG
3 a	CCK-7-G	YMGWMDFG
3 b	5-CCK-7-G	YMGWMD5* ⁵ G
3 c	10-CCK-7-G	10* ¹⁰ MGWMDFG
4	PRPG	SRTHRHSMBERTPDINPAWYASRGIRPVGRFG

* Denotes substitution with the corresponding amino acid.

Peptides are important synthetic targets for both biological and non-biological applications. Most commonly they are obtained by traditional chemical peptide synthesis, but limitations to this approach include the requirement for protecting-group strategies, the production of impure peptides as a result of incomplete amino acid coupling, and the incompatibility of some amino acid sequences with chemical synthesis. Recombinant peptide biosynthesis is an attractive alternative approach that does not require protected or even pure amino acids.^[1] The amino acid sequence is fully controlled by the translation of genetic material.^[2] Native and engineered biosynthetic machinery can be used to incorporate a wide range of amino acids, in addition to those normally found in proteins, thereby expanding the utility of this method.^[3–7]

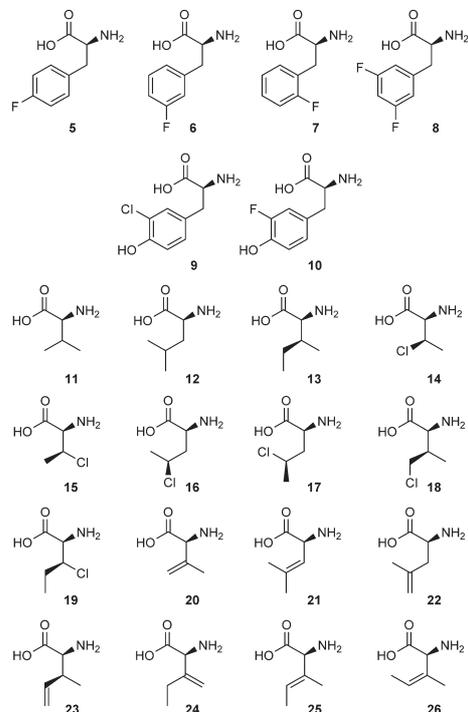
Cell-free expression systems using partially-purified cellular material supplemented with DNA, amino acids, and other ingredients have been developed as powerful tools for peptide and protein biosynthesis.^[8,9] Small peptides can be expressed fused to proteins that are subsequently removed, most commonly by enzymatic cleavage.^[10] Here we report the adaptation of this procedure for the synthesis of calcitonin prohormone CTG (1 a; 33 amino acids; Table 1) and its analogues 1 b–g substituted with the unusual amino acids 5–9 and 14 (Scheme 1), all from a single gene construct (Scheme 2). Ami-

dated peptide hormones have recently been prepared as fusion proteins expressed in Chinese hamster ovary cells,^[11] but that approach is unsuitable for the incorporation of atypical amino acids. Chemical methods have also been reported for peptide fragmentation.^[12,13] These depend on the selective reaction of particular amino acid side-chains, such as those of Met residues reacting with cyanogen bromide.^[12] They generally suffer from lack of specificity and require large excesses of toxic reagents and harsh reaction conditions, and have been largely superseded by enzymatic procedures.

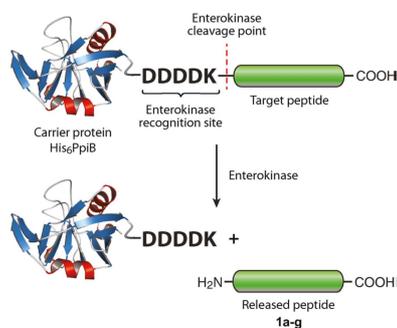
Here we describe an alternative protocol that uses native peptide biosynthetic machinery to incorporate the chlorinated amino acids 16–18, as latent cleavage sites for clean and efficient peptide release from fusion proteins, simply by heating. The advantages of the method over enzymatic cleavage are illustrated by the preparation of the gastrin-releasing peptide and cholecystokinin-7 prohormones GRPG (2 a) and CCK-7-G (3 a), and their analogues 2 b and 3 b, c substituted with the amino acids 5, 10 and [D₁₀]-12 (Scheme 3; Table 1). We also report incorporation of the dehydro amino acids 22 and 23, as latent sites for peptide cleavage, by treatment with iodine under mild conditions. This was used to prepare 2 a and the prolactin-releasing peptide prohormone PRPG (4; Scheme 4, Table 1). A related process with (S)-allylglycine has been reported,^[14] but it required a bioengineered system, which is not necessary with 22 and 23. Although hydroxy acids have also been incorporated into proteins by bioengineering,^[15] where they provide cleavage sites through treatment with hydrazine or

[a] M. Liutkus, S. A. Fraser, K. Caron, Dr. D. J. Stigers, Prof. Dr. C. J. Easton
Research School of Chemistry, Australian National University
Canberra, ACT 2601 (Australia)
E-mail: Chris.Easton@anu.edu.au

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under <http://dx.doi.org/10.1002/cbic.201600091>.

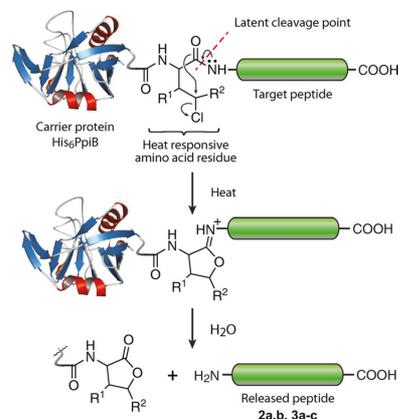


Scheme 1. Amino acids used in this study.

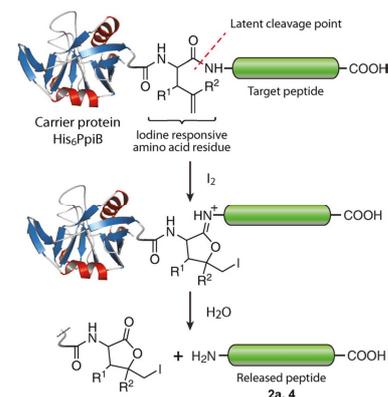


Scheme 2. Enterokinase-catalysed release of peptides 1 a–g from fusion protein.

alkaline hydrolysis,^[16] the hydroxy acid remains attached to the released material so free peptides are not produced with this procedure.



Scheme 3. Release of the peptides 2 a, b and 3 a–c by heat treatment of fusion protein. For peptides 2 a, b, R¹ = Me, R² = H; for peptides 3 a–c, R¹ = H, R² = Me.



Scheme 4. Release of peptides 2 a and 4 by treatment of fusion protein with iodine. For peptide 2 a, R¹ = Me, R² = H; for peptide 4, R¹ = H, R² = Me.

Cell-free expression using the supernatant of the lysate of *Escherichia coli* BL21 star (DE3) after centrifugation at 30000 g^[6–8] was used to produce 1 a attached to the His₆-tagged carrier protein peptidyl-prolyl *cis-trans* isomerase B (His₆PpiB). After meta-Ion affinity chromatography, the purified fusion protein was incubated with enterokinase at 24 °C for 18 h to release 1 a, which was isolated through HPLC (Scheme 2). The expression system can be biased towards the incorporation of unusual amino acids by withholding the normal analogues with which they compete for translation into protein. Accordingly, 1 b–e were produced when Phe was

withheld and the corresponding fluorides **5–8** were added, with all three Phe residues replaced in each case. Similarly, adding chloride **9** instead of Tyr resulted in substitution of that residue to give **1f**. These syntheses illustrate how multiple peptides may be produced from a single gene.

Previously we reported incorporation of chlorides **14–18** as substitutes for the corresponding aliphatic amino acids **11–13** during cell-free expression of His₆PpiB.^{6,7} The earlier studies were performed with a mixture of **16** and **17**, but these compounds have now been studied separately, and each is incorporated into protein as a substitute for Leu (**12**). Chloride **19** had not been investigated, but it has now been found to efficiently substitute for Ile (**13**) during protein expression, thus showing that all possible analogues of the aliphatic amino acids **11–13** with a chlorine in place of a methyl group are translated during protein biosynthesis. The Ile analogue **18** lactonizes with a half-life of 15 min under the conditions of protein expression (30 °C, pH 7.5) and is therefore insufficiently stable for direct use. Instead, the corresponding methyl ester is utilized because it is more stable, before being enzymatically hydrolyzed in situ to maintain a steady supply of the deprotected species.⁶ Chlorides **14–17** and **19** can be supplied to the expression system as free amino acids or as methyl esters, even though esterification of **16** and **17** is not required because they lactonize relatively slowly (half-lives ~6 and 4 h, respectively). Incorporation of **14** as a substitute for Val (**11**) was used with the method described above to obtain CTG analogue **1g** (Scheme 2).

During these studies, it was observed that heating to denature the protein for SDS-PAGE analysis results in extensive fragmentation of His₆PpiB with each of the γ -chlorides **16–18** (Figure 1), but no cleavage with the β -chlorides **14, 15, and 19**. The fragmentation occurs within minutes at 100 °C, takes over an hour at 80 °C, and is even slower at lower temperatures, being negligible under the conditions required for protein expression and purification (Supporting Information). Fragmentation is consistent with the propensity of **16–18** to lactonize. It is also analogous to the process brought about by cyanogen bromide and other chemical reagents, except that no external agent is required. Instead, when present in peptide chains, the

residues of **16–18** possess the latent functionality for heat-activated cleavage.

Incorporation of chlorides **16–18** and subsequent peptide cleavage were exploited to develop a new protocol for peptide synthesis (Scheme 3). Accordingly, peptides fused to His₆PpiB were prepared by cell-free expression, followed by purification of the fusion proteins through metal-ion affinity chromatography, before heating at 100 °C for 5 min for peptide release. GRPG (**2a**) was prepared by using chloride **18** instead of Ile to introduce the latent cleavable link, while CCK-7-G (**3a**) was obtained by using chlorides **16** and **17** in place of Leu. The yield of each peptide was >70% based on the corresponding fusion protein. Labeled GRPG **2b** was prepared by simultaneous incorporation of **18** and replacement of **12** with [D₁₀]-**12**. Similarly, the fluorinated forms of CCK-7-G (**3b,c**) were obtained by using chlorides **16** and **17** as well as by substituting fluorides **5** and **10** for Phe and Tyr, respectively. The method is thus not limited to the preparation of natural peptides. The use of **16–18** to release peptides from fusion proteins compares favorably with enzyme-catalyzed and exogenous chemical methods. It is rapid and specific to the position of incorporation of the chloride, and only requires heating (instead of the addition of external reagents). It also occurs efficiently with precipitated protein (Supporting Information), whereas enzyme-catalyzed cleavage requires soluble substrates with exposed cleavage sites.

We also evaluated misincorporation of dehydro analogues of Val, Leu, and Ile in cell-free protein expression. (S)-4,5-Dehydroleucine (**22**) and (2S,3S)-4,5-dehydroisoleucine (**23**) had been reported to be translational replacements for **12**⁴ and **13**,⁵ respectively, in auxotrophic cells, but we investigated each member of the complete set of the side-chain olefinic amino acids **20–26**. None of the dehydroisoleucines **24–26** substituted for **13**, but the other dehydro amino acids **20–23** are readily incorporated in place of their saturated counterparts **11–13**. His₆PpiB produced with either **22** or **23** undergoes extensive fragmentation upon treatment with iodine, through iodocyclization followed by hydrolysis, consistent with the rapid iodolactonization of the free amino acids under these conditions. These observations were exploited to develop the other protocol for peptide synthesis (Scheme 4). Accordingly, GRPG (**2a**) fused to His₆PpiB was prepared by expression using alkene **23** instead of Ile. After purification through metal-ion affinity chromatography, the fusion protein was treated with iodine for peptide release. Similarly, PRPG (**4**) was prepared by incorporation of **22** in place of Leu as the latent cleavable link. Following treatment of the corresponding fusion protein with five equivalents of iodine at 4 °C for 10 min, **4** was obtained in 40% yield. Higher temperatures, longer reaction times, and larger excesses of iodine decreased the yield, presumably because iodine also causes side reactions. The production of **4** using **22** as the cleavable link is nevertheless remarkable because this peptide is not released in any detectable quantity from the corresponding fusion proteins formed with chlorides **16** and **17**. The failure to produce **4** in the latter cases is attributable to the N-terminal serine residue in the peptide that would be released interfering with the

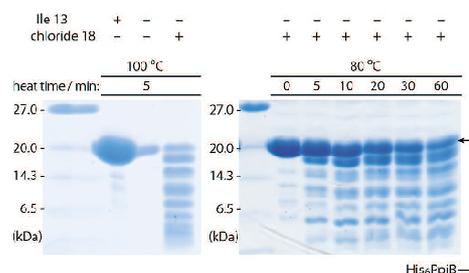


Figure 1. SDS-PAGE analysis of the effect of heat treatment of His₆PpiB expressed with chlorinated amino acid **18**, and of the control His₆PpiB prepared by using Ile.

fragmentation, as has been reported previously with cyanogen bromide and related chemical cleavage methods.^[12] This interference is apparently greatly reduced with the short reaction time and low temperature required when using iodine and the dehydroleucine **22**.

The new methods of peptide synthesis (Schemes 3 and 4) are complementary. The former requires only heat for the site-specific cleavage to release the peptide from fusion protein. The latter is generally more prone to side reactions, but occurs at lower temperatures and thus is more suitable for the preparation of thermosensitive targets. Both methods demonstrate the utility of modified amino acids to introduce functional groups into peptides at unactivated positions. Both can be evaluated by using the same gene, to incorporate either chloride **18** or alkene **23** in place of Ile, as the latent cleavable residue, or chloride **16** or **17** or alkene **22** in place of Leu. Peptides that have neither **12** nor **13**, such as CCK-7-G (**3a**), and peptides with just one of the two, such as GRPG (**2a**) and PRPG, can be prepared by appropriate selection of chloride **16–18** and alkene **22** or **23**, all of which are readily incorporated into proteins by the native biosynthetic machinery. The methods are unsuitable for the preparation of peptides that contain both **12** and **13**, as one or other of these would inevitably be substituted by one of the chlorides **16–18**, or the olefins **22** or **23**, as a cleavable link, but generally this limitation could be addressed by making inconsequential changes to the amino acid sequence. For example, encoding a different aliphatic amino acid in place of the problematic one would normally have little or no effect on peptide structure and function,^[17] particularly in non-biochemical applications. The methods are not limited to the synthesis of peptides composed of the amino acids normally found in proteins, as diverse other amino acids are also translated.^[3–7] Furthermore, amino acids **16–18**, **22**, and **23** are readily available for use in these procedures, commercially in the case of **22**, by synthesis of the racemate of **23**,^[5] and through side-chain halogenation of (S)-2-aminopentanoic acid, Val and their methyl esters for the chlorides **16–18** and the corresponding methyl esters.^[18] As stated above, methyl esters hydrolyze in situ during cell-free protein expression, so they can be used directly as sources of the corresponding free amino acids.^[6] Stereoisomers such as the racemate of **23**, and chloride **18** and its (3S)-diastereomer halogenation by-product can be separated, but this is not necessary as the expression system is highly stereoselective and mixtures of amino acids may be used.

In conclusion, these new protocols for the synthesis of peptides are versatile and straightforward. They constitute the first examples of cleavage of scissile amino acid residues in peptides, incorporated by using the innate promiscuity of native biological cell-free protein expression systems. They provide rapid and robust alternatives to enzymatic digestion and chemical treatment of natural peptide sequences for the release of peptides from fusion proteins. They therefore significantly enhance the utility of biosynthetic processes for peptide production.

Acknowledgements

We acknowledge funding from the Australian Research Council and the assistance of Drs. Apostolos Alissandratos and Hideki Onagi with this work.

Keywords: amino acids · cell-free synthesis · cleavage reactions · lactamization · peptides · protein expression

- [1] S. Maicas, I. Moukadiri, A. Nieto, E. Valentin, *Mol. Biotechnol.* **2013**, *55*, 150–158.
- [2] a) T. M. Schmeing, V. Ramakrishnan, *Nature* **2009**, *461*, 1234–1242; b) J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **2013**, *341*, 1238149.
- [3] a) T. L. Hendrickson, V. de Crécy-Lagard, P. Schimmel, *Annu. Rev. Biochem. Chem.* **2004**, *73*, 147–176; b) S. Lepthien, L. Merkel, N. Budisa, *Angew. Chem. Int. Ed.* **2010**, *49*, 5446–5450; *Angew. Chem.* **2010**, *122*, 5576–5581; c) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444; d) J. A. Johnson, Y. Y. Lu, J. A. Van Deventer, D. A. Tirrell, *Curr. Opin. Chem. Biol.* **2010**, *14*, 774–780; e) M. G. Hoels, N. Budisa, *Angew. Chem. Int. Ed.* **2011**, *50*, 2896–2902; *Angew. Chem.* **2011**, *123*, 2948–2955; f) P. O'Donoghue, J. Ling, Y.-S. Wang, D. Soll, *Nat. Chem. Biol.* **2013**, *9*, 594–598; g) K. Lang, J. W. Chin, *Chem. Rev.* **2014**, *114*, 4764–4806; h) S. A. Fraser, C. J. Easton, *Aust. J. Chem.* **2015**, *68*, 9–12.
- [4] Y. Tang, P. Wang, J. A. Van Deventer, A. J. Link, D. A. Tirrell, *ChemBioChem* **2009**, *10*, 2188–2190.
- [5] M. L. Mock, T. Michon, J. C. M. van Hest, D. A. Tirrell, *ChemBioChem* **2006**, *7*, 83–87.
- [6] I. N. Arthur, J. E. Hennessy, D. Padmakshan, D. J. Stigers, S. Lesturgez, S. A. Fraser, M. Liutkus, G. Otting, J. G. Oakeshott, C. J. Easton, *Chem. Eur. J.* **2013**, *19*, 6824–6830.
- [7] D. J. Stigers, Z. I. Watts, J. E. Hennessy, H.-K. Kim, R. Martini, M. C. Taylor, K. Ozawa, J. W. Keillor, N. E. Dixon, C. J. Easton, *Chem. Commun.* **2011**, *47*, 1839–1841.
- [8] a) L. Guignard, K. Ozawa, S. E. Pursglove, G. Otting, N. E. Dixon, *FEBS Lett.* **2002**, *524*, 159–162; b) K. Ozawa, M. J. Headlam, D. Mouradov, S. J. Watt, J. L. Beck, K. J. Rodgers, R. T. Dean, T. Huber, G. Otting, N. E. Dixon, *FEBS J.* **2005**, *272*, 3162–3171.
- [9] a) *Cell-Free Translation Systems* (Ed.: A. S. Spirin), Springer, Berlin, **2002**; b) *Cell-Free Protein Synthesis: Methods and Protocols* (Eds.: K. Alexandrov, W. A. Johnson), Humana, Totowa, **2014**.
- [10] a) J. Arnau, C. Lauritzen, G. E. Petersen, J. Pedersen, *Protein Expression Purif.* **2006**, *48*, 1–13; b) D. S. Waugh, *Protein Expression Purif.* **2011**, *80*, 283–293; c) C. L. Young, Z. T. Britton, A. S. Robinson, *Biotechnol. J.* **2012**, *7*, 620–634.
- [11] K. R. Carlson, S. C. Pomerantz, J. L. Li, O. Vafa, M. Naso, W. Strohl, R. E. Mains, B. A. Eipper, *BMC Biotechnol.* **2015**, 15:95.
- [12] a) D. L. Crimmins, S. M. Mische, N. D. Denslow, *Chemical Cleavage of Proteins in Solution in Current Protocols in Protein Science Supplement 41, Unit 11.4* (Eds.: J. E. Coligan, B. M. Dunn, D. W. Speicher, P. T. Wingfield), Wiley, Weinheim, **2005**, pp. 11.4.1–11.4.11; b) R. Kaiser, L. Metzka, *Anal. Biochem.* **1999**, *266*, 1–8.
- [13] a) B. Witkop, *Adv. Protein Chem.* **1962**, *16*, 221–321; b) V. Rahali, J. Gueguen, *J. Protein Chem.* **1999**, *18*, 1–12; c) N. M. Milović, L. M. Dutcă, N. M. Kostić, *Chem. Eur. J.* **2003**, *9*, 5097–5106; d) P. M. Hwang, J. S. Pan, B. D. Sykes, *FEBS Lett.* **2014**, *588*, 247–252; e) Y. Seki, K. Tanabe, D. Sasaki, Y. Sohma, K. Oisaki, M. Kanai, *Angew. Chem. Int. Ed.* **2014**, *53*, 6501–6505; *Angew. Chem.* **2014**, *126*, 6619–6623.
- [14] a) B. Wang, M. Lodder, J. Zhou, T. T. Baird, K. C. Brown, C. S. Craik, S. M. Hecht, *J. Am. Chem. Soc.* **2000**, *122*, 7402–7403; b) B. Wang, K. C. Brown, M. Lodder, C. S. Craik, S. M. Hecht, *Biochemistry* **2002**, *41*, 2805–2813.
- [15] a) J. Guo, J. Wang, J. C. Anderson, P. G. Schultz, *Angew. Chem. Int. Ed.* **2008**, *47*, 722–725; *Angew. Chem.* **2008**, *120*, 734–737; b) T. Kobayashi, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *J. Mol. Biol.* **2009**, *385*, 1352–1360.
- [16] a) Y.-M. Li, M.-Y. Yang, Y.-C. Huang, Y.-T. Li, P. R. Chen, L. Liu, *ACS Chem. Biol.* **2012**, *7*, 1015–1022; b) N. A. Bindman, S. C. Bobeica, W. R. Liu, W. A. van der Donk, *J. Am. Chem. Soc.* **2015**, *137*, 6975–6978.

Chlorinated Amino Acids in Peptide Production

Chapter 3. Cleavage of Peptide Bonds



CHEMBIOCHEM
Communications

- [17] a) J. U. Bowie, J. F. Reidhaarolson, W. A. Lim, R. T. Sauer, *Science* **1990**, *247*, 1306–1310; b) S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, M. H. Hecht, *Science* **1993**, *262*, 1680–1685; c) D. S. Riddle, J. V. Santiago, S. T. Bray-Hall, N. Doshi, V. P. Grantcharova, Q. Yi, D. Baker, *Nat. Struct. Biol.* **1997**, *4*, 805–809; d) I. Ladunga, R. F. Smith, *Protein Eng.* **1997**, *10*, 187–196; e) C. E. Schafmeister, S. L. LaPorte, L. J. W. Miercke, R. M. Stroud, *Nat. Struct. Biol.* **1997**, *4*, 1039–1046; f) L. R. Murphy, A. Wallqvist, R. M. Levy, *Protein Eng.* **2000**, *13*, 149–152; g) T. Li, K. Fan, J. Wang, W. Wang, *Protein Eng.* **2003**, *16*, 323–330; h) L. Y. Yampolsky, A. Stoltzfus, *Genetics* **2005**, *170*, 1459–1472; i) C. Etchebest, C. Benros, A. Bornot, A.-C. Camproux, A. G. de Brevern, *Eur. Biophys. J.* **2007**, *36*, 1059–1069; j) S. Hormoz, *Sci. Rep.* **2013**, *3*, 2919; k) G. Zhang, K. Wang, B. Zheng, M. Cheng, Y. Li, K. Liu, L. Cai, *Chem. Commun.* **2013**, *49*, 11086–11088.
- [18] Z. I. Watts, C. J. Easton, *J. Am. Chem. Soc.* **2009**, *131*, 11323–11325.

Manuscript received: February 15, 2016
Accepted article published: February 25, 2016
Final article published: ■ ■ ■ ■ 0000

CHEMBIOCHEM

Supporting Information

Peptide Synthesis through Cell-Free Expression of Fusion Proteins Incorporating Modified Amino Acids as Latent Cleavage Sites for Peptide Release

Mantas Liutkus, Samuel A. Fraser, Karine Caron, Dannon J. Stigers, and Christopher J. Easton^{*(a)}

cbic_201600091_sm_miscellaneous_information.pdf

Table of Contents

1. General	2
2. Amino Acid Synthesis	3
2.1 NMR Spectra of Synthesised Compounds	6
3. Gene construction	19
4. Cell-free protein expression	21
5. Production of CTG (1a) and analogues 1b-g	22
6. Analysis of His ₆ PpiB produced using chlorinated amino acids	29
7. Production of GRPG (2a), CCK-7-G (3a) and the analogues 2b and 3b,c through heat treatment of fusion proteins	35
8. Analysis of His ₆ PpiB produced using dehydro amino acids	40
9. Production of GRPG (2a) and PRPG (4) through treatment of fusion proteins with iodine	44
10. References	45

1. General

NMR spectra were recorded using either a Varian Mercury 300 MHz spectrometer, a Varian MR 400 MHz spectrometer, a Bruker 400 MHz Ascend spectrometer, a Varian INOVA 500 MHz spectrometer, or a Bruker Avance 800 spectrometer, at 298 K. NMR solvents CDCl₃ with an isotopic purity of 99.8%, D₂O with an isotopic purity of 99.75% and CD₃OD with an isotopic purity of 99.8% were purchased from Cambridge Isotope Laboratories, Inc. High resolution ESI mass spectra were recorded using a Waters LCT premier TM XE oa-TOF mass spectrometer.

Semi-preparative and preparative HPLC was performed on either a HP Agilent 1100 series HPLC with a quaternary pump, analytical automatic liquid sampler, column compartment and diode array detector running with Agilent Chemstation software, a HP Agilent 1100 series HPLC with a preparative binary pump system, preparative automatic liquid sampler and a diode array detector with a preparative flow cell running with Agilent Chemstation software, or a Waters 600 Controller with a Waters 717 plus Autosampler and a Waters 2996 Photodiode Array Detector running with Empower Pro Empower 2 software. Eluting fractions with the latter system were collected using a Waters Fraction Collector III. Flash column chromatography was run on Merck KGaA Silica 60 (240-400 mesh).

Melting points were determined using a Stanford Research Systems MP100 Optimelt Automated Melting Point System in open capillaries in air and are uncorrected. Elemental analyses were performed by the Research School of Chemistry Microanalytical Service at the Australian National University, using a Carlo Erba 1106 automatic analyser.

The amino acids used for normal cell-free protein synthesis, including (*S*)-valine (**11**), (*S*)-leucine (**12**) and (*2S,3S*)-isoleucine (**13**), were Bio-Ultra grade from Sigma-Aldrich, Inc. (*S*)-3-Fluorophenylalanine (**6**) was obtained from PepTech Corporation, whereas the other aromatic amino acids **5** and **7-10** were supplied as racemates by Sigma-Aldrich, Inc., (*S*)-4,5-dehydroleucine (**22**) was purchased from Chem-Impex International, Inc., and perdeuterated (*S*)-leucine (**d₁₀-12**) was bought from Cambridge Isotope Laboratories, Inc. The chlorides **14** and **15** were synthesised from (*2S,3R*)-threonine and (*2S,3S*)-*allo*-threonine using literature procedures.^[1-3] Using the same method, *threo*-β-hydroxynorvaline^[4,5] was used to prepare the chloride **19** as a racemate. The methyl ester of the chloride **18** was prepared by chlorination of the methyl ester of (*S*)-valine (**11**) as reported.^[6] Using the same approach, the chlorides **16** and **17** were prepared as a mixture from (*S*)-norvaline.^[6] Alternatively, the methyl esters of the individual chlorides **16** and **17** were prepared from the methyl ester of (*S*)-norvaline, as described below. The dehydroisoleucines **24-26** were prepared as racemates by treatment of *N*-phthaloyl-3-bromoisoleucine methyl ester with silver nitrate, followed by deprotection using hydrochloric acid.^[7,8] Samples of CTG (**1a**), GRPG (**2a**) and PRPG (**4**) for use as standards were purchased from GL Biochem (Shanghai) Ltd, while CCK-7-G (**3a**) was prepared through solid phase peptide synthesis using a CEM Liberty Microwave Peptide Synthesizer on WANG resin.

2. Amino Acid Synthesis

Methyl (2*S*,4*R*)-2-amino-4-chloropentanoate and methyl (2*S*,4*S*)-2-amino-4-chloropentanoate: (*S*)-Norvaline methyl ester (3.5 g, 21 mmol) was dissolved in glacial trifluoroacetic acid in a quartz flask and the mixture was exposed to a 300 W sunlamp while chlorine gas was bubbled through the solution. The solvent was then evaporated to dryness. The residue was dissolved in DCM and the solution was treated with triethylamine and di-*tert*-butyl dicarbonate (4.6 g, 21 mmol) before being stirred at RT overnight. The solution was then concentrated to remove all volatiles and a portion of the residual material was subjected to HPLC (Grace Alltima C₁₈ 250 x 22 mm, 5µm), using a mobile phase of 60% MeOH in H₂O at a flow rate of 10 mL min⁻¹. Lyophilization of the collected fractions gave methyl (2*S*,4*R*)-2-[(*tert*-butoxycarbonyl)amino]-4-chloropentanoate as a colourless crystalline solid (55 mg) (*t*_R = 30.4 min; M.p. 68.5-69.5 °C; ¹H NMR (400 MHz, CDCl₃): δ = 5.03 (br, 1H), 4.52 (br, 1H), 4.13 (m, 1H), 3.76 (s, 3H), 2.20 (m, 1H), 2.00 (m, 1H), 1.56 (d, *J* = 7 Hz, 3H), 1.46 ppm (s, 9H); ¹³C NMR (100 MHz, CDCl₃): δ = 172.8, 155.4, 80.2, 54.4, 52.4, 51.8, 43.0, 28.3, 25.5 ppm; HRMS (ESI): *m/z* calcd for C₁₁H₂₀³⁵ClNNaO₄: 288.0979 [M+Na]⁺; found: 288.0981; calcd for C₁₁H₂₀³⁷ClNNaO₄: 290.0949 [M+Na]⁺; found: 290.0947) whereas methyl (2*S*,4*S*)-2-[(*tert*-butoxycarbonyl)amino]-4-chloropentanoate was obtained as a colourless oil (80 mg) (*t*_R = 33.0 min; ¹H NMR (400 MHz, CDCl₃): δ = 5.18 (br, 1H), 4.44 (br, 1H), 4.18-4.08 (m, 1H), 3.77 (s, 3H), 2.32-2.11 (m, 2H), 1.57 (d, *J* = 7 Hz, 3H), 1.45 ppm (s, 9H); ¹³C NMR (100 MHz, CDCl₃): δ = 172.5, 155.2, 80.2, 54.0, 52.6, 51.5, 42.8, 28.3, 25.1 ppm; HRMS (ESI): *m/z* calcd for C₁₁H₂₀³⁵ClNNaO₄: 288.0979 [M+Na]⁺; found: 288.0980; calcd for C₁₁H₂₀³⁷ClNNaO₄: 290.0949 [M+Na]⁺; found: 290.0949).

Methyl (2*S*,4*R*)-2-[(*tert*-butoxycarbonyl)amino]-4-chloropentanoate and the (2*S*,4*S*)-diastereomer were separately treated with trifluoroacetic acid (5 ml) for 1 h at RT. All volatiles were then removed to give methyl (2*S*,4*R*)-2-amino-4-chloropentanoate (¹H NMR (400 MHz, D₃COD): δ = 4.30-4.18 (m, 2H), 3.86 (s, 3H), 2.36 (m, 1H), 2.21 (m, 1H), 1.60 (d, *J* = 6 Hz, 3H); ¹³C NMR (100 MHz, D₃COD): δ = 170.7, 54.9, 54.1, 52.3, 41.9, 25.7 ppm; HRMS (ESI): *m/z* calcd for C₆H₁₃³⁵ClNO₂: 166.0635 [M+H]⁺; found: 166.0638; calcd for C₆H₁₃³⁷ClNO₂: 168.0605 [M+H]⁺; found: 168.0607) and methyl (2*S*,4*S*)-2-amino-4-chloropentanoate (¹H NMR (400 MHz, D₃COD): δ = 4.31 (m, 1H), 4.23 (br, 1H), 3.86 (s, 3H), 2.38 (m, 1H), 2.13 (m, 1H), 1.57 (d, *J* = 7 Hz, 3H); ¹³C NMR (100 MHz, D₃COD): δ = 170.6, 55.3, 53.9, 52.0, 42.1, 26.0 ppm; HRMS (ESI): *m/z* calcd for C₆H₁₃³⁵ClNO₂: 166.0635 [M+H]⁺; found: 166.0635; calcd for C₆H₁₃³⁷ClNO₂: 168.0605 [M+H]⁺; found: 168.0601). The stereochemistry of these chlorides was assigned by conversion to the corresponding lactones and γ-hydroxyamino acids, and comparison of their NMR data with those of authentic samples.^[9]

3,4-Dehydrovaline (20): *N*-Phthaloyl-3,4-dehydrovaline methyl ester^[7] (40 mg, 0.15 mmol) was added to a solution of 4 M HCl (1 mL) and AcOH (0.2 mL). The mixture was heated at reflux for 7.5 h, after which it was concentrated under reduced pressure to give a colourless solid. The solid was purified by dissolution in EtOH/HCl followed by the addition of propylene oxide,^[10] which produced a fine colourless precipitate of the title compound that was collected by centrifugation. Yield 13 mg (73%); M.p. >217 °C (decomp.); ¹H NMR (300 MHz, D₂O): δ=5.21 (m, 2H), 4.25 (s, 1H), 1.77 ppm (s, 3H); HRMS

(ESI): m/z calcd for $C_5H_{10}NO_2$: 116.0712 $[M+H]^+$; found: 116.0712. These spectral data are consistent with reported values.^[11]

3,4-Dehydroleucine (21): NBS (0.86 g, 4.8 mmol) was added to a solution of (*E*)-4-methylpent-2-enoic acid (0.52 mL, 4.4 mmol) in α,α,α -trifluorotoluene (10 mL) and the resultant mixture was heated under irradiation with a 300 W sunlamp at reflux for 0.5 h, before it was cooled and filtered, and the filtrate was concentrated under reduced pressure. Crystallization of the residue from α,α,α -trifluorotoluene afforded (*E*)-4-bromo-4-methylpent-2-enoic acid as a colourless solid. Yield 0.70 g (82%); M.p. 101-102 °C; 1H NMR (300 MHz, $CDCl_3$): δ =7.27 (d, J =15.3 Hz, 1H), 5.90 (d, J =15.3 Hz, 1H), 1.93 ppm (s, 6H); HRMS (ESI): m/z calcd for $C_6H_8^{79}BrO_2$: 190.9708 $[M-H]^-$; found: 190.9709. These spectral data are consistent with reported values.^[12]

Anhydrous NH_3 (50 mL) was added to a solution of (*E*)-4-bromo-4-methylpent-2-enoic acid (75 mg, 0.39 mmol) in dry THF (2 mL) and the mixture was let stand for 16 h before it was concentrated under reduced pressure to give an off-white solid. HPLC (Waters Spherisorb SAX 250 x 1.6 mm, 5 μ m) of this material using a mobile phase of 0.25% AcOH in H_2O at a flow rate of 0.7 mL min^{-1} , followed by HPLC (Alltech Prevail C_{18} 250 x 10 mm, 5 μ m) using a mobile phase of 0.1% TFA in H_2O at a flow rate of 4 mL min^{-1} , gave the TFA salt of 3,4-dehydroleucine (**21**) as a colourless solid. Yield 7 mg (7%); M.p. 216-218 °C (decomp.); 1H NMR (400 MHz, D_2O): δ =5.22 (d, J =10.0 Hz, 1H), 4.70 (d, J =10.0 Hz, 1H), 1.81 ppm (s, 6H); ^{13}C NMR (101 MHz, D_2O): δ =173.2, 145.5, 115.1, 52.4, 25.5, 18.2 ppm; HRMS (ESI): m/z calcd for $C_6H_{12}NO_2$: 130.0868 $[M+H]^+$; found: 130.0868. These spectral data are consistent with reported values.^[12]

(2RS,3RS)-4,5-Dehydroisoleucine (23): A solution of *N*-Cbz-glycine (1.0 g, 4.8 mmol), but-2-yn-1-ol (0.36 mL, 4.8 mmol) and DCM (11 mL) was added to a solution of DMAP (58 mg, 0.48 mmol) and DCC (0.99 g, 4.8 mmol) in DCM (8 mL) maintained at -20 °C, then the mixture was stirred and allowed to warm to RT over 3 h. The colourless precipitate that formed during this time was separated by filtration and washed with DCM. The combined filtrates were washed with 1 N HCl and saturated aqueous $NaHCO_3$, and the washings were extracted with DCM. The combined organic phases were dried ($MgSO_4$) and concentrated under reduced pressure. Flash chromatography of the residual oil on silica, eluting with EtOAc/hexane, provided *N*-Cbz-glycine but-2-ynyl ester as a colourless solid. Yield 1.1 g (91%); M.p. 52-54 °C; 1H NMR (300 MHz, $CDCl_3$): δ =7.39-7.29 (m, 5H), 5.26 (br s, 1H), 5.13 (s, 2H), 4.72 (d, J =2.3 Hz, 2H), 4.03 (m, 2H), 1.85 ppm (t, J =2.3 Hz, 3H); ^{13}C NMR (75 MHz, $CDCl_3$): δ =169.6, 156.4, 136.1, 128.5, 128.1, 128.0, 83.8, 72.5, 67.0, 53.6, 42.6, 3.6 ppm; HRMS (ESI): m/z calcd for $C_{14}H_{15}NaNO_4$: 284.0899 $[M+Na]^+$; found: 284.0899.

Lindlar's catalyst (100 mg) was added to a solution of *N*-Cbz-glycine but-2-ynyl ester (1.1 g, 4.2 mmol) in dry MeOH (70 mL) and the mixture was stirred at RT under an atmosphere of H_2 for 4 h, before it was filtered through celite, and the celite was washed with MeOH. The combined MeOH solutions were concentrated under reduced pressure to give (*Z*)-*N*-Cbz-glycine but-2-enyl ester as a colourless oil. Yield 1.1 g (97%); 1H NMR (300 MHz, $CDCl_3$): δ =7.39-7.29 (m, 5H), 5.82-5.65 (m, 1H), 5.60-5.45 (m, 2H), 5.11 (s, 2H), 4.70 (d, J =6.9 Hz, 2H), 3.95 (d, J =5.6 Hz, 2H), 1.69 ppm (d, J =6.6 Hz, 3H); ^{13}C NMR (75 MHz, $CDCl_3$): δ =170.1, 156.4, 136.2, 130.3, 128.5, 128.1, 128.1, 123.5, 67.0, 60.9, 42.7, 13.1 ppm; HRMS (ESI): m/z calcd for $C_{14}H_{17}NaNO_4$: 286.1055 $[M+Na]^+$; found: 286.1055.

An analytically pure sample was obtained by HPLC (Grace Alltima C₁₈ 250 x 22 mm, 5 μm) using a mobile phase of 0.1% TFA in MeOH/H₂O (65:35) at a flow rate of 10 mL min⁻¹. HPLC t_R=19.8 min; elemental analysis calcd (%) for C₁₄H₁₇NO₄: C 63.87, H 6.51, N 5.32; found: C 64.18, H 6.82, N 5.59.

*n*BuLi (3.6 mL, 5.8 mmol, 1.6 M in hexanes) was added slowly to a stirred solution of diisopropylamine (0.8 mL, 5.8 mmol) in dry THF (20 mL) maintained at -20 °C under an atmosphere of N₂. After 20 min, the solution was cooled to -78 °C and (*Z*)-*N*-Cbz-glycine but-2-enyl ester (0.51 g, 1.9 mmol) in dry THF (30 mL) was added. After 5 min, ZnCl₂ (0.32 g, 2.3 mmol) was added in portions, and the resulting mixture was allowed to warm to RT over 6 h, before 1 M HCl was added and the mixture was concentrated under reduced pressure. The oily residue was dissolved in Et₂O and the solution was washed with 1 M HCl and then twice with 1 M NaOH. The aqueous washings were combined, acidified to pH 1 with HCl, and extracted twice with Et₂O. The combined organic phases were dried (MgSO₄) and concentrated under reduced pressure. Flash chromatography of the residual yellow oil on silica, eluting with AcOH/EtOAc/hexane, afforded a *ca.* 4:1 mixture of the diastereomers of *N*-Cbz-4,5-dehydroisoleucine as a colourless solid. Yield 0.35 g (68%); ¹H NMR (300 MHz, CDCl₃): δ=9.35 (br s, 1H), 7.39-7.29 (m, 5H), 5.70 (m, 1H), 5.24-5.12 (m, 5H), 4.41 (dd, *J*=8.7, *J*=4.4 Hz, 0.8 x 1H), 4.24 (m, 0.2 x 1H), 2.86 (m, 1H), 1.14 ppm (d, *J*=6.9 Hz, 3H); HRMS (ESI): *m/z* calcd for C₁₄H₁₈NO₄: 264.1236 [M+H]⁺; found: 264.1234. These spectral data are consistent with reported values.^[13]

Iodotrimethylsilane (0.46 mL, 3.3 mmol) was added to a stirred solution of a *ca.* 4:1 mixture of the diastereomers of *N*-Cbz-4,5-dehydroisoleucine (0.29 g, 1.1 mmol) in dry CHCl₃ (10 mL). After 20 min, MeOH (10 mL) was added and the solution was concentrated under reduced pressure, to give a colourless solid that was washed with Et₂O. The solid was purified by dissolution in EtOH/HCl followed by the addition of propylene oxide, to give colourless crystals of (2*RS*,3*RS*)-4,5-dehydroisoleucine (**23**) together with *ca.* 5% of the diastereomer. Yield 60 mg (47%); M.p. 216-217 °C (decomp.); ¹H NMR (400 MHz, D₂O): δ=5.77 (m, 1H), 5.28-5.24 (m, 2H), 3.77 (d, *J*=5.6 Hz, 0.05 x 1H), 3.63 (d, *J*=5.6 Hz, 0.95 x 1H), 2.85-2.75 (m, 1H), 1.03 (d, *J*=7.2 Hz, 0.95 x 3H), 0.97 (d, *J*=7.2 Hz, 0.05 x 3H); HRMS (ESI): *m/z* calcd for C₆H₁₀NO₂: 128.0712 [M-H]⁻; found: 128.0713. These spectral data are consistent with reported values.^[14]

2.1 NMR Spectra of Synthesised Compounds

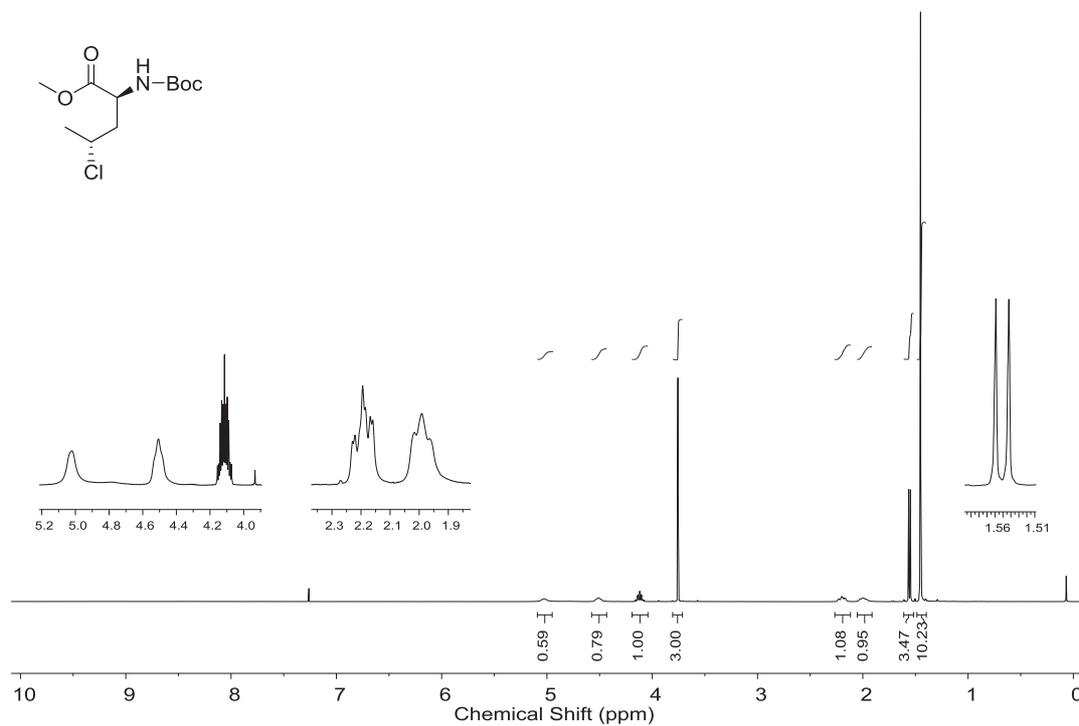
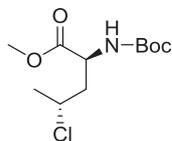
Spectra of samples recorded in D₂O have been recorded with and without acetonitrile or alternatively, sodium [D₄]3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionate. When sodium [D₄]3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionate was used as a reference, it was introduced by using a coaxial insert with a standard 5 mm NMR tube. The manual and automated shimming for these samples is problematic resulting in broadened peaks, therefore both spectra have been provided for clarity.

Chlorinated Amino Acids in Peptide Production

Chapter 3. Cleavage of Peptide Bonds

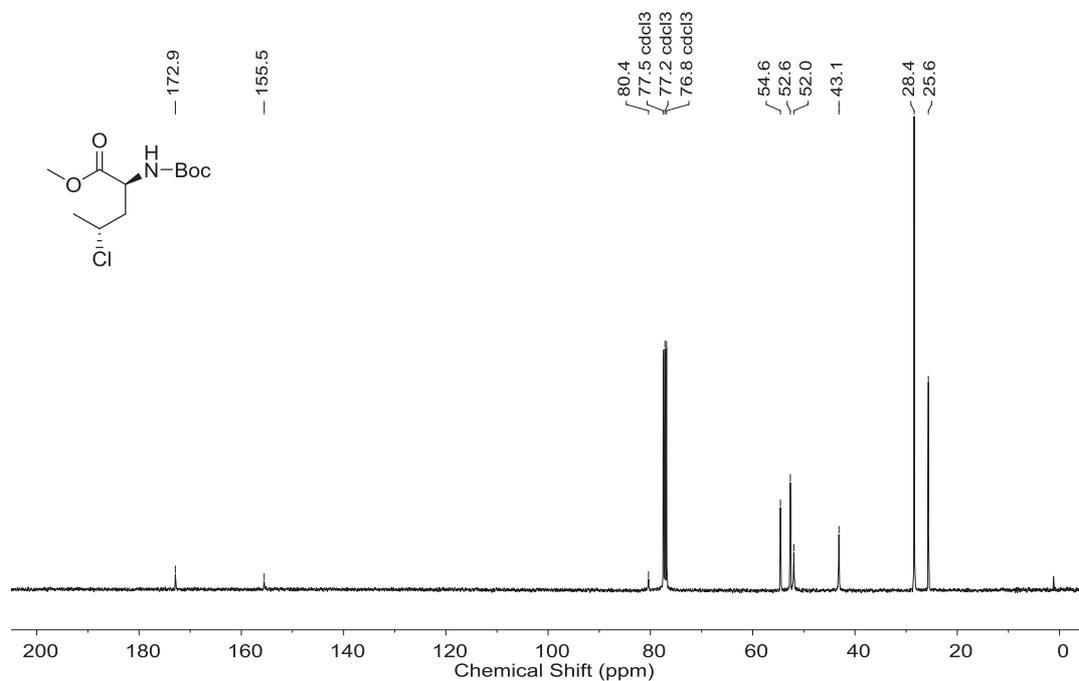
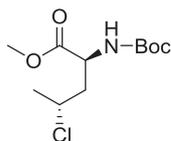
^1H NMR, 400 MHz, CDCl_3

Methyl (2*S*,4*R*)-2-[(*tert*-butoxycarbonyl)amino]-4-chloropentanoate



^{13}C NMR, 100 MHz, CDCl_3

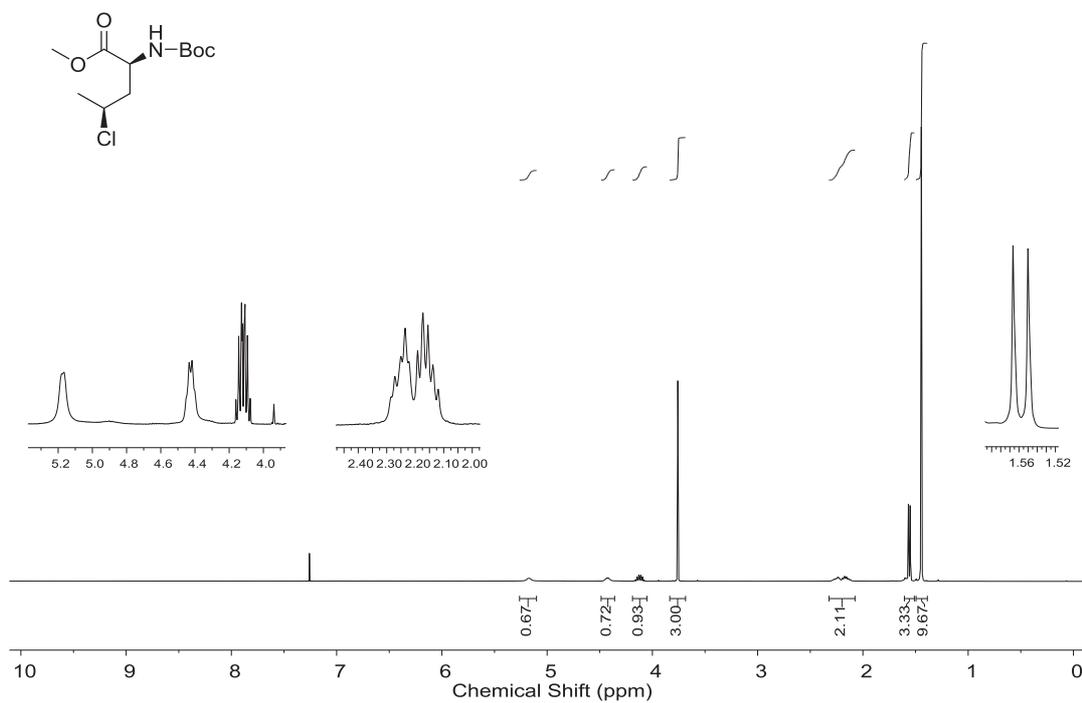
Methyl (2*S*,4*R*)-2-[(*tert*-butoxycarbonyl)amino]-4-chloropentanoate



Chlorinated Amino Acids in Peptide Production
Chapter 3. Cleavage of Peptide Bonds

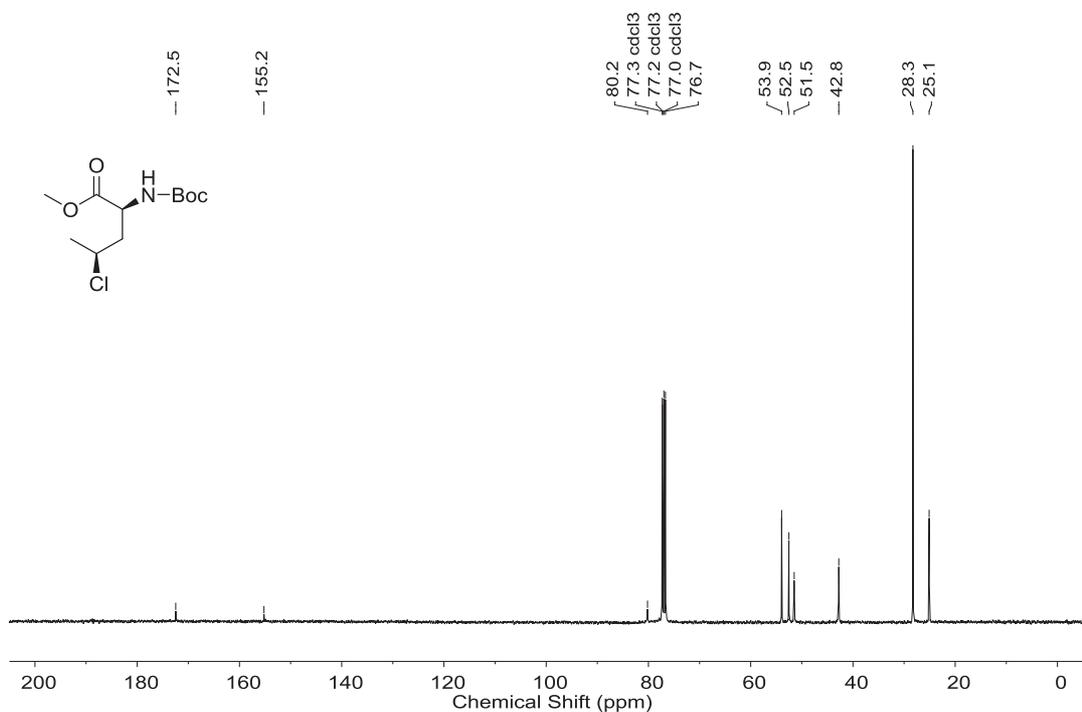
^1H NMR, 400 MHz, CDCl_3

Methyl (2S,4S)-2-[(*tert*-butoxycarbonyl)amino]-4-chloropentanoate



^{13}C NMR, 100 MHz, CDCl_3

Methyl (2S,4S)-2-[(*tert*-butoxycarbonyl)amino]-4-chloropentanoate

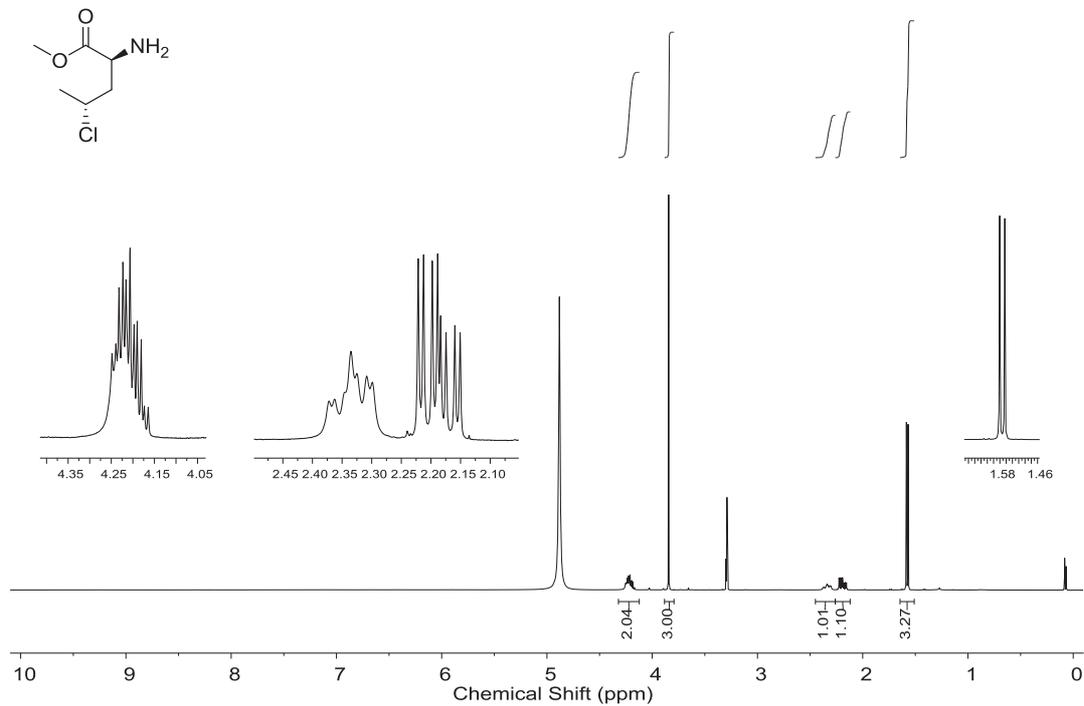


Chlorinated Amino Acids in Peptide Production

Chapter 3. Cleavage of Peptide Bonds

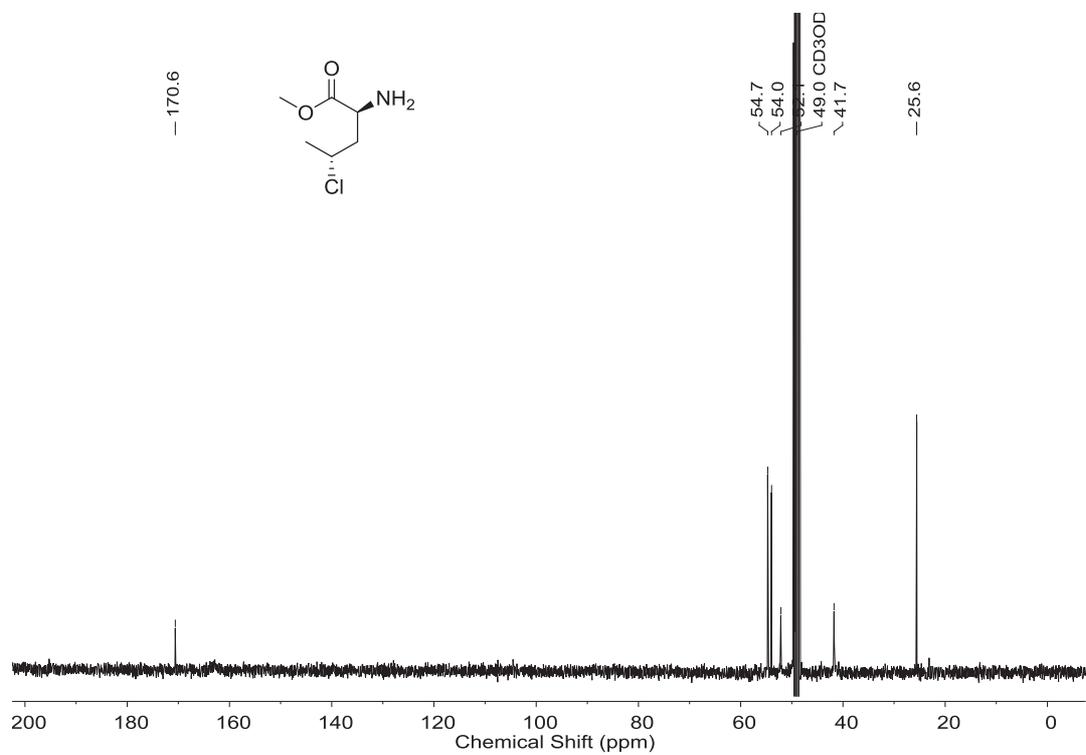
^1H NMR, 400 MHz, CD_3OD

Methyl (2*S*,4*R*)-2-amino-4-chloropentanoate



^{13}C NMR, 100 MHz, CD_3OD

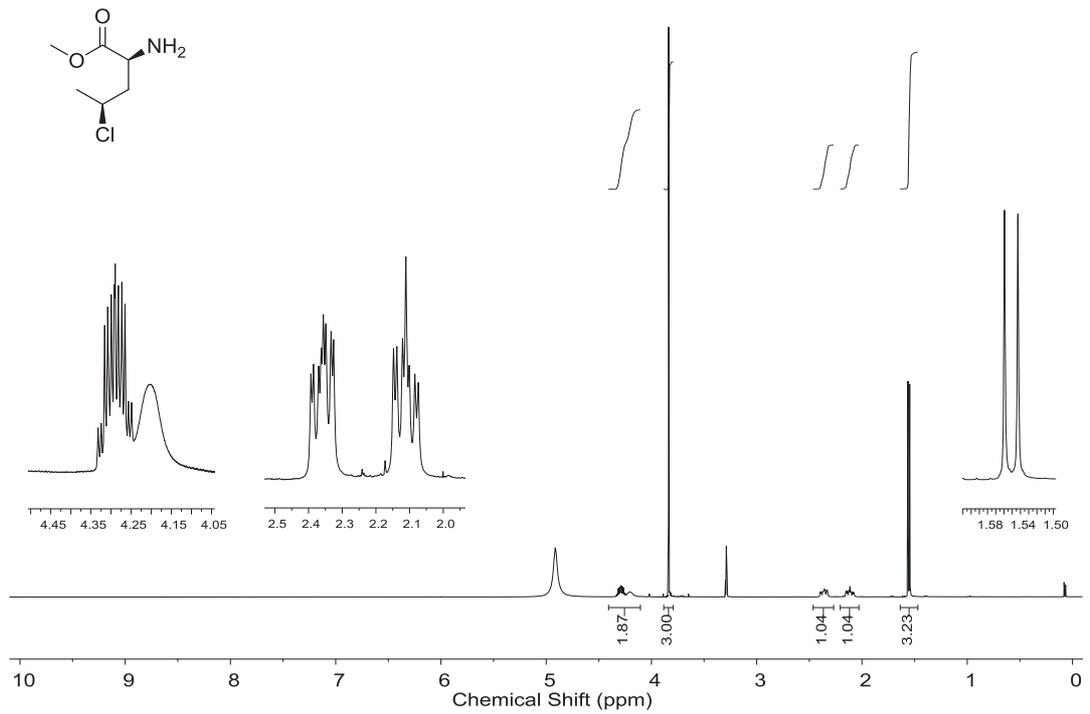
Methyl (2*S*,4*R*)-2-amino-4-chloropentanoate



Chlorinated Amino Acids in Peptide Production
Chapter 3. Cleavage of Peptide Bonds

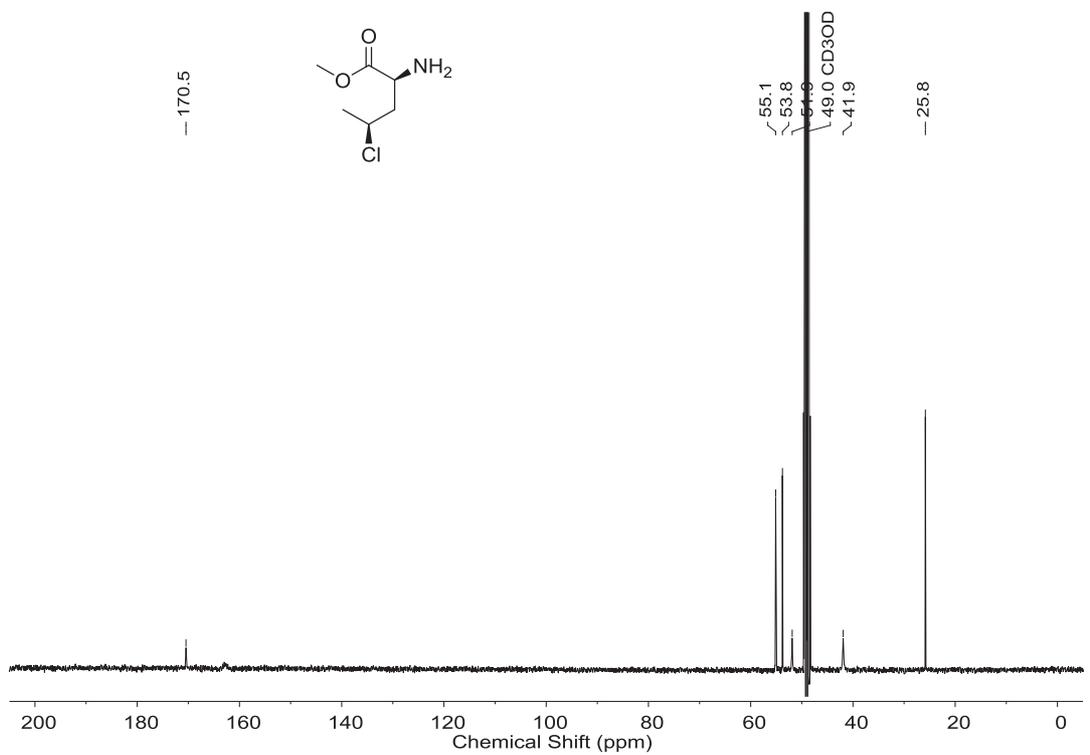
^1H NMR, 400 MHz, CD_3OD

Methyl (2S,4S)-2-amino-4-chloropentanoate



^{13}C NMR, 100 MHz, CD_3OD

Methyl (2S,4S)-2-amino-4-chloropentanoate

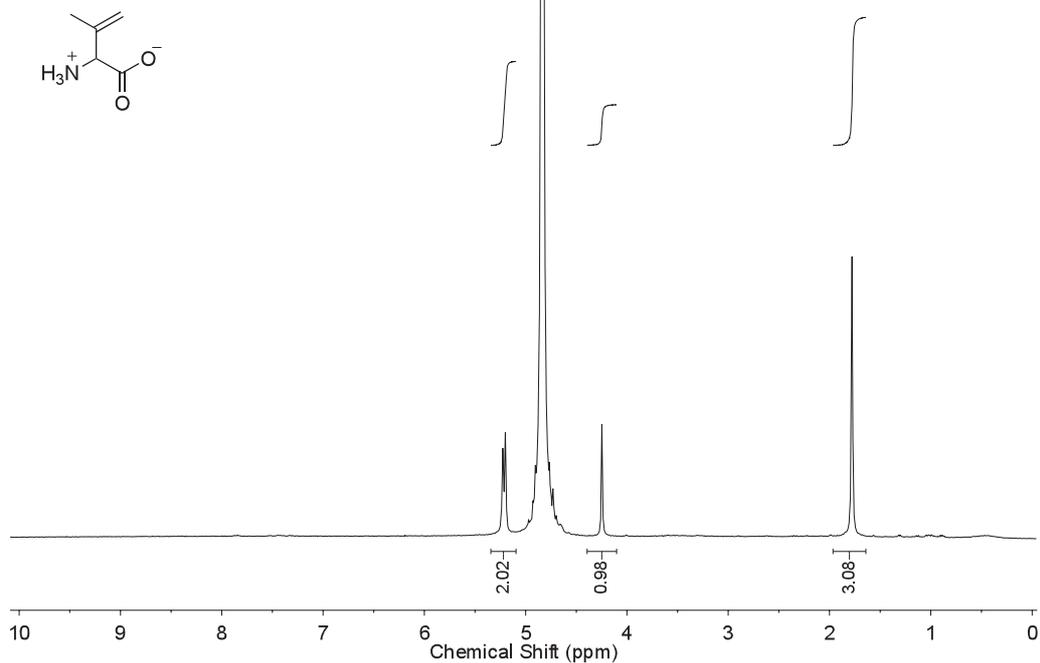


Chlorinated Amino Acids in Peptide Production

Chapter 3. Cleavage of Peptide Bonds

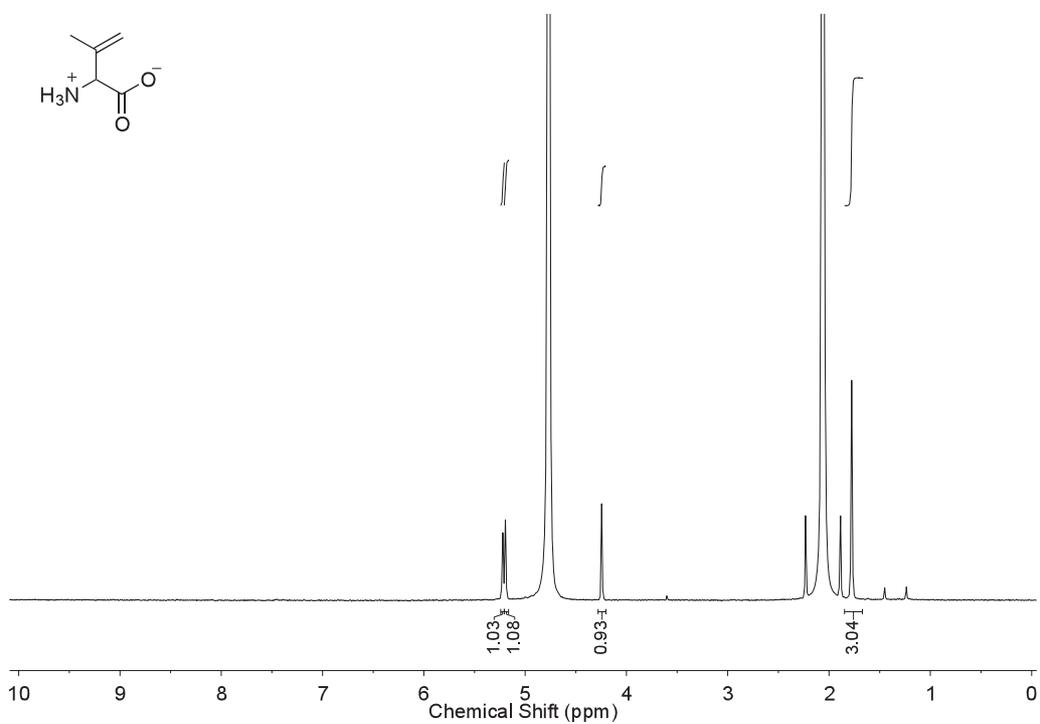
^1H NMR, 300 MHz, D_2O

3,4-Dehydrovaline (20)



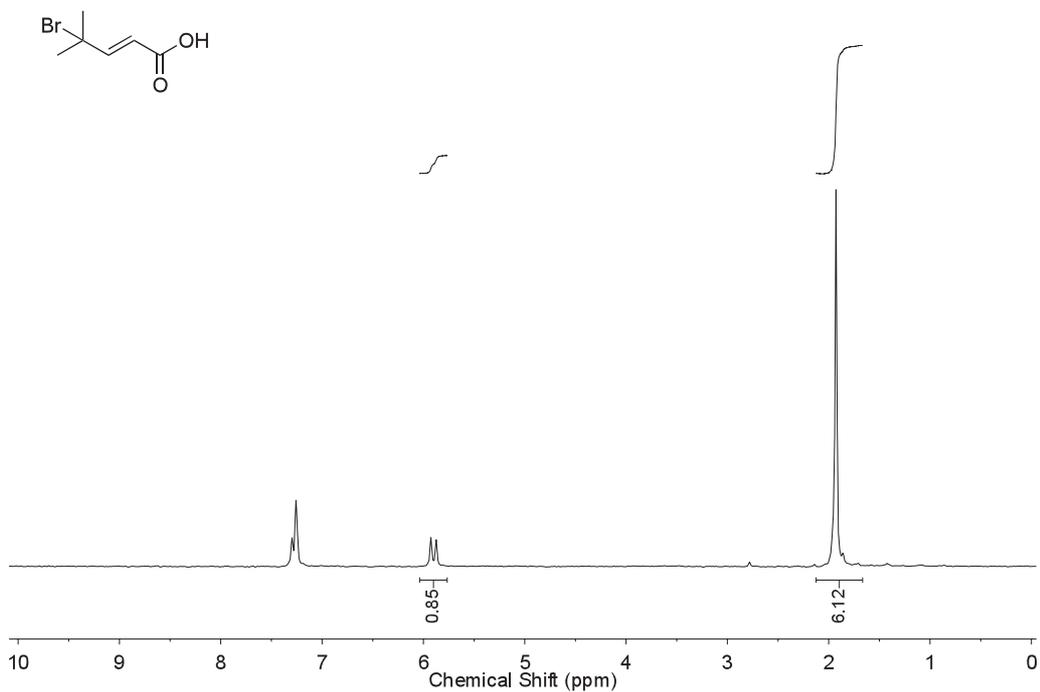
^1H NMR, 400 MHz, D_2O

3,4-Dehydrovaline (20) & MeCN



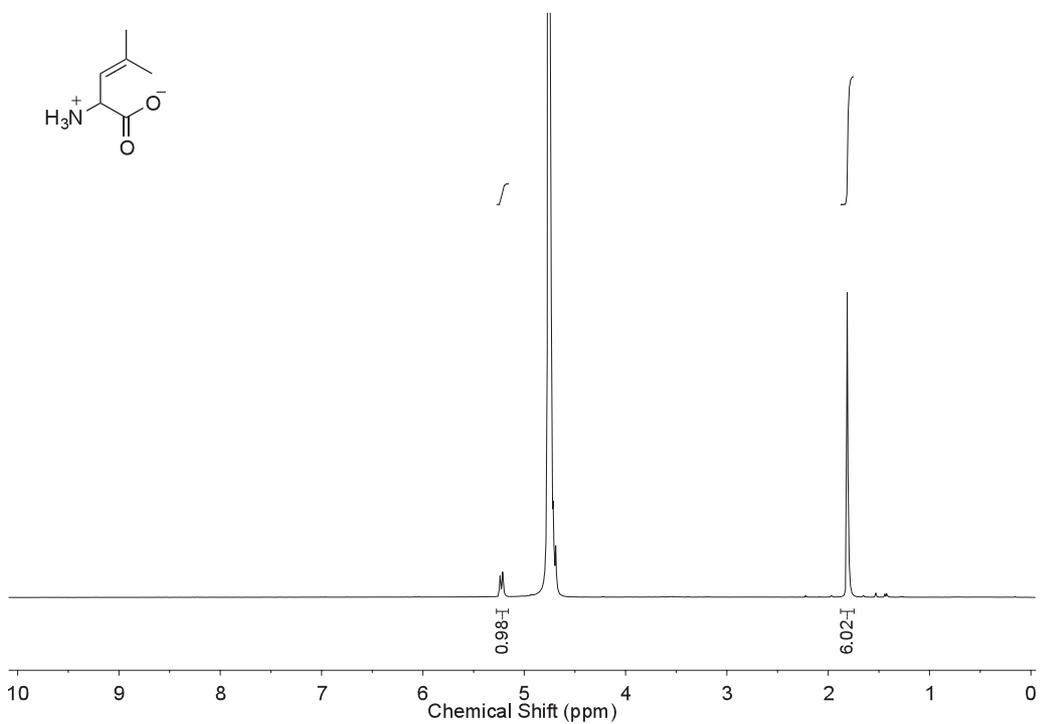
^1H NMR, 300 MHz, CDCl_3

(*E*)-4-Bromo-4-methylpent-2-enoic Acid



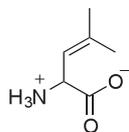
^1H NMR, 400 MHz, D_2O

3,4-Dehydroleucine (21)

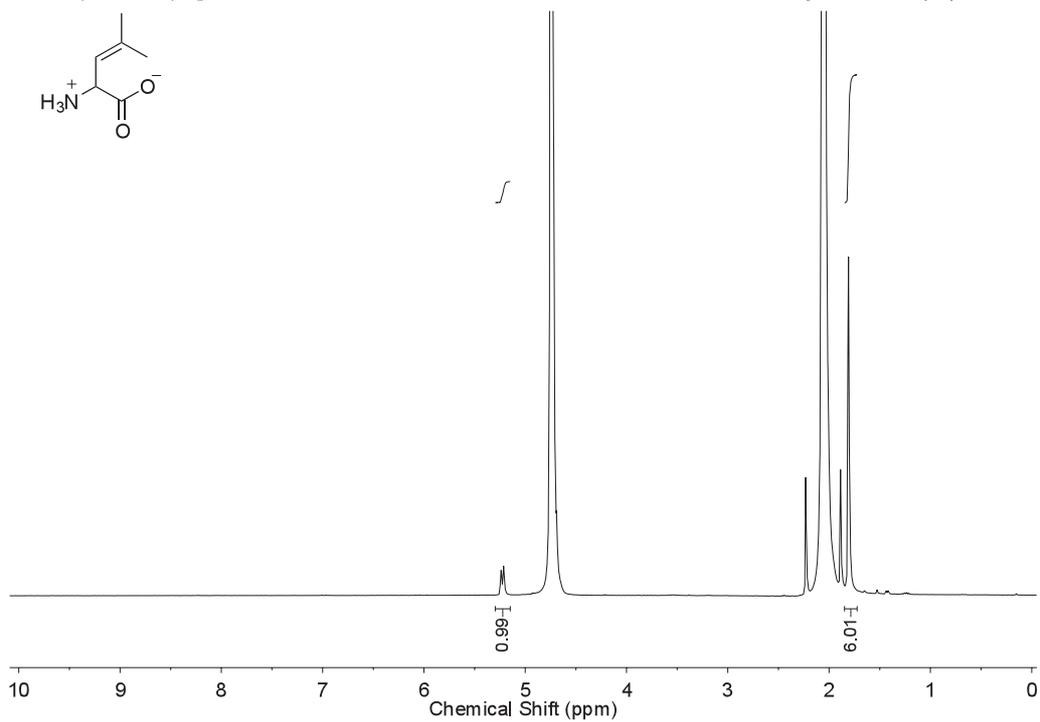


Chlorinated Amino Acids in Peptide Production
Chapter 3. Cleavage of Peptide Bonds

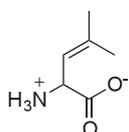
^1H NMR, 400 MHz, D_2O



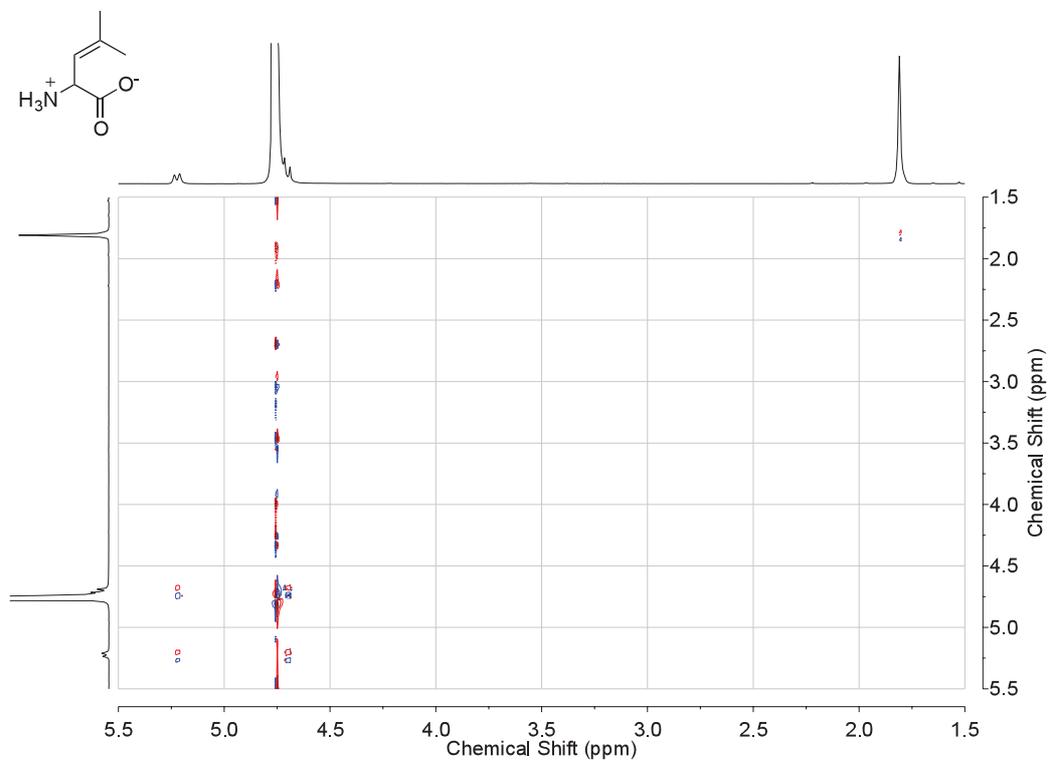
3,4-Dehydroleucine (21) & MeCN



^1H - ^1H COSY, 400 MHz, D_2O

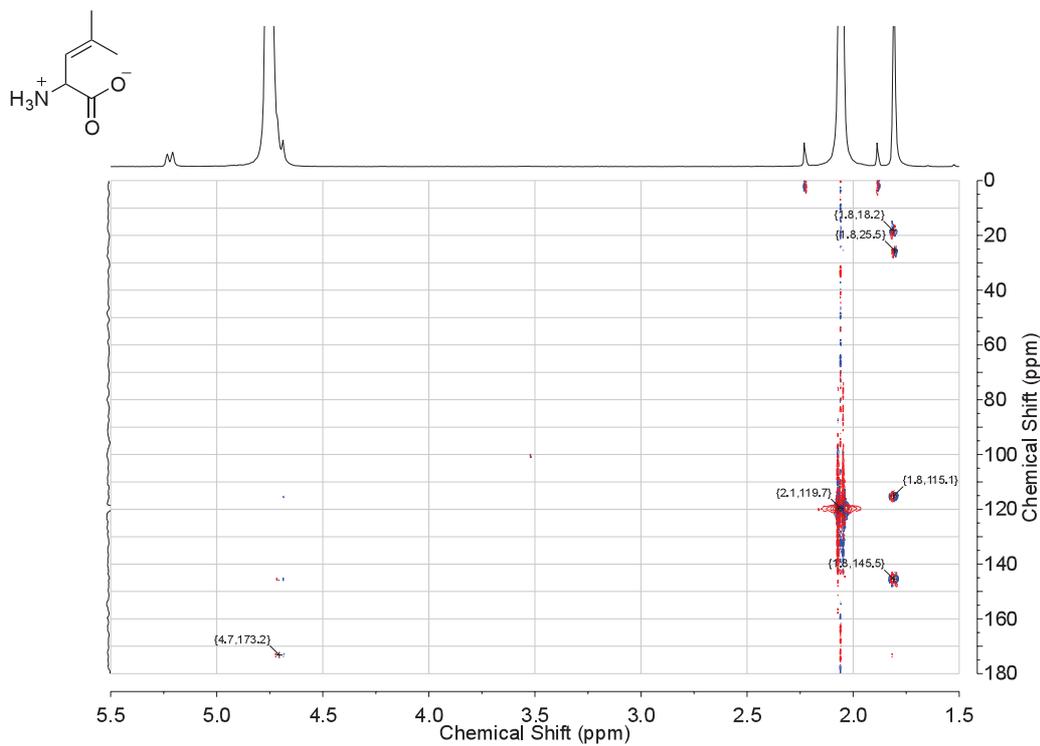


3,4-Dehydroleucine (21)



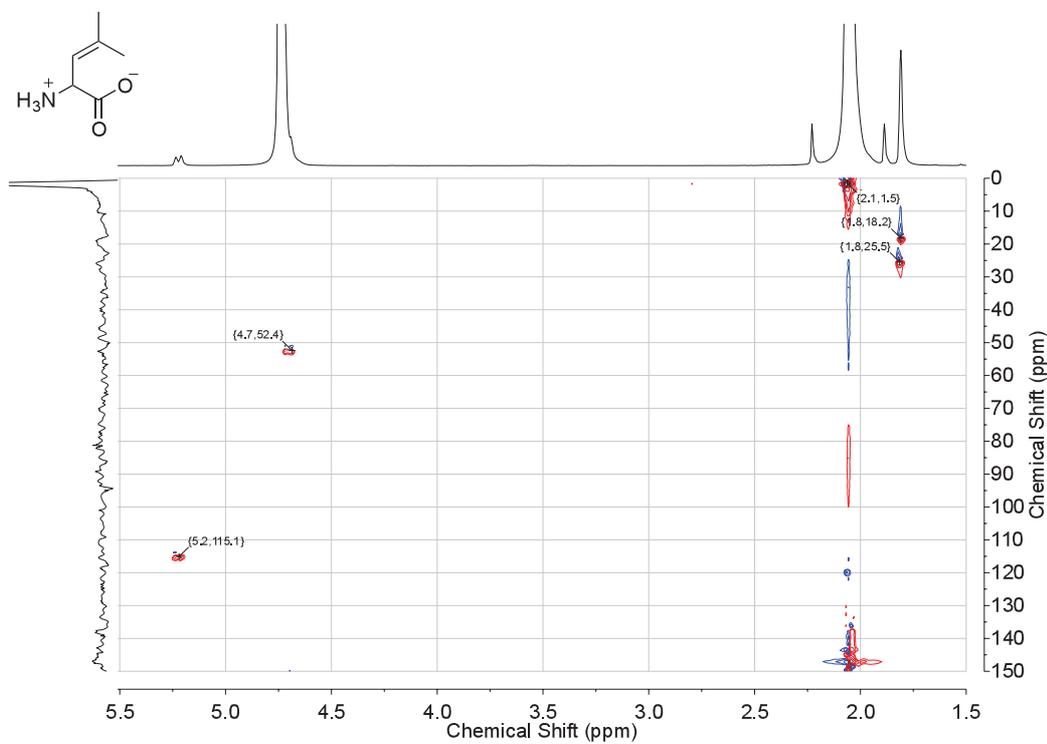
^1H - ^{13}C HMBC, 400 MHz, D_2O

3,4-Dehydroleucine (21) & MeCN



^1H - ^{13}C HSQC, 400 MHz, D_2O

3,4-Dehydroleucine (21) & MeCN

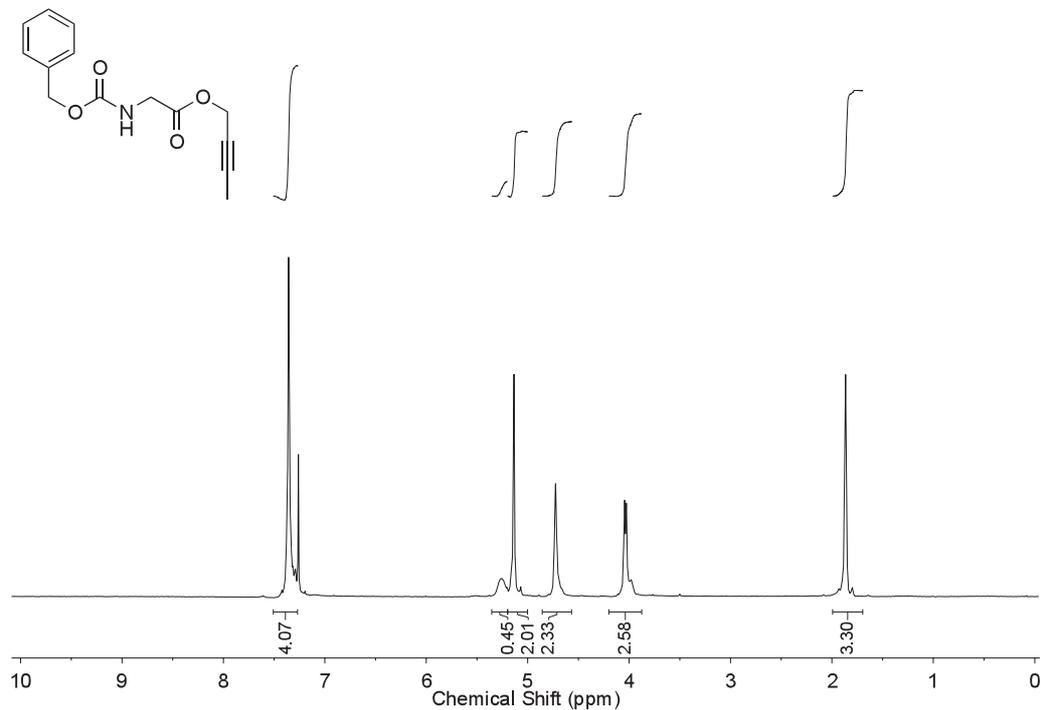


Chlorinated Amino Acids in Peptide Production

Chapter 3. Cleavage of Peptide Bonds

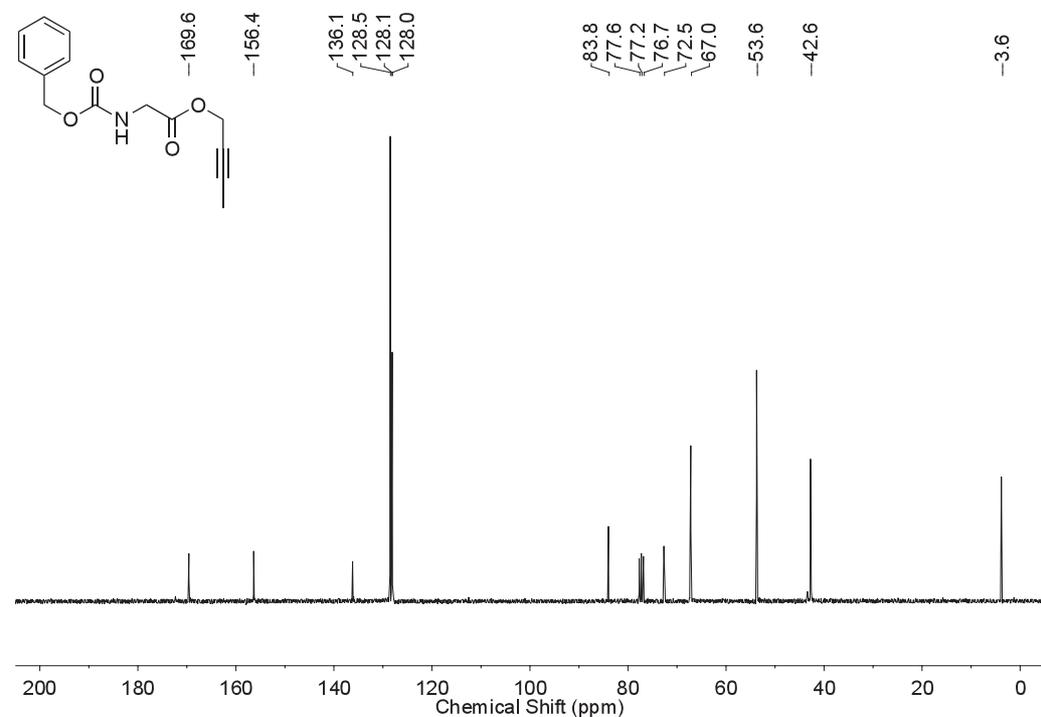
¹H NMR, 300 MHz, CDCl₃

N-Cbz-glycine But-2-ynyl Ester



¹³C NMR, 75 MHz, CDCl₃

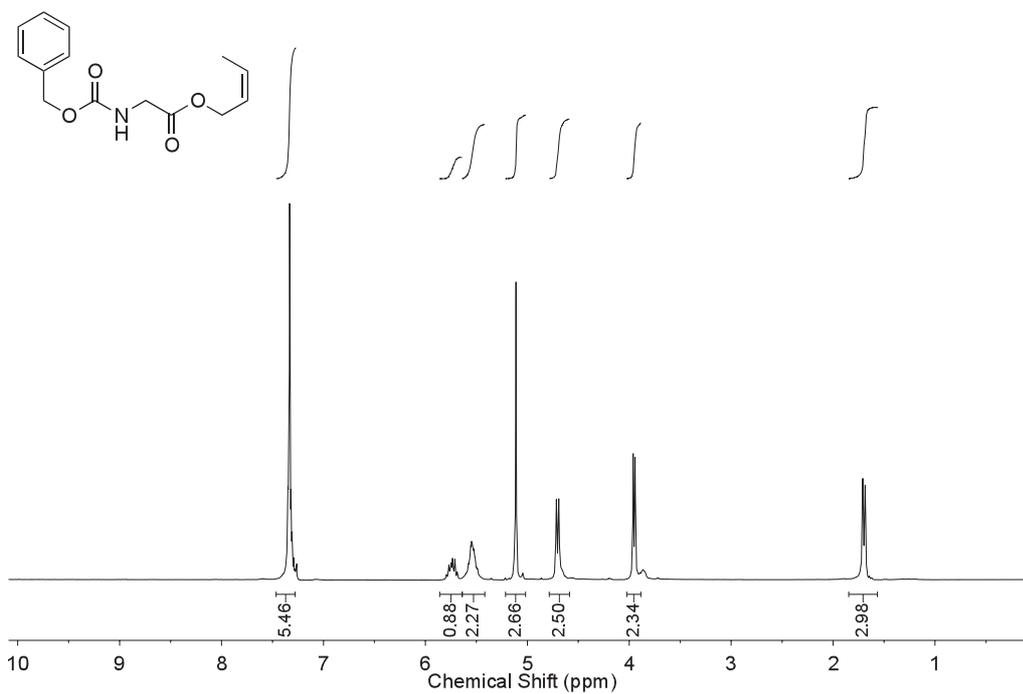
N-Cbz-glycine But-2-ynyl Ester



Chlorinated Amino Acids in Peptide Production
Chapter 3. Cleavage of Peptide Bonds

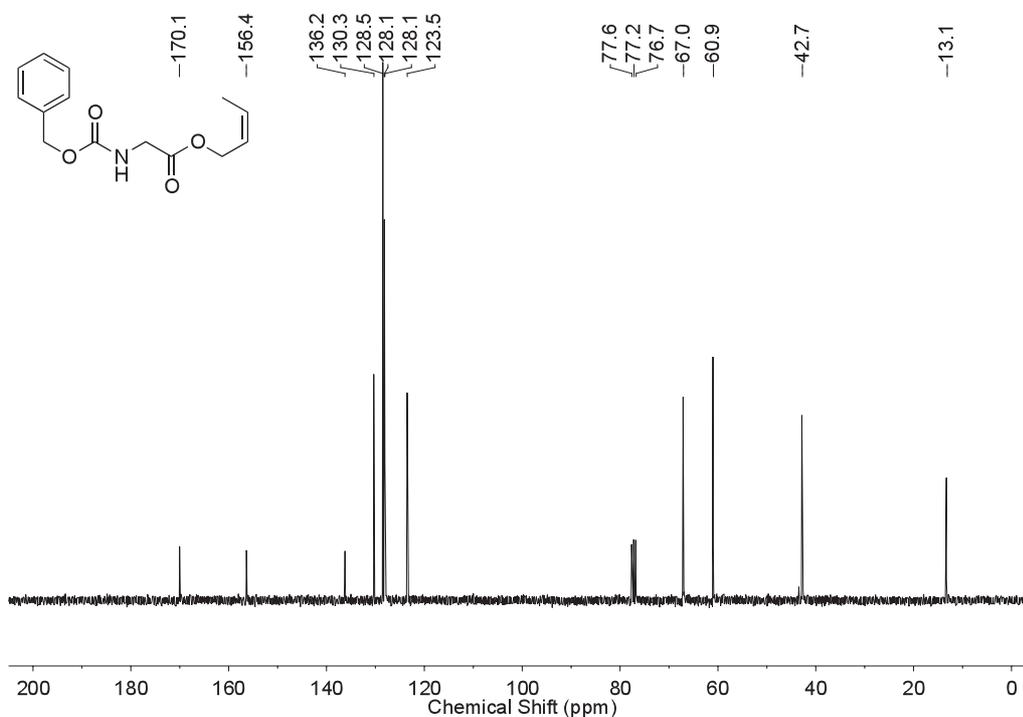
¹H NMR, 300 MHz, CDCl₃

(Z)-N-Cbz-glycine But-2-enyl Ester



¹³C NMR, 75 MHz, CDCl₃

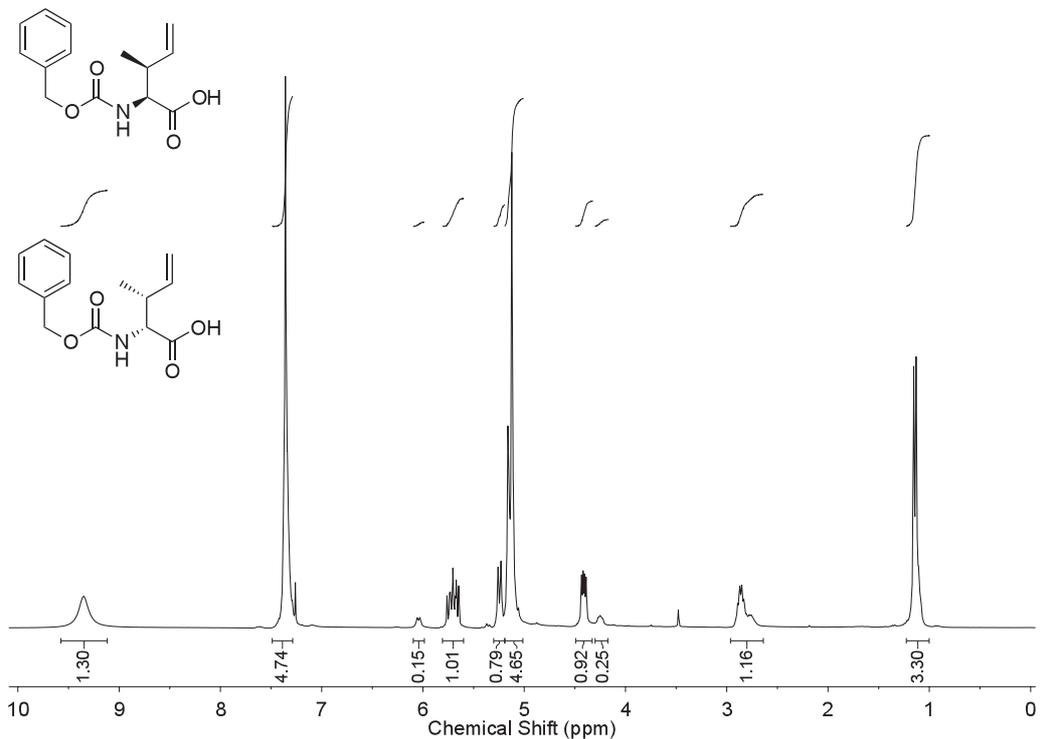
(Z)-N-Cbz-glycine But-2-enyl Ester



Chlorinated Amino Acids in Peptide Production
Chapter 3. Cleavage of Peptide Bonds

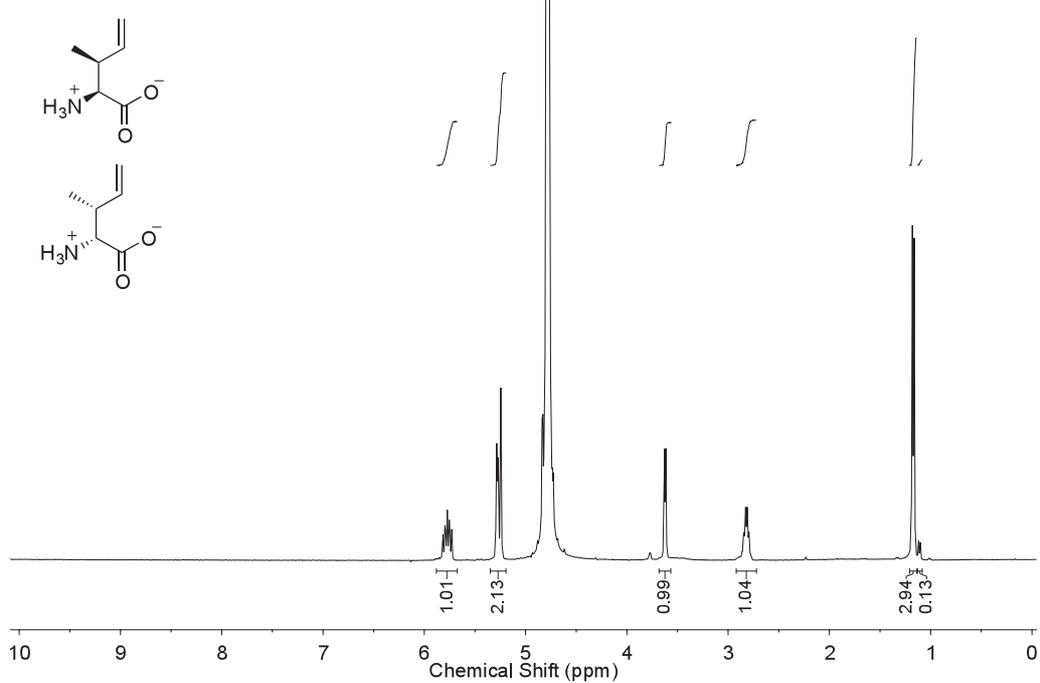
^1H NMR, 300 MHz, CDCl_3

(2*RS*,3*RS*)-*N*-Cbz-4,5-dehydroisoleucine



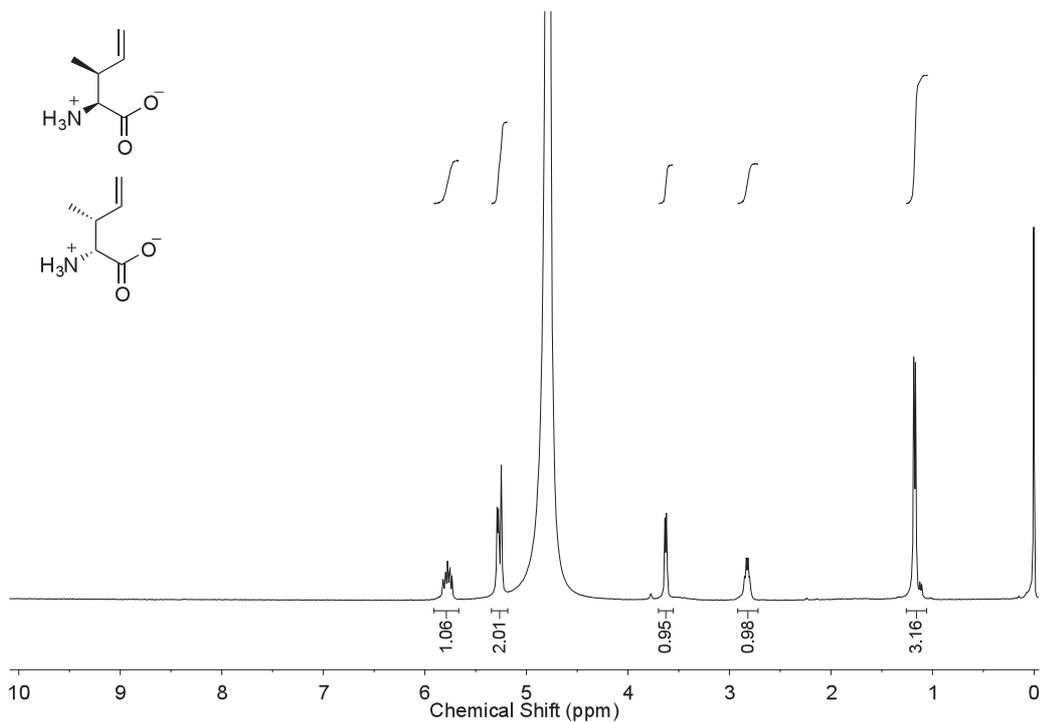
^1H NMR, 400 MHz, D_2O

(2*RS*,3*RS*)-4,5-Dehydroisoleucine (23)



^1H NMR, 400 MHz, D_2O

(2*RS*,3*RS*)-4,5-Dehydroisoleucine (23) & $[\text{D}_4]\text{TSP}$



3. Gene construction

Phusion® DNA polymerase, restriction enzymes and a Quick DNA Ligation™ kit were obtained from New England BioLabs® Inc. Alkaline phosphatase was from Roche Diagnostics GmbH. A Bio-RAD C-1000 Thermal Cycler was used for PCR. DNA electrophoresis was carried out in a Bio-RAD Wide Mini-Sub® Cell GT electrophoresis kit, electrophoresis-grade agarose was purchased from Invitrogen™, and a QIAquick® Gel extraction kit was used to recover DNA. For protein expression, plasmid DNA was amplified in *E. coli* AN1459 cells and isolated using Qiagen® Maxi or Mega kits. The PpiB_His₆_CTG gene was constructed by the fusion of PpiB and His₆_CTG genes using the two-step PCR mutagenesis approach. The PpiB gene was available as described previously,^[6] the His₆Gly₁₀Asp₄Lys_CTG gene was ordered from GeneArt AG, and DNA primers were ordered from IDT®. The PCR amplicon was then digested with restrictases NdeI and EcoRI (HF) and ligated into the expression vector^[15] with matching sticky overhangs. The same approach was followed for the construction of the His₆PpiB_GRP, His₆PpiB_CCK-7-G and His₆PpiB_P RPG genes, except the GRP, CCK-7-G and PRP gene segments were obtained from IDT® in the form of single-strand DNA and the hexahistidine tag was encoded by the vector.^[15]

Gene integrity was corroborated by nucleotide sequence determination (Biomolecular Resource Facility, Australian National University). Below are the sequences of His₆PpiB and the fusion proteins. The hormones are displayed in bold, while the cleavage sequences are shown in red.

His₆PpiB (19.2 kDa)

```
MHHHHHHMVT FHTNHGDIVI KTFDDKAPET VKNFLDYCRE GFYNNTIFHR
VINGFMIQGG GFEPGMKQKA TKEPIKNEAN NGLKNTRGTL AMARTQAPHS
ATAQFFINNV DNDFLNFSGE SLQGWGYCVF AEVVDGMDVV DKIKGVATGR
SGMHQDVPKE DVIIIESVTVS EN
```

PpiB_His₆_CTG (23.6 kDa)

```
MVTFHTNHGD IVIKTFDDKA PETVKNFLDY CREGFYNN TI FHRVINGFMI
QGGGFEPGMK QKATKEPIKN EANGLKNTR GTLAMARTQA PHSATAQFFI
NVVDNDFLNF SGESLQGWGY CVFAEVVDGM DVVDKIKGVA TGRSGMHQDV
PKEDVIIESV TVSEHHHHHH GGGGGGGGGG DDDDKCGNLS TCMLGTYTQD
FNKFHTFPQT AIGVGAPG
```

His₆PpiB_GRP (22.7 kDa)

```
MHHHHHHMVT FHTNHGDIVI KTFDDKAPET VKNFLDYCRE GFYNNTIFHR
VINGFMIQGG GFEPGMKQKA TKEPIKNEAN NGLKNTRGTL AMARTQAPHS
ATAQFFINNV DNDFLNFSGE SLQGWGYCVF AEVVDGMDVV DKIKGVATGR
SGMHQDVPKE DVIIIESVTVS ENGGGGGGGA IVPLPAGGGT VLTKMYPRGN
HWAVGHLMG
```

His₆PpiB_CCK-7-G (20.8 kDa)

MHHHHHMMVT FHTNHGDIVI KTFDDKAPET VKNFLDYCRE GFYNNTIFHR
VINGFMIQGG GFEPGMKQKA TKEPIKNEAN NGLKNTRGTL AMARTQAPHS
ATAQFFINVV DNDFLNFSGE SLQGWGYCVF AEVVDGMDVV DKIKGVATGR
SGMHQDVPKE DVIIESVTVS ENGGGGGGGA **LYMGWMDFG**

His₆PpiB_PRP (23.6 kDa)

MHHHHHMMVT FHTNHGDIVI KTFDDKAPET VKNFLDYCRE GFYNNTIFHR
VINGFMIQGG GFEPGMKQKA TKEPIKNEAN NGLKNTRGTL AMARTQAPHS
ATAQFFINVV DNDFLNFSGE SLQGWGYCVF AEVVDGMDVV DKIKGVATGR
SGMHQDVPKE DVIIESVTVS ENGGARGGGA **LSRTHRHSME IRTPDINPAW**
YASRGIRPVG RFG

4. Cell-free protein expression

Cell-free protein synthesis was carried out at 30 °C for 18 h, or at 37 °C for 4 h, following a literature procedure,^[2,6,16-19] with a 1 mM concentration of each of the twenty amino acids normally found in proteins, or by: withholding (*S*)-phenylalanine and instead using a 2 mM concentration of (*S*)-3-fluorophenylalanine (**6**) or a 4 mM concentration of the racemate of one of the amino acids **5**, **7** and **8**; withholding (*S*)-tyrosine and instead using a 4 mM concentration of the racemate of the chloride **9** or the fluoride **10**; withholding (*S*)-valine (**11**) and instead using a 2 mM concentration of one of the chlorides **14** and **15** or a 4 mM concentration of the racemate of the dehydrovaline **20**; withholding (*S*)-leucine (**12**) and instead using a 2 mM concentration of perdeuterated (*S*)-leucine (**d₁₀-12**), the chlorides **16** and **17**, one of the corresponding methyl esters, or the (*S*)-dehydroleucine **22**, a 4 mM concentration of the racemate of the amino acid **21**, or a mixture comprising 4, 16 and 4 mM concentrations of the racemates of the dehydroisoleucines **24–26**; and/or withholding (*2S,3S*)-isoleucine (**13**) and instead using a 2 mM concentration of the methyl ester of the chloride **18**, or a 4 mM concentration of the racemate of the chloride **19** or the dehydroisoleucine **23**. Racemic materials are suitable for use in incorporation experiments simply by doubling the quantity used, because there is negligible misincorporation of (*2R*)-amino acids.^[14,20-21]

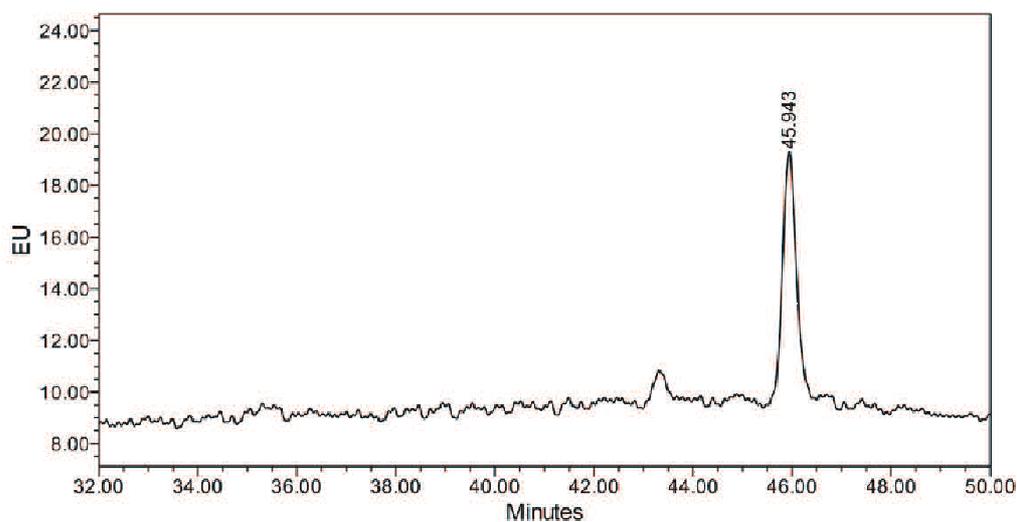
Insoluble protein was collected by centrifugation and the remaining contents of the expression chamber were purified using His GraviTrap™, HisTrap™ HP or His SpinTrap™ columns from GE Healthcare using conditions recommended by the manufacturer. Amicon® Ultra-4 (YM-3,000) centrifugal filter devices from Millipore were used for sample concentration and buffer removal. Protein concentrations were measured on a NanoDrop ND-1000 spectrophotometer from ThermoScientific, with the extinction coefficients being estimated using online ProtParam tool. Gels (20% acrylamide) for SDS-PAGE analysis were cast using acrylamide and bis-acrylamide solutions from Bio-RAD and run on a Mini-PROTEIN® Tetra system. Bio-Safe® Coomassie Blue was used for gel staining. Protein markers were purchased from New England Biolabs® Inc.

5. Production of CTG (1a) and analogues 1b-g

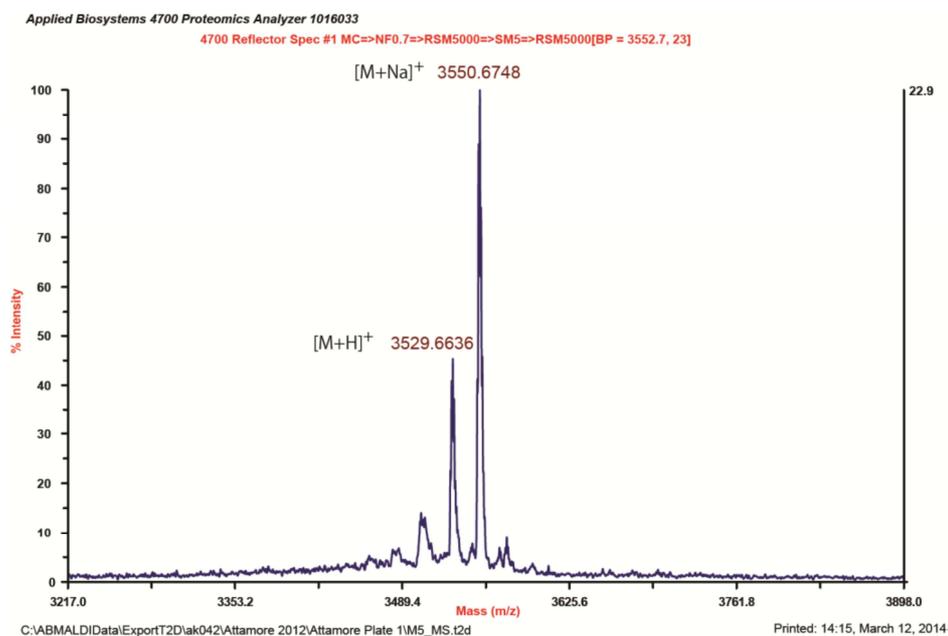
Soluble PpiB_His₆_CTG fusion protein, expressed using normal and modified amino acids and purified as described above (section 4), was incubated with enterokinase light chain, purchased from New England Biolabs® Inc., at 24 °C for 18 h, in buffer provided by the manufacturer. The digestion mixture was fractionated by HPLC using a Waters Alliance Separation Module 2695 with a Waters 2996 Photodiode Array Detector (Synergi C₁₂ Max-RP 250 x 4.6 mm, 4 μm), and a linear gradient mobile phase of 20-30% MeCN in buffer over 40 min, followed by 10 min using 30% MeCN, at a flow rate of 1 mL min⁻¹. Collected fractions were desalted for analysis using an Oasis® HLB 2.1 × 20 mm 25 μm cartridge column. Mass spectra of the peptides were recorded on an Agilent 1100 series LC/MSD TOF instrument (direct injection) or an AB MDS Sciex 4800 MALDI-TOF-TOF Mass Analyzer on a 4-hydroxyazobenzene-2-carboxylic acid (8 mg mL⁻¹ in 70% MeCN + 0.1% TFA) matrix. CTG (**1a**) was identical with a commercial standard. Characterization data for the CTG analogues **1b-g** are provided below.

5₃-CTG (1b): HPLC

SAMPLE INFORMATION			
Sample Name:	CTG-4F-Phe_1_pure_12May2013		
Acq. Method Set:	Calcitonin Analysis 50min	Sample Set Name:	
Vial:	16	Processing Method:	Calcitonin Analysis
Injection #:	1	Channel Name:	2475ChA ex390/em470
Injection Volume:	200.00 ul	Proc. Chnl. Descr.:	2475ChA ex390/em470
Run Time:	50.0 Minutes		
Date Acquired:	11/1/2013 11:45:09 PM EST		
Date Processed:	1/7/2016 7:07:32 PM EST		

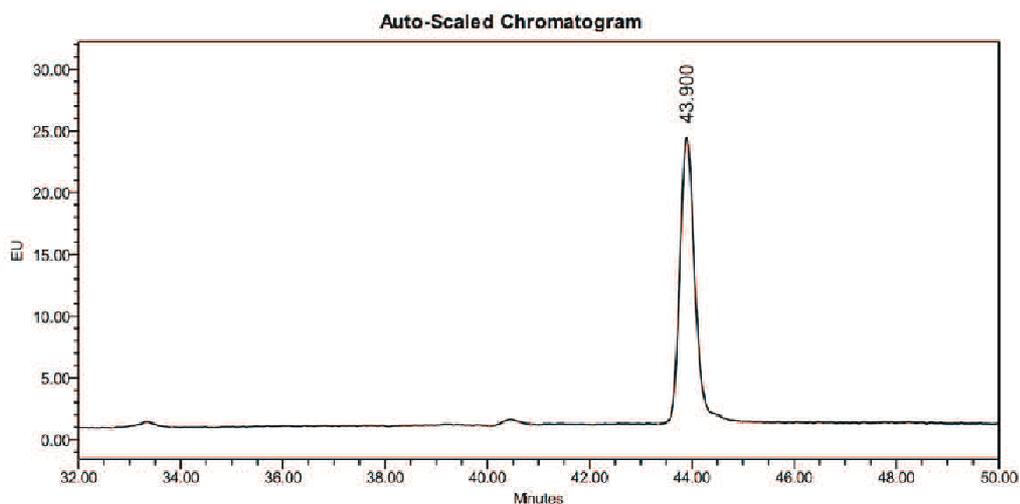


5₃-CTG (1b): MALDI-TOF-MS

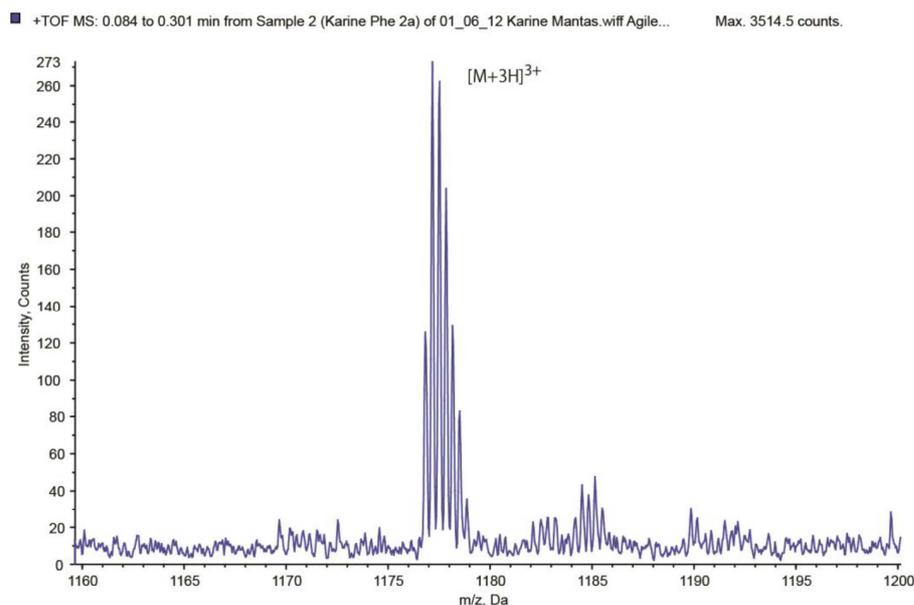


6₃-CTG (1c): HPLC

SAMPLE INFORMATION		
Sample Name:	CTG PheF 3 27Nov2012	Sample Set Name:
Sample Type:	Unknown	Acq. Method Set: Calcitonin Analysis 50min
Vial:	80	Processing Method: Calcitonin analysis
Injection #:	1	Channel Name: 2475ChA ex390/em470
Injection Volume:	50.00 ul	Proc. Chnl. Descr.: 2475ChA ex390/em470
Run Time:	50.0 Minutes	
Date Acquired:	11/28/2012 3:39:28 AM EST	
Date Processed:	1/7/2016 5:47:34 PM EST	

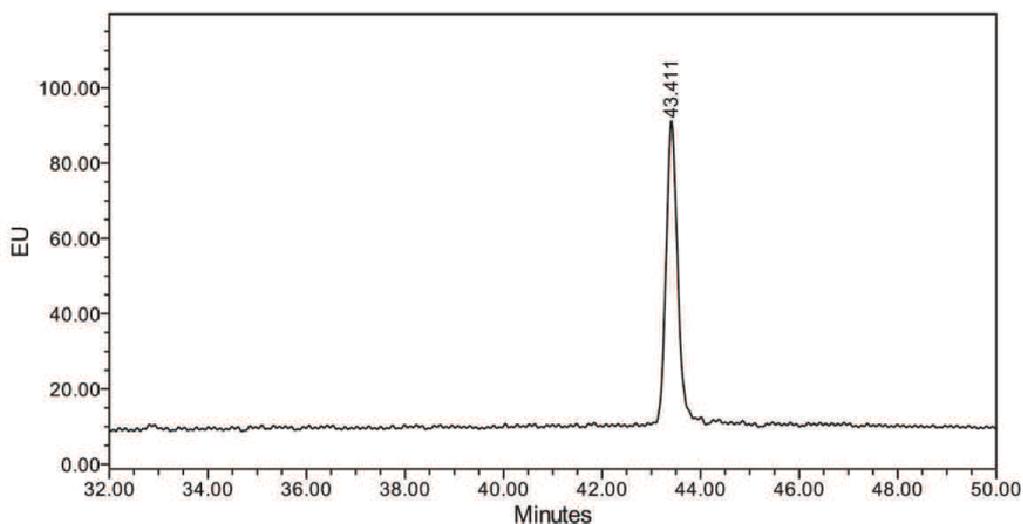


6₃-CTG (1c): ESI-MS

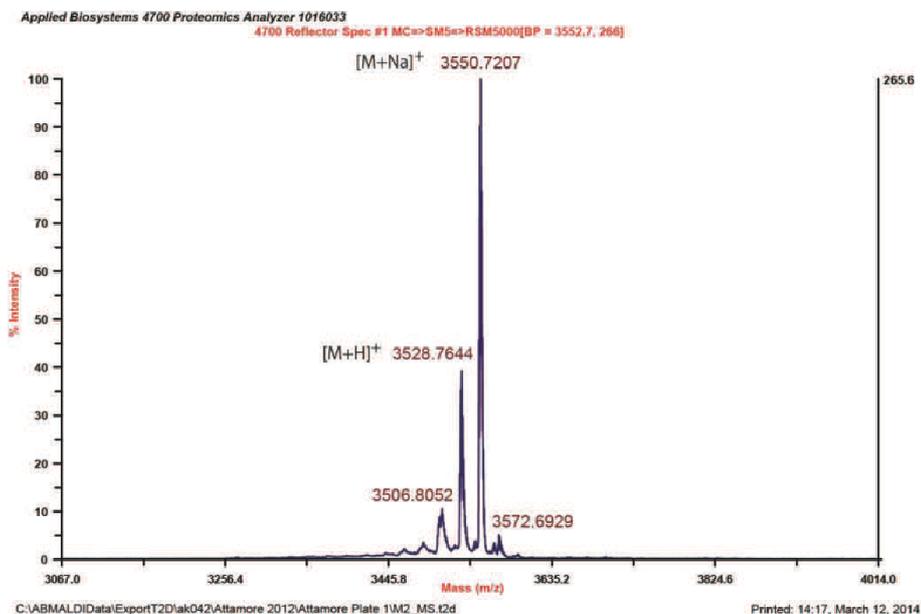


7₃-CTG (1d): HPLC

SAMPLE INFORMATION			
Sample Name:	CTG-2F-Phe_2_pure_12May2013		Sample Set Name:
Acq. Method Set:	Calcitonin Analysis 50min	Processing Method:	Calcitonin Analysis
Vial:	14	Channel Name:	2475ChA ex390/em470
Injection #:	1	Proc. Chnl. Descr.:	2475ChA ex390/em470
Injection Volume:	200.00 ul		
Run Time:	50.0 Minutes		
Date Acquired:	11/1/2013 8:57:37 PM EST		
Date Processed:	1/7/2016 7:08:21 PM EST		

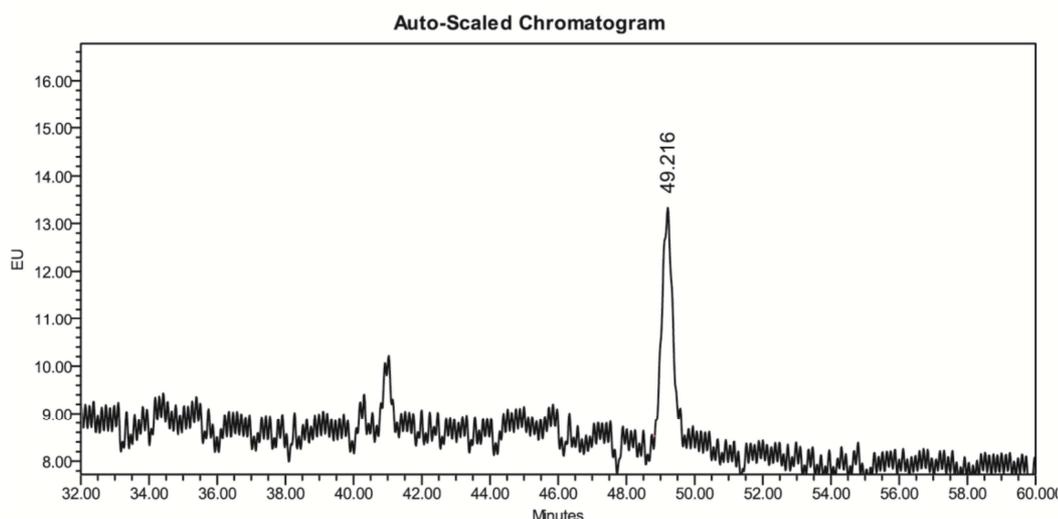


7₃-CTG (1d): MALDI-TOF-MS



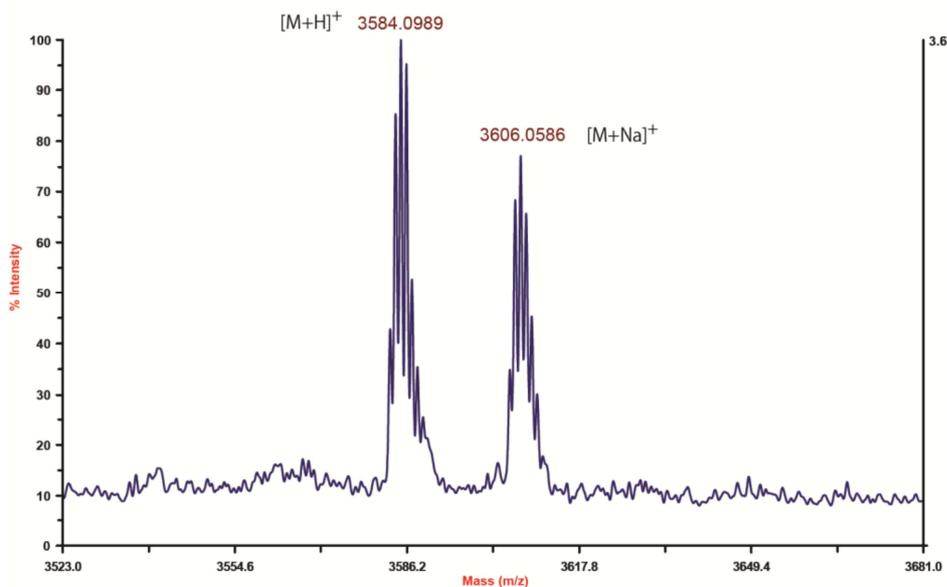
8₃-CTG (1e): HPLC

SAMPLE INFORMATION			
Sample Name:	PpiBCTG_Phe3-5F_pure2_May2013		Sample Set Name:
Sample Type:	Unknown	Acq. Method Set:	Calcitonin Analysis 60min
Vial:	1	Processing Method:	Calcitonin analysis
Injection #:	1	Channel Name:	2475ChA ex390/em470
Injection Volume:	50.00 ul	Proc. Chnl. Descr.:	2475ChA ex390/em470
Run Time:	60.0 Minutes		
Date Acquired:	3/10/2014 2:27:04 PM EST		
Date Processed:	1/7/2016 5:39:17 PM EST		



8₃-CTG (1e):MALDI-TOF-MS

Applied Biosystems 4700 Proteomics Analyzer 1016033
4700 Reflector Spec #1 MC=>BC=>SM5=>RSM5000=>NF0.7[BP = 1643.8, 6]



C:\ABMALDIData\ExportT2\lak042\Attamore 2012\Attamore Plate 1\1D15_MS.t2d

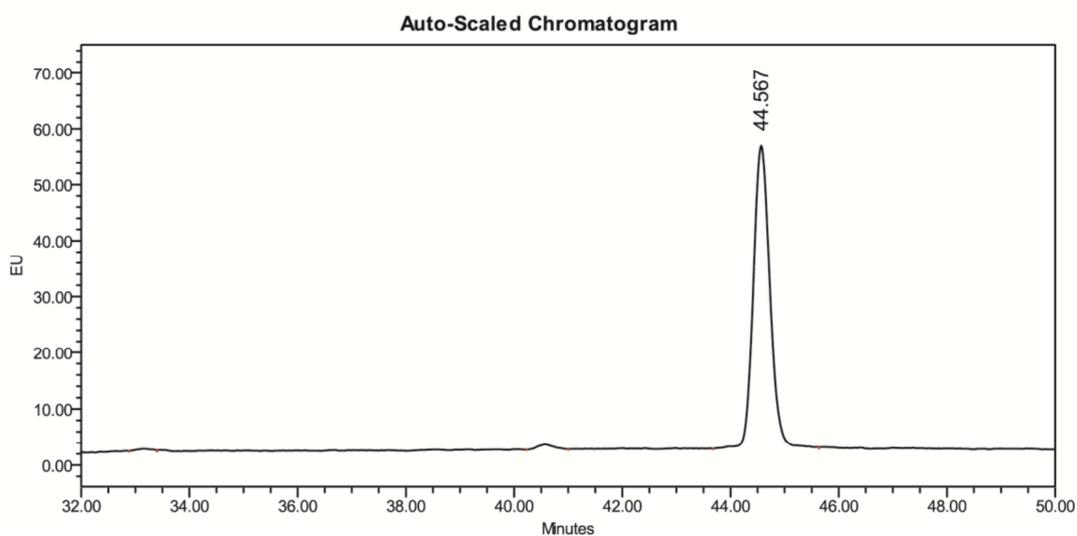
Printed: 13:49, March 12, 2014

Chlorinated Amino Acids in Peptide Production

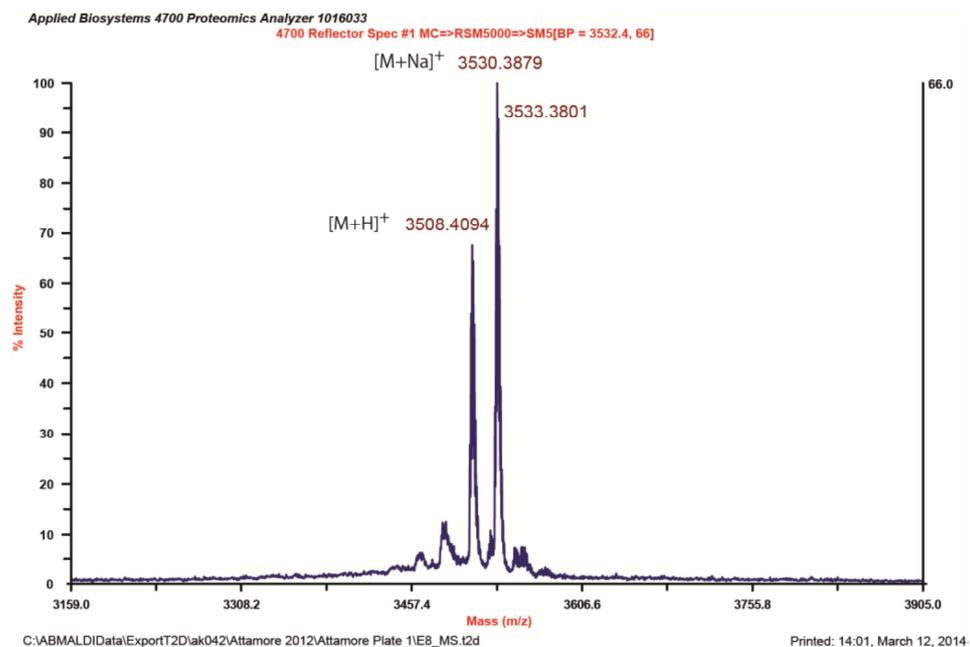
Chapter 3. Cleavage of Peptide Bonds

9-CTG (1f): HPLC

SAMPLE INFORMATION			
Sample Name:	CTG TyrCl 4 27Nov2012	Sample Set Name:	
Sample Type:	Unknown	Acq. Method Set:	Calcitonin Analysis 50min
Vial:	76	Processing Method:	Calcitonin analysis
Injection #:	1	Channel Name:	2475ChA ex390/em470
Injection Volume:	50.00 ul	Proc. Chnl. Descr.:	2475ChA ex390/em470
Run Time:	50.0 Minutes		
Date Acquired:	11/27/2012 10:43:53 PM EST		
Date Processed:	1/7/2016 5:49:05 PM EST		

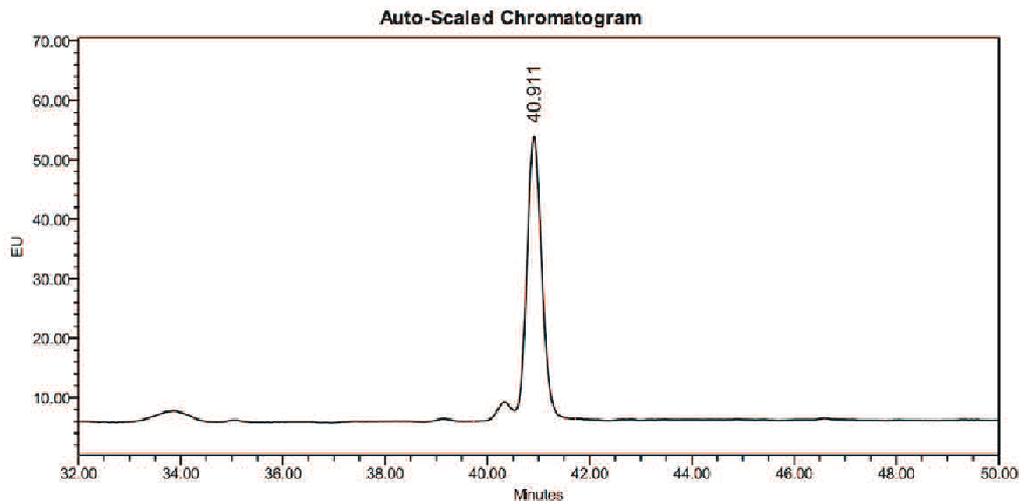


9-CTG (1f): MALDI-TOF-MS



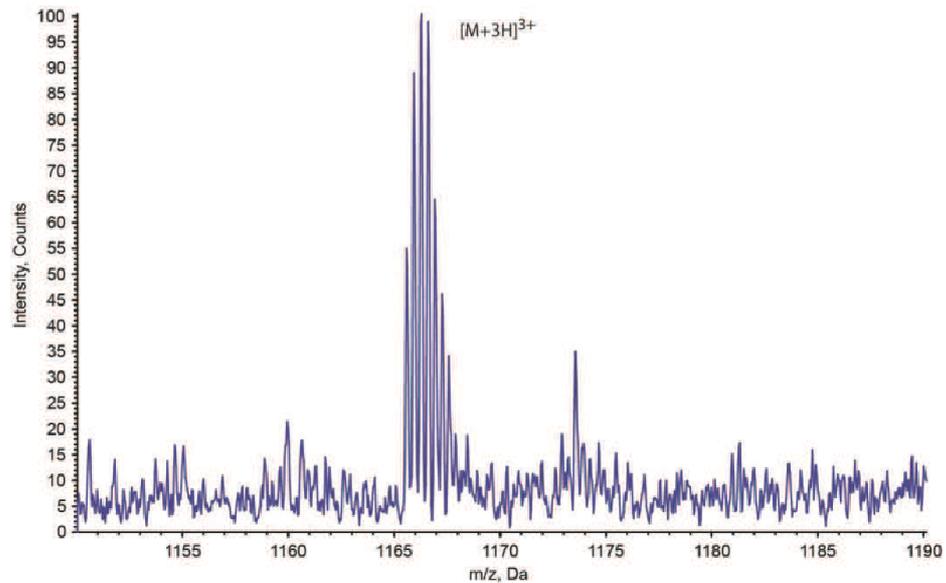
14-CTG (1g): HPLC

SAMPLE INFORMATION		
Sample Name:	CTG-R-ValCl 1 17Jan2012	Sample Set Name:
Sample Type:	Unknown	Acq. Method Set: Calcitonin Analysis 50min
Vial:	28	Processing Method: Calcitonin analysis
Injection #:	1	Channel Name: 2475ChA ex390/em470
Injection Volume:	100.00 ul	Proc. Chnl. Descr.: 2475ChA ex390/em470
Run Time:	50.0 Minutes	
Date Acquired:	1/23/2012 9:35:39 PM EST	
Date Processed:	1/7/2016 5:36:51 PM EST	



14-CTG (1g): ESI-MS

■ +TOF MS: 0.096 to 0.259 min from Sample 10 (Karine 3.1) of 23_01_12 Sam Karine.wiff Agilent Max. 2.8e4 counts.



6. Analysis of heat treatment of His₆PpiB produced using chlorinated amino acids

His₆PpiB expressed using the methyl esters of the chlorides **16** and **17** as described above (section 4) was analysed by SDS-PAGE, before and after metal-ion affinity chromatography (Figure S1a). Mass spectra of the purified proteins are shown in Figure S1b.

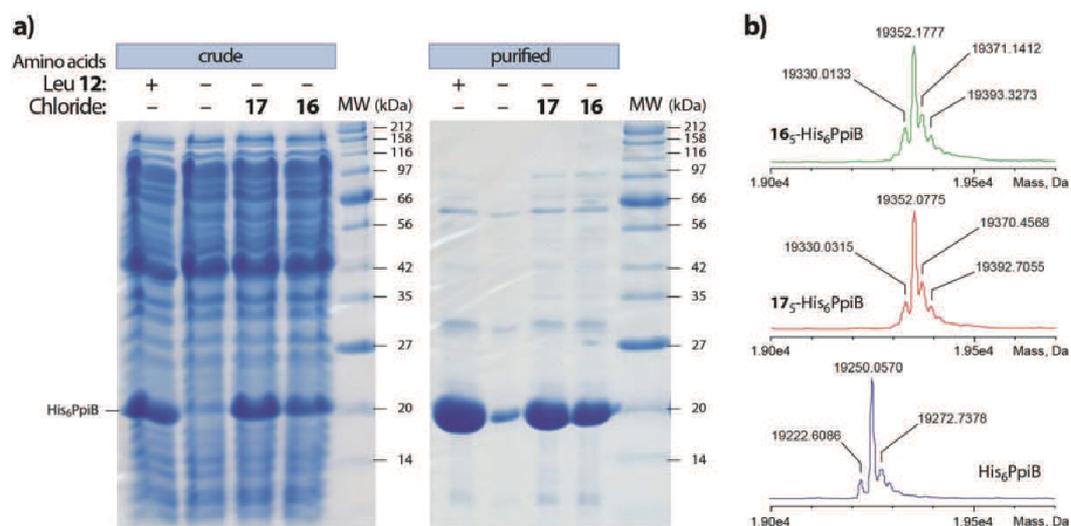


Figure S1. a) SDS-PAGE analysis of His₆PpiB expressed with the methyl esters of the chlorides **16** and **17** instead of (S)-leucine (**12**), and b) deconvoluted ESI-MS spectra of the corresponding proteins, showing full substitution (5/5) of leucine (**12**) with the chlorides **16** and **17** generated in situ from the esters.

His₆PpiB expressed using the chlorides **14**, **15** and **19** as described above was analysed by SDS-PAGE, before and after metal-ion affinity chromatography (Figure S2a). Mass spectra of corresponding purified proteins are shown in Figure S2b.

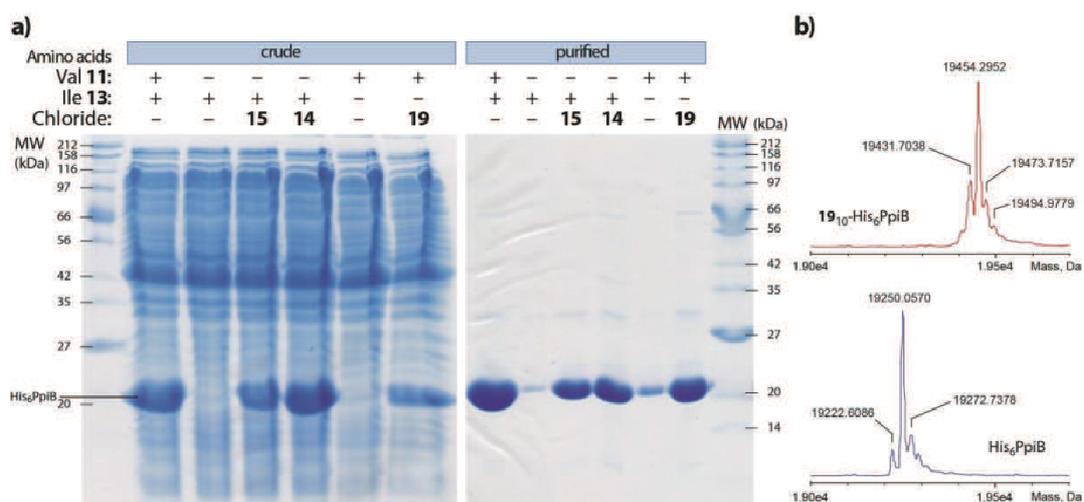


Figure S2. a) SDS-PAGE analysis of His₆PpiB expressed with the chlorinated amino acids **14**, **15** and **19** instead of (S)-valine (**11**) or (2S,3S)-isoleucine (**13**), and b) deconvoluted ESI-MS spectra of His₆PpiB

expressed with all normal amino acids (blue; 19250 Da *N*-formylated, 19222 Da deformedylated) and the chloride **19** instead of (2*S*,3*S*)-isoleucine (**13**) (red; 19454 Da *N*-formylated, 19431 Da deformedylated). The increase of mass by ca. 200 Da indicates complete (10/10) substitution of (2*S*,3*S*)-isoleucine **13** with the chloride **19**.

His₆PpiB expressed using the chlorides **14-19** and the methyl esters of the chlorides **16-18**, and purified using metal-ion affinity chromatography, as described above, was premixed with Laemmli sample buffer (80 mM Tris, pH = 6.8, 1% SDS, 100 mM DTT, 5% glycerol) in mini-centrifuge tubes which were then heated in a heat block for up to 1 h. After cooling, the samples were analysed by SDS-PAGE (Figures S3-S6).

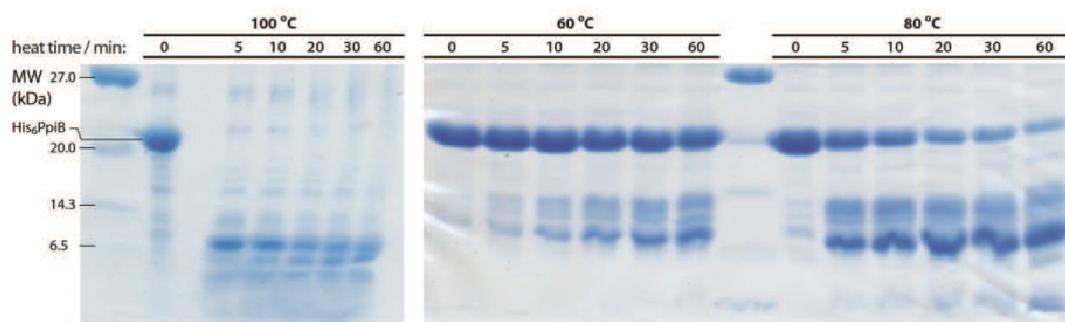


Figure S3. SDS-PAGE analysis of the effect of heat treatment on His₆PpiB expressed using the chlorides **16** and **17** instead of (*S*)-leucine (**12**).

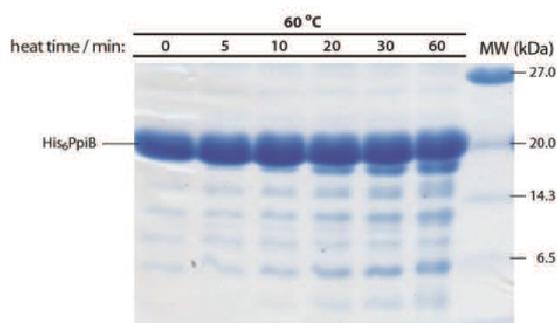


Figure S4. SDS-PAGE analysis of the effect of heat treatment on His₆PpiB expressed using the chloride **18** instead of (2*S*,3*S*)-isoleucine (**13**).

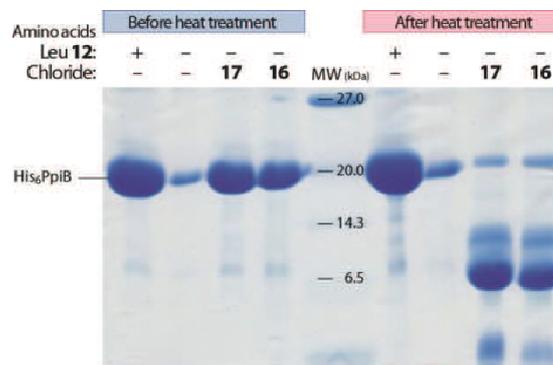


Figure S5. SDS-PAGE analysis of His₆PpiB expressed using the methyl esters of the chlorides **16** and **17** instead of (*S*)-leucine (**12**), before (left) and after (right) heat treatment for 5 min at 100 °C.

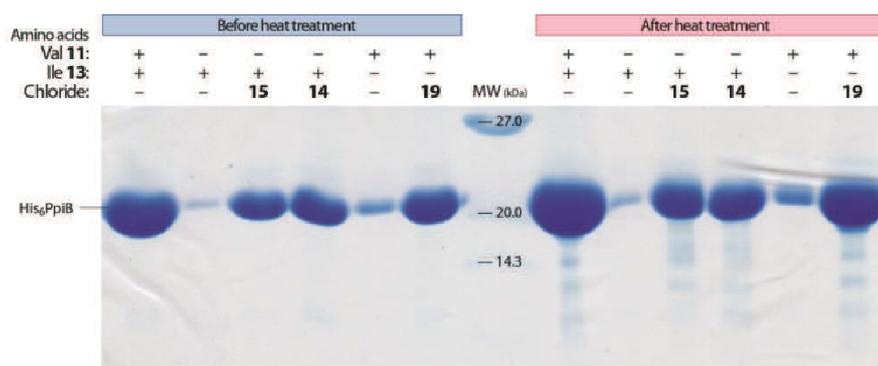


Figure S6. SDS-PAGE analysis of His₆PpiB expressed with the chlorides **14**, **15** and **19** instead of (*S*)-valine (**11**) and (2*S*,3*S*)-isoleucine (**13**), before (left) and after (right) heat treatment for 5 min at 100 °C.

His₆PpiB contains five residues of (*S*)-leucine (**12**). When substituted with the chlorides **16** and **17**, cyclization at each of these followed by hydrolysis to cleave the proximal carboxyl peptide bond according to the process illustrated in Scheme 2 would be expected to lead to the six fragments indicated in Figure S7 and Table S1. Portions of the ESI-MS spectrum of His₆PpiB expressed using the chlorides **16** and **17**, after heat treatment at 100 °C for 5 min, are shown in Figures S8 and S9. Peaks observed at *m/z* 773 and 737 correspond to [M+H]⁺ ions of KNTRGTλ and NFSGESλ, respectively, those around *m/z* 904 and 775 correspond to [M+6H]⁶⁺ and [M+7H]⁷⁺ ions of the fragment of mass 5420 Da, those around *m/z* 917 correspond to [M+3H]³⁺ ions of the fragment of mass 2749 Da, and those around *m/z* 698 correspond to [M+6H]⁶⁺ ions of the fragment of mass 4185 Da. Only the C-terminal fragment of mass 5396 Da was not detected. No fragments resulting from incomplete cleavage were detected in the sample.

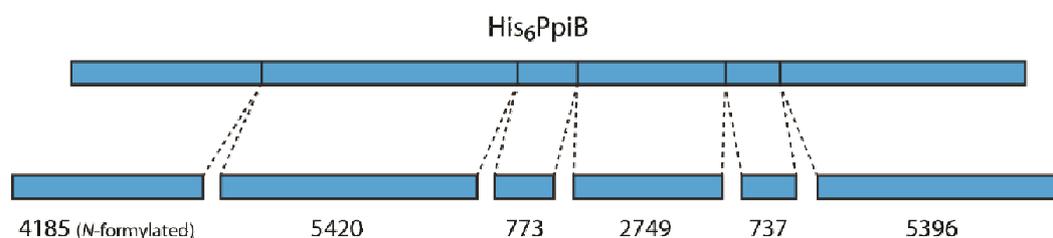


Figure S7. Schematic representation of peptide fragments generated by cleavage of His₆PpiB expressed using the chlorides **16** and **17**. The masses are listed in Daltons.

Table S1. Peptide fragments expected to be generated by cleavage of His₆PpiB expressed using the chlorides **16** and **17**.

Fragment	Mass, Da
MHHHHHMMVT FHTNHGDIVI KTFDDKAPET VKNFλ*	4157,
	4185 (<i>N</i> -formylated)
DYCREGFYNN TIFHRVINGF MIQGGGFEPG MKQKATKEPI KNEANNGλ*	5420
KNTRGTλ*	773
AMARTQAPHS ATAQFFINNV DNDFλ*	2749
NFSGESλ*	737
QGWGYCVFAE VVDGMDVVDK IKGVATGRSG MHQDVPKEDV IIESVTVSEN	5396

*λ denotes the lactone produced by the fragmentation process

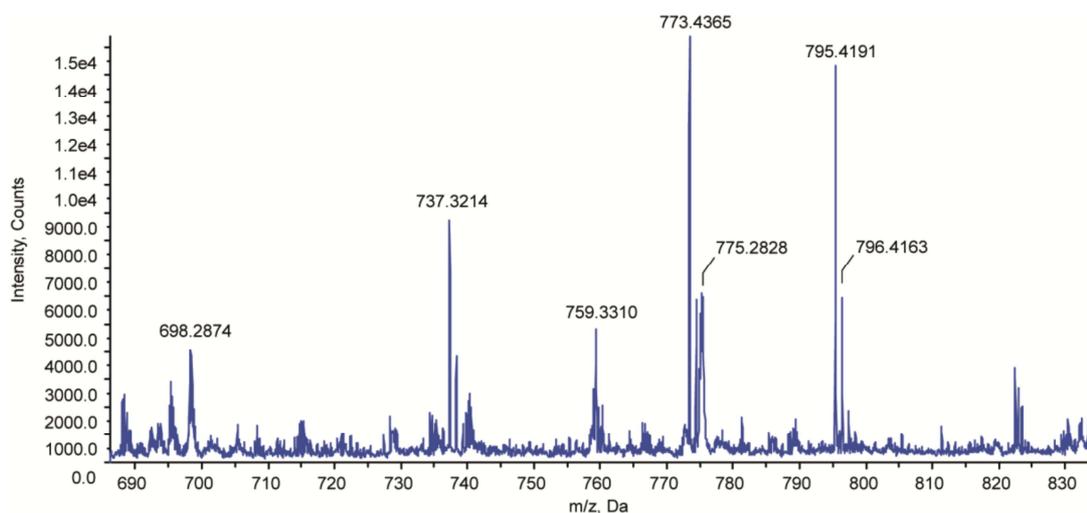


Figure S8. A portion of the ESI-MS spectrum of His₆PpiB expressed with the chlorides **16** and **17**, after heat treatment for 5 min at 100 °C.

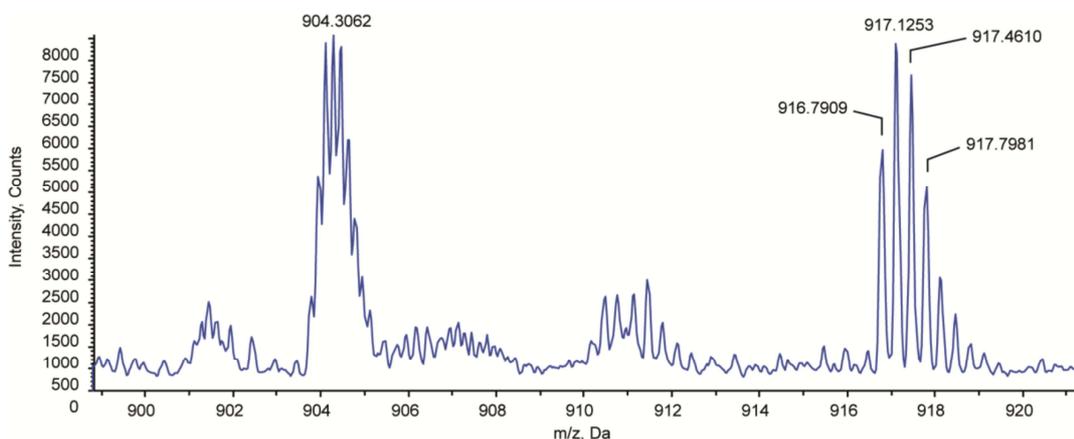


Figure S9. Another portion of the ESI-MS spectrum of His₆PpiB expressed with the chlorides **16** and **17**, after heat treatment for 5 min at 100 °C.

His₆PpiB contains ten residues of (2*S*,3*S*)-isoleucine (**13**). When substituted with the chloride **18**, cyclization at each of these followed by hydrolysis to cleave the proximal carboxyl peptide bond according to the process illustrated in Scheme 2 would be expected to lead to the eleven fragments indicated in Table S2. Portions of the ESI-MS spectrum of His₆PpiB expressed using the chloride **18**, after heat treatment at 100 °C for 5 min, are shown in Figures S10-S12. The peak observed at *m/z* 864 corresponds to the [M+H]⁺ ion of the C-terminal fragment of mass 863 Da, those around *m/z* 737 and 747 correspond to [M+3H]³⁺ ions of the N-terminal fragment (2210 Da) and the *N*-formylated derivative, those around *m/z* 655 correspond to [M+H]⁺ ions of the fragment of mass 654 Da, and those around *m/z* 862 and 690 correspond to [M+4H]⁴⁺ and [M+5H]⁵⁺ ions of the fragment of mass 3442 Da.

Chlorinated Amino Acids in Peptide Production

Chapter 3. Cleavage of Peptide Bonds

Table S2. Peptide fragments expected to be generated by cleavage of His₆PpiB expressed using the chloride **18**.

Fragment	Mass, Da
MHHHHHMMVT FHTNHGDλ*	2210
Vλ*	214
KTFDDKAPET VKNFLDYCRE GFYNNNTλ*	3214
FHRVλ*	654
NGFMλ*	564
QGGGFEPGMK QKATKEPλ*	1887
KNEANGLKN TRGTLAMART QAPHSATAQF Fλ*	3442
NVVDNDFLNF SGESLQGWGY CVFAEVVDGM DVVDKλ*	3964
KGVATGRSGM HQDVPKEDVλ*	2107
λ*	115
ESVTVSEN	863

*λ denotes the lactone produced by the fragmentation process

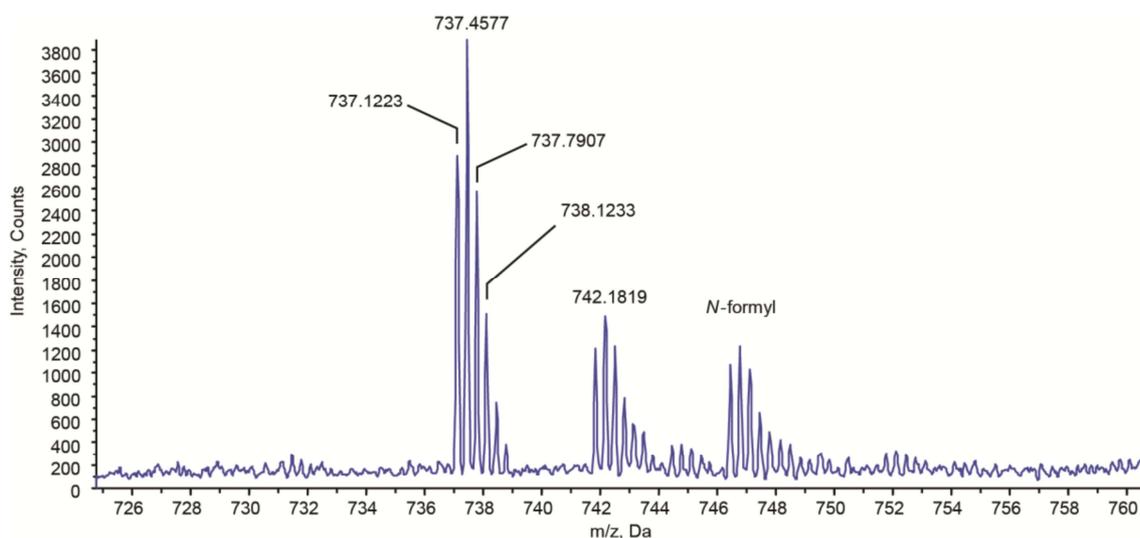


Figure S10. A portion of the ESI-MS spectrum of His₆PpiB expressed with the chloride **18**, after heat treatment for 5 min at 100 °C.

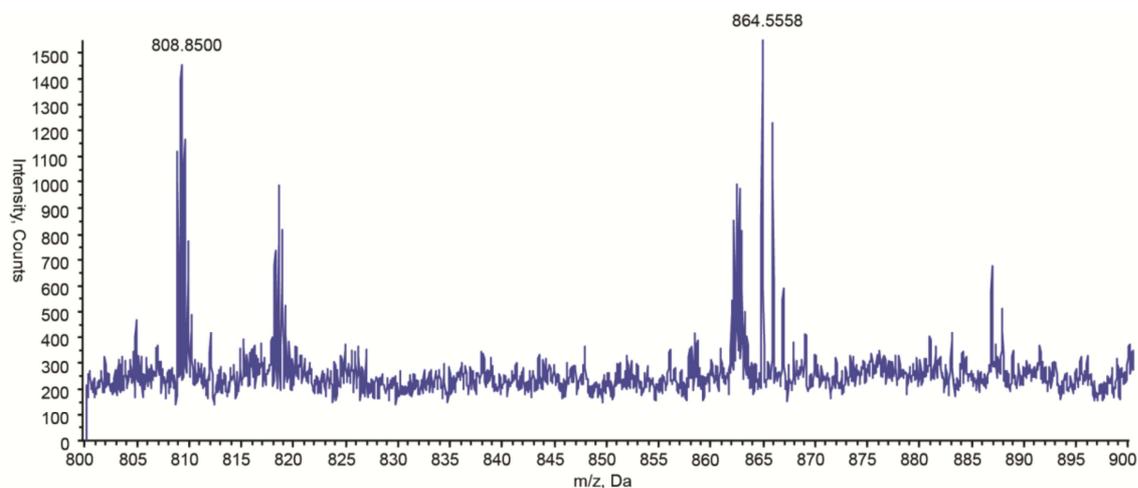


Figure S11. Another portion of the ESI-MS spectrum of His₆PpiB expressed with the chloride **18**, after heat treatment for 5 min at 100 °C.

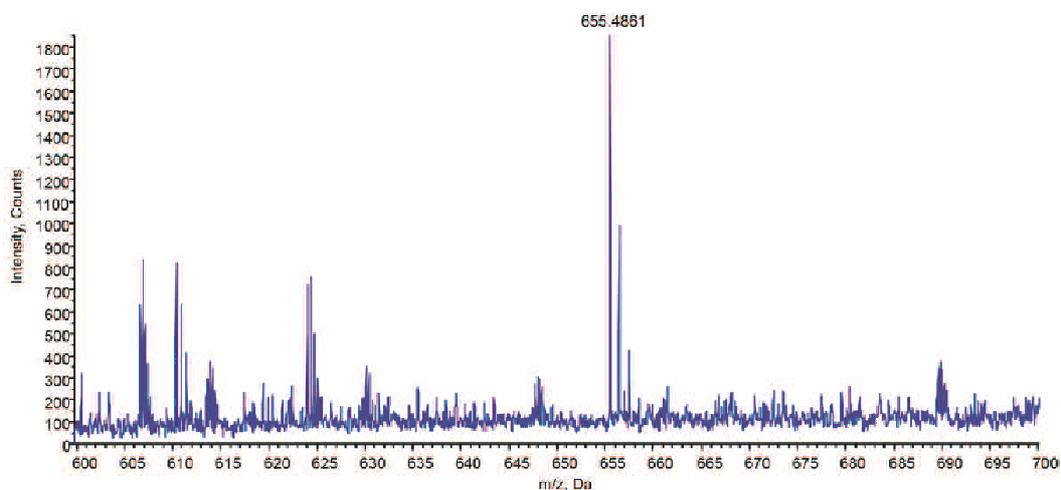


Figure S12. Another portion of the ESI-MS spectrum of His₆PpiB expressed with the chloride **18**, after heat treatment for 5 min at 100 °C.

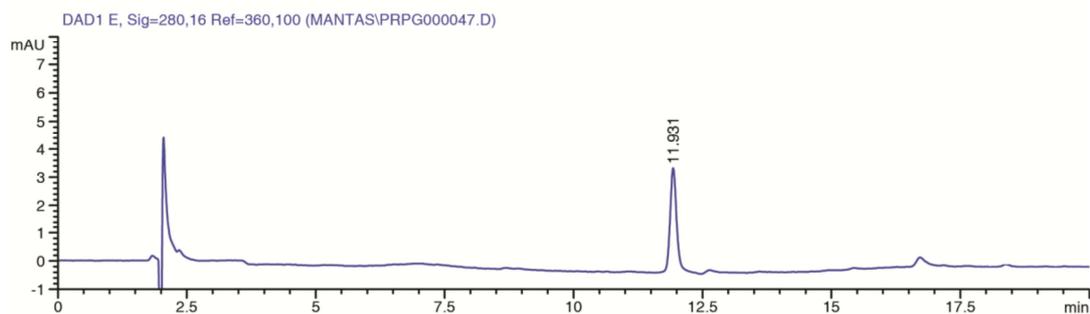
7. Production of GRPG (2a), CCK-7-G (3a) and the analogues 2b and 3b,c through heat treatment of fusion proteins

His₆PpiB_GRPG and His₆PpiB_CCK-7-G fusion proteins, expressed using normal and modified amino acids including the chlorides **16-18**, and isolated as described above (section 4), were heated in water at 100 °C for 5 min. Soluble and insoluble proteins were treated in the same way. The resultant mixtures were fractionated using a HP Agilent 1100 series HPLC with Chemstation software (Symmetry300™ C₁₈ 150 x 4.6 mm, 5 μm), and a linear gradient mobile phase of 0-100% MeCN in 0.1% TFA over 30 min, at a flow rate of 1 mL min⁻¹. Mass spectra of the peptides were recorded on an Agilent 1100 series LC/MSD TOF instrument (direct injection) or a Waters ACQUITY TQ Detector. GRPG (**2a**) was identical with a commercial standard and CCK-7-G (**3a**) was the same as material prepared through solid phase peptide synthesis. Characterization data for the GRPG analogue **2b**, and CCK-7-G (**3a**) and its analogues **3b,c**, are provided below.

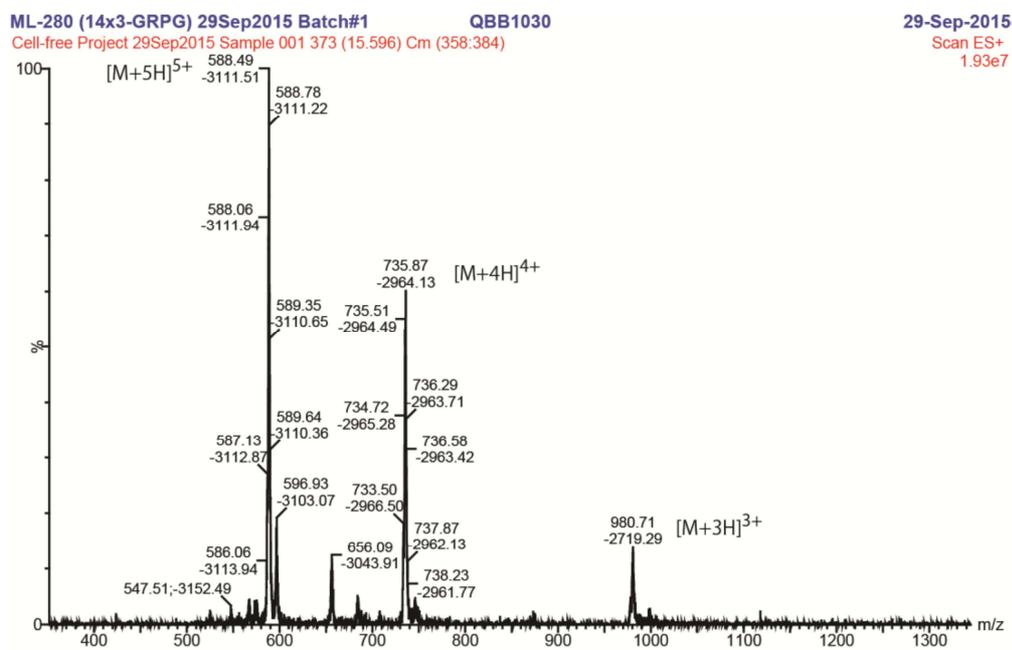
d_{10-12_3} -GRPG (2b): HPLC

ML280 X

=====
Injection Date : 9/14/2015 12:21:31 PM
Sample Name : GRPG Location : Vial 95
Acq. Operator : MANTAS
Acq. Instrument : PADDY Inj Volume : 100 μ l
Acq. Method : C:\CHEM32\1\METHODS\HORMONE.M
Last changed : 9/14/2015 12:19:15 PM by MANTAS
(modified after loading)



d_{10-12_3} -GRPG (2b): ESI-MS



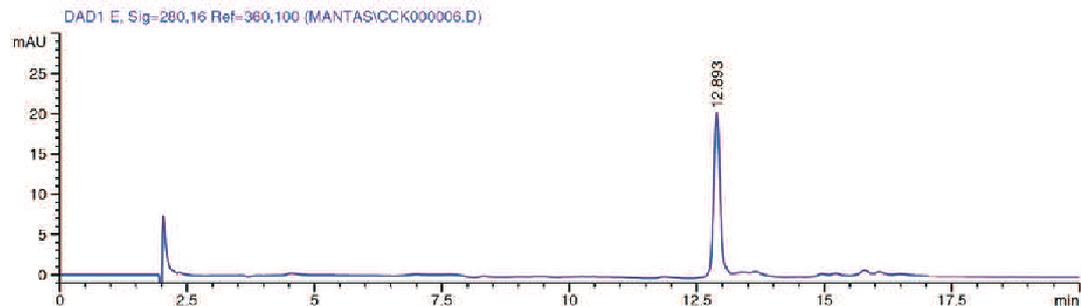
Chlorinated Amino Acids in Peptide Production

Chapter 3. Cleavage of Peptide Bonds

CCK-7G (3a): HPLC

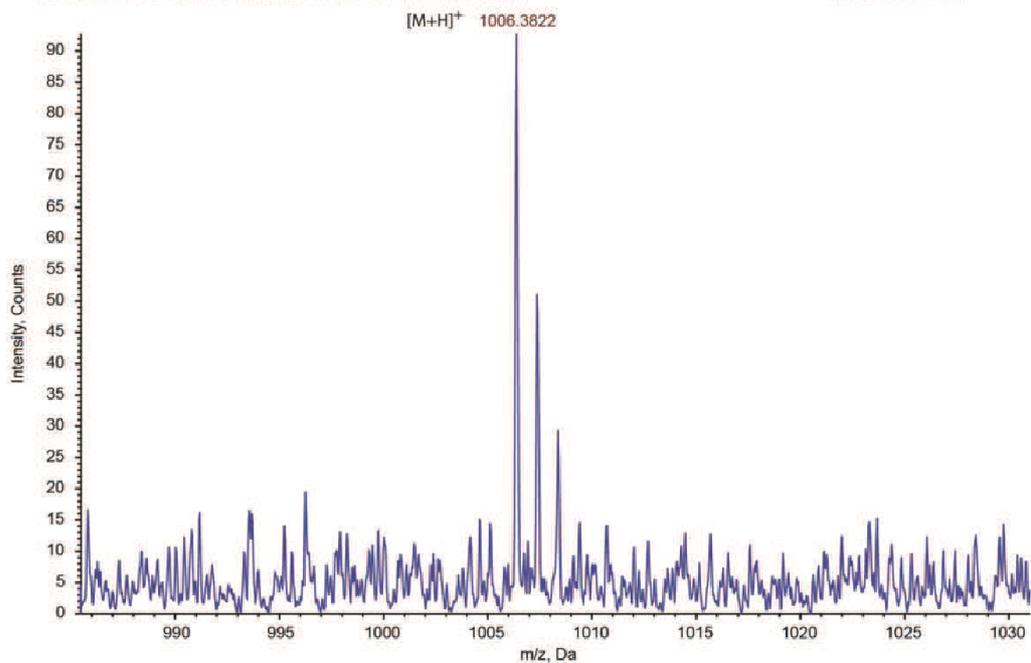
ML276B2 100u

```
=====
Injection Date : 2/24/2015 4:46:15 PM
Sample Name    : CCK                               Location : Vial 63
Acq. Operator  : MANTAS
Acq. Instrument: PADDY                             Inj Volume: 100 µl
Acq. Method    : C:\CHEM32\1\METHODS\HORMONE.M
Last changed   : 2/24/2015 4:42:46 PM by MANTAS
                (modified after loading)
=====
```



CCK-7G (3a): HPLC

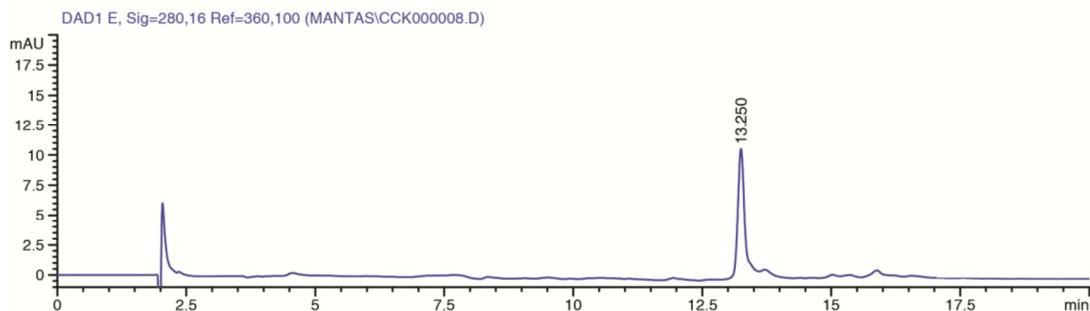
■ +TOF MS: 0.226 to 0.642 min from 12032015 mANTASb.wiff Agilent Max. 844.8 counts.



5-CCK-7-G (3b): HPLC

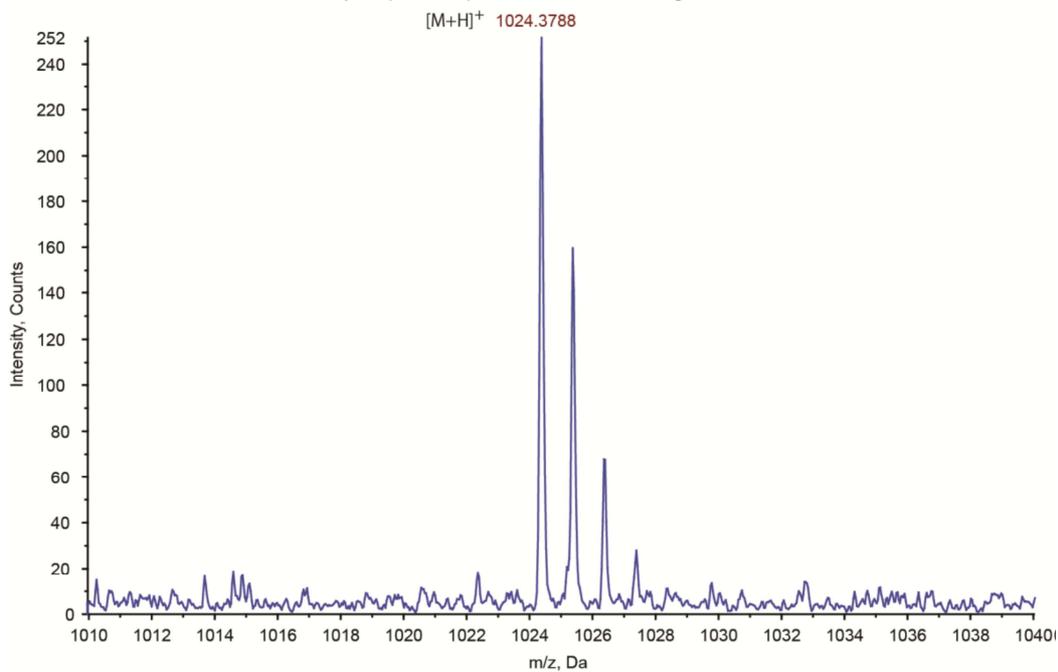
ML276C2 100u

```
=====
Injection Date : 2/24/2015 6:29:36 PM
Sample Name    : CCK                               Location : Vial 63
Acq. Operator  : MANTAS
Acq. Instrument: PADDY                             Inj Volume : 100 µl
Acq. Method    : C:\CHEM32\1\METHODS\HORMONE.M
Last changed   : 2/24/2015 6:26:18 PM by MANTAS
                (modified after loading)
=====
```



5-CCK-7-G (3b): ESI-MS

■ +TOF MS: 0.045 to 0.425 min from Sample 6 (ML276C2) of 05032015Mantas.wiff Agilent Max. 2118.3 counts.



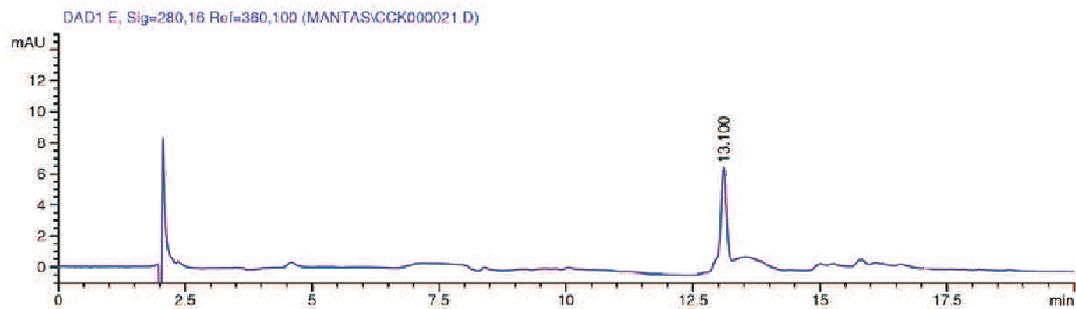
Chlorinated Amino Acids in Peptide Production

Chapter 3. Cleavage of Peptide Bonds

10-CCK-7-G (3c): HPLC

ML279D2 100ul

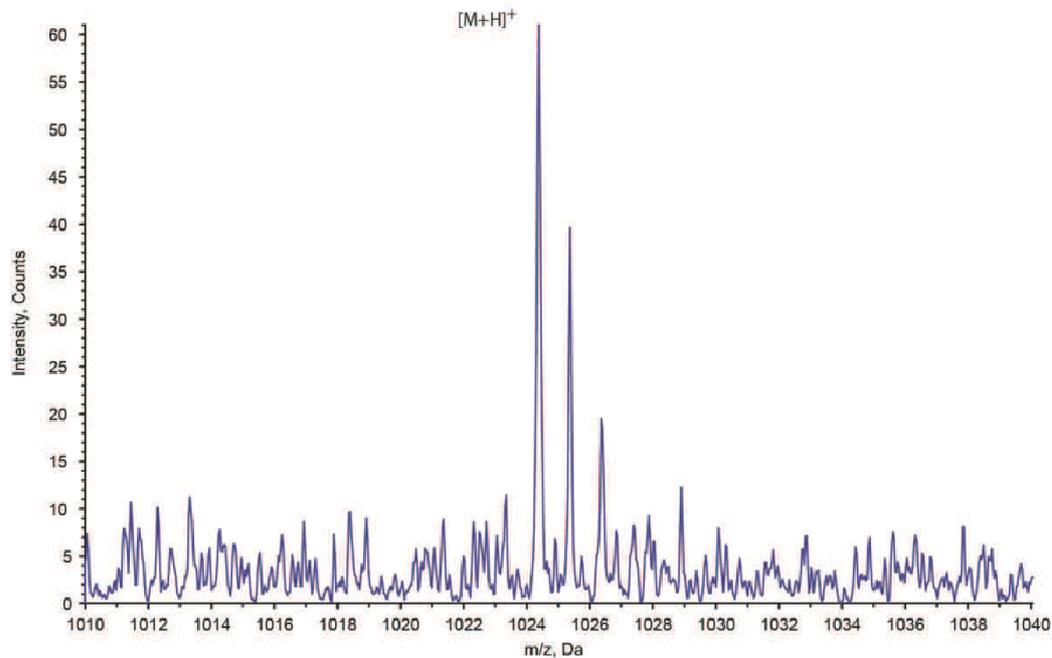
=====
Injection Date : 3/27/2015 7:25:58 PM
Sample Name : CCKG Location : Vial 53
Acq. Operator : MANTAS
Acq. Instrument : PADDY Inj Volume : 100 ul
Acq. Method : C:\CHEM32\1\METHODS\HORMONE.M
Last changed : 3/27/2015 7:22:54 PM by MANTAS
(modified after loading)



10-CCK-7-G (3c): ESI-MS

■ +TOF MS: 0.260 to 0.567 min from Sample 4 (ML 276D4) of 12032015 MANTASb.wiff Agilent

Max. 1002.6 counts.



8. Analysis of His₆PpiB produced using dehydro amino acids

Soluble His₆PpiB expressed using the dehydro amino acids **20-26** as described above (section 4) was analysed by SDS-PAGE, after metal-ion affinity chromatography (Figure S13). Mass spectra of the purified proteins are shown in Figures S14-S18. Insoluble protein was only observed when the dehydrovaline **20** was used, and this material was also analysed using SDS-PAGE.

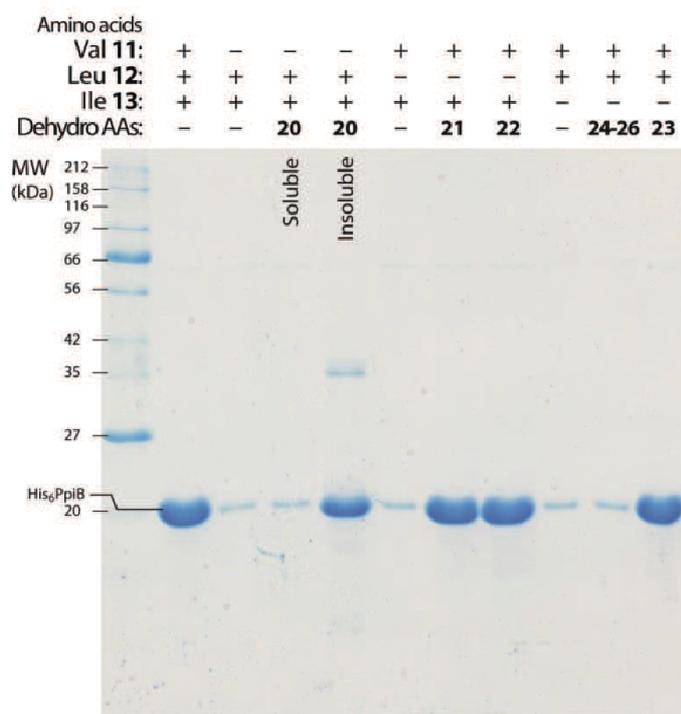


Figure S13. SDS-PAGE analysis of His₆PpiB expressed with the dehydro amino acids **20-26**.

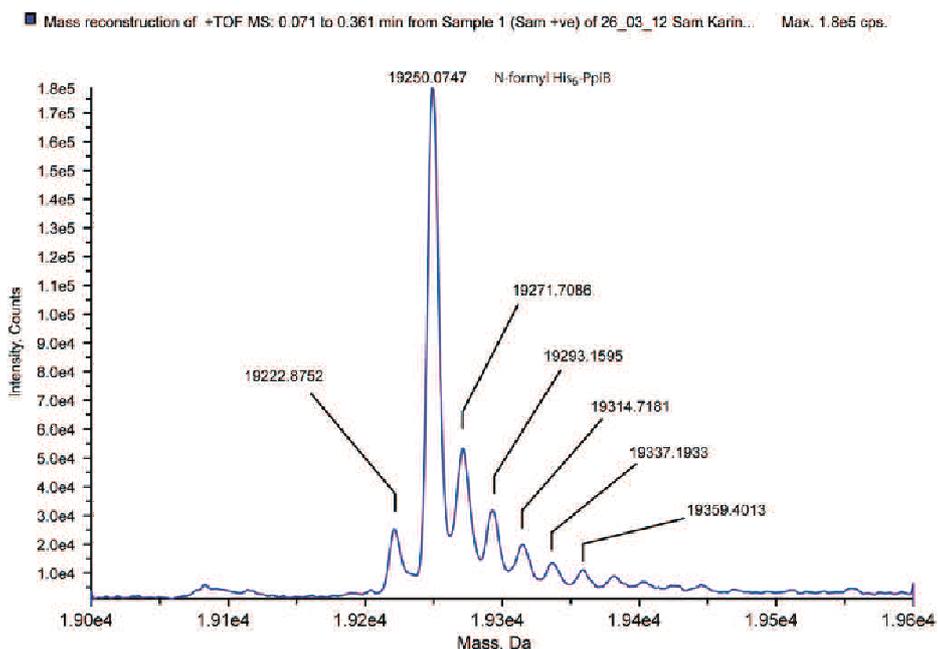


Figure S14. Deconvoluted mass spectrum of PpiB formed in the presence of 1 mM of all twenty usual amino acids.

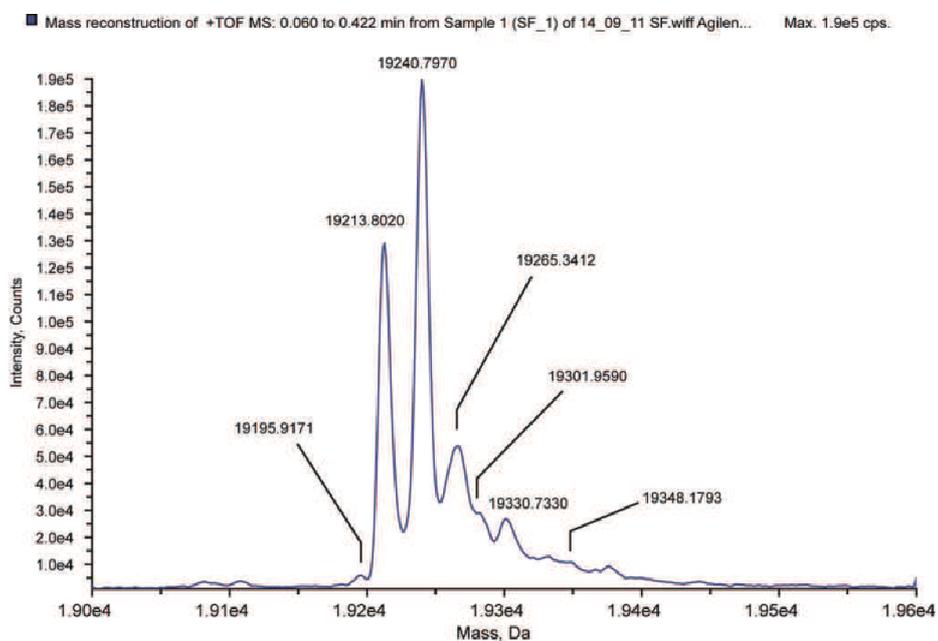


Figure S15. Deconvoluted mass spectrum of PpiB formed in the presence of 2 mM (*S*)-4,5-dehydroleucine (**22**) instead of 1 mM (*S*)-leucine (**12**), and with 1 mM of each of the other nineteen usual amino acids.

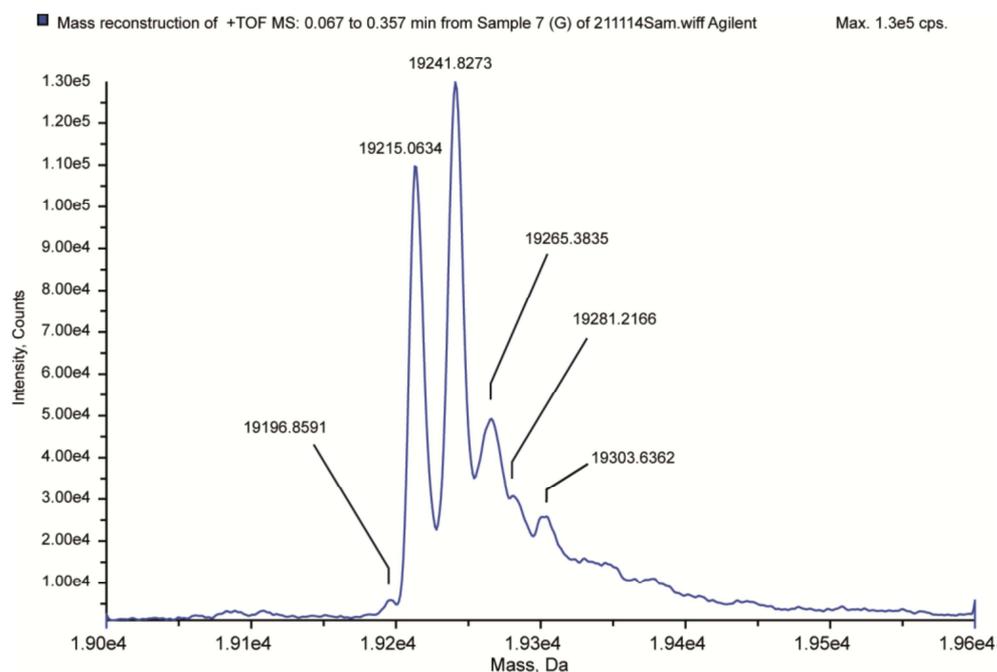


Figure S16. Deconvoluted mass spectrum of PpiB formed in the presence of 4 mM 3,4-dehydroleucine (**21**) instead of 1 mM (*S*)-leucine (**12**), and with 1 mM of each of the other nineteen usual amino acids.

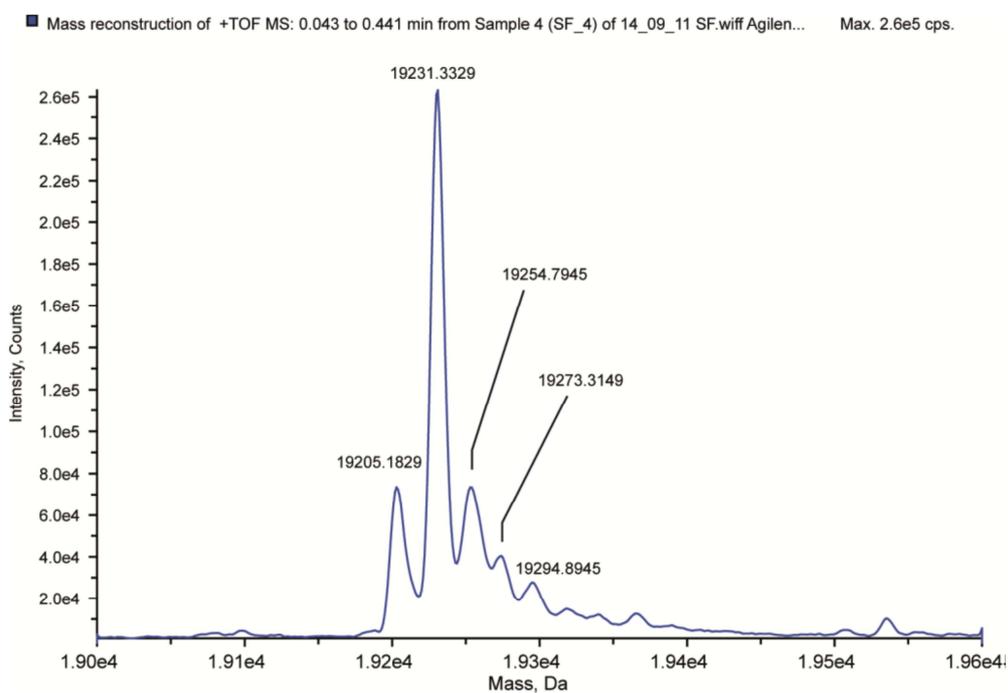


Figure S17. Deconvoluted mass spectrum of PpiB formed in the presence of 4 mM (*2R,3R*)-4,5-dehydroisoleucine (**23**) instead of 1 mM (*2S,3S*)-isoleucine (**13**), and with 1 mM of each of the other nineteen usual amino acids.

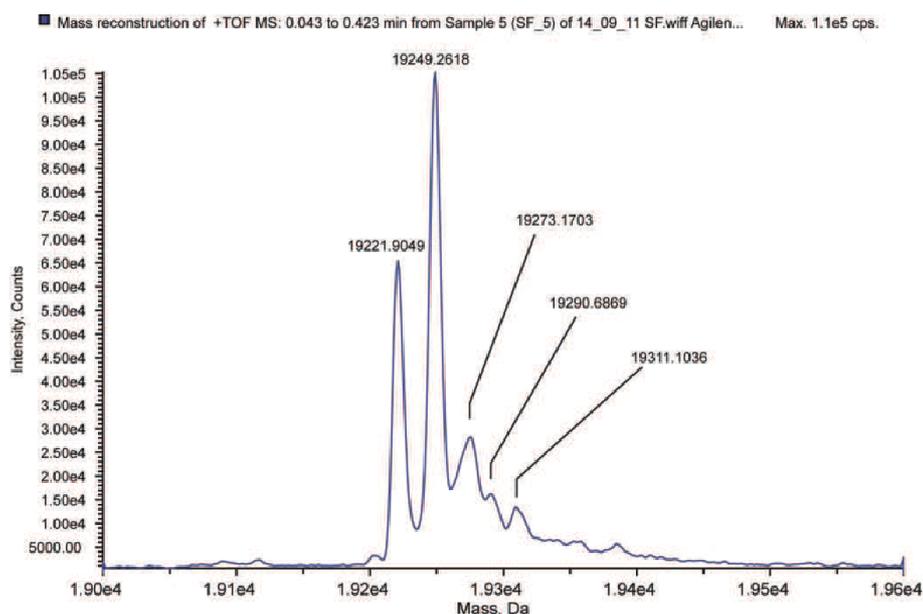


Figure S18. Deconvoluted mass spectrum of PpiB formed in the presence of 4.4 mM, 16.0 mM and 4.0 mM of the dehydroisoleucines **24**, **25** and **26**, respectively, instead of 1 mM (2*S*,3*S*)-isoleucine (**13**), and with 1 mM of each of the other nineteen usual amino acids.

These analyses show that withholding either (*S*)-valine (**11**), (*S*)-leucine (**12**) or (2*S*,3*S*)-isoleucine (**13**) results in a marked decrease in PpiB synthesis. Mass spectrometry shows that the small amount of protein that does form is unmodified and its production is attributable to low levels of residual amino acids being present in the S30 extract.^[2] Only this background level of unmodified PpiB is produced in the presence of the dehydroisoleucines **24-26**, establishing that none of them acts as a substitute for (2*S*,3*S*)-isoleucine (**13**). Addition of either of the dehydro amino acids **20-23** results in efficient protein synthesis, comparable to that observed with all twenty of the usual amino acids. This demonstrates that these compounds are easily incorporated in place of (*S*)-valine (**11**), (*S*)-leucine (**12**) or (2*S*,3*S*)-isoleucine (**13**). It was impractical to obtain a good mass spectrum of the PpiB formed using the dehydrovaline **20**, due to its insolubility, but the spectrum of the protein produced in each of the other cases shows a dominant peak corresponding to full misincorporation, of the dehydroleucines **21** and **22** in place of all five of the residues of (*S*)-leucine (**12**) normally present in PpiB, and of the dehydroisoleucine **23** instead of all ten residues of (2*S*,3*S*)-isoleucine (**13**).

9. Production of GRPG (2a) and PRPG (4) through treatment of fusion proteins with iodine

His₆PpiB_GRPG and His₆PpiB_PRPG fusion proteins, expressed using normal and modified amino acids including the dehydro amino acids **22** and **23**, and isolated as described above (section 4), were treated with a solution of iodine (7.9 mM) in water/THF (1:1, v:v). The resultant mixtures were fractionated using a HP Agilent 1100 series HPLC with Chemstation software (Symmetry300TM C₁₈ 150 x 4.6 mm, 5 μm), and a linear gradient mobile phase of 0-100% MeCN in 0.1% TFA over 30 min, at a flow rate of 1 mL min⁻¹. Mass spectra of the peptides were recorded on an Agilent 1100 series LC/MSD TOF instrument (direct injection) or a Waters ACQUITY TQ Detector. Reaction conditions were optimised by varying the incubation time, temperature and amount of iodine. Best results (38% yield) were obtained after incubation of the protein with 5 molar equivalents of iodine for 10 min at 4 °C. Adding DMSO (10% by volume) facilitated the production of GRPG (**2a**). GRPG (**2a**) and PRPG (**4**) obtained in this manner were identical with commercial standards.

10. References

- [1] N. Valls, M. Borregan, J. Bonjoch, *Tetrahedron Lett.* **2006**, *47*, 3701-3705.
- [2] D. J. Stigers, Z. I. Watts, J. E. Hennessy, H. K. Kim, R. Martini, M. C. Taylor, K. Ozawa, J. W. Keillor, N. E. Dixon, C. J. Easton, *Chem. Commun.* **2011**, *47*, 1839-1841.
- [3] K. Yamashita, K. Inoue, K. Kinoshita, Y. Ueda, H. Murao, WO 1999/033785, **1999**.
- [4] L. Panella, A. M. Aleixandre, G. J. Kruidhof, J. Robertus, B. L. Feringa, J. G. de Vries, A. J. Minnaard, *J. Org. Chem.* **2006**, *71*, 2026-2036.
- [5] Y. Ozaki, S. Maeda, M. Miyoshi, K. Matsumoto, *Synthesis* **1979**, 216-217.
- [6] I. N. Arthur, J. E. Hennessy, D. Padmakshan, D. J. Stigers, S. Lesturgez, S. A. Fraser, M. Liutkus, G. Otting, J. G. Oakeshott, C. J. Easton, *Chem.-Eur. J.* **2013**, *19*, 6824-6830.
- [7] C. J. Easton, M. C. Merrett, *Tetrahedron* **1997**, *53*, 1151-1156.
- [8] C. J. Easton, A. J. Edwards, S. B. McNabb, M. C. Merrett, J. L. O'Connell, G. W. Simpson, J. S. Simpson, A. C. Willis, *Org. Biomol. Chem.* **2003**, *1*, 2492-2498.
- [9] a) M. Jacob, M. L. Roumestant, P. Viallefont, J. Martinez, *Synlett* **1997**, 691-692; b) B. D. Dangel, J. A. Johnson, D. Sames, *J. Am. Chem. Soc.* **2001**, *123*, 8149-8150.
- [10] P. Bey, J. P. Vevert, V. Vandorsselaer, M. Kolb, *J. Org. Chem.* **1979**, *44*, 2732-2742.
- [11] J. E. Baldwin, S. B. Haber, C. Hoskins, L. I. Kruse, *J. Org. Chem.* **1977**, *42*, 1239-1241.
- [12] R. D. Allan, *Aust. J. Chem.* **1979**, *32*, 2507-2516.
- [13] H. C. Qu, X. Y. Gu, Z. H. Liu, B. J. Min, V. J. Hruby, *Org. Lett.* **2007**, *9*, 3997-4000.
- [14] M. L. Mock, T. Michon, J. C. M. van Hest, D. A. Tirrell, *ChemBioChem* **2006**, *7*, 83-87.
- [15] C. Neylon, S. E. Brown, A. V. Kralicek, C. S. Miles, C. A. Love, N. E. Dixon, *Biochemistry* **2000**, *39*, 11989-11999.
- [16] K. Ozawa, M. J. Headlam, D. Mouradov, S. J. Watt, J. L. Beck, K. J. Rodgers, R. T. Dean, T. Huber, G. Otting, N. E. Dixon, *FEBS J.* **2005**, *272*, 3162-3171.
- [17] M. A. Apponyi, K. Ozawa, N. E. Dixon, G. Otting, Cell-Free Protein Synthesis for Analysis by NMR Spectroscopy, in *Methods in Molecular Biology*, Vol. 426 (Eds.: B. Kobe, M. Guss, T. Huber), Humana Press, Totowa, N. J., **2008**, pp. 257-268.
- [18] K. Ozawa, M. J. Headlam, P. M. Schaeffer, B. R. Henderson, N. E. Dixon, G. Otting, *Eur. J. Biochem.* **2004**, *271*, 4084-4093.
- [19] L. Guignard, K. Ozawa, S. E. Pursglove, G. Otting, N. E. Dixon, *FEBS Lett.* **2002**, *524*, 159-162.
- [20] J. K. Montclare, S. Son, G. A. Clark, K. Kumar, D. A. Tirrell, *ChemBioChem* **2009**, *10*, 84-86.
- [21] J. A. Van Deventer, J. D. Fisk, D. A. Tirrell, *ChemBioChem* **2011**, *12*, 700-702.

Chapter 4. Engineering with Chlorinated Amino Acids without Sequence Restrictions

The discernment of the propensity of γ -chlorinated amino acids for lactonisation has led to the development of a heat-responsive proteolysis trigger that has been successfully employed for the preparation of peptides, as described in the previous chapter. A potential limitation of the method is that the target peptide must lack one of the amino acids that can be substituted by a latent proteolysis-inducing analogue, isoleucine or leucine, thus seemingly limiting the method to the production of specifically selected targets. As a potential solution, interchange of the structurally similar amino acids from the BCAA family was considered. To test the feasibility of this approach, a small inhibitor of proteases aprotinin containing both two isoleucines and two leucines in the sequence was chosen as a target.

Aliphatic halogenated amino acids have been successfully employed in the past for diverse protein modifications, including protein labelling,^[135] formation of cyclic peptides,^[136] protein cross-linking^[137] and protein chemical ligation.^[138] In contrast, amino acids with functionalised side chains for bioorthogonal reactions on proteins can only react with specially tailored chemical agents.^[94,139] The resistance of the β -chlorinated amino acids to heat treatment, as demonstrated in the previous chapter, suggested the reactive functionalities would be retained in the produced peptides and

could be used for subsequent elaboration. However, the reported loss of protein function after expression with 2-amino-3-chlorobutyric acid,^[78,86] posed another serious limitation to the potential use of the chlorinated amino acids in protein engineering.

4.1. Expression of aprotinin

Aprotinin, also known as bovine pancreatic trypsin inhibitor, or BPTI, was identified as the most suitable target for the investigation. The 58 amino acid protein with extensively studied structure^[140] contained both two isoleucine and two leucine residues, making it a representative protein in terms of amino acid composition. Due to the smaller size, aprotinin could not be expressed individually in a cell-free protein expression system, thus requiring the use of a carrier protein. Additionally, the protein contains a single valine in the sequence that can be substituted by the chlorinated analogues **15a** and **15b**, meaning any restructuring would originate from a single point in the protein, allowing for a clearer evaluation of the effects of the chlorides.^[141] Finally, as an inhibitor of serine proteases, aprotinin readily crystallises as a complex with trypsin,^[140] thus allowing for the direct observation of protein activity by X-ray crystallography.

As described in the previous chapter, aprotinin was fused to the gene of the carrier protein peptidyl-prolyl *cis-trans* isomerase B (PpiB), with leucine encoded between the two proteins for heat controlled separation through the use of the chloride **25**. To prevent the heat labile analogue from incorporation into aprotinin, the leucine residues in aprotinin were mutated to isoleucine (L6I,L29I), and the hexahistidine tag was attached directly to the mutated aprotinin, at the C-terminus, to enable purification of

aprotinin through metal affinity chromatography (Figure 4.1). The cell-free reaction to express the fusion protein was carried out in 10 mL of inner reaction mixture with the 4-chloronorleucine **25** in place of leucine (Figure 4.2). Due to the selectivity afforded by the biological machinery, the chloride **25** could be used as a mixture with its precursor (*S*)-norvaline (and other by-products); only the two diastereomers of the chloride **25** (both of which have similar reactivity) were recognised and incorporated into expressed protein. This substantially simplified the preparation of materials, particularly on a larger scale. To further reduce the number of steps, the heat-labile fusion protein was not purified after expression, but instead the entire reaction mixture was subjected to 100°C for 5 min. After the heat shock, 6 M guanidinium chloride was added to solubilise the precipitated proteins and the tagged aprotinin was extracted through Ni(II)-affinity chromatography under denaturing conditions. Refolding of aprotinin was not impeded by either the histidine-tag or the homologous substitutions, and active inhibitor was obtained after overnight incubation of the purified peptide in Tris buffer. Activity was confirmed through a trypsin inhibition assay.

```
MVTFHTNHGD  IVIKTFDDKA  PETVKNFLDY  CREGFYNN TI  FHRVINGFMI
QGGGFEPGMK  QKATKEPIKN  EANNGLKNTR  GTLAMARTQA  PHSATAQFFI
NVVDNDFLNF  SGESLQGWGY  CVFAEVVDGM  DVVDKIKGVA  TGRSGMHQDV
PKEDV I IESV  TVSENGGGAL  RPDFCIEPPY  TGPCKARIIR  YFYNAKAGIC
QTFVYGGCRA  KRNNEKSAED  CMRTC GGAHH  HHHH
```

Figure 4.1. The sequence of PpiB_BPTI_His₆ fusion protein. The BPTI segment is in bold, the mutated L6I and L29I residues are italicised, and the cleavage point is underlined.

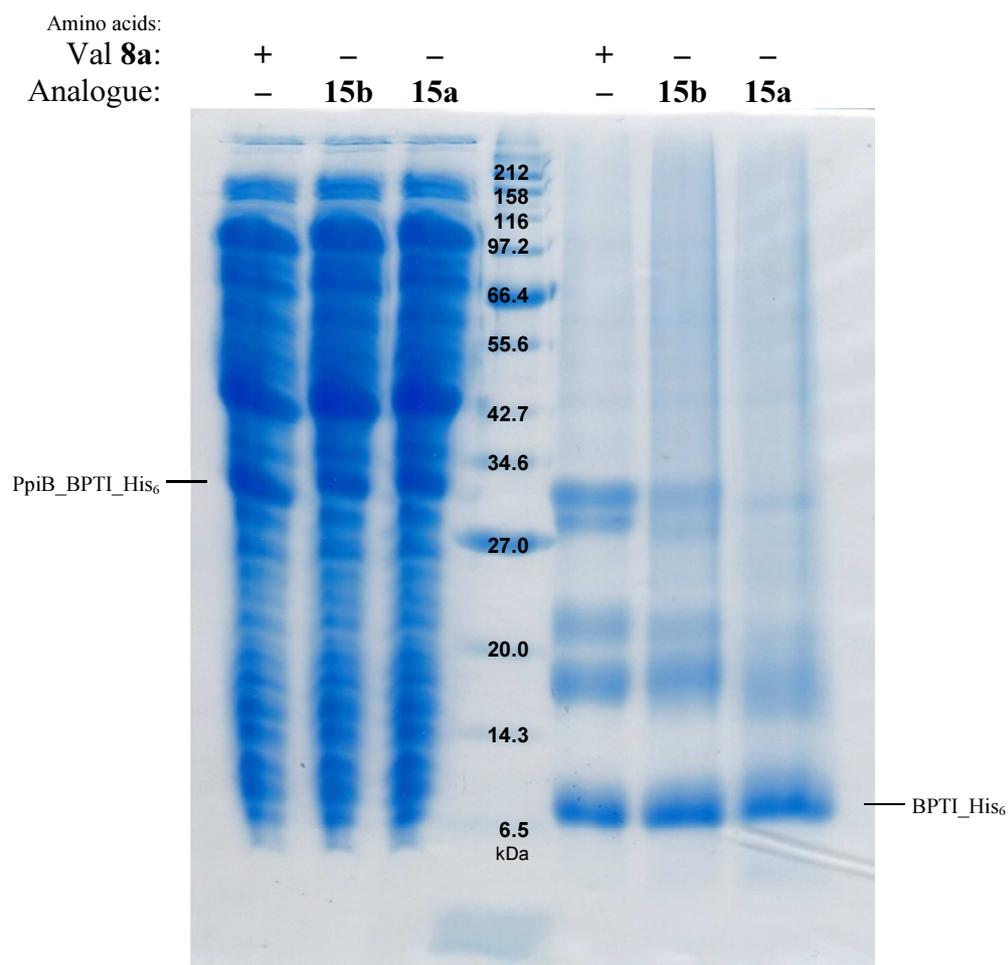


Figure 4.2. Production of modified aprotinin. Left: crude reaction mixtures containing PpiB-aprotinin fusion proteins; right: aprotinin obtained after Ni(II)-affinity purification of heat treated (100°C for 5 min) reaction mixtures.

The fusion protein was then expressed with a double substitution. Next to the analogue of leucine **25**, as described above, (2*R*,3*S*) and (2*R*,3*R*)-2-amino-3-chlorobutyric acids **15b** and **15a** were separately added to the expression mixtures in place of valine **8a**. Expression levels of the doubly-substituted fusion proteins were similar with the singly-substituted material (Figure 4.2). After heat treatment to induce bond cleavage, the

chlorinated analogues of aprotinin were purified and refolded like the peptide composed solely from the proteogenic amino acids. Over 1 mg of aprotinin was obtained in all three cases, despite the carrier protein PpiB accounting for three quarters of expressed material. The scalability of cell-free protein expression makes large-scale production of proteins feasible.

The production process of aprotinin highlights the stark differences in the chemical properties of β - and γ -functionalised amino acids. Similar reactivity towards external reagents had been observed with both classes of halogenated amino acids,^[136] but the γ -functionalised amino acids show high propensity for cyclisation, driven by the formation of 5-membered rings, with elevated temperatures needed to compensate for the reduced nucleophilicity of the carboxy-group in a peptide chain, whereas the halide at the β -position is unaffected by raised temperatures.

4.2. Crystallisation and crystal analysis

The produced proteins were complexed with trypsin. Due to the propensity of trypsin to self-digest, pure active trypsin is unattainable, with partially digested protein present even in highest grade material. To ensure these fragments do not interfere with the crystallisation process, the complexes were re-purified through Ni(II)-affinity chromatography, this time under native conditions. As the C-terminus of aprotinin is on the face opposite to the trypsin binding region, the hexahistidine-tag was not sequestered by the complexation and thus free to bind to immobilised metal ions. The complexes were crystallised by hanging drop method using conditions previously used

for the aprotinin-trypsin complex.^[140] Hexagonal crystals suitable for X-ray diffraction were obtained within a week (Figure 4.3).

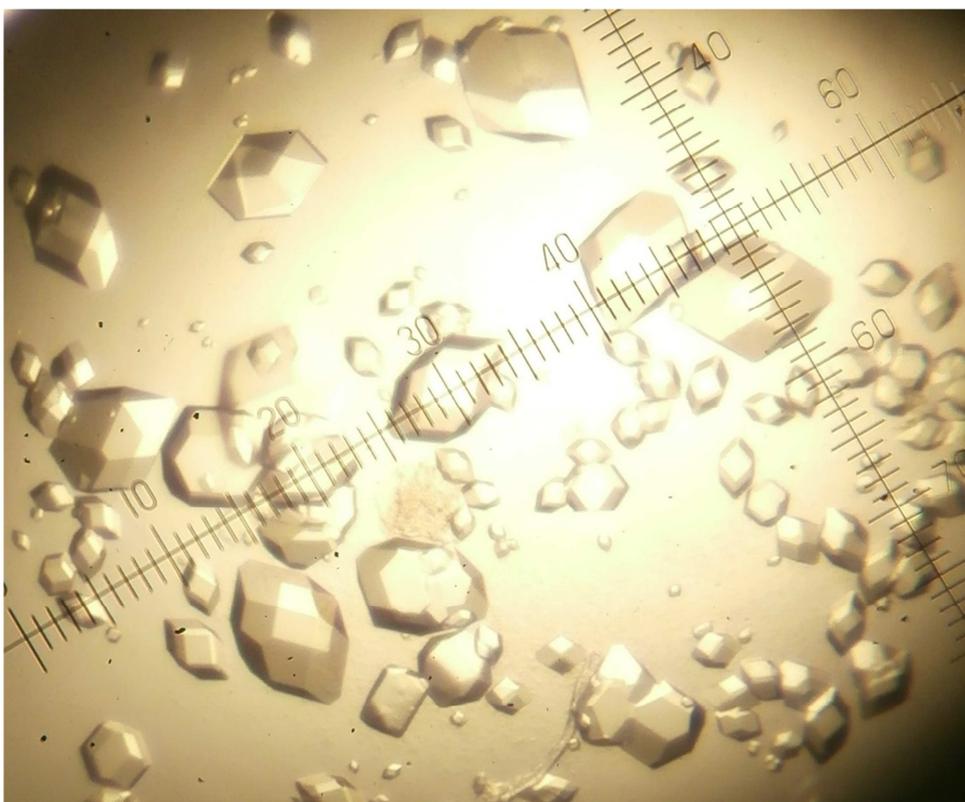


Figure 4.3. Crystals of aprotinin-trypsin complexes.

The complex composed of natural amino acids was solved to 1.7 Å resolution (**C1**). The same resolution was achieved for the complex containing (2*R*,3*R*)-2-amino-3-chlorobutyric acid **15a** (**C3**), whereas an even greater resolution of 1.6 Å was attained for the complex with the (2*R*,3*S*)-diastereomer **15b** (**C2**, Table 4.1). The positions of the electron-rich chlorine atoms were clearly visible in the electron density maps during refinement. The structures were refined using the published model (PDB id: 3OTJ) as a

starting structure. In all three solved structures the electron density for the C-terminal his-tag on aprotinin was not visible, suggesting the tag did not adopt an ordered structure. The first two N-terminal residues Arg¹ and Pro² were also not observed, as is often the case with the flexible termini protruding from globular structures.

In the structure containing (2*R*,3*S*)-2-amino-3-chlorobutyric acid **15b**, the modelled chlorinated side-chain contained a surplus of electron density around the chlorine atom than was observed from the diffraction pattern, indicating the position was partly occupied by valine **8a** that was in competition with the chloride **15b** during protein expression. A good fit with the detected electron density was observed after the occupancy of the chlorine atom was reduced to 80%; accounting for the contribution of valine residue, about 70% incorporation of the chloride was achieved. In contrast, the modelled side-chain of (2*R*,3*R*)-2-amino-3-chlorobutyric acid **15a** fit the experimental electron density, indicating good (>90%) replacement of valine.

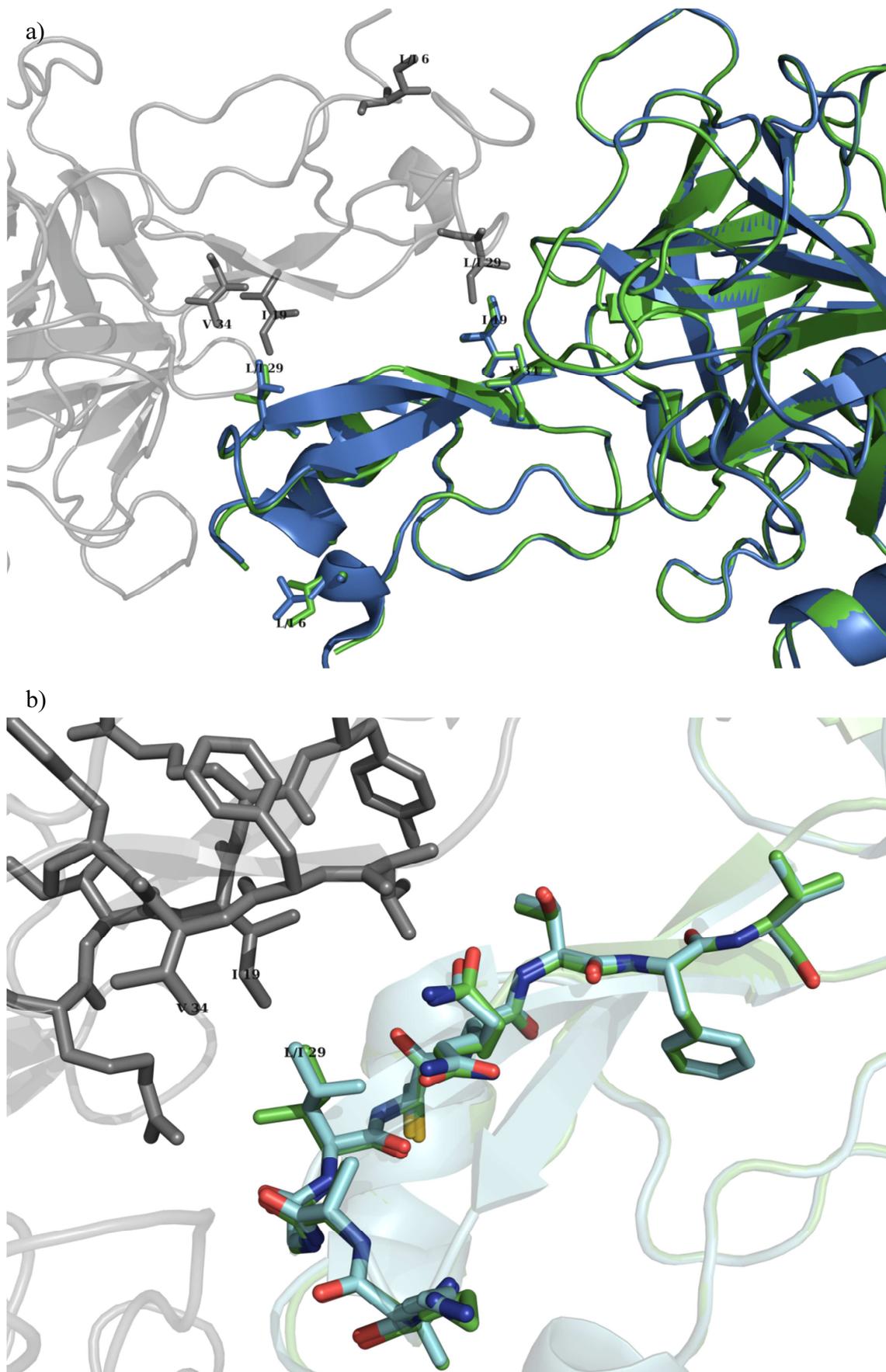
Table 4.1. Crystallographic data and refinement statistics.

Structure	C1	C2	C3
Data collection^[a]			
Unit cell parameters			
a / Å	74.6	74.6	74.9
b / Å	80.6	81.0	80.7
c / Å	124.4	124.1	124.2
α / °	90.0	90.0	90.0
β / °	90.0	90.0	90.0
γ / °	90.0	90.0	90.0
Space group	I222	I222	I222
Resolution / Å	37.28-1.70 (1.73-1.70)	32.69-1.60 (1.62-1.60)	37.42-1.70 (1.73-1.70)
Observed reflections	282892 (12434)	351682 (16263)	202246 (8828)
Unique reflections	39898 (1973)	50182 (2455)	41146 (1986)
Completeness / %	96.6 (91.4)	100.0 (100.0)	99.1 (88.6)
Multiplicity	7.1 (6.3)	7.0 (6.6)	4.9 (4.4)
R_{merge} ^[b]	0.084 (0.359)	0.060 (0.316)	0.090 (0.733)
$\langle I/\sigma(I) \rangle$	12.8 (4.2)	17.3 (4.7)	8.1 (1.6)
Structure refinement^[a]			
Resolution / Å	1.70 (1.75-1.70)	1.60 (1.64-1.60)	1.70 (1.75-1.70)
R_{work} ^[c]	0.154 (0.208)	0.163 (0.223)	0.163 (0.300)
R_{free} ^[c]	0.190 (0.251)	0.193 (0.291)	0.201 (0.326)
No. reflections	37893 (2618)	47649 (3435)	39065 (2644)
No. free R reflections	1997 (142)	2520 (195)	2071 (148)
No. non-hydrogen atoms	2430	2427	2449
No. protein atoms	2084	2085	2084
No. water molecules	320	316	339
Average B factor / Å ²	20.9	18.3	24.2
R.m.s.d. bonds / Å	0.023	0.026	0.020
R.m.s.d. angles / °	2.170	2.393	1.913
Ramachandran plot / %			
favoured regions	98.9	98.2	98.9
allowed regions	1.1	1.8	1.1
disallowed regions	0	0	0

[a] The number in parentheses in the value in the highest resolution shell. [b] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_{hkl_i} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_i \langle I_{hkl_i} \rangle$, where I_{hkl_i} is the intensity of the i th observation of reflection hkl and $\langle I_{hkl} \rangle$ is the average intensity. [c] $R_{\text{work}} = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$; R_{free} was calculated using 5% of randomly selected reflections for cross-validation.

The observed extent of incorporation of the chlorinated amino acids is in good agreement with earlier reports of valyl-tRNA synthetase displaying twenty-fold^[142] greater affinity for the (2*R*,3*R*) diastereomer **15a**, that is even able to compete with the natural substrate (*S*)-valine **8a**,^[102] than the (2*R*,3*S*) chloride **15b** that is highly suppressed by the natural substrate. The enzyme appears to tolerate replacement of the methyl groups of (*S*)-valine **8a** with more polar groups differently, with modification of the *pro-S* position significantly more acceptable.^[142] This is reflected by the recognition of (2*S*,3*R*)-threonine as a substrate by the synthetase, requiring subsequent editing to prevent misincorporation of the hydroxylated amino acid into proteins,^[69,77] whereas no activity is detectable with (2*S*,3*S*)-*allo*-threonine.^[142]

The substitution of the two endogenous leucine residues Leu⁶ and Leu²⁹ with isoleucine did not cause any restructuring in the conformation of protein structure, as compared to the structure of wild-type protein (Figure 4.4). Both of the residues are positioned on the surface of aprotinin, facing away from trypsin in the protein complex. In the crystal lattice, the side-chain of residue⁶ is protruding into a solvent channel, whereas residue²⁹ is close to a crystallographic two-fold screw axis that occurs as a result of the presence of both a crystallographic dyad and body centring in the I222 space group; the residue is positioned *ca.* 4 Å from Ile¹⁹ and Val³⁴ of symmetry-related aprotinin unit and likely forms a stabilising hydrophobic interaction during crystallisation. No rearrangement of protein backbone or neighbouring residues is evident as a result of amino acid replacement, which can likely be explained by the aliphatic amino acids exhibiting similar properties, at least with respect to the maintenance of protein folding.



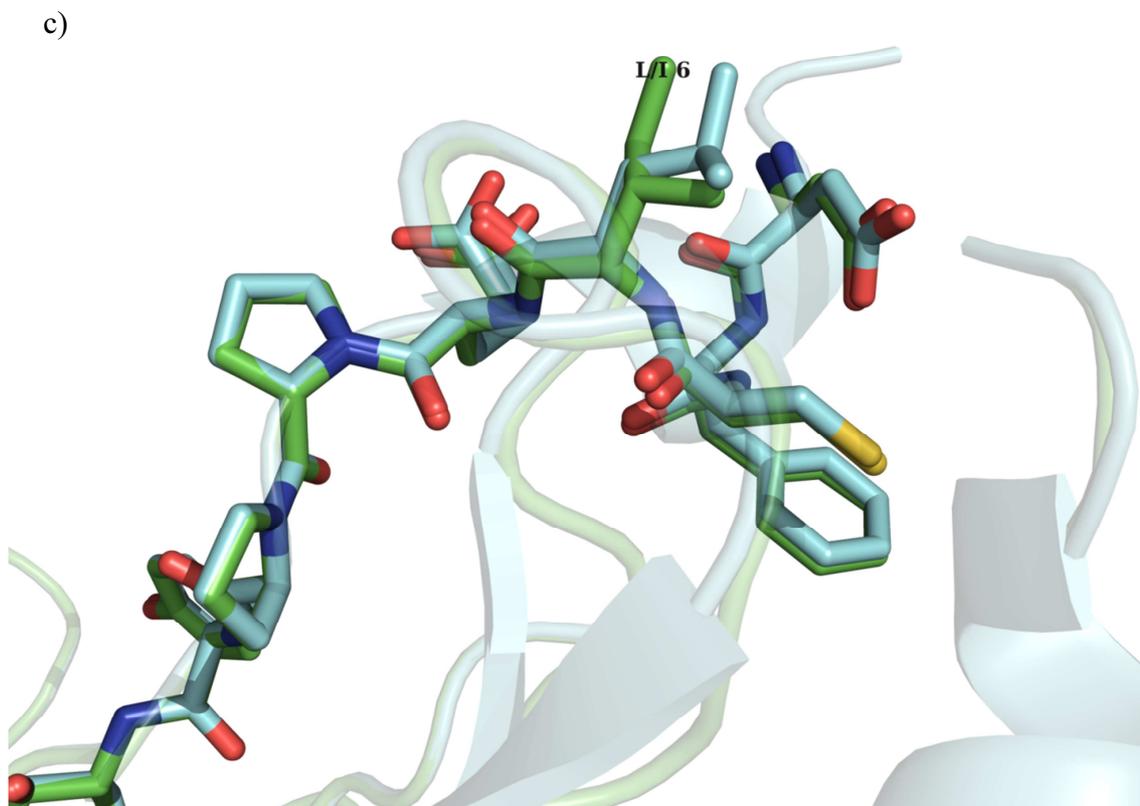
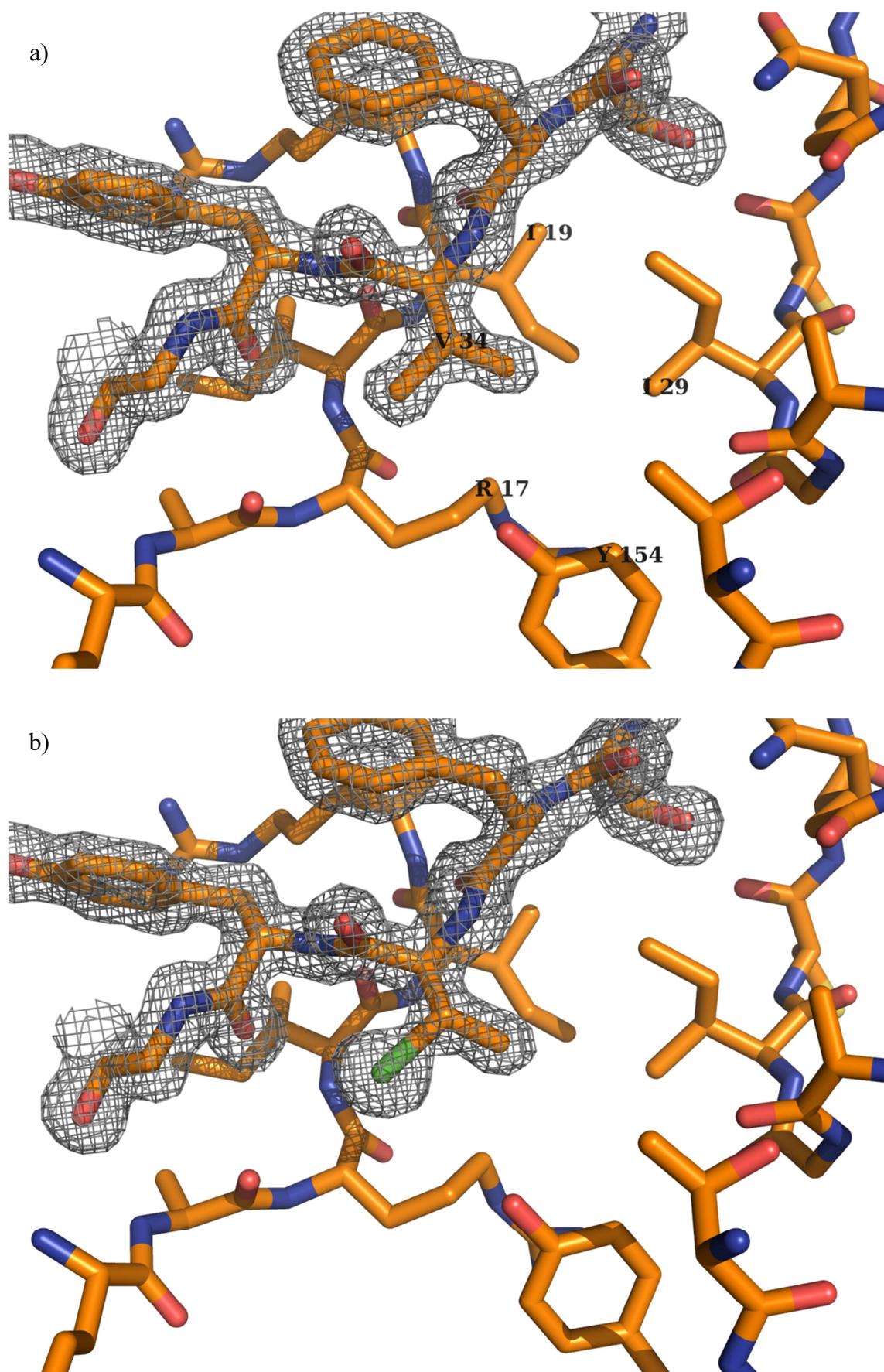


Figure 4.4. An overlay of the crystal structures of (L6I,L29I)-aprotinin/trypsin complex (green) and wild-type aprotinin/trypsin complex (PDB id: 4Y0Y; blue). Residue²⁹ is in close proximity to a hydrophobic patch created by Val³⁴ and Ile¹⁹ of a symmetry partner (grey) and occupies the space of the replaced residue, whereas residue⁶ is exposed to a solvent channel. Neither the overall structure of the protein (a) nor the local environments of the substituted residues (b,c) are affected by the mutation.

Similarly, the protein structure was not influenced by the inclusion of the chlorinated analogues of valine (Figure 4.5). The sole valine residue in aprotinin Val³⁴ is located on the surface of the protein and comes in close contact with trypsin in the inhibitory complex, with the methyl groups of Val³⁴ less than 4 Å from O^ζ of Tyr¹⁵⁴ in the trypsin chain, but is otherwise included in a hydrophobic pocket created by the hydrocarbon chain of Arg¹⁷ in aprotinin and the hydrophobic patch formed with the symmetry-related unit, as described above. Both of the chlorinated amino acids adopted the exact same position and conformation as valine in the original structure, with the chlorine atoms assuming the respective positions occupied by the corresponding methyl groups of valine.

The complete overlap of the chlorinated side-chains and the background (*S*)-valine in the crystal structures further reiterates the capacity of the chlorides to functionally replace the proteogenic isostere. After adjustment of the occupancies, the *B*-factors of the halogenated side-chains matched the average *B*-factors of the respective overall structures, as was the case with the native structure, indicating the positions of the modified side-chains were well defined.



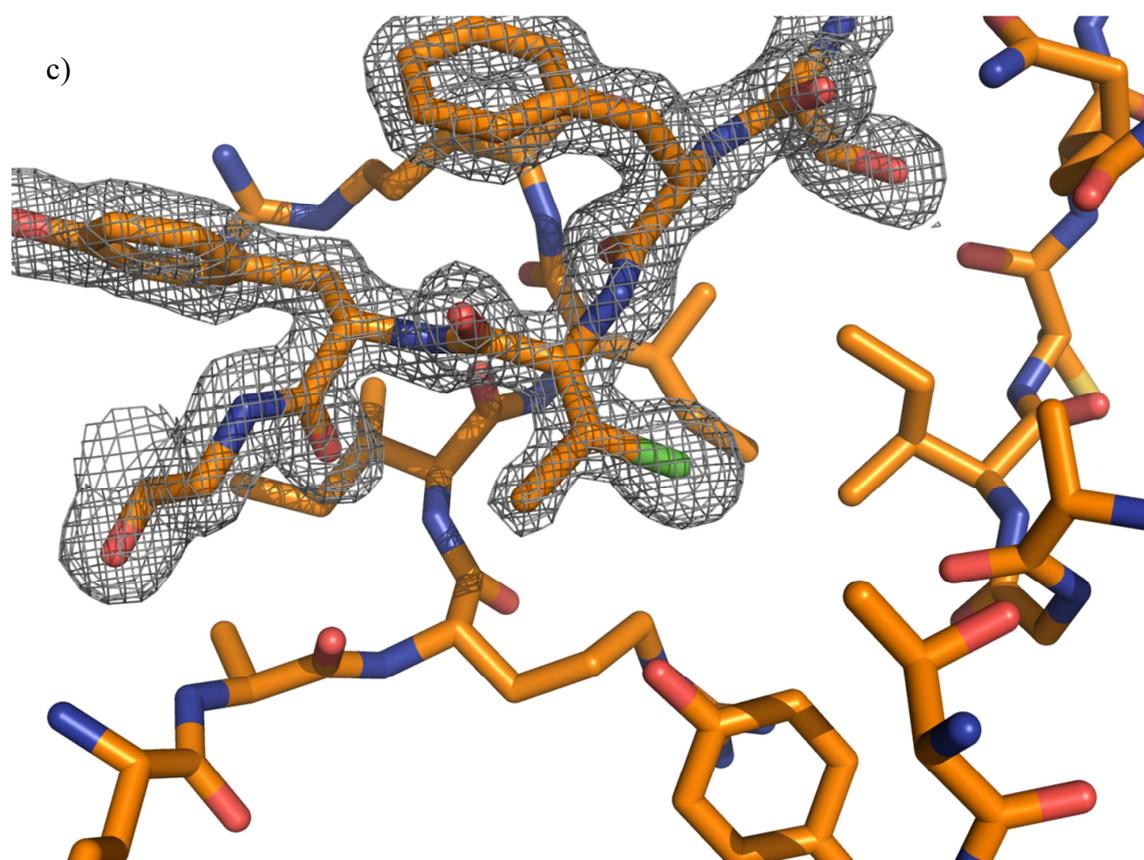


Figure 4.5. The conformation of aprotinin chain residue³⁴ in the aprotinin-trypsin complex containing a) valine, b) (2*R*,3*S*)-2-amino-3-chlorobutyric acid **15b** and c) (2*R*,3*R*)-2-amino-3-chlorobutyric acid **15a**. Carbon is orange, nitrogen - blue, oxygen - red and chlorine is green. The three structures are presented from the same perspective. Electron density maps are contoured at 1.5 σ .

4.3. HPLC purification of aprotinin

To demonstrate the wider feasibility of the approach, free un-tagged aprotinin was also sought. Mutants of aprotinin devoid of isoleucine (I18L,I19L) and leucine (L6I,L29I) were attached to PpiB *via* the corresponding residues and the hexahistidine-tags were attached at the N-termini of the fusions (Figure 4.6). The constructs were expressed

with the chlorides **26c** and **25**, respectively, and purified by Ni(II)-ion affinity chromatography. After exposure to 100°C, the samples were analysed by reverse phase HPLC. To ensure that the produced aprotinin was not attached to other peptide fragments through disulfide bridges after the exposure to high temperatures (aprotinin contains six cysteines, forming three disulfide bridges when folded), the samples were reduced with tris(2-carboxyethyl)phosphine (TCEP) prior to analysis. This resulted in greater retention of the peptides on the reverse phase resin, but isolation of aprotinin was nevertheless successful (Figure 4.7). While the purification through immobilised metal ion affinity was clearly advantageous for larger proteins, the isolation of mutants of aprotinin through HPLC served as a clear proof of concept that peptide production through heat treatment is not limited to short peptides.

```
MHHHHHMMVT FHTNHGDIVI KTFDDKAPET VKNFLDYCRE GFYNNTIFHR
VINGFMIQGG GFEPGMKQKA TKEPIKNEAN NGLKNTRGTL AMARTQAPHS
ATAQFFINVV DNDFLNFSGE SLQGWGYCVF AEVVDGMDVV DKIKGVATGR
SGMHQDVPKE DVIIIESVTVS ENGGGAIRPD FCLEPPYTGP CKARLLRYFY
NAKAGLCQTF VYGGCRAKRN NFKSAEDCMR TCGGA
```

```
MHHHHHMMVT FHTNHGDIVI KTFDDKAPET VKNFLDYCRE GFYNNTIFHR
VINGFMIQGG GFEPGMKQKA TKEPIKNEAN NGLKNTRGTL AMARTQAPHS
ATAQFFINVV DNDFLNFSGE SLQGWGYCVF AEVVDGMDVV DKIKGVATGR
SGMHQDVPKE DVIIIESVTVS ENGGGALRPD FCLEPPYTGP CKARIIRYFY
NAKAGICQTF VYGGCRAKRN NFKSAEDCMR TCGGA
```

Figure 4.6. The sequences of His₆PpiB_I_BPTI(I18L,I19L) and His₆PpiB_L_BPTI(L6I,L29I) fusion proteins. The BPTI segments are in bold, the mutated residues are italicised, and the cleavage points are underlined.

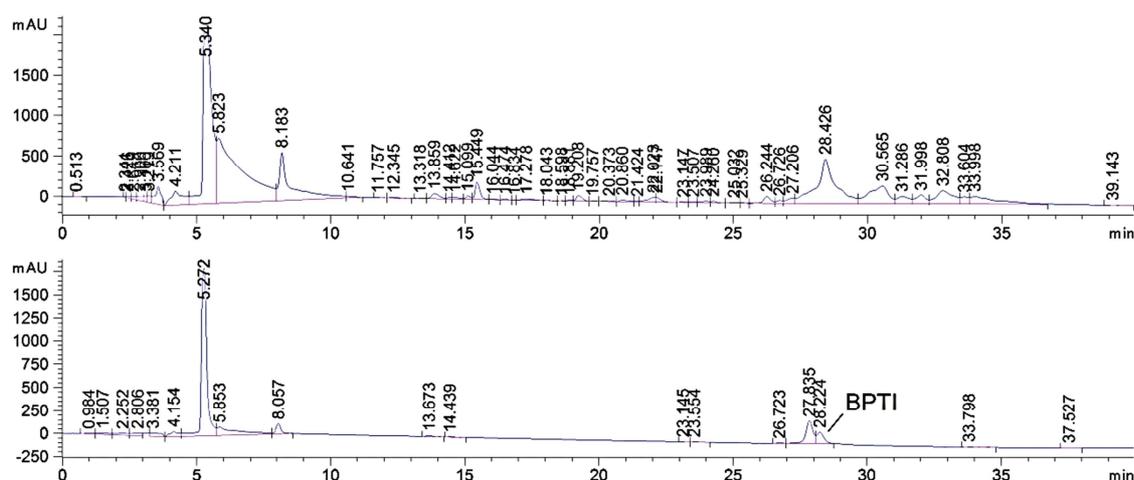


Figure 4.7. Top: isolation of aprotinin-I18L,I19L produced after heat treatment of His₆PpiB_I_BPTI fusion protein expressed with the chloride **26a** by HPLC; bottom: aprotinin standard.

4.4. Use of trypsin column

Finally, to contrast the newly described route with earlier methods, aprotinin was produced following an earlier established procedure.^[143] A trypsin cleavage site (lysine) was introduced at the boundary between aprotinin and the carrier protein PpiB (Figure 4.8). As described above, aprotinin is not digested by trypsin, but instead forms an inhibitory 1:1 complex with trypsin. Thus, when the fusion protein was exposed to trypsin in solution, the carrier protein was digested and free aprotinin was observed by SDS-PAGE (Figure 4.9). Following established protocols,^[143] trypsin in HEPES buffer (pH 7.5) was incubated on a shaker with *N*-hydroxysuccinamide-activated agarose beads for 4 h at 4°C for immobilisation. The beads were then blocked with 1 M ethanolamine for 1 h at 4°C and packed in a column. The fusion protein was then loaded, and the column was incubated on a shaker for 16 h at 4°C, allowing the carrier

protein to be separated and completely destroyed, along with other contaminants, while aprotinin was bound to the resin and subsequently eluted with an acidic buffer (Figure 4.9, right). While aprotinin was obtained in the end, the lengthy procedure highlights the advantages of the rapid heat-induced cleavage.

```
MHHHHHMMVT FHTNHGDIVI KTFDDKAPET VKNFLDYCRE GFYNNTIFHR
VINGFMIQGG GFEPGMKQKA TKEPIKNEAN NGLKNTRGTL AMARTQAPHS
ATAQFFINVV DNDFLNFSGE SLQGWGYCVF AEVVDGMDVV DKIKGVATGR
SGMHQDVPKE DVIIIESVTVS ENGGGGGGGA KAKRPDFCLE PPYTGPCKAR
IIRYFYNAKA GLCQTFVYGG CRAKRNNFKS AEDCMRTCGG A
```

Figure 4.8. The sequence of His₆PpiB_AKAK_BPTI fusion protein. BPTI segment is in bold, the trypsin cleavage site is underlined.

The observed behaviour of the chlorinated amino acids is somewhat surprising, given their greater polarity and reactivity. β and γ -Chlorinated amino acid residues had been shown to react similarly with heteroatom nucleophiles,^[136] but the rapid reaction of γ -chlorinated side-chains during exposure to 100°C in water without affecting the β -modified residues enables regiospecific functionalisation of the two classes of compounds. Additionally, in the past the chlorides were used in smaller peptides,^[144] while the crystal structures of aprotinin demonstrated that the chlorinated amino acids could be introduced into larger proteins as well, without compromising the folding of the protein. The complete preservation of protein structure emphasises that chlorinated amino acids are not inherently detrimental to protein functionality and are potential substitutes for the aliphatic amino acids. This is in accordance with a recent demonstration that peptidyl-prolyl *cis-trans* isomerase B (PpiB) was soluble when

expressed globally substituted with the chlorinated isosteres, and, in the case of (2*R*,3*R*)-2-amino-3-chlorobutyric acid, the enzyme was even shown to be active.^[102]

Furthermore, as established with aprotinin in this study, structurally related amino acids can be easily exchanged without affecting protein structure or function, and the replacement can be used to limit the occurrence of codons used for the incorporation of modified amino acids, essentially putting the analogues under genetic control.

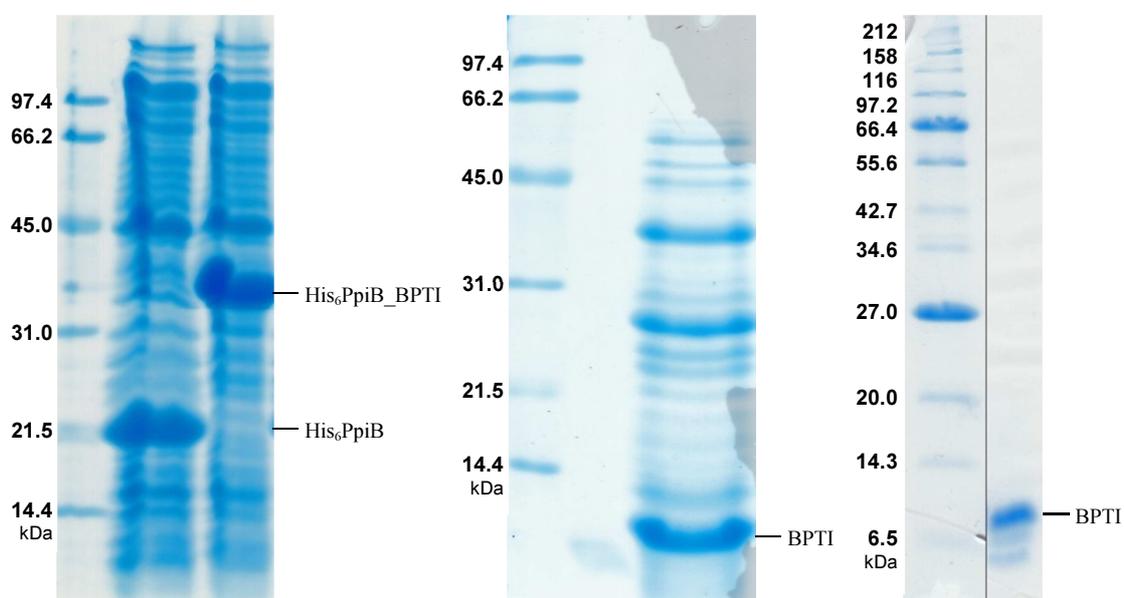


Figure 4.9. The production of aprotinin (BPTI) through trypsin digestion of a fusion protein.

Left: expression of His₆PpiB_BPTI fusion protein; middle: trypsin digestion of His₆PpiB_BPTI fusion protein in solution; right: pure aprotinin eluted from the trypsin column.

4.5. Conclusions

The similar structures and similar roles of aliphatic amino acids allow them to be readily interchanged in proteins without affecting structure or function of the protein. The interchange can be used as a tool to reduce the incidence of sites of incorporation of unnatural amino acids supplied during protein expression, making the incorporation of the analogues site-specific. This enables the application of heat-labile γ -chlorinated amino acid-mediated fusion protein fragmentation to the production of more diverse (generally bigger) proteins. This is particularly pertinent when functionalised peptides are desired, as their production through other means can be problematic. The stability of β -chlorinated amino acids under the conditions of heat-controlled proteolysis makes this method particularly suitable for the production of halogenated peptides, in milligram quantities. The crystal structures demonstrate that β -chlorinated amino acids are able to fulfil the structural roles of aliphatic amino acids, making them highly suitable in protein design.

Chapter 5. Labile Analogues of Methionine

Following the development of the method for peptide production through the expression of fusion proteins with readily cleavable linkers, an expansion of the set of latent substrates capable of bond cleavage was considered. While, the exchangeability^[145] of the non-polar amino acids made the production of most targets possible with just the two analogues **25** and **26c**, with seamless substitutions not affecting target structure or function, as demonstrated in the previous chapter, a wider range of latent analogues could provide more alternatives for amino acid substitutions, potentially requiring fewer modifications, and introduce more versatility to the method.

A pursuit of a latent analogue of methionine **28** was deemed particularly worthy. As one of the least frequently occurring amino acid in proteins,^[146] methionine was expected to be less frequently encountered in smaller peptide targets, thus requiring a lesser modification, and the infamous promiscuity of methionyl-tRNA synthetase (MRS) (see Introduction) made the endeavour more promising. The long library of known accepted (and rejected) unnatural analogues (Table 1.3) made it possible to make generalisations about the selection criteria of the synthetase and make predictions about the acceptance of new structures more reliably.

The primary commonality between all amino acids accepted by MRS appears to be a non-branching side-chain (Table 1.3). This is an added advantage as chemical synthesis of linear analogues is significantly simpler using more ‘routine’ bench chemistry. Additionally, multiple diastereomers are not possible without branching, making stereochemical considerations much simpler.

The majority of the readily accepted analogues of methionine appear to have the same chain length as methionine **28**, thus suggesting size exclusion by the active site is the dominant mechanism of substrate recognition, and the hydrogen bonding with the side-chain (Figure 1.7) is not essential. This is largely supported by the activation of the analogues with hydrocarbon side-chains **35-37** that are not capable of forming hydrogen bonds.

The bonding with the side-chain does play a role, as exemplified by the recognition of ethionine **31** and trifluoromethionine **39**, both sterically bigger than methionine **28**, whereas their respective non-sulfur-containing isosteres homonorleucine **45** and trifluoronorleucine **46** are rejected. Additionally, homocysteine **50** is known to be turned over (proof-read) by MRS,^[69,75] while hydrocarbon-chain bearing amino acids of similar size are not under normal circumstances.^[107] The bond with the sulfur atom formed in the active site could potentially facilitate proper positioning of the residues.

The observed trends in the recognition patterns of methionine analogues suggested that of the two selection mechanisms employed by MRS, size exclusion and hydrogen bonding, only one needs to be satisfied for an amino acid to be recognised as a substrate. Hence, amino acids with linear chains bearing terminal halogens could be potential substrates for MRS (Figure 5.1). 5-Chloro and 5-bromonorvalines **82a** and

82b, with the same chain length as methionine, were anticipated to be recognised by the synthetase, much like the other isosteres of methionine **28**. Although cyclisation of the residues would lead to 6-membered rings, peptide bond cleavage was still expected, and the analysis of both chlorinated and brominated amino acids would enable to compare the propensities of the two halogens for bond cleavage.

Analogues **83a** and **83b** that matched the chain length of previously explored isosteres of isoleucine and leucine but were one carbon unit shorter than methionine **28**, on the other hand, relied on the ability of the halogens to mimic the interactions of the sulfur. It was hoped that the amino acids could be activated like homocysteine **50**, but the proof-reading mechanism employed by the enzyme (Scheme 1.20) would be ineffective due to the electrophilic nature of the halogens.

Finally, 2-aminohex-5-enoic acid **35** was considered for proteolysis through iodolactonisation. Allylglycine **43** that was originally used to develop protein fragmentation through exposure to iodine (Scheme 1.6)^[52,53] is rejected by MRS under normal circumstances,^[107] but it was anticipated that the homologue **35**, a known substrate,^[107,108] could also lead to peptide bond cleavage when treated with iodine.

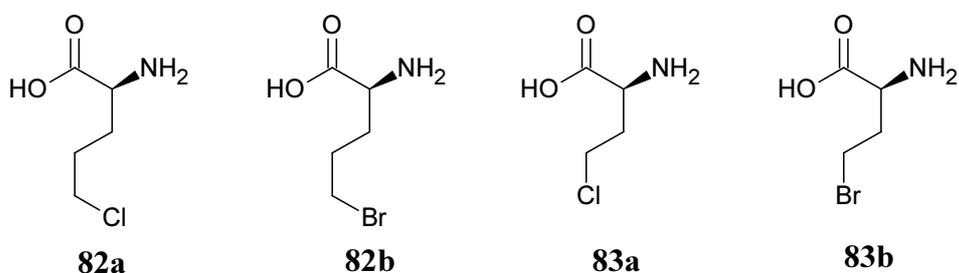
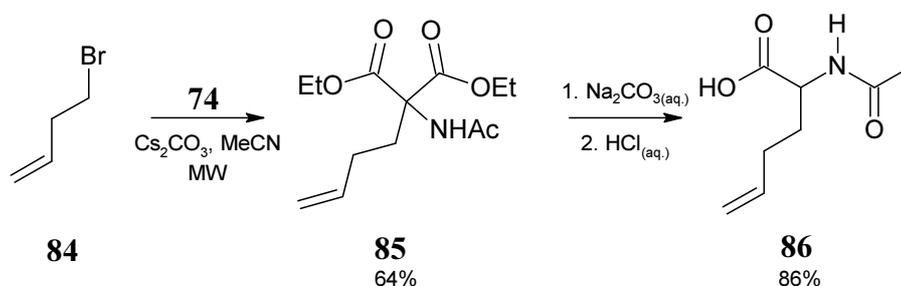


Figure 5.1. Intended halogenated analogues of methionine.

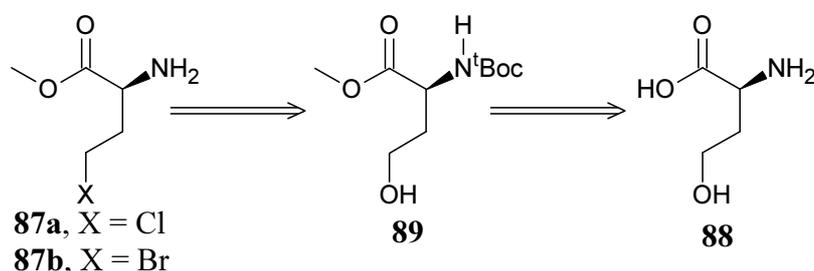
5.1. Preparation of the analogues of methionine

Literature conditions^[108,133] were followed in the preparation of the alkene containing amino acid (Scheme 5.1). Diethyl acetamidomalonate **74** was alkylated with 4-bromobut-1-ene **84**. After saponification and acid decarboxylation of the resultant malonate **85**, 2-acetamidohex-5-enoate **86** was directly used for protein expression, taking advantage of the ability of S30 to deprotect amino acids *in situ* (*vide supra*).



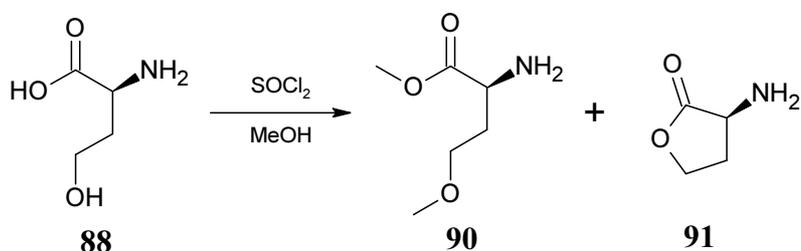
Scheme 5.1. The synthesis of 2-acetamidohex-5-enoic acid **86**.

The analogues **83a** and **83b** were predicted to be too unstable for use in protein expression as free acids, based on work with γ -chlorovaline **26a** (*vide supra*). Hence, the corresponding methyl esters **87a** and **87b**, able to generate the free acids **83a** and **83b** *in situ*, were desired for use in protein expression. Homoserine **88** was chosen as the starting point for the synthesis (Scheme 5.2). As applied during the purification of the analogues of branched-chain amino acids (*vide supra*), the ^tBoc protecting group was chosen for the protection of the amino group. The halogen would be introduced at a final stage, making divergent synthesis was possible and leaving the possibility for the exploration of other potential leaving groups in the future.



Scheme 5.2. Retrosynthetic analysis of the halogenated analogues.

Initially installation of the methyl ester was attempted through esterification with methanol and thionyl chloride.^[147] Unfortunately, the desired product was not observed. Instead, methoxinine methyl ester **90**, a product of over-methylation, and homoserine lactone **91** were isolated (Scheme 5.3), indicating the conditions were too harsh.

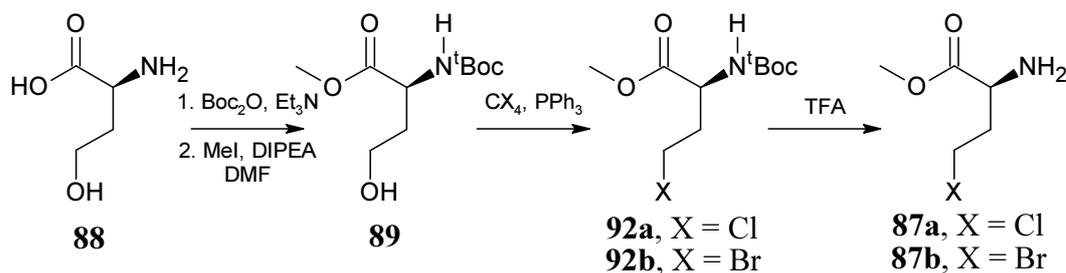


Scheme 5.3. Attempted esterification of homoserine **88**.

It was then decided to introduce the ^tBoc group first and then selectively alkylate the carboxylate with iodomethane under mildly basic conditions, exploiting the different acidities of the hydroxy and the carboxy groups (Scheme 5.4).^[148] The esterified homoserine **89** was isolated by flash silica chromatography, but the yields were low after initial trials. The poor yields were attributed to the propensity of the ester **89** to

cyclise, likely catalysed by the acidic silica. Purification by chromatography was then abandoned, and the diprotected homoserine **89** was taken to the next step without purification or prolonged storage.

An Appel reaction was used to introduce the halogens. Bromination in dichloromethane with 1 equivalent of carbon tetrabromide proceeded rapidly at room temperature and went to near completion with hours. The product **92b**, however, was very prone to decomposition, as signified by gradual emergence of the colour of bromine in the sample. After purification by flash silica chromatography, the bromide **92b** was kept at -80°C to prevent break-down.



Scheme 5.4. Preparation of the halogenated esters.

Chlorination, on the other hand, only proceeded at reflux in carbon tetrachloride (77°C) and required longer reaction times. Complete conversion was not achieved, despite the large excess of the chlorinating reagent, and competing lactonisation was observed during the reaction. After purification, however, the product **92a** was a stable oil and could be stored on bench.

The halogens **92a** and **92b** were treated with glacial TFA to give the methyl esters **87a** and **87b**, respectively.

The halogenated analogues of methionine **28** were then tested for incorporation into the test protein His₆PpiB. 2-Acetamidohex-5-enoate **86** and several other commercially available amino acids were used as controls for the expression. Consistent with expectations, both analogues **87a** and **87b** led to protein production (Figure 5.2). Wild-type levels of protein were produced with norleucine **37** and 2-acetamidohex-5-enoate **86**, both known competitors of methionine **28**,^[107,108,111,114] and the chlorinated analogue **87a**, whereas slightly lower amounts were obtained with the bromide **87b**. This, however, is in stark contrast to the complete rejection of the other short-chained amino acids norvaline **42**, allylglycine **43**, homoserine **88** and homocysteine **50**, where only background levels of protein expression were seen. The recognition of the halides **87a** and **87b** by MRS as substrates strongly supports the idea that the halogens are forming a hydrogen bond in the active site of the enzyme to mimic the thioether of methionine.

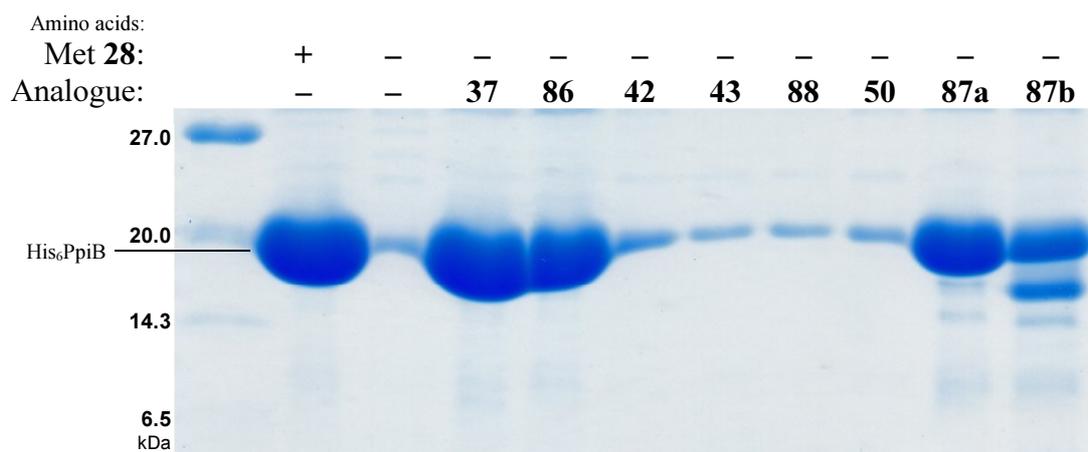


Figure 5.2. SDS-PAGE analysis of the incorporation of analogues of methionine **28** into His₆PpiB.

Incorporation of the analogues was confirmed by ESI-MS analysis of the produced proteins (Figure 5.3). Complete replacement (7/7) of methionine was achieved with norleucine **37**, 2-aminohex-5-enoic **35** acid and the chlorinated analogue **83a** (generated *in situ* from the ester **87a**), whereas only wild-type background protein was obtained with the shorter analogues. The brominated analogue **87b** presented a curiosity in that fully brominated protein was not visible in the spectrum (expected mass 19480 Da); instead, the main peak was consistent with the replacement of 5/7 methionine residues with the bromide **83b** and the remaining two sites occupied by homoserine **88**. Hydrolysis of the bromide in solution is not surprising, particularly given the instability observed during the preparation of the intermediate **92b**, and homoserine is a likely product of degradation. However, attempted protein expression in the presence homoserine **88** clearly established that the hydroxylated amino acid is not directly incorporated during protein expression (Figure 5.2), suggesting the hydroxylated residues detected in the proteins are a result of post-translational damage to brominated

residues. Furthermore, the presence of a clearly dominant signal suggested that the damage was occurring in a sequence-specific manner and the susceptibility of the brominated residues could be influenced by the location in the protein (*e.g.* buried vs. exposed on the surface). However, no conclusion could be drawn from the available data.

More peculiarly, an additional weaker protein band at about *ca.* 16 kDa was observed by the SDS-PAGE analysis of the brominated protein sample (Figure 5.2), several kilodaltons smaller than His₆PpiB (*ca.* 19 kDa). However, no material in that mass range was detectable by MS. Over several months at 4°C, the '16 kDa' band became more intense, whereas the 19 kDa His₆PpiB weakened. Concomitantly, the original 19353 Da signal in the MS became weaker, and the satellite peak at 19272 Da became the dominant peak. The decrease of mass by 80 Da suggested elimination of bromide (isotopes of bromine could not be resolved) and the emergence of the satellite peak matched the intensification of the '16 kDa' protein band on SDS-PAGE, thus it was hypothesised that one of the remaining brominated side-chains in the protein was slowly reacting with an internal nucleophile, *e.g.* a cysteine thiol or a histidine imidazole, and the strain introduced to the chain by this reaction led to the aberrant migration of the protein on the gel. Bromoalkyl chains have been shown to readily react with proximal sulfur and nitrogen nucleophiles,^[137] but no conclusive evidence to support this hypothesis is available, although the results of the fragmentation analysis are consistent with such reactions taking place (*vide infra*).

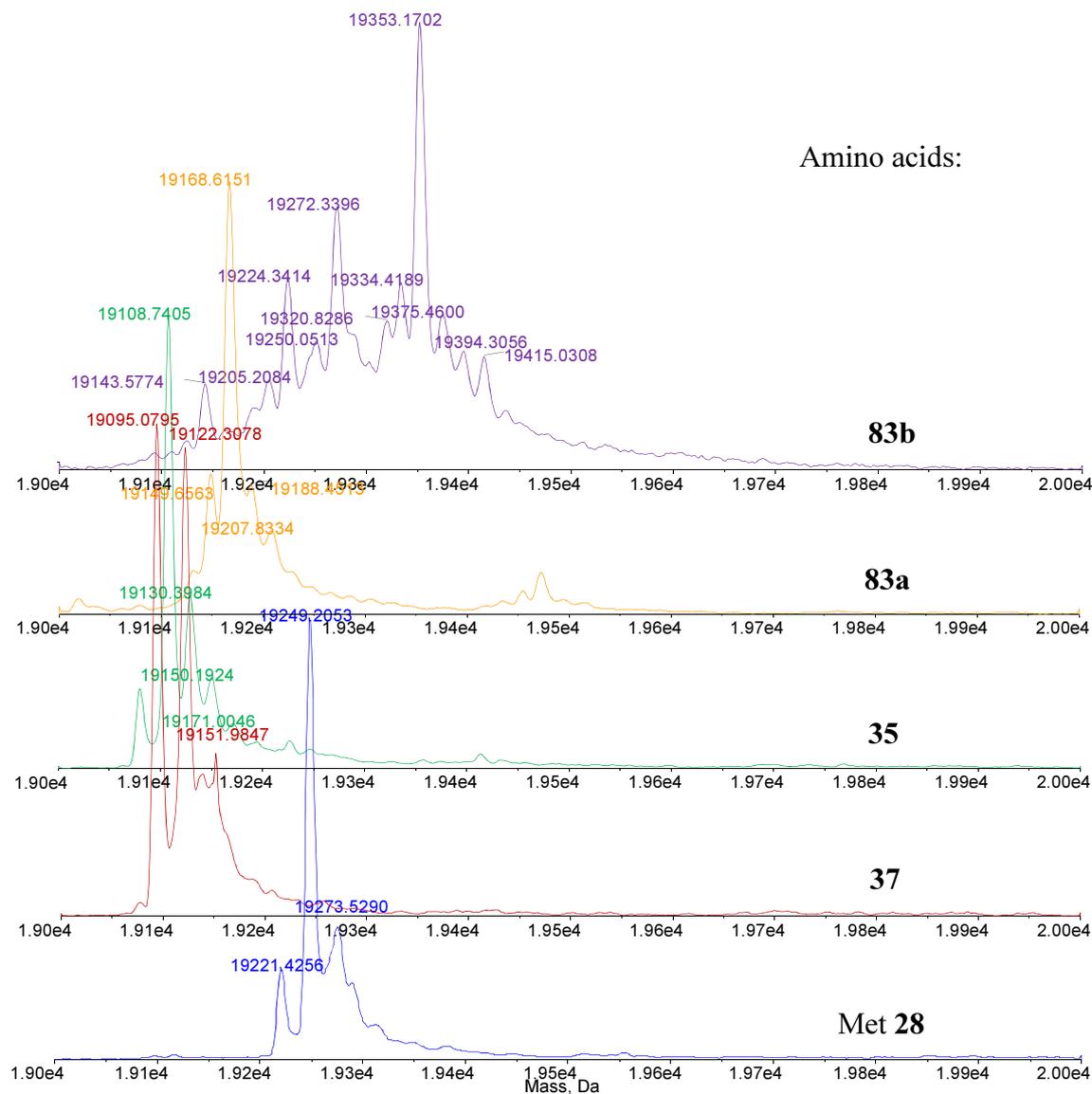
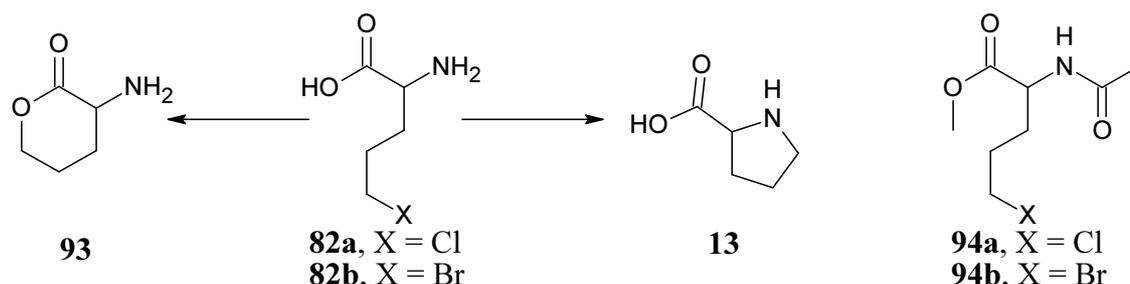


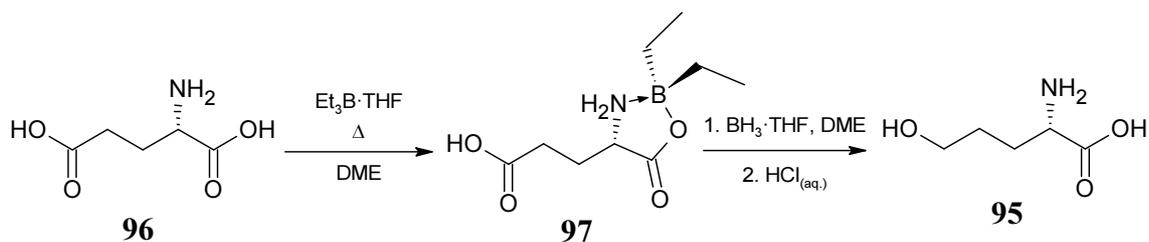
Figure 5.3. Deconvoluted ESI-MS of His₆PpiB expressed with the analogues of (*S*)-methionine **28**. The spectra show complete replacement (7/7) of (*S*)-methionine **28** (blue) with (*S*)-norleucine **37** (19122 Da; red), (*S*)-2-amino-hex-5-enoic acid **35** (19109 Da; green) and (*S*)-2-amino-5-chlorobutyric acid **83a** (19169 Da; orange), but not (*S*)-2-amino-5-bromobutyric acid **83b** (purple; see text). A second (-28 Da) signal (particularly strong in the sample containing (*S*)-norleucine **37**; red) corresponds to *N*-deformylated protein.

The halides **82a** and **82b** were also expected to be unstable with respect to lactonisation, similar to the γ -chlorinated amino acids, forming the lactone **93** (Scheme 5.5). In addition, participation of the free amine to form proline **13** would also be favourable due to the longer chain, as observed with chelate-directed functionalisation of (*S*)-norvaline.^[125] Thus, to avoid decomposition, it was decided to attempt protein expression with di-protected amino acids **94a** and **94b**, relying on *in situ* deprotection. *O*-Methyl and *N*-acetyl protecting groups were chosen as (*S*)-*N*-acetylleucine methyl ester was shown to be deprotected by the S30 (Chapter 2).



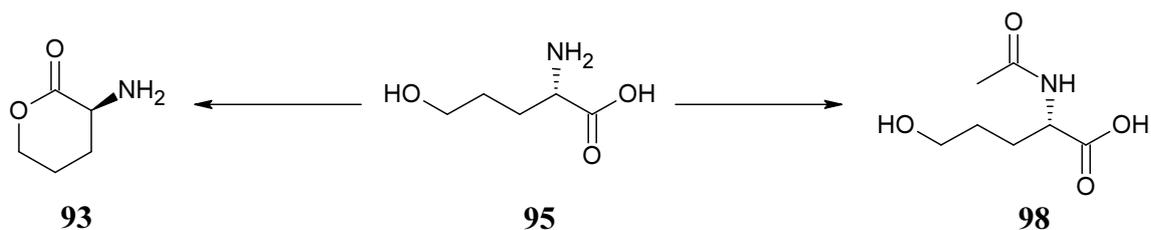
Scheme 5.5. The potential routes of decomposition of the halides **82a** and **82b** (left) and the protected stable analogues **94a** and **94b** (right).

(*S*)-5-Hydroxynorvaline **95**, the key intermediate in the synthesis of the halides **94a** and **94b**, was prepared by reduction of (*S*)-glutamic acid **96** (Scheme 5.6).^[149] Triethylborane was used to regioselectively protect the main-chain carboxy-group from reduction, producing the boroxazolidone **97** almost quantitatively. The side-chain was then selectively reduced with borane, and the boroxazolidone was hydrolysed with aqueous acid.



Scheme 5.6. Synthesis of (*S*)-5-hydroxynorvaline **95**.

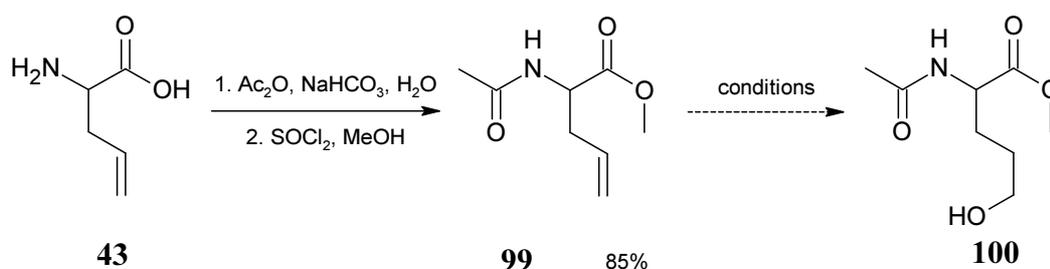
Conversion of (*S*)-glutamic acid **96** to (*S*)-5-hydroxynorvaline **95** was not complete, with some starting material still present in the sample, but the hydroxynorvaline **95** was used in subsequent steps without purification due to the propensity for lactonisation (Scheme 5.7). Formation of the lactone **93** was not extensive during the preparation of 5-hydroxynorvaline **95**, but became a strongly competing reaction during the attempted acylation. Only small amounts of *N*-acetyl-5-hydroxynorvaline **98** were obtained, thus a different approach to prepare the desired compounds was attempted.



Scheme 5.7. Derivatisation of 5-hydroxynorvaline **95**.

Hydroboration of protected allylglycine **99** was considered (Scheme 5.8).^[150] After methylation and acetylation of allylglycine **43**, the alkene **99** was subjected to a variety of hydroborating agents, like 9-borabicyclo[3.3.1]nonane (9-BBN) and

dicyclohexylborane **101**, followed by hydrogen peroxide, but the desired product was not detected. Dicyclohexylborane **101** was produced *in situ* from borane and dicyclohexene and the formation was confirmed by EI-MS (low (Figure 5.4) and high resolution); the allylglycine-borane adduct, however, was never observed. The bulkiness of the borane reagents may have been a hinderance. Hydroxylated borane **102** was also detected in the mixture (Figure 5.4), suggesting trace amounts of moisture may have interfered with the reaction.



Scheme 5.8. Hydroboration of allylglycine.

By this stage of the synthesis, the homoserine derivatives **87a** and **87b** were produced and shown to substitute methionine **28** during cell-free protein expression. Therefore, as a set of halogenated analogues of methionine was available, it was decided to no longer pursue the compounds **94a** and **94b**.

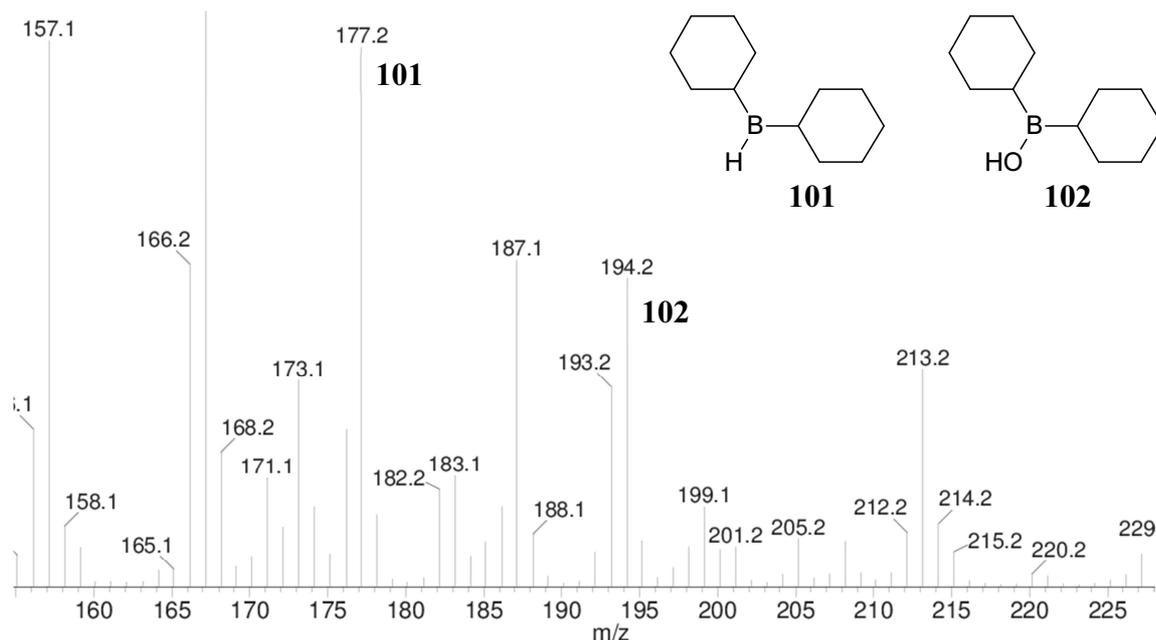


Figure 5.4. EI-MS of dicyclohexylborane **101** intermediate.

5.2. Fragmentation analysis of modified His₆PpiB

To validate the analogues of methionine could be used for peptide production, fragmentation of substituted protein was investigated. His₆PpiB expressed with the halides **87a** and **87b** was subjected to 100°C in water for up to 1 h, and the samples were analysed by SDS-PAGE (Figure 5.5). As observed with analogues of BCAAs (Chapter 3), the reaction was complete within minutes, with little change visible after the first 5 min. Both the chlorinated and the brominated protein showed similar results, with no distinct protein bands clearly discernable in the heated samples.

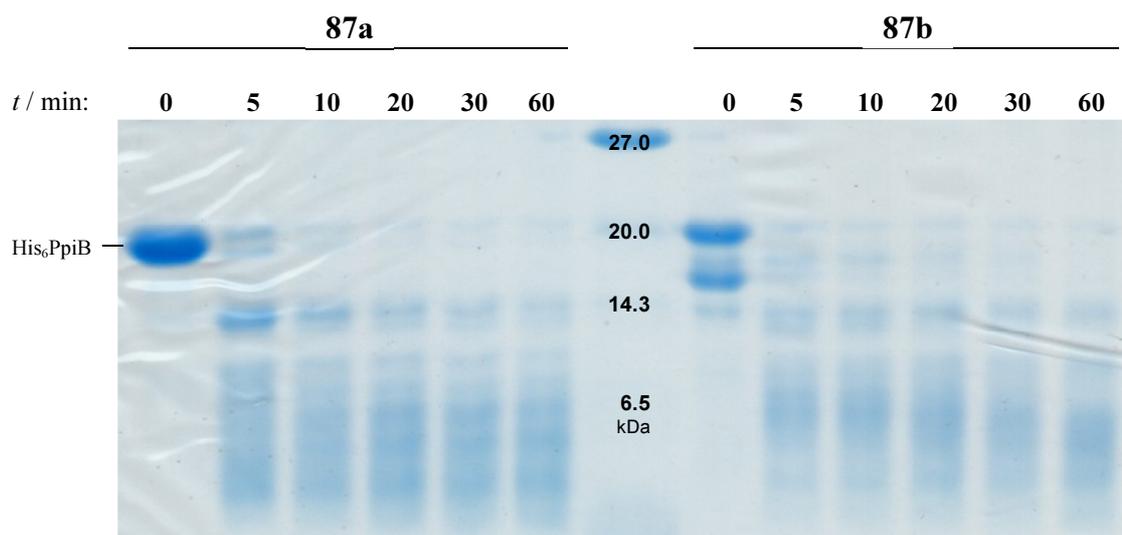


Figure 5.5. SDS-PAGE analysis of the fragmentation of His₆PpiB expressed with the chloride **87a** (left) and the bromide **87b** (right) instead of (*S*)-methionine **28** due to exposure to 100°C.

The heat-treated samples were then analysed by MS. As His₆PpiB contains seven methionine residues in the sequence, a total of eight peptides are expected upon complete fragmentation of the modified protein, listed in Table 5.1. In the sample of His₆PpiB expressed with the chloride **87a**, four of the predicted fragments, 2806 Da (Figure 5.6; [M+4H]⁴⁺, m/z = 703 Da), 2137 Da (Figure 5.6; [M+3H]³⁺, m/z = 713 Da), 1584 Da (Figure 5.6; [M+2H]²⁺, m/z = 793 Da) and 943 Da (Figure 5.7; [M+H]⁺, m/z = 944 Da), were clearly identified by MS, observed in both protonated and partly sodiated states. As with previously analysed samples (Chapter 3), some of the predicted fragments could not be detected, likely due to poor ionisability, and some of the detected signals could not be assigned to any compounds resulting for the collapse of His₆PpiB, likely caused by impurities.

Table 5.1. Predicted peptide fragments generated by cleavage of His₆PpiB at methionine positions. The N-terminal fragment may be *N*-formylated.

Peptide fragment	Mass, Da
λ*	101, 129
HHHHHHλ*	923
VTFHTNHGDI VIKTFDDKAP ETVKNFLDYC REGFYNNITF HRVINGFλ*	5614
IQGGGFEPGλ*	943
KQKATKEPIK NEANGLKNT RGTLAλ*	2806
ARTQAPHSAT AQFFINVVDN DFLNFSGESL QGWGYCVFAE VVDGλ*	4889
DVVDKIKGVA TGRSGλ*	1584
HQDVPKEDVI IESVTVSEN	2137

*λ denotes lactone produced by the fragmentation process

A clearly discernable signal of a 3704 Da peptide (Figure 5.6; $[M+5H]^{5+}$, $m/z = 742$ Da), however, was identified as matching the combined size of the two C-terminal-most predicted fragments of His₆PpiB. The detected mass was consistent with elimination of the chloride from the halogenated residue at the interface between the two fragments, resulting in the 5-membered intermediate. However, the intermediate lactone-imine is unlikely to have survived the exposure to 100°C. It was therefore theorised that the chlorinated side-chain reacted with the imidazole ring of the neighbouring histidine (Figure 5.8). However, the exact structure of the fragment could not be deduced from the mass spectrum, as the chlorinated residue could have reacted with any other side-chain.

Chlorinated Amino Acids in Peptide Production
Chapter 5. Labile Analogues of Methionine

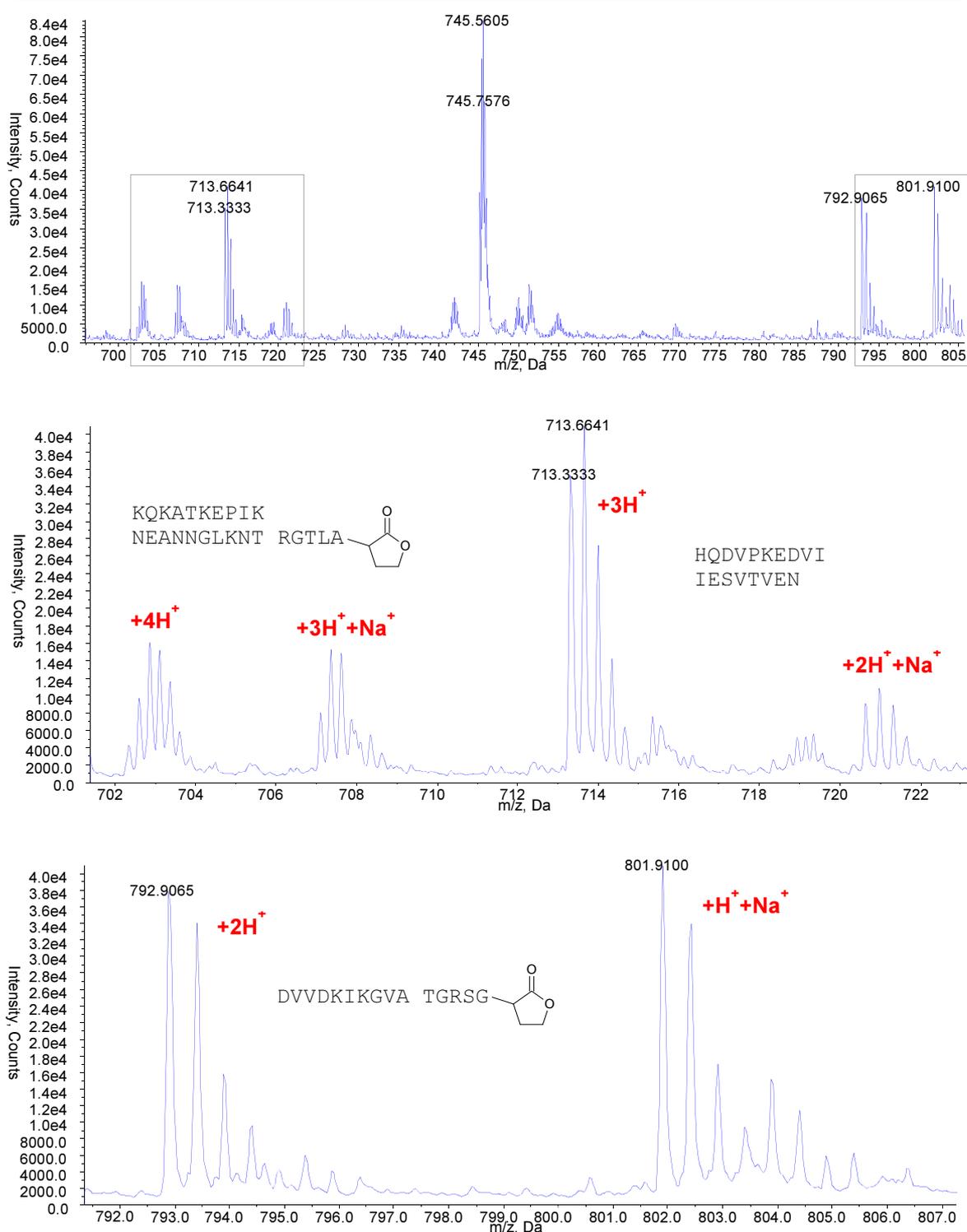


Figure 5.6. Representative ESI-MS signals of the 2806 Da, 2137 Da and 1584 Da peptide fragments identified in the sample of heat-treated His₆PpiB expressed with the chloride **87a** instead of (*S*)-methionine **28**. Broad window (top) and enlarged images shown.

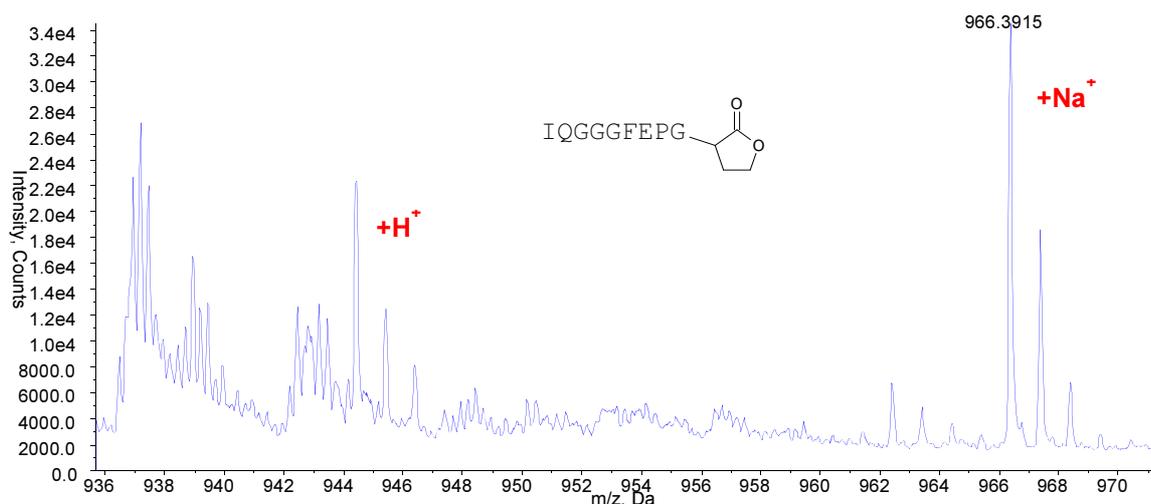


Figure 5.7. Representative ESI-MS signals of the 943 Da (IQGGGFEPGλ^{*}) peptide fragment identified in the sample of heat-treated His₆PpiB expressed with the chloride **87a** instead of (*S*)-methionine **28**.

While the exact nature of the peptide could not be identified, it was clear that the fragment was a result of incomplete fragmentation of His₆PpiB. The two C-terminal fragments of His₆PpiB comprising the larger peptide were detected individually (1584 Da and 2137 Da; Figure 5.6), thus confirming that peptide bond cleavage at the intended cleavage site was not completely precluded, but the modification of the halogenated residue leading to the formation of the larger 3704 Da fragment appeared to be in competition with bond scission. No other fragments resulting from incomplete bond cleavage were identified, however, suggesting the side reaction is sequence dependent, and the extent to which each of the pathways is followed could not be determined from the mass spectra as the peptides detected in Figure 5.6 are all in different charged states and structurally different, thus the intensities of the signals are not a true reflection of the relative amounts of the fragments.

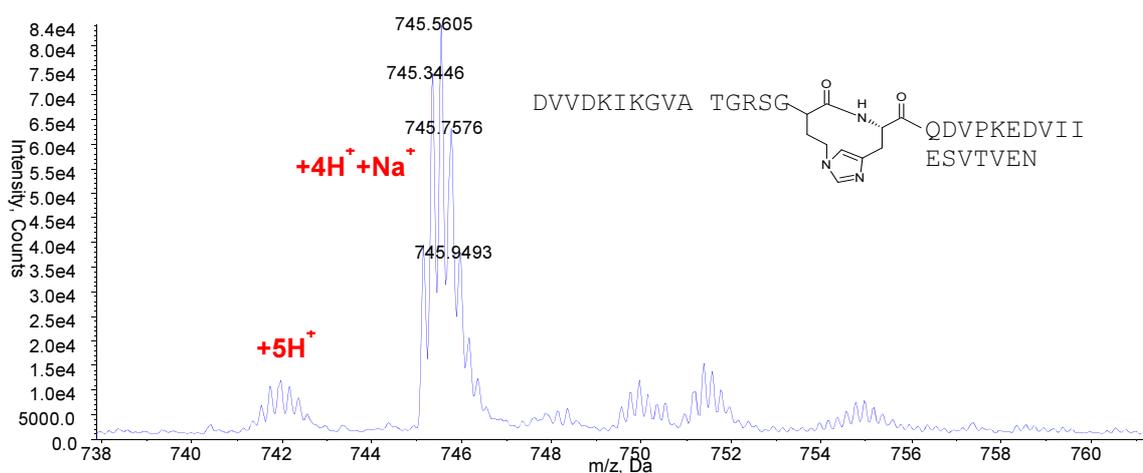


Figure 5.8. Representative ESI-MS signal of the 3704 Da peptide fragment identified in the sample of heat-treated His₆PpiB expressed with the chloride **87a** instead of (*S*)-methionine **28**.

In the sample of the heat treated His₆PpiB expressed with the bromide **87b** the strongest signals in the mass spectrum belonged to the 3704 Da fragment (Figure 5.9). None of the predicted eight fragments (Table 5.1) were clearly discernable. This could be explained by the damage to the brominated side-chains observed by mass spectrometry in full-length His₆PpiB (*vide supra*). The detection of the single fragment indicated the brominated amino acid **83b** is capable of facilitating bond fragmentation, but the susceptibility to side reactions posed a serious hindrance to the use as a cleavable linker.

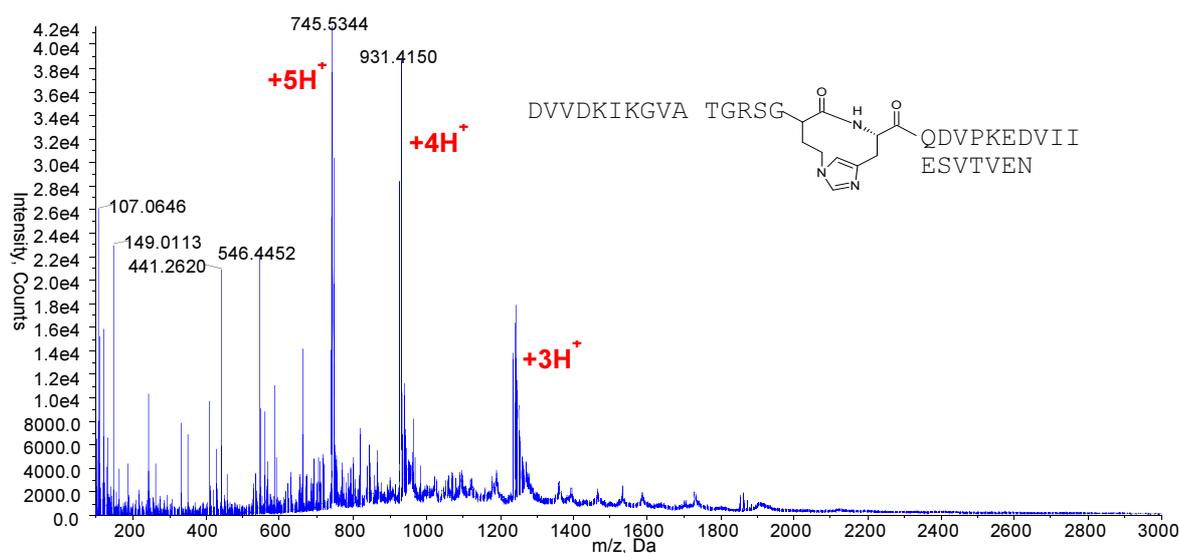


Figure 5.9. ESI-MS of His₆PpiB expressed with the bromide **87b** instead of (*S*)-methionine **28** after exposure to 100°C for 5 min.

To test whether the unsaturated but-3-enyl side-chain can be exploited for proteolysis like the olefins described in Chapter 3, His₆PpiB containing 2-aminohex-5-enoic acid **35** was incubated with excess amounts of molecular iodine in tetrahydrofuran from 5 min to 1 h on ice. Protein expressed with norleucine **37** was used as control, allowing the comparison between the reactivities of the alkene and the corresponding alkane, rather than a thioether of methionine **28**. The samples retained the colour of iodine throughout the incubation, thus indicating that iodine was not prematurely consumed. After the desired time the reaction was quenched with tris(2-carboxyethyl)phosphine (TCEP) and the samples were analysed by SDS-PAGE.

No intact protein was detectable in samples with unsaturated side-chains after 5 min, whereas the saturated control was not degraded after 1 h (Figure 5.10), suggesting the alkene-bearing chain was fully capable of mediating peptide bond cleavage. However, a

number of additional protein bands, attributable to damaged His₆PpiB whose migration was affected by modifications, appeared in the control sample, highlighting the susceptibility of the protein to side reactions with iodine. Furthermore, no change was seen between the 5 min and 1 h samples, indicating the rate of non-specific reactions occurring on the protein rivals the rate of bond cleavage, thus explaining why yields of iodine-mediated peptide production are lower than the yields of heat cleavage (see Chapter 3).

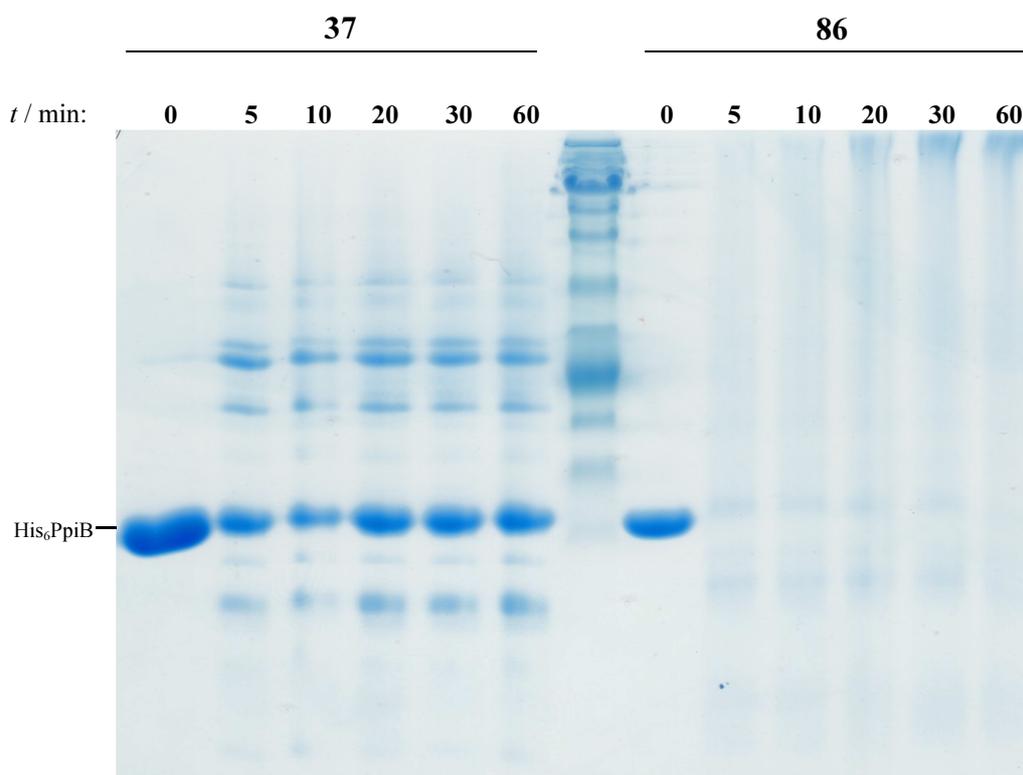


Figure 5.10. SDS-PAGE analysis of the reaction of His₆PpiB expressed with norleucine **37** (left) and 2-acetamidohex-5-enoic acid **86** (right) in place of (*S*)-methionine **28** with iodine at 0°C. After the indicated time the reactions were quenched with TCEP.

ESI-MS was then used to probe the side-reactions occurring on the protein during iodination. Wild-type His₆PpiB was treated with small amounts of iodine for 5 min. The sample was then stored at -78°C to stop the reaction, and only thawed immediately prior to the analysis. No single species was clearly identifiable in the spectrum, but instead an array of peaks, roughly clustered into five groups, ranging from 20.0 kDa to 20.6 kDa (Figure 5.11; *cf.* 19.2 kDa for unmodified His₆PpiB). The masses of the five peak clusters roughly correspond to His₆PpiB with 7-11 added iodine atoms. The aromatic residues are most susceptible to the addition of iodine. As His₆PpiB contains one tryptophan, three tyrosine and twelve phenylalanine residues, the MS pattern suggests fairly rapid double iodination of tyrosine and the tryptophan, with potentially slower modification of phenylalanine. This reactivity of the aromatic systems toward iodine can account for the lower yields of recovered peptide hormones after iodine-mediated protein fragmentation, as compared to the yields of heat-responsive fragmentation.

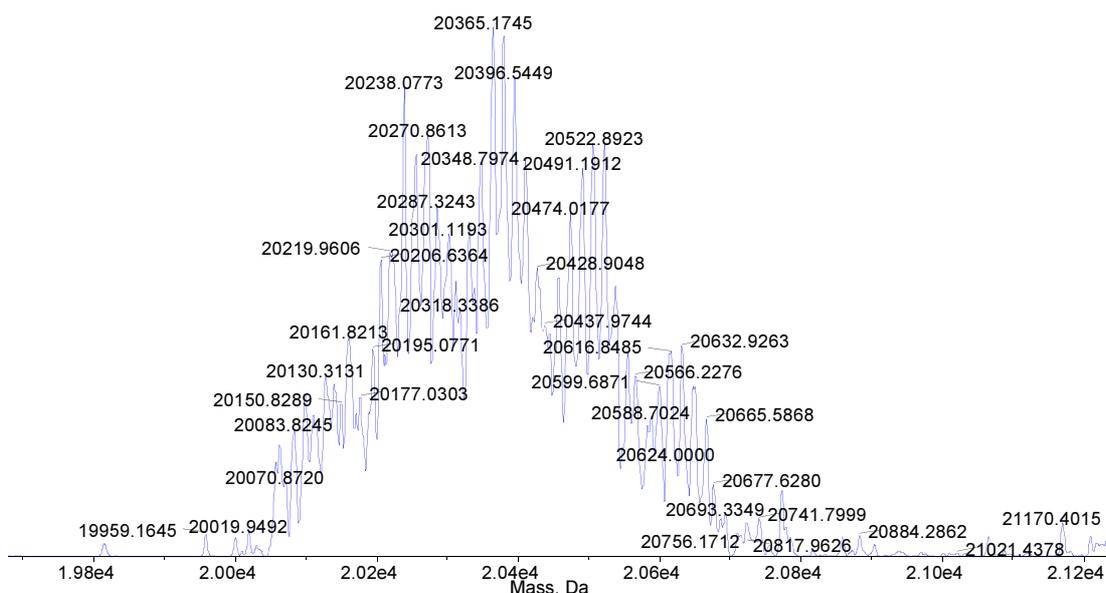


Figure 5.11. Deconvoluted ESI-MS spectrum of His₆PpiB after exposure to iodine.

5.3. Production of oxytocin

Following the success of the preliminary fragmentation analysis, the utility of latent analogues of methionine was tested in the preparation of a model peptide, using the previously described model (Chapter 3). Oxytocin was chosen as it naturally lacks methionine in the sequence (CYIQNCPLG),* making it an ideal candidate.

The gene for oxytocin was obtained as a DNA primer and attached to the C-terminus of His₆PpiB *via* a methionine residue (Figure 5.12) by PCR amplification, as described earlier. The fusion protein was expressed in the cell-free system with the analogues of methionine **87a**, **87b** and **86** (Figure 5.13).

```
MHHHHHMMVT FHTNHGDIVI KTFDDKAPET VKNFLDYCRE GFYNNTIFHR  
VINGFMIQGG GFEPGMKQKA TKEPIKNEAN NGLKNTRGTL AMARTQAPHS  
ATAQFFINVV DNDFLNFSGE SLQGWGYCVF AEVVDGMDVV DKIKGVATGR  
SGMHQDVPKE DVIIESVTVS ENGGGGGGGA MCYIQNCPLG
```

Figure 5.12. The sequence of His₆PpiB_OXT fusion protein. The hormone segment is in bold, the cleavage point is underlined.

Essentially complete substitution was achieved with the chloride **87a**, as confirmed by ESI-MS (Figure 5.14). The chlorinated protein was heated to 100°C for 5 min and then analysed by reverse phase HPLC (Figure 5.15). To ensure any non-specific disulfide bonds established between the produced oxytocin and other fragments during the heat treatment could not sequester the target protein, the sample was reduced with TCEP

* – the dash denotes a disulfide bridge

prior to the analysis. Much to the anticipation, the oxytocin peptide was detected and isolated from the sample (Figure 5.16), eluting at 14.5 min.

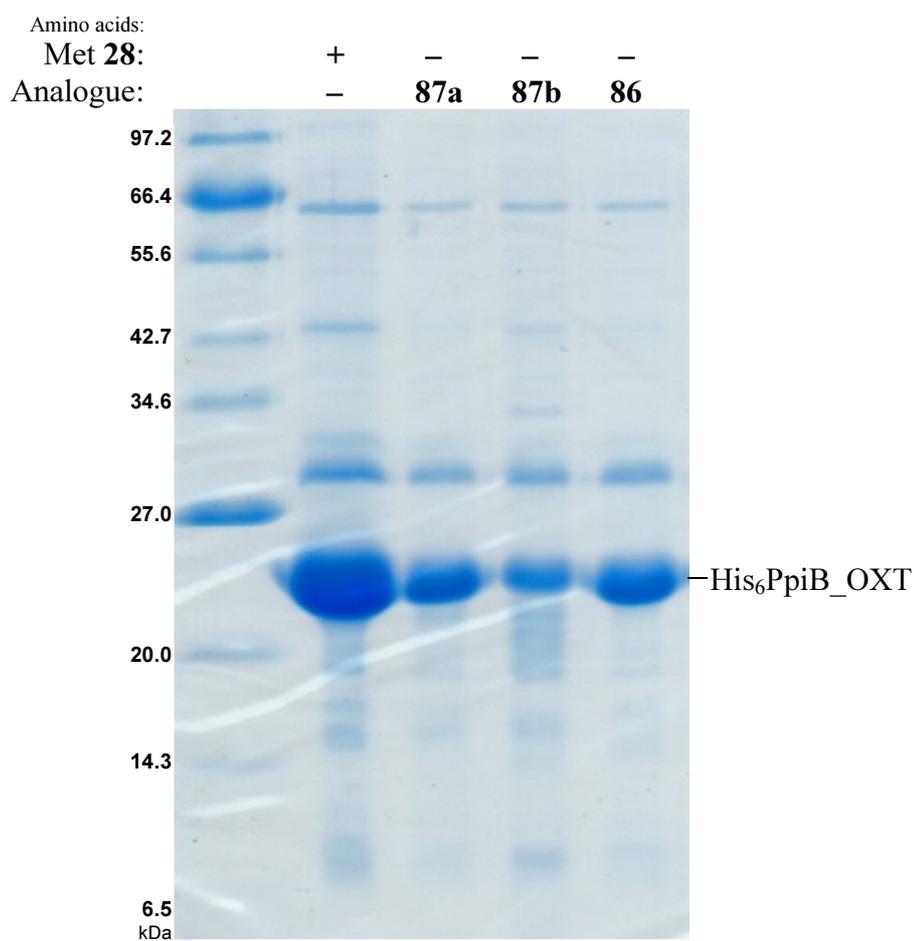


Figure 5.13. SDS-PAGE analysis of cell-free protein expression of His₆PpiB_OXT fusion protein with the analogues **87a**, **87b** and **86** in place of (*S*)-methionine **28**.

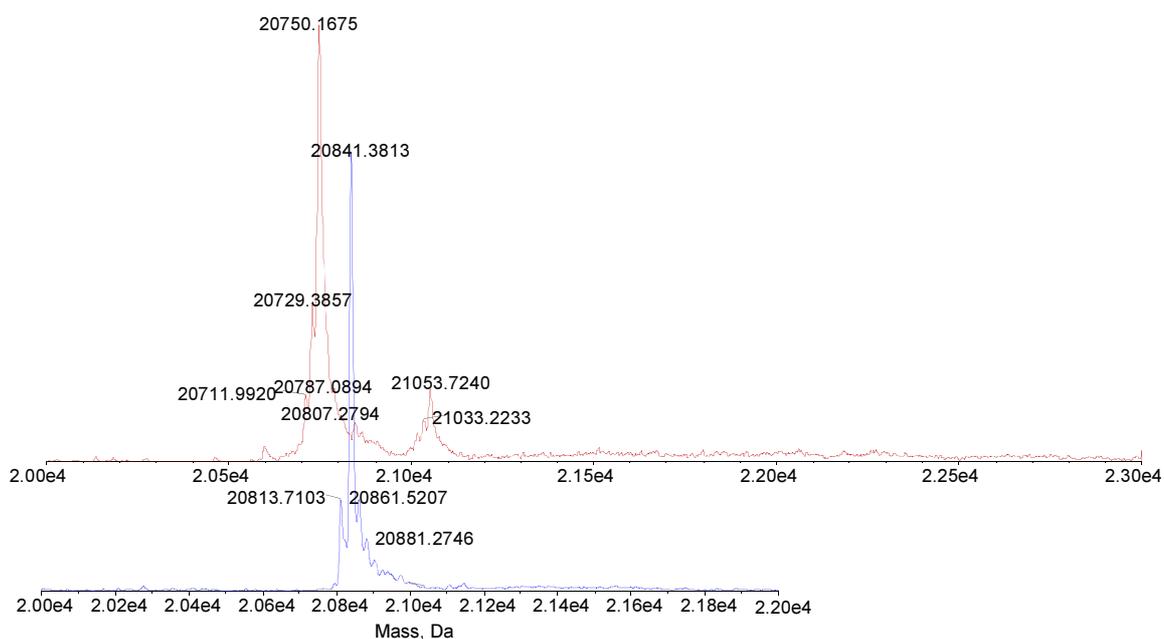


Figure 5.14. Deconvoluted ESI-MS spectra of His₆PpiB_OXT fusion proteins expressed with all natural amino acids (below; blue) and with methyl 2-amino-4-chlorobutanoate **87a** instead of (*S*)-methionine **28** (above; red). The mass reduction of 92 Da corresponds to complete replacement (8/8) of methionine (8×11.5 Da; the apparent change of 91 Da is the result of unresolved peaks due to isotopic distribution).

Aside from the target oxytocin, other fragments from the carrier protein His₆PpiB, detected in the heated sample (*vide supra*), were also isolated and clearly identified by ESI-MS (Figure 5.15). Even more satisfactorily, the largest of the fragments from the carrier protein (5614 Da) that could not be detected in the mixture, likely due to poorer ionisability, eluted as a single species (27.8 min) and was clearly identified (Figure 5.17). Interestingly, no fragments resulting from incomplete fragmentation were detected, although they could have been masked by co-eluting impurities, as not all of the peaks in the chromatogram were identified.

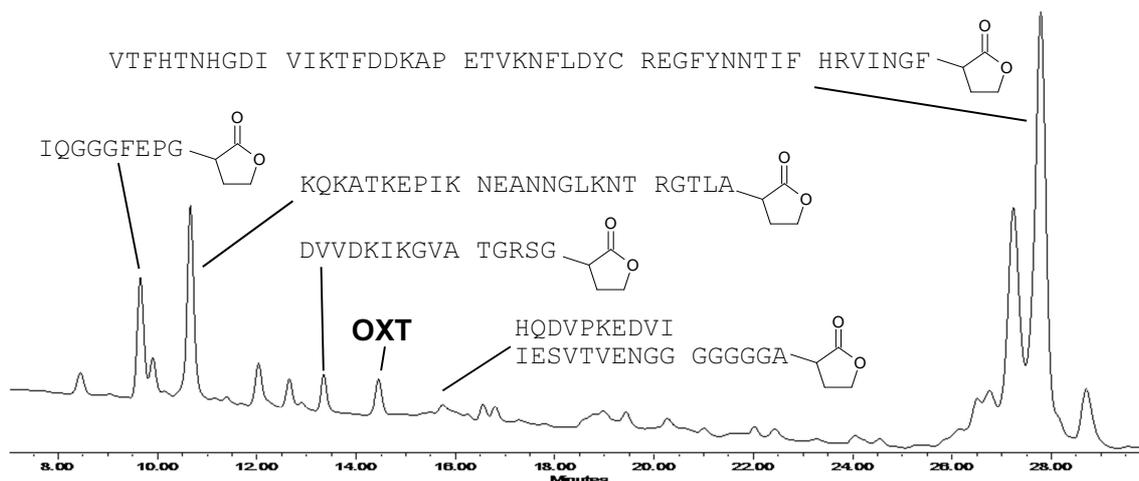


Figure 5.15. HPLC chromatogram of His₆PpiB_OXT fusion protein expressed with the chloride **87a** instead of methionine **28** after exposure to 100°C for 5 min, analysed on a Symmetry300TM C18 column (4.6×150 mm, 5μ), eluting with acetonitrile gradient from 10% to 60% in 0.1% TFA over 40 min, 1 mL/min.

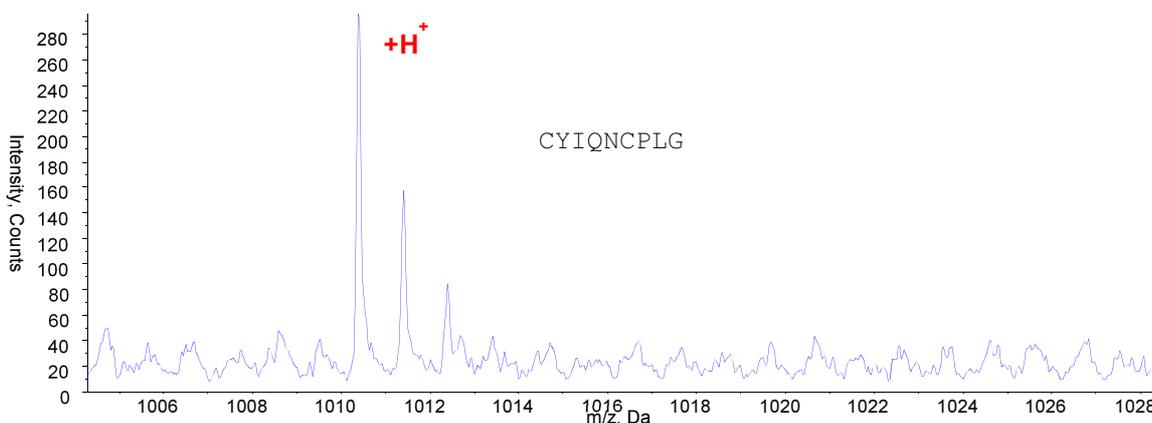


Figure 5.16. The MS signal of the isolated oxytocin fragment, produced after heat-treatment (100°C for 5 min) of His₆PpiB_OXT fusion protein expressed with the chloride **87a** instead of (*S*)-methionine **28**.

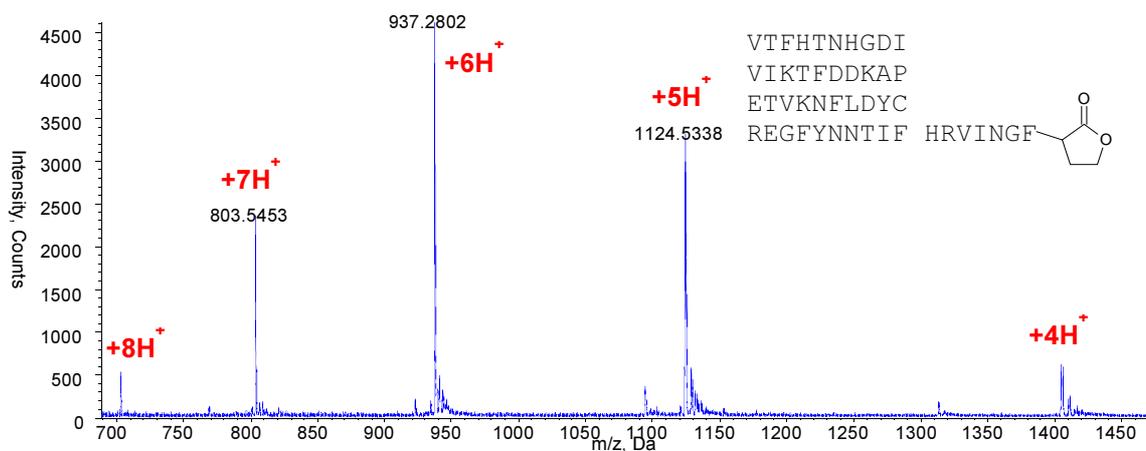


Figure 5.17. ESI-MS of the 5614 Da peptide fragment eluted at 27.8 min.

To estimate the yield of peptide bond fragmentation, the concentration of the purified chlorinated fusion protein was measured by absorbance at 280 nm. The extinction coefficient was calculated based on the amino acid sequence using online ProtParam tool; the chlorinated side-chains were judged not to influence the absorbance at 280 nm. To account for the impurities that were non-specifically retained during nickel affinity chromatography, the intensities of protein bands on the SDS-PAGE gel were integrated using ImageJ image processing software; based on the gel in Figure 5.13 *ca.* 20% of the protein in the sample were contaminants.

HPLC traces were recorded at 210 nm. A calibration curve was used to measure the amount of isolated oxytocin.

The efficiency of the bond cleavage to produce oxytocin was found to be about 40%. More accurate estimates were not possible due to the difficulty of measuring exact concentrations of the fusion protein. However, the 40% yield is significantly smaller than the >70% observed in the preparation of GRPG and CCKG (Chapter 3). The

cysteine at the N-terminus of oxytocin is likely responsible for the drop in the yield of bond cleavage. The nucleophile in the proximity of the scissile bond can interfere with the chemical cleavage, either by mediating the collapse of the cyclic intermediate in a way that does not lead to bond cleavage (Scheme 1.3, right) or directly displacing the chloride leaving group as observed with the 3704 Da fragment (Figure 5.8), and the high temperatures are likely to exacerbate the side-reactions. This was particularly prominent during the attempted preparation of prolactin releasing peptide prohormone PRPG that starts with serine (Chapter 3). While problems associated with the serine at the cleavage point can be partly circumvented during cyanogen bromide-mediated cleavage, no PRPG was produced using heat responsive bond cleavage. Although cysteine is not reported as interfering with the cyanogen bromide-mediated cleavage reaction, the increased reactivity of the thiol group at higher temperatures is not surprising.

Neighbouring residue participation can also potentially explain the intensity of the 5614 Da fragment in the chromatogram (Figure 5.15). Besides being the biggest peptide fragment and hence having the largest extinction co-efficient, the cleavage points at both termini of the peptide are next to amino acids with non-nucleophilic side-chains (valine at the N-terminus, isoleucine at the C-terminus), thus bond cleavage would have less interference and the peptide is expected to be produced with the highest possible yields.

Although production of oxytocin was not achieved with as high a yield as the previously reported peptides (Chapter 3, Chapter 4), the reduction in yield is a consequence of the sequence of the target peptide, specifically, the N-terminal residue.

As was demonstrated by the fragmentation analysis (*vide supra*), the bond cleavage is efficient when the chlorinated side-chain does not precede a serine, threonine or cysteine. Hence, the chloride **87a** presents an efficient method for cleavage of peptide bonds at methionine residues.

Expression level of the brominated His₆PpiB_OXT fusion protein was lower (Figure 5.13), likely due to more rapid decomposition of the brominated amino acid **83b** generated *in situ* from the ester **87b**. As was the case with brominated His₆PpiB (*vide supra*), some of the brominated side-chains hydrolysed after incorporation into protein, based on ESI-MS (Figure 5.18). The major peaks at 20846.5 Da and 20913.7 Da seen in the spectrum approximately matched the masses of proteins with four and three out of eight damaged brominated side-chains, respectively.

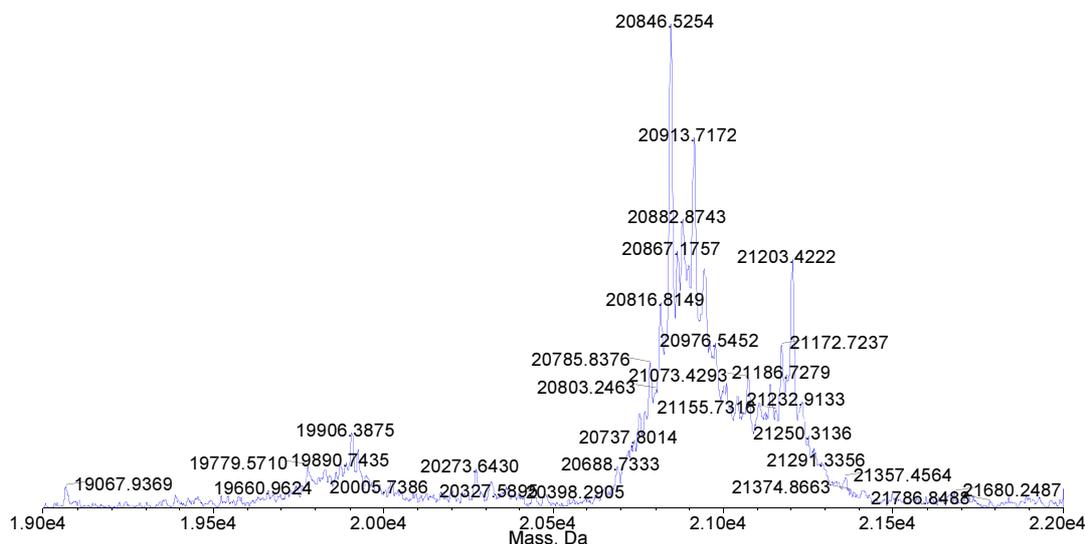


Figure 5.18. Deconvoluted ESI-MS spectrum of His₆PpiB_OXT fusion protein expressed with methyl 2-amino-4-bromobutanoate **87b** instead of methionine **28**.

Oxytocin was detected in the sample of the brominated protein exposed to 100°C for 5 min by HPLC (Figure 5.19), but only about 5% of the expected amount was recovered. This is not surprising given the propensity of the brominated side-chain for undergoing side reactions. Additionally, the brominated residue bridging the carrier protein and oxytocin is most likely exposed to the solvent, making the halide particularly susceptible to external reagents. As such, the brominated amino acid **87b** may not be particularly well suited for the preparation of small peptides *via* the described approach, although production of peptides with less reactive N-termini may have substantially higher yields. Regardless, the brominated side-chains have been utilised in the post-translational functionalisation of proteins,^[137] and the capacity to introduce the functionality into proteins through protein expression can significantly simplify the preparation procedures.

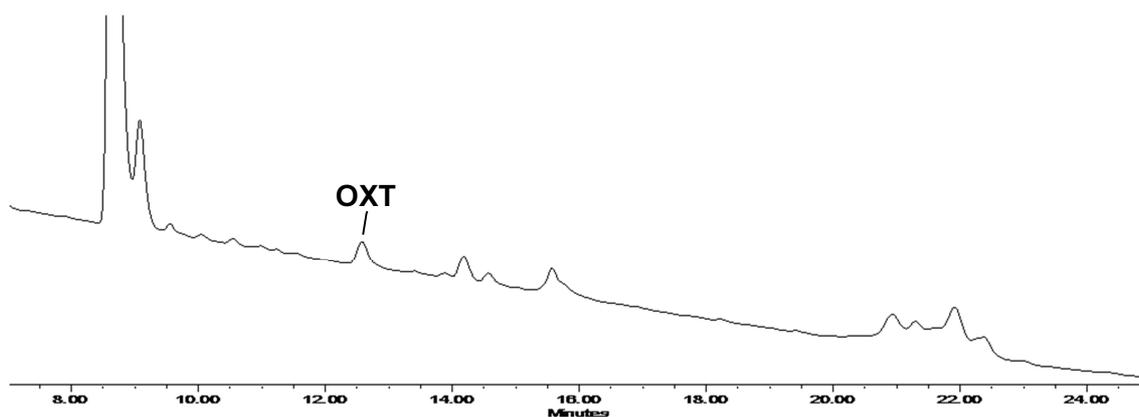


Figure 5.19. HPLC chromatogram of His₆PpiB_OXT fusion protein expressed with the bromide **87b** instead of (*S*)-methionine **28** after exposure to 100°C for 5 min, analysed on a Symmetry300TM C18 column (4.6×150 mm, 5μ), eluting with acetonitrile gradient from 10% to 60% in 0.1% TFA over 30 min, 1 mL/min.

Finally, the unsaturated protein was investigated. Based on MS analysis, near complete substitution of methionine **28** with the alkene **35** was achieved (Figure 5.20). The protein was treated with molecular iodine at 0°C from 2 to 10 min, following a previously described method (Chapter 3). However, oxytocin was not detected.

One possible explanation for the lack of oxytocin production could be a different mechanism of protein fragmentation. While protein substituted with the alkene **35** clearly underwent fragmentation upon exposure to iodine (Figure 5.10), the longer chain of the labile amino acid could lead to cleavage of an alternate peptide bond. Alternatively, the side-reactions associated with iodination could be consuming the target peptide too rapidly. Finally, the cysteine could be interfering with the fragmentation. Iodination was successfully used to produce prolactin releasing peptide prohormone (PRPG), containing N-terminal serine; however, while the lower temperature prevented the reaction between the hydroxyl group and the cyclic intermediate in the case of PRPG, the thiol on oxytocin could still capture the iodinated alkene, thus preventing the bond cleavage. Regardless, production of oxytocin *via* this approach was not successful.

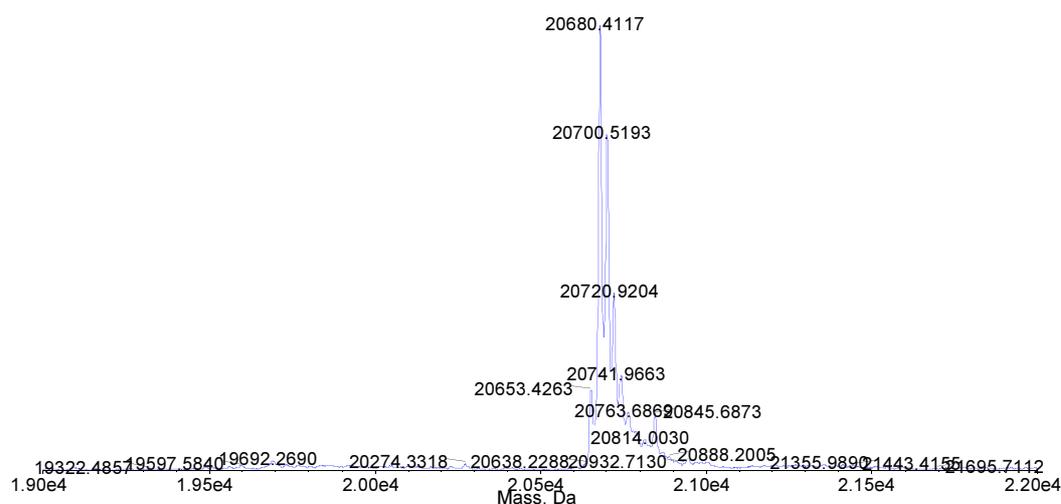


Figure 5.20. Deconvoluted ESI-MS spectrum of His₆PpiB_OXT fusion protein expressed with 2-acetamidohex-5-enoic acid **86** instead of (*S*)-methionine **28**. The 20680.4 Da peak corresponds to 8/8 replacements of methionine, 20700.5 Da – 7/8 replacements, 20720.9 Da – 6/8 replacements.

5.4. Conclusions

Two halogenated analogues of methionine were successfully synthesised and incorporated into proteins during cell-free protein expression. The ability of the halogens to form a hydrogen bond the active site of MRS, mimicking the thioether group of methionine, has enabled the recognition of the unnatural amino acids despite their shorter side-chains, as compared to methionine.

The chlorinated analogue leads to reliable fragmentation of proximal peptide bonds in response to heat exposure; the utility of the analogue was demonstrated through the production of peptide hormone oxytocin. The yield of the oxytocin was lower than the yields observed with other peptide hormones, most likely a consequence of a cysteine residue in the vicinity of the scissile bond.

The brominated analogue could not be effectively utilised for the production of small peptides through fragmentation of peptide bonds due to undergoing prominent side reactions. However, the facile incorporation of this amino acid into proteins could be exploited as a method of incorporating reactive functional groups into proteins.

The unsaturated analogue of methionine 2-aminohex-5-enoic acid was shown to lead to protein fragmentation when exposed to iodine, but the production of oxytocin was unsuccessful.

Chapter 6. Conclusions and Future Directions

The inability of the native protein synthetic machinery to discriminate against unnatural amino acids, due to lack of evolutionary pressure to evolve such ability, is extremely valuable in the production of functionalised proteins. Understanding the selection criteria of the synthetases enables more reliable development of potential substrates. With the demonstration that γ -chlorovaline **26a** is a substrate of IRS, which had previously been masked by instability of the analogue, it was concluded AARSs of the BCAA family are not able to distinguish between a chlorine atom and a methyl group, enabling facile replacement of largely unreactive aliphatic amino acids with electrophilic chlorinated analogues. The survey of the known substrates of MRS enabled the development of two new analogues of methionine **28**, the chloride **83a** and the bromide **83b**, showing that the halogens are able to mimic the bonding interactions of (thio)ethers in the active site of MRS.

The γ -chlorinated amino acids **25**, **26a** and **83a**, owing to the favourable formation of 5-membered rings, had a latent capacity, promoted by elevated temperatures, to facilitate the cleavage of proximal peptide bonds. This capacity was utilised in the preparation of peptides, expressed by a cell-free system as fusion proteins. Three peptide prohormones, each lacking leucine, isoleucine or methionine, were produced by utilising the available γ -chlorinated analogues **25**, **26a** and **83a**, respectively. The nature

of cell-free expression systems enabled simultaneous incorporation of two unnatural amino acids into the proteins, allowing preparation of modified peptides.

Proteins expressed with the β -chlorinated amino acid of the series **15a**, **15b** and **27** were not fragmented by temperature. The radical differences in reactivity between the β - and γ -chloroamino acids were exemplified by the preparation of chlorinated aprotinin; protein chains containing the chlorides **15a** and **15b** were separated from the carrier protein using the analogue of leucine **25**. Crystal structures of the chlorinated aprotinin revealed no rearrangement of the peptidic chain had occurred as a consequence of the substitution, thus indicating that the halogenated residues are not inherently detrimental to proteins and can be incorporated into functional proteins for subsequent modification.

Despite the dramatic differences in response to temperature, both β - and γ -chloroamino acids would be expected to react with nucleophiles. As genetically-encoded halogenated amino acids have been employed for protein cross-linking through cross-reaction with nucleophilic side-chains,^[137] the chlorinated isosteres of proteogenic amino acids described in this Thesis have the potential to be applied for the same purpose, avoiding the need for engineered systems.

This might be particularly true with brominated amino acids. Exploration of the reactivity of the γ -brominated residue **83b** showed that while the residue is able to promote peptide bond cleavage like the corresponding chlorinated residues, it appeared predisposed to react with external nucleophiles under more tame conditions, making the analogue more suitable for use for labelling of proteins rather than fragmentation. Similarly, brominated isosteres of the aliphatic amino acids would be expected to be more reactive than the chlorinated counterparts. However, as incorporation of the said

brominated amino acids has not yet been demonstrated, the full potential of these residues remains to be explored.

The potential applications of the halogenated analogues of the proteogenic amino acids are particularly pertinent in light of the interchangeability of leucine and isoleucine with no consequences to protein structure and function, as demonstrated through the production of aprotinin. While heat sensitive analogues of isoleucine, leucine and methionine are available, this exchangeability of hydrophobic amino acids extended the applicability of the method to preparation of targets naturally containing all three residues, and provided a means to control the site of incorporation of the residue through modification of the gene sequences.

The production of aprotinin also showcases the scalability of the system. The ability to produce large quantities of material is critical for practical applications, like protein crystallography. The potential to express modified proteins on a large scale can further facilitate the application of functionalised amino acids.

The developments in the use of halogenated amino acids can also simplify the use other functionalised amino acids, like 2-amino-3-methylpent-4-ynoic acid **77a**. As an analogue of isoleucine **7**, the amino acids substitutes a biologically inert saturated hydrocarbon chain with an alkynyl side-chain that is equally unreactive towards biological systems, with only some loss in flexibility of the chain. While alkyne analogues of methionine also exist (Table 1.3), the isoleucine analogue **77a** is less likely to perturb protein function, as the thioether group of methionine is sometimes involved in biological interactions.^[151] The alkyne side chain provides a more specific tag for protein modification, compared to the halogenated side-chains, and can be used for

protein labelling using bioorthogonal chemistry under mild biological conditions. Similar chemistry is already employed in protein studies, relying on engineered systems to genetically encode functionalised amino acids,^[93,94] and the analogue of isoleucine could simplify the experimental design using the strategy for strategic positioning of the residue in the protein.

Despite the capacity to control the substitution patterns by amino acid replacement, at times proteins containing all twenty amino acids may be desired. Thus, for a successful use of heat-induced proteolysis for site-specific bond cleavage a route for incorporation of a lactonisation-prone amino acid as the 21st amino acid might be a worthy pursuit in the future. The inherent instability of γ -chlorinated amino acids is likely to present additional challenges to those encountered in the expansion of genetic code, but the development of such a system enable the use of heat-controlled proteolysis without any sequence restrictions in the target peptide.

Chapter 7. Experimental

7.1. General experimental

NMR spectra were recorded on a Varian MR400 (400 MHz for ^1H , 376 MHz for ^{19}F and 100 MHz for ^{13}C), Varian Mercury 300 (300 MHz for ^1H and 75 MHz for ^{13}C), Varian Inova 500 (500 MHz for ^1H and 125 MHz for ^{13}C), Bruker AscendTM 400 (400 MHz for ^1H and 100 MHz for ^{13}C), Bruker Avance 600 (600 MHz for ^1H and 150 MHz for ^{13}C) and Bruker Avance 800 (800 MHz for ^1H and 200 MHz for ^{13}C) spectrometers. The residual solvent peak was used as reference. Chloroform-D (99.8% D), DMSO-D₆ (99.9% D) and acetone-D₆ (99.9% D) were purchased from Cambridge Isotope Laboratories Inc. and methanol-D₄ (99.8% D) and D₂O (99.8% D) were from Sigma-Aldrich Co.

Low resolution (LR) electron spray ionisation (ESI) mass spectrometry was carried out using a Micromass LC-ZMD single quadrupole liquid chromatograph mass spectrometer, electron impact (EI) mass spectrometry was conducted on Waters AutoSpec Premier spectrometer. High resolution mass spectrometry was carried out on a Waters LCT PremierTM XE orthogonal acceleration time-of-flight (oa-TOF) mass spectrometer. Mass spectrometry of short peptides was conducted on an Agilent 1100 series LC/MSD TOF instrument (direct injection) or Waters Acquity spectrometer with TQ detector. Mass spectra of large proteins were collected on an Agilent 1100 series LC/MSD TOF instrument (direct injection).

Preparative HPLC was performed on a Waters 600 controller with Waters 717plus autosampler and a Waters 2996 photodiode array detector, running Empower 2 software. Analytical HPLC was performed on a Waters Alliance Separation module 2695 with a Waters 2996 photodiode array detector, running Empower 2 software, and an Agilent Hewlett-Packard series 1100 separation system with Chemstation.

Silica chromatography was performed with Silica gel 60 (0.040-0.063 mm) from Merck. Thin layer chromatography was performed on TLC plates Silica gel 60 F₂₅₄ from Merck, visualising with irradiation under UV lamp and potassium permanganate, vanillin or ninhydrin dip.

Melting point analyses were carried out using OptiMelt Automated Melting Point System from Stanford Research Systems, operated with Meltview software. All melting points are uncorrected.

Optical rotation was measured on a PerkinElmer Model 343 Polarimeter at 589 nm using a 1 dm pathlength cuvette.

All samples were freeze-dried using FreeZone 4.5 freeze-drier from Labconco.

Chemicals for synthesis were purchased from Sigma-Aldrich Co., TCI, Alfa-Aesar and Merck. Inorganic chemicals were from AJAX or Scharlau. Solvents for synthesis were obtained from Merck, VWR International and Sigma-Aldrich Co. Solvents for HPLC were obtained from Sigma-Aldrich Co. or Scharlau. Water was purified with a MilliQ-Reagent water system.

DNA primers for construction of the fusion were ordered from IDT[®]. Phusion[®] DNA polymerase, restriction enzymes and T7 quick DNA Ligation[™] kit were obtained from

New England BioLabs[®] Inc. Alkaline phosphatase was from Roche Diagnostics GmbH. Qiagen[®] Gel extraction kit was used to recover DNA after agarose gel electrophoresis. For protein expression DNA plasmid was amplified in *E.coli* AN1459 cells and isolated using Qiagen[®] Maxi or Mega kits.

7.2. Cell-free protein expression

Cell-free protein expression was carried out as described before.^[16,102,123] Spectra/Por[®] dialysis membrane (#2, MWCO: 12-14 kDa) was purchased from Spectrum Laboratories Inc. T7 RNA polymerase was obtained from New England BioLabs[®] Inc. The reactions were carried out for 14 h at 30°C.

The proteogenic amino acids were supplied in 1 mM concentration. Non-proteogenic amino acids were used in 2 mM concentration and the corresponding natural amino acid was withheld. All initial incorporation experiments were carried out on His₆PpiB.

Sequence:

```
MHHHHHMVT FHTNHGDIVI KTFDDKAPET VKNFLDYCRE GFYNNTIFHR  
VINGFMIQGG GFEPGMKQKA TKEPIKNEAN NGLKNTRGTL AMARTQAPHS  
ATAQFFINVV DNDFLNFSGE SLQGWGYCVF AEVVDGMDVV DKIKGVATGR  
SGMHQDVPKE DVIIESVTVS EN
```

The contents of the inner reaction were then purified using HisTrap[™] HP columns from GE Healthcare using conditions recommended by the manufacturer. Amicon Ultra-4 (YM-3,000) centrifugal filter devices from Millipore were used for sample concentration and buffer removal. Protein concentrations were measured on NanoDrop ND-1000 spectrophotometer from ThermoScientific. Gels (20% acrylamide) for SDS-

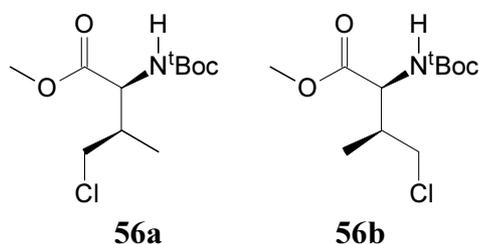
PAGE analysis were cast using acrylamide and bis-acrylamide solutions from Bio-RAD and run on Mini-PROTEIN[®] Tetra system. Bio-Safe[®] Coomassie Blue was used for gel staining. Protein markers were purchased from New England Biolabs[®] Inc. Mass spectra of the proteins were collected on an Agilent 1100 series LC/MSD TOF instrument (direct injection).

7.3. Synthesis of compounds from Chapter 2.

Chlorination of (*S*)-valine and derivatives **8a-c**

(*S*)-valine **8a**, *N*-acetyl-(*S*)-valine **8b** or (*S*)-valine methyl ester **8c** were dissolved in TFA in a quartz flask in batches of 500 mg, and chlorine gas was bubbled through the solution under irradiation by a 300 W sunlamp.^[119] The solvent was then evaporated *in vacuo* to dryness, and the residue was used for analysis or subsequent experiments as described below or in the main text.

Methyl 2-[(*tert*-butoxycarbonyl)amino]-4-chloro-3-methylbutanoates **56a,b**

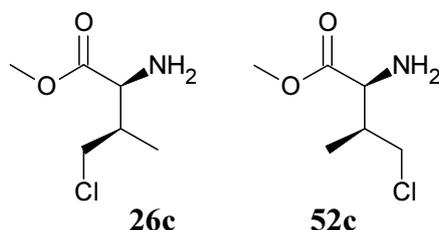


The dried residue of (*S*)-valine methyl ester **8c** (4 g, 24 mmol) chlorination mixture was dissolved in dry dichloromethane and treated with triethylamine until pH \geq 7. Di-*tert*-butyl dicarbonate (4.6 g, 21 mmol) was then added and the mixture was stirred

at room temperature for 16 h. All volatiles were then removed *in vacuo* and a portion of the material was subjected to purification by HPLC (Alltima C18 5 μ 250 \times 22 mm column, 10 mL min⁻¹, 60% methanol). After lyophilisation of the collected fractions 2-[(*tert*-butoxycarbonyl)amino]-4-chloro-3-methylbutanoate **56a,b** (155 mg) was isolated as a mixture of (2*S*,3*S*) and (2*S*,3*R*) diastereomers as a colourless oil. The ratio of the diastereomers in the mixture varied slightly from batch to batch, but was on average 1:1. NMR peaks were assigned to the individual diastereomers with the help of COSY, ¹³C-HSQC and ¹³C-HMBC.

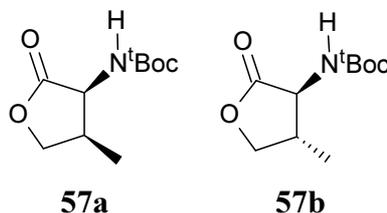
t_R = 29.0 min; **diastereomer 1**: ¹H NMR (400 MHz, CDCl₃): δ = 5.21 (d, J = 7 Hz, 1 H; *NH*), 4.40 (dd, J = 8, 5 Hz, 1 H; *CHN*), 3.74/3.75 (s, 3 H; *OCH*₃), 3.59-3.53 (m, 1 H; *CHHCl*), 3.42 (dd, J = 11, 7 Hz, 1H; *CHHCl*), 2.45-2.29 (m, 1 H; *CHCH*₃), 1.43 (s, 9 H; *C(CH*₃)₃), 1.03 ppm (d, J = 7 Hz, 3 H; *CHCH*₃); ¹³C NMR (100 MHz, CDCl₃): δ = 172.1/171.8, 155.6/155.3, 80.1, 55.9, 52.5/52.3, 46.6, 39.1, 28.2, 14.6 ppm; **diastereomer 2**: ¹H NMR (400 MHz, CDCl₃): δ = 5.09 (d, J = 7 Hz, 1 H; *NH*), 4.58 (dd, J = 9, 3 Hz, 1 H; *CHN*), 3.74/3.75 (s, 3 H; *OCH*₃), 3.59-3.53 (m, 1 H; *CHHCl*), 3.33 (dd, J = 11, 7 Hz, 1 H; *CHHCl*), 2.45-2.29 (m, 1 H; *CHCH*₃), 1.43 (s, 9 H; *C(CH*₃)₃), 0.94 ppm (d, J = 7 Hz, 3H; *CHCH*₃); ¹³C NMR (100 MHz, CDCl₃): δ = 172.1/171.8, 155.6/155.3, 80.1, 55.0, 52.5/52.3, 47.0, 39.1, 28.2, 12.8 ppm; LRMS (ESI, +ve): m/z : 288.5, 290.5 [*M*⁺+Na]; HRMS (ESI, +ve): m/z : calcd for [C₁₁H₂₀NO₄Cl+Na]⁺ 288.0979 and 290.0949, found 288.0979 and 290.0941.

Methyl 2-amino-4-chloro-3-methylbutanoates 26c and 52c



The mixture of the chlorides **56a,b** was treated with trifluoroacetic acid (5 mL) and the solution was stored for 1 h at room temperature. All volatiles were then removed *in vacuo* to give methyl 2-amino-4-chloro-3-methylbutanoate as a mixture of (2*S*,3*R*)-**26c** and (2*S*,3*S*)-**52c** diastereomers as an amorphous colourless solid.

m.p. 75.5-77.5°C; (**2*S*,3*R***)-**26c**: ^1H NMR (400 MHz, D_3COD): δ = 4.29 (d, J = 4 Hz, 1 H; *CHN*), 3.86 (s, 3 H; *OCH*₃), 3.73 (dd, J = 11, 8 Hz, 1 H; *CHHCl*), 3.66 (obscured dd, J = 11, 6 Hz, 1 H; *CHHCl*), 2.43 (dddq, J = 8, 7, 6, 4 Hz, 1 H; *CHCH*₃), 1.10 ppm (d, J = 7 Hz, 3 H; *CCH*₃); ^{13}C NMR (100 MHz, D_3COD): δ =169.6, 55.7, 54.0/53.8, 47.0, 39.3, 13.4 ppm; (**2*S*,3*S***)-**52c**: ^1H NMR (400 MHz, D_3COD): δ = 4.26 (d, J = 4 Hz, 1 H; *CHN*), 3.87 (s, 3 H; *OCH*₃), 3.69-3.66 (obscured, 2 H; *CH*₂Cl), 2.59 (dp, J = 7, 4 Hz, 1 H; *CHCH*₃), 1.13 ppm (d, J = 7 Hz, 3 H; *CCH*₃); ^{13}C NMR (100 MHz, D_3COD): δ =170.2, 55.8, 54.0/53.8, 46.5, 38.5, 13.6 ppm; LRMS (ESI, +ve): m/z : 166.1, 168.1 [$\text{M}^+\text{+H}$]; HRMS (ESI, +ve): m/z : calcd for [$\text{C}_{11}\text{H}_{20}\text{NO}_4\text{Cl+H}$]⁺ 166.0635 and 168.0605, found 166.0637 and 168.0610.

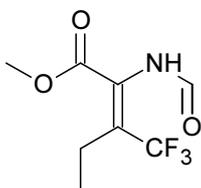
tert*-Butyl (4-methyl-2-oxotetrahydrofuran-3-yl)-carbamates **57a** and **57b*

The dried residue of (*S*)-valine **8a** (5 g, 42.7 mmol) chlorination mixture was dissolved in water and Na₂CO_{3(sat.)} was added and until pH ≥ 7. The solution was stirred at room temperature for 4 h. Di-*tert*-butyl dicarbonate (9.3 g, 42.7 mmol) and β-cyclodextrin (2.4 g, 2.1 mmol) were then added to the mixture, and the solution was stirred for 16 h a room temperature. The solution was then extracted with ethyl acetate, the organic phase was dried (MgSO₄) and evaporated *in vacuo*. A portion of the residue was purified by HPLC (Alltima C18 5μ 250×22 mm column, 12 mL min⁻¹, 40% methanol) to give the title compounds **57a** (120 mg) and **57b** (170 mg) as white crystalline solids. Spectroscopic data are consistent with literature reports.^[125]

(*2S,3R*)-**57a**: $t_R = 30.4$ min; m.p. 123-215°C; $[\alpha]_D^{20} = -54.2$ (c = 0.02 in methanol); ¹H NMR (300 MHz, CDCl₃): δ = 5.12 (d, br, $J = 8$ Hz, 1 H; NH), 4.38 (t, $J = 9$ Hz, 1 H; OCHH), 4.10-3.97 (m, 1 H; NCH), 3.79 (t, $J = 10$ Hz, 1 H, OCHH), 2.58-2.40 (m, 1 H; NCCH), 1.43 (s, 9 H; C(CH₃)₃), 1.19 ppm (d, $J = 6$ Hz, 3 H; CCCH₃).

(*2S,3S*)-**57b**: $t_R = 28.0$ min; m.p. 110.5-112°C; $[\alpha]_D^{20} = +41.6$ (c = 0.02 in methanol); ¹H NMR (300 MHz, CDCl₃): δ = 5.12 (s, br, 1 H; NH), 4.48 (t, $J = 6$ Hz, 1 H; NCH), 4.37 (dd, $J = 9, 5$ Hz, 1 H; OCHH), 4.06 (t, $J = 9$ Hz, 1 H, OCHH), 2.98-2.84 (m, 1 H; NCCH), 1.42 (s, 9 H; C(CH₃)₃), 0.97 ppm (d, $J = 6$ Hz, 3 H; CCCH₃).

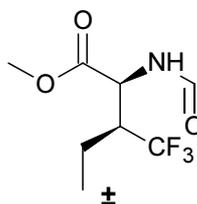
Methyl (Z)-2-formamido-3-(trifluoromethyl)pent-2-enoate 61



Sodium *tert*-butoxide (1.15 g, 12 mmol) was dissolved in dry tetrahydrofuran and the solution was cooled to -40°C . Technical grade methyl isocyanoacetate **58** (1 mL, 11 mmol) was then added dropwise and the solution was stirred for 30 min under nitrogen atmosphere. 1,1,1-Trifluorobutanone **59** (1.9 mL, 14 mmol) in dry tetrahydrofuran was then added, maintaining the temperature of the reaction at -40°C , and the mixture was stirred for 1 h. The mixture was then allowed to warm to room temperature and 1 M $\text{HCl}_{(\text{aq.})}$ was added until $\text{pH} = 4$, and the mixture was stirred for an additional 1 h. The volatiles were then removed *in vacuo*, the residue was dissolved in water and extracted with ethyl acetate. The organic layer was dried (MgSO_4) and evaporated. The residue was purified by flash silica chromatography, eluting with petroleum ether / ethyl acetate = 3:1, to give the target compound as a yellowish oil that solidified upon standing (610 mg, 2.7 mmol, 25%). Spectroscopic data are consistent with literature reports.^[126]

$R_f = 0.4$; $^1\text{H NMR}$ (300 MHz, $(\text{CD}_3)_2\text{CO}$): $\delta = 9.28$ (s, br, 1 H; NH), 8.21 (s, 1 H; NCH), 3.79 (s, 3 H; OCH_3), 2.30 (q, $J = 8$ Hz, 2 H; CH_2), 1.11 ppm (t, $J = 8$ Hz, 3 H; CCH_3).

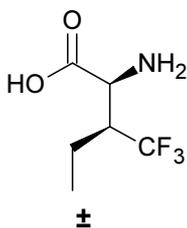
Methyl 2-formamido-3-(trifluoromethyl)pentanoate



Methyl (*Z*)-2-formamido-3-(trifluoromethyl)pent-2-enoate **61** (462 mg, 2.05 mmol) was dissolved in methanol and 10% Pd/C was added to the solution. The mixture was stirred for 16 h at room temperature under hydrogen atmosphere. The catalyst was then filtered off with celite, the solvent was evaporated *in vacuo* and the residue was subjected to flash silica chromatography, eluting with petroleum ether / ethyl acetate = 1:1. The target compound was isolated as a white amorphous solid (365 mg, 1.61 mmol, 78%). Spectroscopic data are consistent with literature reports.^[126]

$R_f = 0.5$; $^1\text{H NMR}$ (300 MHz, CDCl_3): $\delta = 8.24$ (s, 1 H; NCH), 6.36 (s, br, 1 H; NH), 5.11 (dd, $J = 6, 3$ Hz, 1 H; O_2CCH), 3.81 (s, 3 H; OCH_3), 2.75-2.58 (m, 1 H; F_3CCH), 1.89-1.61 (m, 2 H; CH_2), 1.11 ppm (t, $J = 8$ Hz, 3 H; CCH_3).

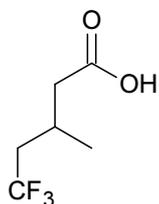
2-Amino-3-(trifluoromethyl)pentanoic acid **24a**



Methyl 2-formamido-3-(trifluoromethyl)pentanoate (353 mg, 1.56 mmol) was dissolved in acetone (2 mL) and 6 M HCl_(aq.) (20 mL) was added. The solution was stirred with reflux for 16 h. The mixture was then cooled and all volatiles were removed *in vacuo* to give the amino acid as a white solid (280 mg, 1.51 mmol, 97%). Spectroscopic data are consistent with literature reports.^[126]

m.p. 175-178°C; ¹H NMR (300 MHz, D₃COD): δ = 4.27 (br, 1 H; NCH), 3.12-2.93 (m, 1 H; F₃CCH), 1.94-1.60 (m, 2 H; CH₂), 1.11 ppm (t, *J* = 8 Hz, 3 H; CCH₃); ¹⁹F NMR (376 MHz, D₃COD): δ = 71.07 ppm (d, *J*_{HF} = 10 Hz, 3 F).

5,5,5-Trifluoro-3-methylpentanoic acid **67**



Method 1: Cyanoacetic acid **63** (50 g, 0.59 mol), acetone **64** (35 g, 0.6 mol) and ammonium acetate (4 g, 52 mmol) were dissolved in benzene (100 mL) and the solution was stirred with reflux in a Dean-Stark apparatus until water condensation ceased (*ca.*

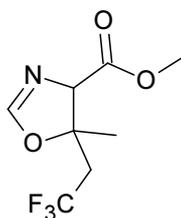
72 h). The Dean-Stark apparatus was then replaced with distillation equipment and the solution was distilled at atmospheric pressure. After benzene was removed, the fraction containing 3-methylbut-3-enenitrile **65** (80 - 115°C) and the isomerised by-product 3-methylbut-2-enenitrile was collected as a clear liquid (29.7 g) and used without further purification.

Impure 3-methylbut-3-enenitrile **65** (29.7 g, 0.35 mol), TFA (47.3 mL, 0.62 mol) and sodium acetate (1.37 g, 16.7 mmol) were dissolved in methanol (200 mL) and water (30 mL) was added. The solution was electrolysed with platinum wires with stirring until 62 kC were passed through the solution (ca. 44 h). The mixture was then diluted with water (700 mL) and extracted with dichloromethane. Dichloromethane was distilled off at atmospheric pressure and the remaining solution was vacuum-distilled. 10% Pd/C was added to the collected volatiles and the mixture was stirred for 16 h at room temperature under hydrogen atmosphere. The catalyst was then filtered off with celite and the flow-through was distilled through a Vigreux column. The fraction collected between 165°C and 170°C containing 5,5,5-trifluoro-3-methylpentannitrile **66** was mixed with 6 M HCl_(aq.) (70 mL) and the biphasic mixture was stirred with reflux overnight. The phases were then separated and the organic layer was mixed with 10% (w/w) NaOH_(aq.). The two layers were then separated and the aqueous layer was acidified with 6 M HCl_(aq.). A new layer of organic material separated from the aqueous solution that was collected and redistilled to give the title compound as a colourless liquid (2.08 g, 12.2 mmol, 2%). Spectroscopic data are consistent with literature reports.^[129]

Method 2: Benzyl(triphenylphosphoranylidene)acetate **69** (40 g, 97 mmol) and 4,4,4-trifluorobutanone (4.4 g, 35 mmol) were dissolved in dichloromethane (200 mL) and the solution was stirred with reflux for 16 h. The mixture was then cooled and all volatiles were evaporated *in vacuo*. The residue was purified by flash silica chromatography, eluting with petroleum ether / ethyl acetate = 19:1, to give the products of the Wittig reaction as a mixture of *cis* and *trans* alkenes. The alkenes were then dissolved in ethanol, 10% Pd/C was added and the mixture was incubated under hydrogen atmosphere (50 psi) with shaking for 3 h. The catalyst was then filtered off through celite and the flow-through was distilled to give the title compound as a colourless liquid (1.1 g, 6.6 mmol, 19%).

^1H NMR (300 MHz, CDCl_3): δ = 10.94 (s, br, 1 H; COOH), 2.54-1.96 (m, 5 H), 1.13 ppm (d, J = 6 Hz, 3 H; CH_3); ^{19}F NMR (376 MHz, CDCl_3): δ = 63.56 ppm (t, J_{HF} = 11 Hz, 3 F).

Methyl 5-methyl-5-(2,2,2-trifluoroethyl)-1,3-oxazolidine-4-carboxylate 72

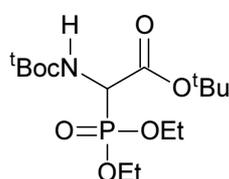


Methyl isocyanoacetate **58** (0.2 mL, 2.2 mmol) and 4,4,4-trifluorobutanone **68** (0.3 mL, 2.64 mmol) were dissolved in dry ether (5 mL) and Cu(I) oxide (16 mg, 0.1 mmol) was added. The mixture was stirred under nitrogen atmosphere for 16 h. The mixture was then subjected to flash silica chromatography, eluting with petroleum ether / ethyl

acetate = 1:1, to give the title compound, a mixture of two diastereomers with a ratio of *ca.* 3:7, as a colourless oil (397 mg, 1.8 mmol, 80%). The compound was highly moisture sensitive; analytical purity was not attainable.

$R_f = 0.45$; *major diastereomer*: ^1H NMR (400 MHz, CDCl_3): $\delta = 6.91$ (obs, 1 H; N=CH), 4.56 (d, $J = 2$ Hz, 1 H; O_2CCH), 3.74 (s, 3 H; OCH_3), 2.67-2.53 (m, 2 H; CH_2), 1.36 ppm (s, 3 H; CCH_3); ^{13}C NMR (100 MHz, CDCl_3): $\delta = 169.0, 155.8, 124.6$ (q, $J_{\text{CF}} = 278$ Hz; CF_3), 82.8, 75.1, 52.3, 43.5 (q, $J_{\text{CF}} = 28$ Hz; CCF_3), 20.5 ppm (q, $J_{\text{CF}} = 2$ Hz); ^{19}F NMR (376 MHz, CDCl_3): $\delta = 61.00$ ppm (t, $J_{\text{HF}} = 10$ Hz, 3 F); *minor diastereomer*: ^1H NMR (400 MHz, CDCl_3): $\delta = 6.90$ (obs, 1 H; N=CH), 4.42 (d, $J = 2$ Hz, 1 H; O_2CCH), 3.74 (s, 3 H; OCH_3), 2.51-2.41 (m, 2 H; CH_2), 1.59 ppm (s, 3 H; CCH_3); ^{13}C NMR (100 MHz, CDCl_3): $\delta = 169.2, 155.9, 125.0$ (q, $J_{\text{CF}} = 278$ Hz; CF_3), 82.7, 77.2, 52.4, 37.5 (q, $J_{\text{CF}} = 29$ Hz; CCF_3), 25.6 ppm (q, $J_{\text{CF}} = 2$ Hz); ^{19}F NMR (376 MHz, CDCl_3): $\delta = 60.54$ ppm (t, $J_{\text{HF}} = 11$ Hz, 3 F); LRMS (ESI, +ve): m/z : 226.2 [$\text{M}^+\text{+H}$]; HRMS (ESI, +ve): m/z : calcd for $[\text{C}_8\text{H}_{10}\text{NO}_3\text{F}_3\text{+H}]^+$ 226.0691, found 226.0691.

***tert*-Butyl (diethoxyphosphoryl)[(*tert*-butoxycarbonyl)amino]acetate 73**

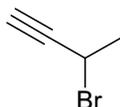


tert-Butyl (*tert*-butoxycarbonyl)glycinate (2 g, 8.7 mmol) was dissolved in dry carbon tetrachloride, *N*-bromosuccinimide (1.54 g, 8.7 mmol) was added and the mixture was stirred for 4 h with irradiation under a 300W sunlamp, until the starting material was

consumed. The solution was then filtered and the filtrate was evaporated *in vacuo*. The resultant yellow oil was redissolved in dichloromethane and triethyl phosphite (1.7 mL, 9.6 mmol) was added. The solution was stirred under nitrogen atmosphere for 16 h. The solvent was then evaporated *in vacuo* and the residue was purified by flash silica chromatography, eluting with petroleum ether / ethyl acetate = 1:1, to give the title compound as a colourless oil (2.86 g, 7.8 mmol, 90%). Spectroscopic data are consistent with literature reports.^[152]

$R_f = 0.38$; $^1\text{H NMR}$ (300 MHz, CDCl_3): $\delta = 5.32$ (d, br, $J = 8$ Hz, 1 H; NH), 4.70 (dd, $J_{\text{HP}} = 22$ Hz, $J_{\text{HH}} = 9$ Hz, 1 H; NCH), 4.26-4.05 (m, 4 H; CH_2), 1.48 (s, 9 H; $\text{C}(\text{CH}_3)_3$), 1.43 (s, 9 H; $\text{C}(\text{CH}_3)_3$), 1.32 ppm (t, $J = 7$ Hz, 6 H; POCCH_3).

3-Bromobut-1-yne **79**

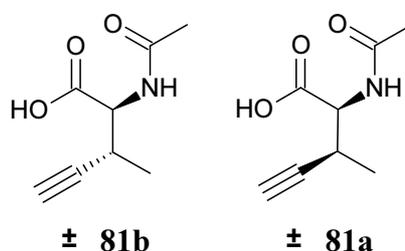


Triphenyl phosphite (23 mL, 87 mmol) was dissolved in dry ether (50 mL) in a glove box and bromine (4.4 mL, 86 mmol) was added dropwise. The precipitate was repeatedly washed with ether and decanted in the glove box, until the washings were colourless. After the final wash the powder was dried at 50°C under vacuum for 3 h. Premixed but-3-yn-2-ol **78** (3 mL, 38 mmol) and pyridine (3.8 mL, 48 mmol) were added dropwise to the solid with stirring at 0°C and the slurry was stirred at room temperature under nitrogen atmosphere for 2 h. The mixture was then vacuum-distilled, collecting the volatiles at -196°C. The yellow liquid was redistilled at atmospheric

pressure to give the title compound as an opaque colourless liquid (3.3 g, 25 mmol, 66%). Spectroscopic data are consistent with literature reports.^[153]

¹H NMR (300 MHz, CDCl₃): δ = 4.58 (dq, *J* = 7, 2 Hz, 1 H; BrCH), 2.65 (d, *J* = 2 Hz, 1 H; C≡CH), 1.92 ppm (d, *J* = 7 Hz, 3 H; CH₃).

2-Acetamido-3-methylpent-4-ynoic acids **81a** and **81b**



3-Bromobut-1-yne **79** (1 g, 7.5 mmol), diethyl acetamidomalonate **74** (1.78 g, 8.2 mmol) and caesium carbonate Cs₂CO₃ (2.7 g, 8.25 mmol) were added to dry acetonitrile (4 mL) and the mixture was subjected to microwave irradiation (100°C, 300 W) for 1 h with stirring. The reaction mixture was then filtered to remove inorganic salts and the filtrate was evaporated to dryness.

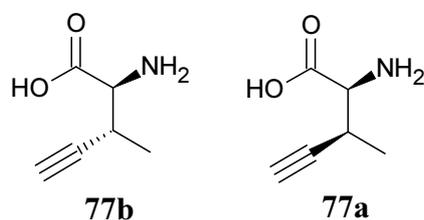
The residual oil was suspended in 10% (w/w) Na₂CO_{3(aq.)} solution and the suspension was stirred with reflux for 16 h. The reaction mixture was then cooled down and extracted with chloroform. The aqueous layer was then acidified with HCl_(aq.) until pH = 3. All volatiles were then evaporated *in vacuo* and the crystalline residue was extracted with hot (55°C) acetone. The solvent was then evaporated *in vacuo* and the residue was purified by reverse phase HPLC, eluting with 15% methanol in 0.1% TFA, 10 mL/min. The minor (2*S*,3*R*)-(±) diastereomer **81b** eluted at 27 min (108 mg, 0.64 mmol, 8.5%),

the major (2*S*,3*S*)-(±) diastereomer **81a** was collected at 31 min (305 mg, 1.8 mmol, 24%). Spectroscopic data are consistent with literature reports.^[133]

(2*S*,3*S*)-(±)-**81a**: ¹H NMR (300 MHz, D₃COD): δ = 4.55 (d, *J* = 5 Hz, 1 H; NCH), 3.22-3.10 (m, 1 H; C≡CCH), 2.49 (d, *J* = 2 Hz, 1 H; C≡CH), 2.05 (s, 3 H; NCCH₃), 1.22 ppm (d, *J* = 7 Hz, 3 H; CCCH₃).

(2*S*,3*R*)-(±)-**81b**: ¹H NMR (300 MHz, D₃COD): δ = 4.57 (d, *J* = 6 Hz, 1 H; NCH), 3.08-2.96 (m, 1 H; C≡CCH), 2.46 (d, *J* = 2 Hz, 1 H; C≡CH), 2.02 (s, 3 H; NCCH₃), 1.24 ppm (d, *J* = 7 Hz, 3 H; CCCH₃).

(2*S*,3*S*)- And (2*S*,3*R*)-2-amino-3-methylpent-4-ynoic acids **77a** and **77b**



(2*S*,3*S*)-(±)-2-Acetamido-3-methylpent-4-ynoic acid **81a** (305 mg, 1.8 mmol) was dissolved in water and 0.1 M NaOH_(aq) was added until pH ≈ 7.5. Several milligrams of acylase were added to the solution and the mixture was incubated at 30°C for 48 h. The solution was then filtered through a YM-10 filter and the flow-through was subjected to ion-exchange chromatography, eluting with 10% (w/w) NH_{3(aq)}. The solvent was removed *in vacuo* to give (2*S*,3*S*)-2-amino-3-methylpent-4-ynoic acid **77a** (35.6 mg, 0.28 mmol, 31%) as a colourless solid. (2*S*,3*R*)-(±)-2-Acetamido-3-methylpent-4-ynoic acid **81b** (108 mg, 0.64 mmol) was treated the same way to give (2*S*,3*R*)-2-amino-3-

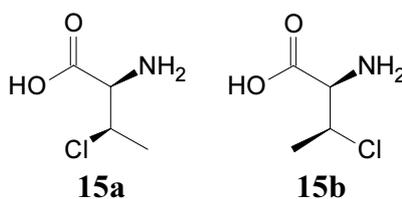
methylpent-4-ynoic acid **77b** (10.0 mg, 0.08 mmol, 25%). Spectroscopic data are consistent with literature reports.^[133]

(2*S*,3*S*)-**77a**: m.p. 198°C (dec.); $[\alpha]_{\text{D}}^{20} = -31.2$ ($c = 0.02$); $^1\text{H NMR}$ (300 MHz, D_3COD): $\delta = 3.42$ (d, $J = 5$ Hz, 1 H; NCH), 3.27-3.16 (m, obs, 1 H; $\text{C}\equiv\text{CCH}$), 2.64 (d, $J = 2$ Hz, 1 H; $\text{C}\equiv\text{CH}$), 1.35 ppm (d, $J = 7$ Hz, 3 H; CH_3).

(2*S*,3*R*)-**77b**: m.p. 205°C (dec.); $[\alpha]_{\text{D}}^{20} = -35.5$ ($c = 0.02$); $^1\text{H NMR}$ (300 MHz, D_3COD): $\delta = 3.68$ (d, $J = 4$ Hz, 1 H; NCH), 3.29-3.22 (m, obs, 1 H; $\text{C}\equiv\text{CCH}$), 2.66 (d, $J = 2$ Hz, 1 H; $\text{C}\equiv\text{CH}$), 1.23 ppm (d, $J = 7$ Hz, 3 H; CH_3).

7.4. Additional experimental data for Chapter 3.

(2*R*,3*R*)- And (2*R*,3*S*)-2-amino-3-chlorobutyric acids **15a** and **15b**



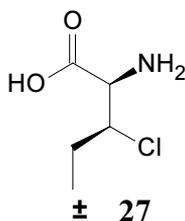
(2*S*,3*R*)-Threonine (1 g, 8.4 mmol) was suspended in HCl-saturated dioxane (5 mL) and thionyl chloride (0.8 mL, 11 mmol, 1.3 eq.) was added.^[154] The mixture was heated to 50°C in a sealed tube with stirring for 12 h. All volatiles were then evaporated *in vacuo* and the residue was redissolved in minimal amount of dry ethanol. Propylene oxide was added dropwise to the solution until precipitation stopped. The white precipitate containing the target compound with small amounts of unreacted starting material (80% conversion) was collected by centrifugation. A portion of the precipitate was subjected

to purification by HPLC (Alltima C18 5 μ 250 \times 22 mm column, 10 mL min⁻¹, 60% methanol in 0.05% HCl_(aq.)). The collected fractions were evaporated, redissolved in dry ethanol and reprecipitated with propylene oxide to give (2*R*,3*S*)-2-amino-3-chlorobutyric acids **15b** a white powder (220 mg). The (2*R*,3*R*)-diastereomer **15a** was prepared from (2*S*,3*S*)-*allo*-threonine using the same procedure. The ¹H NMR spectra are consistent with published data.^[155]

(2*R*,3*R*)-15a: m.p. 122-126°C; [α]_D²⁰ = -26.2 (c = 0.02); ¹H NMR (300 MHz, D₃COD): δ = 4.68 (dq, *J* = 7, 4 Hz, 1 H; CHCl), 3.74 (d, *J* = 5 Hz, 1 H; NCH), 1.66 ppm (d, *J* = 7 Hz, 3 H; CH₃).

(2*R*,3*S*)-15b: m.p. 115°C (dec.); [α]_D²⁰ = 13.6 (c = 0.01); ¹H NMR (300 MHz, D₃COD): δ = 4.71 (dq, *J* = 7, 4 Hz, 1 H; CHCl), 3.88 (d, *J* = 3 Hz, 1 H; NCH), 1.56 ppm (d, *J* = 7 Hz, 3 H; CH₃).

(2*R*,3*S*)-(\pm)-2-Amino-3-Chloropentanoic acid 27



Methyl isocyanoacetate **58** (1 mL, 11 mmol) and propanal (1 mL, 13.2 mmol) were dissolved in dry ether and copper(I) oxide (80 mg, 0.55 mmol, 0.05 eq.) was added.^[130] The suspension was stirred for 16 h at room temperature under nitrogen atmosphere. All volatiles were then evaporated *in vacuo* and the residue was subjected to flash silica

chromatography eluting with light petroleum and ethyl acetate (1:1). Methyl 5-ethyl-4,5-dihydro-1,3-oxazole-4-carboxylate was collected as a colourless oil (860 mg, 5.5 mmol, 50%), exclusively in *anti*-configuration. The material was unstable upon storage (analytical purity was not attainable).

$R_f = 0.4$; $^1\text{H NMR}$ (300 MHz, CDCl_3): $\delta = 6.94$ (s, 1 H, $\text{N}=\text{CH}$), 4.64 (q, $J = 7$ Hz, 1 H; NCHCH), 4.32 (dd, $J = 7$, 1 Hz, 1 H; $\text{NCHC}=\text{O}$), 3.80 (s, 3 H; OCH_3), 1.74 (p, $J = 7$ Hz, 2 H; CH_2), 1.01 ppm (t, $J = 7$ Hz, 3 H; CCH_3); $^{13}\text{C NMR}$ (75 MHz, CDCl_3): $\delta = 171.4$, 156.8, 82.7, 71.7, 52.7, 28.0, 8.9 ppm; LRMS (ESI, +ve): m/z : 158.2 [$\text{M}^+ + \text{H}$]; HRMS (ESI, +ve): m/z : calcd for $[\text{C}_7\text{H}_{11}\text{NO}_3 + \text{H}]^+$ 158.0817, found 158.0814.

Methyl 5-ethyl-4,5-dihydro-1,3-oxazole-4-carboxylate (860 mg, 5.48 mmol) was mixed with 4 M $\text{HCl}_{(\text{aq})}$ and the solution was stirred with reflux for 16 h.^[156] The mixture was then cooled and evaporated to dryness *in vacuo*. The residue was dissolved in a minimal amount of dry ethanol and propylene oxide was added dropwise until precipitation occurred. The precipitate was collected by centrifugation and washed with ethanol to give (2*S*,3*R*)-(\pm)-3-hydroxynorvaline **14** as a white porous powder (690 mg, 5.19 mmol, 95%). The $^1\text{H NMR}$ spectrum is consistent with published data.

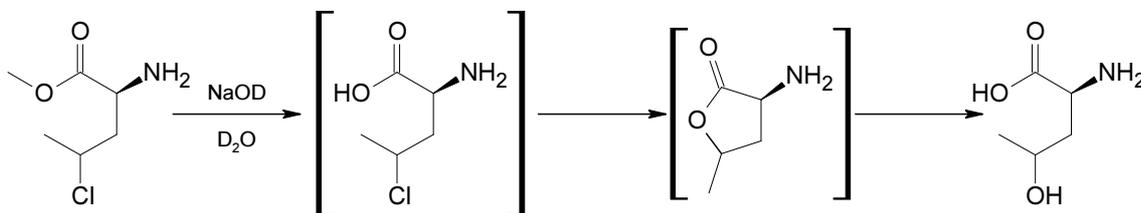
14: m.p. 175-185°C (dec.; lit.^[156] 214-215°C); $^1\text{H NMR}$ (300 MHz, D_3COD): $\delta = 3.95$ -3.86 (m, 1 H; HOCH), 3.41 (d, $J = 4$ Hz, 1 H; NCH), 1.74-1.45 (m, 2 H; CH_2), 1.01 ppm (t, $J = 7$ Hz, 3 H; CH_3).

(2*S*,3*R*)-(\pm)-3-Hydroxynorvaline **14** (400 mg, 3.5 mmol) was converted to (2*R*,3*S*)-(\pm)-3-chloronorvaline **27** following the procedure as described for the chlorides **15a** and **15b**, except HPLC purification was not required. After precipitation with propylene

oxide, the collected precipitate was washed with water to remove any unreacted starting material (the chloride **27** was strikingly insoluble in water, methanol or DMSO) to give (2*R*,3*S*)-(±)-3-chloronorvaline **19** as an off-white powder (420 mg, 2.8 mmol, 78%).

27: m.p. 130.5-135°C; ¹H NMR (300 MHz, D₃COD+TFA): δ = 4.40 (d, *J* = 3 Hz, 1 H; NCH), 4.29 (dt, *J* = 7, 3 Hz, 1 H; CHCl), 1.98 (p, *J* = 7 Hz, 2 H; CH₂), 1.11 ppm (t, *J* = 7 Hz, 3 H; CH₃); ¹³C NMR (75 MHz, D₃COD+TFA): δ=168.4, 63.0, 59.0, 29.2, 12.0 ppm; LRMS (ESI, +ve): *m/z*: 152.0, 154.1 [M⁺+H]; HRMS (ESI, +ve): *m/z*: calcd for [C₅H₁₀NO₂Cl+H]⁺ 152.0478 and 154.0449, found 152.0475 and 154.0437.

Establishment of the stereochemistry at position C-4 of methyl (2*S*,4*S*) and (2*S*,4*R*)-2-amino-4-chloropentanoates



Methyl (2*S*,4*S*) and (2*S*,4*R*)-2-amino-4-chloropentanoates were individually dissolved in D₂O with NaOD and the solutions were incubated at room temperature for 16 h. After the hydrolysis of the methyl esters, the resulting diastereomers of 2-amino-4-chloropentanoic acid **25** lactonised in situ and subsequently hydrolysed to give 2-amino-4-hydroxyacids. The relative stereochemistry at position C-4 of the hydrolysis products was established by ¹H NMR by comparison with literature values.^[157] As the stereocentre at position C-4 inverted once during the transformation and the stereocentre

at position C-2 was not affected, the stereochemistry of the starting chlorides could be determined based on the stereochemistry of the hydrolysis products.

(2*S*,4*S*)-2-amino-4-hydroxypentanoic acid: ¹H NMR (500 MHz, D₂O): δ = 4.03-3.96 (m, 1 H; NCH), 3.73 (dd, *J* = 9, 4 Hz, 1 H; OCH), 2.03 (dt, *J* = 15, 4 Hz, 1 H; CHH), 1.72 (dt, *J* = 15, 9 Hz, 1 H, CHH), 1.17 ppm (d, *J* = 6 Hz, 3 H; CH₃).

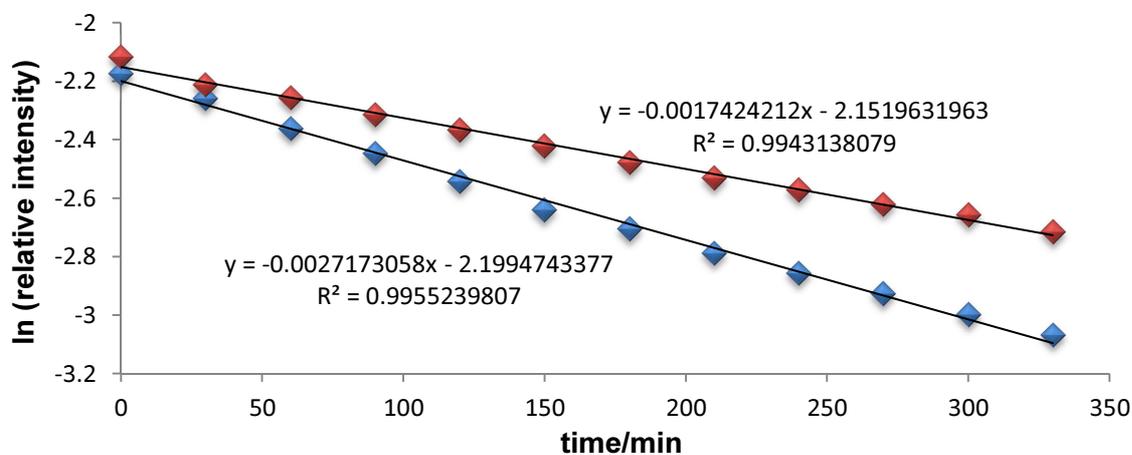
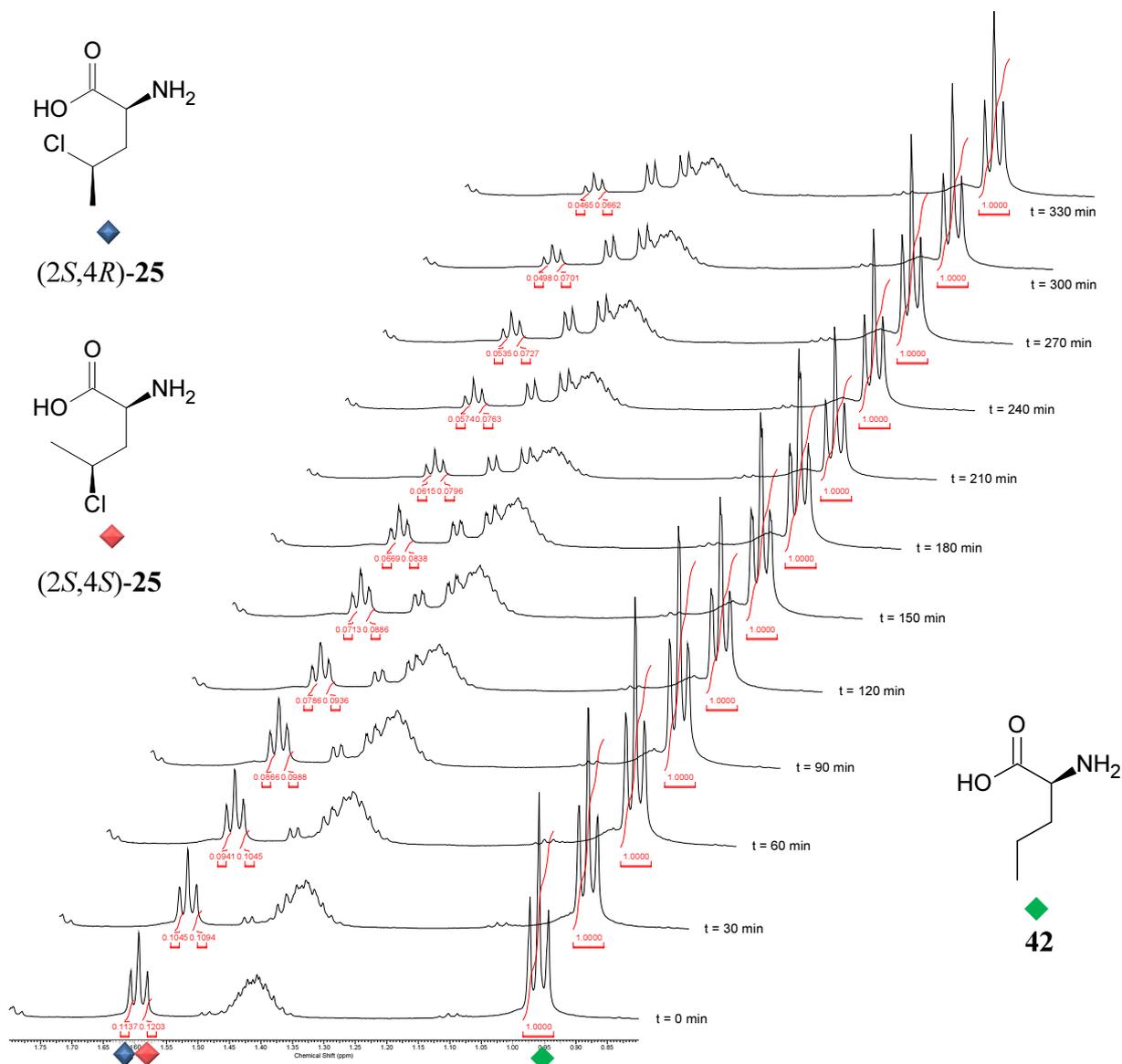
(2*S*,4*R*)-2-amino-4-hydroxypentanoic acid: ¹H NMR (500 MHz, D₂O): δ = 3.96-3.89 (m, 1 H; NCH), 3.89-3.85 (m, 1 H; OCH), 1.98-1.87 (m, 2 H; CH₂), 1.16 ppm (d, *J* = 6 Hz, 3 H; CH₃).

Lactonisation of the diastereomers of 4-chloronorvaline **25**

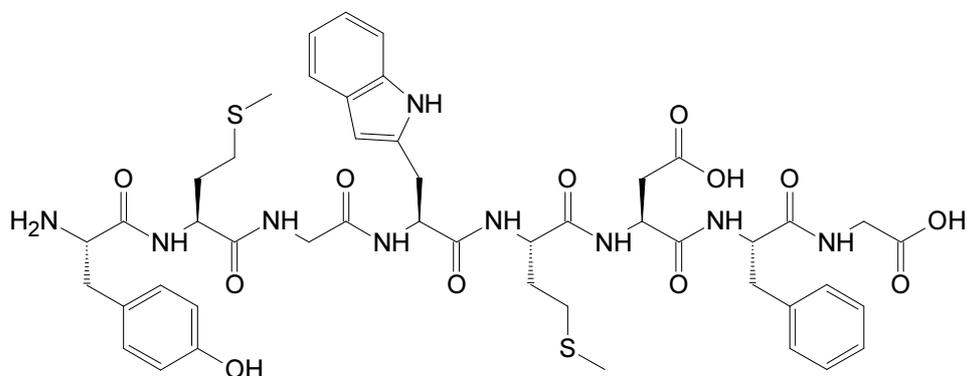
As observed with the chlorinated isosteres of isoleucine (Chapter 2), both diastereomers of the analogue of leucine **25** were prone to decomposition, with the products of hydrolysis positively identified by ¹H NMR as the corresponding 5-membered lactones, by comparison with published data.^[125] To accurately measure the rates of lactonisation, the mixture containing the two diastereomers and their stable precursor (*S*)-norvaline **42** was dissolved in D₂O and pD was adjusted to *ca.* 7 with 1 M NaOD in D₂O. ¹H NMR spectra of the mixture were recorded every 30 min for 6 h at 37°C, and the amounts of the chlorides relative to (*S*)-norvaline were measured.

Logarithmic plots of the amounts of the chlorides confirmed the lactonisation follows 1st order kinetics; the half-lives of decomposition for the two compound were found to be 255 min (4.25 h) for the (2*S*,4*R*)-diastereomer (◆) and 398 min (6.6 h) for the (2*S*,4*S*)-diastereomer (◆), consistent with previous reports.^[102]

Chlorinated Amino Acids in Peptide Production
Chapter 7. Experimental



Cholecystokinin-7-glycine propeptide



The octapeptide was synthesised on a CEM Liberty Microwave Peptide Synthesizer by Fmoc chemistry, on Fmoc-Gly-WANG resin (352 mg, sub. 0.71 mmol g⁻¹, 0.25 mmol) from Auspep. Fmoc protected amino acids (^tBu side-chain protection for Tyr and Asp, ^tBoc for Trp) and other chemicals for coupling were obtained from GL Biochem (Shanghai) Ltd. The coupling was carried out at 75°C for 5 min, using 4 eq. of each protected amino acid with HBTU (4 eq.) and DIPEA (3.6 eq.); deprotection (2×) was at 75°C for 3 min (initial deprotection 30 s) with 20% piperidine in DMF with HOBt (4 eq.). After the synthesis the resin was treated with TFA (1 mL; 2.5% H₂O, 2.5% TIPS) for 24 h and the acid solution was dropwise added to ether (100 mL) at 0°C. The precipitate was collected by centrifugation, air dried and purified by preparative HPLC (Alltima C18 5μ 250×22 mm column, 10 mL min⁻¹, 35% acetonitrile, 0.1% TFA). The fractions were lyophilised to give the target peptide as a white powder (3.5 mg, 3.5 μmol, 1.5%). R_f = 16.3 min.

m.p. 155-165°C; ¹H NMR (800 MHz, DMSO): δ = 10.77 (s, 1 H; Trp-NH_ε), 9.31 (s, br, 1 H; Tyr-OH), 8.71 (s, br, 1 H; Met²-H_N), 8.23-8.20 (m, 2H; Met⁵-H_N + Gly⁸-H_N), 8.18 (t, *J* = 5 Hz, 1 H; Gly³-H_N), 8.11 (d, *J* = 8 Hz, 1 H; Asp-H_N), 8.05 (d, *J* = 7 Hz, 1 H;

Trp-H_N), 7.95 (d, $J = 8$ Hz, 1 H; Phe-H_N), 7.59 (d, $J = 8$ Hz, 1 H; Trp-H_{Ar}), 7.30 (d, $J = 8$ Hz, 1 H; Trp-H_{Ar}), 7.26-7.21 (m, 4 H; Phe-H_{Ar}), 7.16 (tt, $J = 7, 2$ Hz, 1 H; Phe-H_{Ar}), 7.14 (d, $J = 2$ Hz, 1 H; Trp-H_{Ar}), 7.04 (t, $J = 8$ Hz, 1 H; Trp-H_{Ar}), 7.02 (d, $J = 9$ Hz, 2 H; Tyr-H_{Ar}), 6.96 (t, $J = 8$ Hz, 1 H; Trp-H_{Ar}), 6.68 (d, $J = 8$ Hz, 2 H; Tyr-H_{Ar}), 4.56 (dt, $J = 8, 5$ Hz, 1 H; Trp-H_α), 4.51 (q, $J = 7$ Hz, 1 H; Asp-H_α), 4.47 (dt, $J = 9, 5$ Hz, 1 H; Phe-H_α), 4.37 (br, 1 H; Met²-H_α), 4.29 (dt, br, $J = 8, 6$ Hz, 1 H; Met⁵-H_α), 3.86 (br, 1 H; Tyr-H_α), 3.74 (d, $J = 6$ Hz, 2 H; Gly⁸-H_α), 3.71 (dd, $J = 17, 6$ Hz, 1 H; Gly³-H_α), 3.66 (dd, $J = 17, 6$ Hz, 1 H; Gly³-H_α), 3.13 (dd, $J = 15, 5$ Hz, 1 H; Trp-H_β), 3.05 (dd, $J = 14, 5$ Hz, 1 H; Phe-H_β), 2.99-2.93 (m, 2 H; Tyr-H_β + Trp-H_β), 2.82 (dd, $J = 14, 9$ Hz, 1 H; Phe-H_β), 2.73 (dd, $J = 14, 9$ Hz, 1 H; Tyr-H_β), 2.63 (dd, $J = 16, 5$ Hz, 1 H; Asp-H_β), 2.49-2.43 (m, 2 H; Met²-H_γ + Asp-H_β), 2.41-2.36 (m, 2 H; Met²-H_γ + Met⁵-H_γ), 2.35-2.30 (m, 1 H; Met⁵-H_γ), 2.01 (s, 3 H; Met-H_ε), 1.99 (s, 3 H; Met-H_ε), 1.96-1.91 (m, 1 H; Met²-H_β), 1.90-1.84 (m, 1 H; Met⁵-H_β), 1.81-1.73 ppm (m, 2 H; Met²-H_β + Met⁵-H_β); ¹³C NMR (200 MHz, DMSO): $\delta = 172.1, 171.6, 171.0, 170.9, 170.8, 170.7, 170.3, 168.5, 157.7, 157.5, 156.4, 137.6, 136.0, 130.4, 129.1, 128.0, 127.3, 126.2, 125.4, 123.7, 120.8, 118.5, 118.2, 115.3, 111.2, 109.8, 65.8, 54.8, 54.2, 53.9, 53.4, 51.9, 51.9, 49.7, 45.7, 41.8, 40.9, 37.4, 36.7, 36.2, 31.9, 31.7, 29.4, 29.3, 27.6, 14.6, 14.5$ ppm; LRMS (ESI, +ve): m/z : 1006.7 [M⁺+H]; HRMS (ESI, +ve): m/z : calcd for [C₄₇H₅₉N₉O₁₂S₂+H]⁺ 1006.3803, found 1006.3804.

7.5. Experimental procedures of Chapter 4.

Gene construction

All genes for expression were ligated into a pET vector with T7 promoter. The initial fusion protein was prepared by ligation of the aprotinin gene to PpiB gene by using the two-step PCR mutagenesis. The gene for aprotinin was obtained from GeneArt AG. The gene for PpiB was available as described in previous chapters. All modified fusion proteins were obtained by mutagenesis of the original fusion protein *via* the same approach.

Protein production and purification

Cell-free protein expression was carried out at 30°C for 16 h as described in previous chapters. The volume of the inner chamber was 10 mL. After the expression the inner reaction mixture with all insoluble material was incubated at 100°C for 5 min, then guanidinium chloride was added to the final concentration of 6 M. The samples were then purified using HisTrapTM Ni(II)-affinity columns under denaturing conditions, following manufacturer's recommendations. The eluted protein was diluted 1:10 in refolding buffer containing 0.1 M Tris·HCl, 0.2 M KCl, 1 mM EDTA, 5 mM DTT, pH 8.7 and incubated for 16 h at room temperature. Trypsin (20 mg) was then added to the refolded protein and incubated for 1 h; grade I trypsin was obtained from Sigma-Aldrich Co. The complex was then again purified using HisTrapTM Ni(II)-affinity columns under native conditions, following manufacturer recommendations. The eluate was

concentrated and washed through Amicon[®] Ultra (YM-3000) centrifugal filter devices from Millipore.

Crystallisation

The aprotinin-trypsin complex was transferred to 20 mM HEPES buffer, pH 7.5, through repeated wash through Amicon[®] Ultra-4 (YM-3,000) centrifugal filter devices from Millipore and concentrated to *ca.* 10 mg mL⁻¹ protein concentration; protein concentrations were measured by absorbance at 280 nm on a NanoDrop ND-1000 spectrophotometer from ThermoScientific, the extinction coefficients were estimated using online ProtParam tool. Crystallisation was carried out using hanging-drop vapour-diffusion method. 24-well plates and 18 mm siliconised glass circular cover slides were purchased from Hampton Research. 1 mL reservoir solution contained 200 mM HEPES, pH 7.5, 1.45 M (NH₄)₂SO₄ and 10 mM CaCl₂,^[140] 1 μL of which was added to 10 μL of protein solution. Crystals suitable for crystallography were obtained within a week.

Data collection

Crystals were cryoprotected in reservoir solution containing 20% glycerol (v/v). Diffraction data were collected using Cu radiation (wavelength 1.5418 Å) at a temperature of 100 K, recorded as a contiguous sequence of 1° rotations using a MAR345 image plate detector mounted on a mardtb desktop beamline with a GeniX micro-focus X-ray source with cooling supplied by an Oxford Cryostream 700 series.

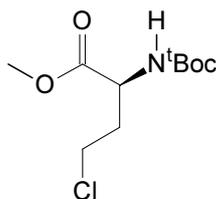
Between 129 and 180 frames were collected for each crystal with exposure time of 2400 or 3600 s per frame.

Structure refinement

Structure refinements were carried out on *refmac* 5.8.0073 as implemented in CCP4.^[158] Model was built with *Coot*,^[159] using a published structure 3otj as an initial model. Coordinates of the chlorinated residues SCB ((2*R*,3*S*)-2-amino-3-chlorobutyl) and RCB ((2*R*,3*R*)-2-amino-3-chlorobutyl) for refinement were constructed using PRODRG server.^[160]

7.6. Synthesis of compounds from Chapter 5.

Methyl (2*S*)-2-[(*tert*-butoxycarbonyl)amino]-4-chlorobutanoate **92a**



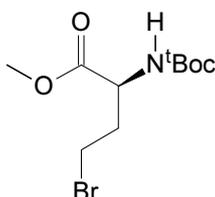
(*S*)-Homoserine **88** (600 mg, 5 mmol) and triethylamine (1 mL, 7.5 mmol) were dissolved in water and di-^tbutyl-dicarbonate (1.2 g, 5.5 mmol) in acetone was added. The solution was stirred with reflux for 3 h, then cooled down and evaporated to dryness *in vacuo*. The residue was redissolved in dimethylformamide and iodomethane (0.4 mL, 6 mmol) was added. Hünig's base was added until pH >> 7 and the reaction was stirred for 16 h at room temperature. All volatiles were then removed *in vacuo* to

give crude methyl (2*S*)-2-[(*tert*-butoxycarbonyl)amino]-4-hydroxybutanoate **89** as a colourless oil that was used in the subsequent step without further purification.

The protected homoserine **89** was dissolved in carbon tetrachloride (50 mL) and triphenylphosphine (1.6 g, 6 mmol) was added with stirring. The reaction was stirred with reflux for 24 h. The solution was then allowed to cool and the solvent was evaporated *in vacuo*. The oily residue was purified by silica chromatography, eluting with light petroleum / ether = 1:1. The target compound was obtained as a colourless oil (460 mg, 1.8 mmol, 37% over three steps).

$R_f = 0.5$; $[\alpha]_D^{20} = -47.6$ (c = 0.05 in methanol); $^1\text{H NMR}$ (400 MHz, CDCl_3): $\delta = 5.13$ (br, 1 H; NH), 4.46 (q, br, $J = 5$ Hz, 1 H; NCH), 3.77 (s, 3 H; OCH_3), 3.60 (t, $J = 7$ Hz, 2 H; CH_2Cl), 2.40-2.26 (m, 1 H; NCCHH), 2.20-2.06 (m, 1 H; NCCHH), 1.45 (s, 9 H; $\text{C}(\text{CH}_3)_3$); $^{13}\text{C NMR}$ (100 MHz, CDCl_3): $\delta = 172.4, 155.3, 80.3, 52.6, 51.4, 40.5, 35.6, 28.3$; LRMS (ESI, +ve): m/z : 274.1, 276.2 [$\text{M}^+ + \text{Na}$]; HRMS (ESI, +ve): m/z : calcd for $[\text{C}_{10}\text{H}_{18}\text{NO}_4\text{Cl} + \text{Na}]^+$ 274.0822 and 276.0793, found 274.0823 and 276.0794.

Methyl (2*S*)-2-[(*tert*-butoxycarbonyl)amino]-4-bromobutanoate **92b**

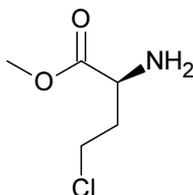


Crude methyl (2*S*)-2-[(*tert*-butoxycarbonyl)amino]-4-hydroxybutanoate **89** (approx. 200 mg, 0.86 mmol) was dissolved in dry dichloromethane and triphenylphosphine (270 mg, 1 mmol) and carbon tetrabromide (332 mg, 1 mmol) were added with stirring.

The reaction was stirred at room temperature under nitrogen atmosphere for 4 h. The solvent was then evaporated *in vacuo* and the residue was purified by silica chromatography, eluting with light petroleum / ether = 1:1, to give the target compound as a colourless oil (205 mg, 0.7 mmol, 80%). The compound was extremely prone to decomposition and was thus stored at -80°C. Spectroscopic data are consistent with literature reports.^[148]

$R_f = 0.5$; $[\alpha]_D^{20} = -42.8$ (c = 0.01 in methanol); $^1\text{H NMR}$ (400 MHz, CDCl_3): $\delta = 5.10$ (br, 1 H; NH), 4.44 (q, br, $J = 4$ Hz, 1 H; NCH), 3.78 (s, 3 H; OCH_3), 3.44 (t, $J = 7$ Hz, 2 H; CH_2Br), 2.49-2.35 (m, 1 H; NCCHH), 2.28-2.15 (m, 1 H; NCCHH), 1.46 (s, 9 H; $\text{C}(\text{CH}_3)_3$).

Methyl (2S)-2-amino-4-chlorobutanoate **87a**

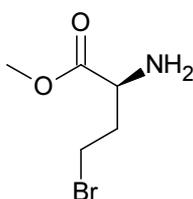


Methyl (2S)-2-[(*tert*-butoxycarbonyl)amino]-4-chlorobutanoate **92a** was dissolved in glacial TFA and stored at room temperature for 1 h. All volatiles were then evaporated *in vacuo* to give the target compound as a viscous colourless oil with essentially quantitative conversion. The compound was used without further processing.

$[\alpha]_D^{20} = +14.7$ (c = 0.02); $^1\text{H NMR}$ (400 MHz, D_3COD): $\delta = 4.23$ (t, $J = 7$ Hz, 1 H; NCH), 3.86 (s, 3 H; OCH_3), 3.76 (t, $J = 7$ Hz, 2 H; CH_2Cl), 2.44 (dq, $J = 15, 7$ Hz, 1 H; NCCHH), 2.30 (dq, $J = 15, 7$ Hz, 1 H; NCCHH); $^{13}\text{C NMR}$ (100 MHz, D_3COD): $\delta =$

170.5, 54.0, 51.6, 40.9, 34.6; LRMS (ESI, +ve): m/z : 152.0, 154.1 $[M^+H]$; HRMS (ESI, +ve): m/z : calcd for $[C_5H_{10}NO_2Cl+H]^+$ 152.0478 and 154.0449, found 152.0478 and 154.0451.

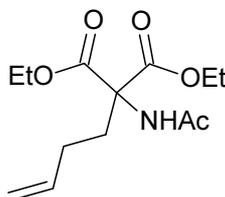
Methyl (2*S*)-2-amino-4-bromobutanoate **87b**



Methyl (2*S*)-2-[(*tert*-butoxycarbonyl)amino]-4-bromobutanoate **92b** was deprotected with glacial TFA in the same way as methyl (2*S*)-2-[(*tert*-butoxycarbonyl)amino]-4-chlorobutanoate **92a**.

$[\alpha]_D^{20} = +10.2$ ($c = 0.01$); 1H NMR (400 MHz, D_3COD): $\delta = 4.23$ (br, 1 H; NCH), 3.87 (s, 3 H; OCH_3), 3.60 (t, $J = 7$ Hz, 2 H; CH_2Br), 2.60-2.46 (m, br, 1 H; $NCC(H)$), 2.40-2.30 (m, 1 H; $NCCH(H)$); ^{13}C NMR (100 MHz, D_3COD): $\delta = 170.5, 54.1, 52.7, 34.9, 28.2$; LRMS (ESI, +ve): m/z : 196.0, 198.0 $[M^+H]$; HRMS (ESI, +ve): m/z : calcd for $[C_5H_{10}NO_2Br+H]^+$ 195.9973 and 197.9953, found 195.9977 and 197.9957.

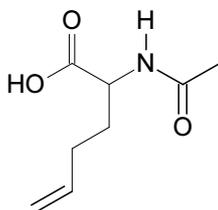
Diethyl (acetylamino)(but-3-en-1-yl)malonate 85



Diethyl acetamidomalonate **74** (136 mg, 0.63 mmol), caesium carbonate Cs_2CO_3 (220 mg, 0.68 mmol) and 4-bromobut-1-ene **84** (100 μg , 0.74 mmol) were added to dry acetonitrile (0.75 mL) in a microwave tube and the mixture was subjected to microwave irradiation (100°C, 200 W) for 1 h with stirring. The reaction mixture was then filtered to remove inorganic salts and the filtrate was evaporated to dryness. The residue was purified by flash silica chromatography, eluting with cyclohexane / ethyl acetate = 1:1, to give the target compound as a colourless oil (110 mg, 0.41 mmol, 64%). Spectroscopic data are consistent with literature reports.^[108]

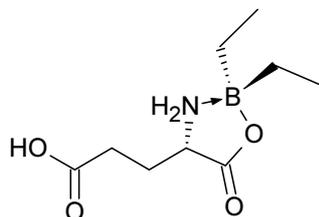
$R_f = 0.47$; $^1\text{H NMR}$ (400 MHz, CDCl_3): $\delta = 6.77$ (s, br, 1 H; NH), 5.80-5.67 (m, 1 H; $\text{CH}=\text{CH}_2$), 5.04-4.92 (m, 2 H; $\text{CH}=\text{CH}_2$), 4.23 (q, $J = 7$ Hz, 4 H; OCH_2), 2.47-2.39 (m, 2 H; $\text{CH}_2\text{CH}=\text{CH}_2$), 2.03 (s, 3 H; NCCH_3), 1.95-1.85 (m, 2 H; $\text{CH}_2\text{CH}_2\text{CH}=\text{CH}_2$), 1.25 (t, $J = 7$ Hz, 6 H; OCCH_3).

2-Aminohept-5-enoic acid **86**



Diethyl (acetylamino)(but-3-en-1-yl)propanedioate **85** (45 mg, 0.17 mmol) was suspended in aqueous sodium carbonate Na_2CO_3 solution (10% w/w) and the mixture was stirred with reflux for 16 h. The clear solution was then cooled down and $\text{HCl}_{(\text{aq})}$ was added until $\text{pH} \approx 3$. The solution was then swirled and all volatiles were removed *in vacuo*. The solid residue was extracted with hot acetone (55°C). The solvent was removed *in vacuo* to give the target compound as a colourless oil (25 mg, 0.15 mmol, 86%). Spectroscopic data are consistent with literature reports.^[108]

^1H NMR (400 MHz, D_3COD): $\delta = 5.89\text{-}5.75$ (m, 1 H; $\text{CH}=\text{CH}_2$), $5.09\text{-}4.97$ (m, 2 H; $\text{CH}=\text{CH}_2$), 4.38 (dd, $J = 9, 5$ Hz, 1 H; NCH), $2.23\text{-}2.06$ (m, 2 H; $\text{CH}_2\text{CH}=\text{CH}_2$), 1.99 (s, 3 H; CH_3), $1.97\text{-}1.84$ (m, 1 H; $\text{CHHCH}_2\text{CH}=\text{CH}_2$), $1.83\text{-}1.70$ (m, 1 H; $\text{CHHCH}_2\text{CH}=\text{CH}_2$).

***B,B*-Diethylboroxazolidone 97**

(*S*)-Glutamic acid **96** (2 g, 13.6 mmol) was suspended in sodium-dried dimethoxyethane (20 mL) and triethylborane (1 M in THF, 16 mL, 16 mmol) was added under nitrogen atmosphere with stirring. The solution was stirred with reflux under nitrogen atmosphere for 20 h, after which time most of the solid solubilised. The mixture was then cooled and filtered. The flow-through was evaporated *in vacuo* to give the title compound as a white gum (quant.) that was used in the next step without further purification. Spectroscopic data are consistent with literature reports.^[149]

¹H NMR (300 MHz, (CD₃)₂SO): δ = 6.51 (dd, J = 11, 8 Hz, 1 H; *NHH*), 5.62 (dd, J = 12, 8 Hz, 1 H; *NHH*), 3.65-3.45 (m, 1 H; *NCH*), 2.44 (t, J = 8 Hz, 2 H; *O₂CCH₂*), 2.12-1.98 (m, 1 H; *NCC₂H*), 1.80-1.64 (m, 1 H; *NCCH₂H*), 0.74-0.64 (m, 6 H; *CH₃*), 0.29-0.10 ppm (m, 4 H; *BCH₂*).

References

- [1] a) H.-D. Jakubke, N. Sewald, *Peptides from A to Z: A Concise Encyclopedia*, Wiley-VCH, Weinheim, **2008**; b) B. Groner, *Peptides as Drugs: Discovery and Development*, **2009**; c) N. Sewald, H.-D. Jakubke, *Peptides: Chemistry and Biology, 2nd edition*, Wiley-VCH, Weinheim, **2009**.
- [2] a) B. Leader, Q. J. Baca, D. E. Golan, *Nat. Rev. Drug Discov.* **2008**, *7*, 21-39; b) A. A. Kaspar, J. M. Reichert, *Drug Discov Today* **2013**, *18*, 807-817; c) L. Otvos, Jr., J. D. Wade, *Frontiers in chemistry* **2014**, *2*, 62-62; d) K. Fosgerau, T. Hoffmann, *Drug Discov Today* **2015**, *20*, 122-128.
- [3] P. Vlieghe, V. Lisowski, J. Martinez, M. Khrestchatisky, *Drug Discov Today* **2010**, *15*, 40-56.
- [4] D. J. Craik, D. P. Fairlie, S. Liras, D. Price, *Chem. Biol. Drug Des.* **2013**, *81*, 136-147.
- [5] T. Uhlig, T. Kyprianou, F. G. Martinelli, C. A. Oppici, D. Heiligers, D. Hills, X. R. Calvo, P. Verhaert, *EuPA Open Proteomics* **2014**, *4*, 58-69.
- [6] J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **2013**, *341*, 1238149.
- [7] a) T. J. Deming, *Adv. Mater.* **1997**, *9*, 299-311; b) "Special Issue: Peptides in Materials Science" (Guest Ed.: J. Schneider), *Biopolymers* **2010**, *94*, 1-155; c) S. Cavalli, F. Albericio, A. Kros, *Chem. Soc. Rev.* **2010**, *39*, 241-263; d) S. Banta, I. R. Wheeldon, M. Blenner, *Annu. Rev. Biomed. Eng., Vol 12* **2010**, *12*, 167-186; e) J. B. Matson, R. H. Zha, S. I. Stupp, *Curr. Opin. Solid St. M.* **2011**, *15*, 225-235; f) L. Liu, K. Busuttill, S. Zhang, Y. Yang, C. Wang, F. Besenbacher, M. Dong, *Phys Chem Chem Phys* **2011**, *13*, 17435-17444; g) *Amino Acids, Peptides and Proteins: Volume 37*, (Eds.: E. Farkas, M. Ryadnov) in Specialist Periodical Reports, RSC Publishing, Cambridge, **2012**; h) A. Lakshmanan, S. Zhang, C. A. E. Hauser, *Trends Biotechnol.* **2012**, *30*, 155-165; i) M. Venanzi, S. Kimura, *Polym. J.* **2013**, *45*, 467-467; j) E. Busseron, Y. Ruff, E. Moulin, N. Giuseppone, *Nanoscale* **2013**, *5*, 7098-7140; k) J. Boekhoven, S. I. Stupp, *Adv. Mater.* **2014**, *26*, 1642-1659; l) G. K. Such, Y. Yan, A. P. R. Johnston, S. T. Gunawan, F. Caruso, *Adv. Mater.* **2015**, *27*, 2278-2297.
- [8] T. W. Overton, *Drug Discov Today* **2014**, *19*, 590-601.
- [9] a) J. F. Morrow, S. N. Cohen, A. C. Y. Chang, H. W. Boyer, H. M. Goodman, R. B. Helling, *P. Nat. Acad. Sci. USA* **1974**, *71*, 1743-1747; b) A. C. Y. Chang, J. H. Nunberg, R. J. Kaufman, H. A. Erlich, R. T. Schimke, S. N. Cohen, *Nature* **1978**, *275*, 617-624.
- [10] a) H. P. Sorensen, K. K. Mortensen, *J. Biotechnol.* **2005**, *115*, 113-128; b) Y. Li, *Protein Expression Purif.* **2011**, *80*, 260-267.
- [11] G. L. Rosano, E. A. Ceccarelli, *Frontiers in microbiology* **2014**, *5*, 172-172.
- [12] S. Graslund, P. Nordlund, J. Weigelt, J. Bray, B. M. Hallberg, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. dhe-Paganon, H.-w. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S.-H. Kim, Z. Rao, Y. Shi, T. C. Terwilliger, C.-Y. Kim, L.-W. Hung, G. S. Waldo, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J. L. Sussman, R. C. Stevens, S. A. Lesley, I. A. Wilson, A. Joachimiak, F. Collart, I. Dementieva, M. I. Donnelly, W. H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, S. K. Burley, J. S. Emtage, J. M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S. C. Almo, J. B. Bonanno, A. Fiser, S. Swaminathan, F. W. Studier, M. R. Chance, A. Sali, T. B. Acton, R. Xiao, L. Zhao, L. C. Ma, J. F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C. K. Ho, D. Wang, H. Wang, M. Jiang, G. T. Montelione, D. I. Stuart, R. J. Owens, S.

- Daenke, A. Schutz, U. Heinemann, S. Yokoyama, K. Bussow, K. C. Gunsalus, C. Struct Genomics, M. Architecture Fonction, C. Berkeley Struct Genomics, C. China Struct Genomics, F. Integrated Ctr Struct, C. Israel Struct Proteomics, G. Joint Ctr Struct, G. Midwest Ctr Struct, X. R. C. New York Struct Genomi, N. E. S. G. Consortium, F. Oxford Prot Prod, F. Prot Sample Prod, M. Max Delbruck Ctr Mol, R. S. G. Proteomics, S. Complexes, *Nat. Methods* **2008**, *5*, 135-146.
- [13] G. J. Gopal, A. Kumar, *Protein J.* **2013**, *32*, 419-425.
- [14] A. S. Spirin, *Cell-Free Translation Systems*, Springer, Berlin Heidelberg, **2002**.
- [15] M. Nirenberg, J. H. Matthaei, *P. Nat. Acad. Sci. USA* **1961**, *47*, 1588-1602.
- [16] a) L. Guignard, K. Ozawa, S. E. Pursglove, G. Otting, N. E. Dixon, *FEBS Lett.* **2002**, *524*, 159-162; b) K. Ozawa, M. J. Headlam, P. M. Schaeffer, B. R. Henderson, N. E. Dixon, G. Otting, *Eur. J. Biochem.* **2004**, *271*, 4084-4093; c) K. Ozawa, M. J. Headlam, D. Mouradov, S. J. Watt, J. L. Beck, K. J. Rodgers, R. T. Dean, T. Huber, G. Otting, N. E. Dixon, *FEBS J.* **2005**, *272*, 3162-3171; d) M. A. Apponyi, K. Ozawa, N. E. Dixon, G. Otting, Cell-Free Protein Synthesis for Analysis by NMR Spectroscopy, in *Methods in molecular biology, Vol 426* (Eds.: B. Kobe, M. Guss, T. Huber), Humana Press, Totowa, N. J., **2008**, 257-268.
- [17] a) O. Kovtun, S. Mureev, W. Jung, M. H. Kubala, W. Johnston, K. Alexandrov, *Methods* **2011**, *55*, 58-64; b) E. D. Carlson, R. Gan, C. E. Hodgman, M. C. Jewett, *Biotechnol. Adv.* **2012**, *30*, 1185-1194; c) M. G. Casteleijn, A. Urtti, S. Sarkhel, *Int. J. Pharm.* **2013**, *440*, 39-47; d) M. T. Smith, K. M. Wilding, J. M. Hunt, A. M. Bennett, B. C. Bundy, *FEBS Lett.* **2014**, *588*, 2755-2761.
- [18] N. Velappan, D. Sblattero, L. Chasteen, P. Pavlik, A. R. M. Bradbury, *Protein Eng. Des. Sel.* **2007**, *20*, 309-313.
- [19] M. Martinez-Alonso, E. Garcia-Fruitos, N. Ferrer-Miralles, U. Rinas, A. Villaverde, *Microb. Cell Fact.* **2010**, *9*.
- [20] C. L. Young, Z. T. Britton, A. S. Robinson, *Biotechnol. J.* **2012**, *7*, 620-634.
- [21] K. Itakura, T. Hirose, R. Crea, A. D. Riggs, H. L. Heyneker, F. Bolivar, H. W. Boyer, *Science* **1977**, *198*, 1056-1063.
- [22] a) K. L. Piers, M. H. Brown, R. E. W. Hancock, *Gene* **1993**, *134*, 7-13; b) K. A. Martemyanov, A. S. Spirin, A. T. Gudkov, *Biotechnol. Lett.* **1996**, *18*, 1357-1362; c) K. A. Martemyanov, V. A. Shirokov, O. V. Kurnasov, A. T. Gudkov, A. S. Spirin, *Protein Expression Purif.* **2001**, *21*, 456-461; d) Y. K. Chae, M. Tonneli, J. L. Markley, *Protein Peptide Lett.* **2012**, *19*, 808-811; e) Y. Liu, L. Ren, L. Ge, Q. Cui, X. Cao, Y. Hou, F. Bai, G. Bai, *Biotechnol. Lett.* **2014**, *36*, 1675-1680.
- [23] R. R. Kosana, C. Bajji, R. M. Kanumuri, K. Panati, L. N. Mangamoori, M. R. Tummuru, V. R. Narala, *Protein Expression Purif.* **2014**, *95*, 136-142.
- [24] S. H. Shen, *P. Nat. Acad. Sci.-Biol.* **1984**, *81*, 4627-4631.
- [25] V. Mäde, S. Els-Heindl, A. G. Beck-Sickinger, *Beilstein J. Org. Chem.* **2014**, *10*, 1197-1212.
- [26] F. Liu, J. P. Mayer, *Top. curr. chem.* **2015**, *362*, 183-228.
- [27] a) J. M. Palomo, *RSC Adv.* **2014**, *4*, 32658-32672; b) B. E. I. Ramakers, J. C. M. van Hest, D. W. P. M. Lowik, *Chem. Soc. Rev.* **2014**, *43*, 2743-2756.
- [28] B. L. Bray, *Nat. Rev. Drug Discov.* **2003**, *2*, 587-593.
- [29] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.
- [30] a) F. Albericio, *Curr. Opin. Chem. Biol.* **2004**, *8*, 211-221; b) A. El-Faham, F. Albericio, *Chem. Rev.* **2011**, *111*, 6557-6602.
- [31] M. Schnolzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, *Int. J. Pept. Res. Ther.* **2007**, *13*, 31-44.
- [32] a) P. White, J. W. Keyte, K. Bailey, G. Bloomberg, *J. Pept. Sci.* **2004**, *10*, 18-26; b) I. Coin, *J. Pept. Sci.* **2010**, *16*, 223-230.
- [33] B. L. Nilsson, M. B. Soellner, R. T. Raines, *Annu. Rev. Bioph. Biom.* **2005**, *34*, 91-118.

- [34] a) L. Raibaut, N. Ollivier, O. Melnyk, *Chem. Soc. Rev.* **2012**, *41*, 7001-7015; b) M. Gongora-Benitez, J. Tulla-Puche, F. Albericio, *ACS Comb. Sci.* **2013**, *15*, 217-228.
- [35] M. C. Jamjian, I. R. McNicholl, *Am. J. Health-Syst. Ph.* **2004**, *61*, 1242-1247.
- [36] a) J. Arnau, C. Lauritzen, G. E. Petersen, J. Pedersen, *Protein Expression Purif.* **2006**, *48*, 1-13; b) D. S. Waugh, *Protein Expression Purif.* **2011**, *80*, 283-293.
- [37] C. Lopez-Otin, J. S. Bond, *J. Biol. Chem.* **2008**, *283*, 30433-30437.
- [38] T. F. Spande, B. Witkop, Y. Degani, A. Patchornik, *Adv. Protein Chem.* **1970**, *24*, 97-260.
- [39] a) N. M. Milovic, L. M. Dutca, N. M. Kostic, *Chem.-Eur. J.* **2003**, *9*, 5097-5106; b) A. Krezel, M. Mylonas, E. Kopera, W. Bal, *Acta Biochim. Pol.* **2006**, *53*, 721-727.
- [40] a) H. M. Kim, B. Jang, Y. E. Cheon, M. P. Suh, J. Suh, *J. Biol. Inorg. Chem.* **2009**, *14*, 151-157; b) J. Prakash, J. J. Kodanko, *Curr. Opin. Chem. Biol.* **2013**, *17*, 197-203; c) P. M. Hwang, J. S. Pan, B. D. Sykes, *FEBS Lett.* **2014**, *588*, 247-252.
- [41] E. Gross, B. Witkop, *J. Am. Chem. Soc.* **1961**, *83*, 1510-1511.
- [42] E. Gross, B. Witkop, *J. Biol. Chem.* **1962**, *237*, 1856-1860.
- [43] W. B. Lawson, B. Witkop, E. Gross, C. M. Foltz, *J. Am. Chem. Soc.* **1961**, *83*, 1509-1510.
- [44] A. S. Inglis, P. Edman, *Anal. Biochem.* **1970**, *37*, 73-80.
- [45] W. A. Schroeder, J. B. Shelton, J. R. Shelton, *Arch. Biochem. Biophys.* **1969**, *130*, 551-556.
- [46] a) K. K. Han, D. Tetaert, Dautreva.M, Moschett.Y, C. Kopeyan, *Eur. J. Biochem.* **1972**, *27*, 585-592; b) R. Kaiser, L. Metzka, *Anal. Biochem.* **1999**, *266*, 1-8.
- [47] R. Joppichkuhn, J. A. Corkill, R. W. Giese, *Anal. Biochem.* **1982**, *119*, 73-77.
- [48] F. H. Carpenter, S. M. Shiigi, *Biochemistry-US* **1974**, *13*, 5159-5164.
- [49] V. Rahali, J. Gueguen, *J. Prot. Chem.* **1999**, *18*, 1-12.
- [50] W. C. Mahoney, P. K. Smith, M. A. Hermodson, *Biochemistry-US* **1981**, *20*, 443-448.
- [51] S. Isoe, L. A. Cohen, *Arch. Biochem. Biophys.* **1968**, *127*, 522-527.
- [52] a) T. Baird, B. X. Wang, M. Lodder, S. M. Hecht, C. S. Craik, *Tetrahedron* **2000**, *56*, 9477-9485; b) B. X. Wang, K. C. Brown, M. Lodder, C. S. Craik, S. M. Hecht, *Biochemistry-US* **2002**, *41*, 2805-2813.
- [53] B. X. Wang, M. Lodder, J. Zhou, T. T. Baird, K. C. Brown, C. S. Craik, S. M. Hecht, *J. Am. Chem. Soc.* **2000**, *122*, 7402-7403.
- [54] Y. Degani, A. Patchornik, *Biochemistry-US* **1974**, *13*, 1-11.
- [55] K. Tanabe, A. Taniguchi, T. Matsumoto, K. Oisaki, Y. Sohma, M. Kanai, *Chem. Sci.* **2014**, *5*, 2747-2753.
- [56] T. J. Holmes, R. G. Lawton, *J. Am. Chem. Soc.* **1977**, *99*, 1984-1986.
- [57] P. Bornstein, G. Balian, *Methods enzymol.* **1977**, *47*, 132-145.
- [58] a) A. Patchornik, M. Sokolovsky, *J. Am. Chem. Soc.* **1964**, *86*, 1206-1212; b) A. Regueiro-Ren, Y. Ueda, *J. Org. Chem.* **2002**, *67*, 8699-8702.
- [59] Y. Seki, K. Tanabe, D. Sasaki, Y. Sohma, K. Oisaki, M. Kanai, *Angew. Chem., Int. Ed.* **2014**, *53*, 6501-6505.
- [60] B. M. Eisenhauer, S. M. Hecht, *Biochemistry-US* **2002**, *41*, 11472-11478.
- [61] S. R. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus, M. Q. Xu, *Gene* **1997**, *192*, 271-281.
- [62] A. L. Adams, D. Macmillan, *J. Pept. Sci.* **2013**, *19*, 65-73.
- [63] a) J. Kang, N. L. Reynolds, C. Tyrrell, J. R. Dorin, D. Macmillan, *Org. Biomol. Chem.* **2009**, *7*, 4918-4923; b) D. Macmillan, A. Adams, B. Premjee, *Isr. J. Chem.* **2011**, *51*, 885-899; c) B. Cowper, L. Shariff, W. Chen, S. M. Gibson, W.-L. Di, D. Macmillan, *Org. Biomol. Chem.* **2015**, *13*, 7469-7476.
- [64] T. M. Schmeing, V. Ramakrishnan, *Nature* **2009**, *461*, 1234-1242.
- [65] M. Ibbá, D. Soll, *Annu. Rev. Biochem.* **2000**, *69*, 617-650.

- [66] C. R. Woese, G. J. Olsen, M. Ibba, D. Soll, *Microbiol. Mol. Biol. R.* **2000**, *64*, 202-236.
- [67] S. S. Yadavalli, M. Ibba, *Advances in Protein Chemistry and Structural Biology, Vol 86: Fidelity and Quality Control in Gene Expression* **2012**, *86*, 1-43.
- [68] a) J. Ling, N. Reynolds, M. Ibba in *Aminoacyl-tRNA Synthesis and Translational Quality Control, Vol. 63* **2009**, pp. 61-78; b) H. Gingold, Y. Pilpel, *Mol. Syst. Biol.* **2011**, *7*; c) L. Ribas de Pouplana, *EMBO J* **2014**, *33*, 1619-1620.
- [69] H. Jakubowski, *Wiley Interdiscip. Rev. RNA* **2012**, *3*, 295-310.
- [70] a) A. R. Fersht, J. S. Shindler, W. C. Tsui, *Biochemistry-US* **1980**, *19*, 5520-5524; b) P. Brick, T. N. Bhat, D. M. Blow, *J. Mol. Biol.* **1989**, *208*, 83-98; c) T. Kobayashi, T. Takimura, R. Sekine, K. Vincent, K. Kamata, K. Sakamoto, S. Nishimura, S. Yokoyama, *J. Mol. Biol.* **2005**, *346*, 105-117; d) T. Kobayashi, T. Takimura, R. Sekine, V. P. Kelly, K. Kamata, K. Sakamoto, S. Nishimura, S. Yokoyama, *J. Mol. Biol.* **2005**, *354*, 739-739.
- [71] A. R. Fersht, C. Dingwall, *Biochemistry-US* **1979**, *18*, 1245-1249.
- [72] a) D. Moras, *P. Nat. Acad. Sci. USA* **2010**, *107*, 21949-21950; b) M. Dulic, N. Cvetesic, J. J. Perona, I. Gruic-Sovulj, *J. Biol. Chem.* **2010**, *285*, 23799-23809.
- [73] a) J. Flossdorf, H. J. Pratorius, M. R. Kula, *Eur. J. Biochem.* **1976**, *66*, 147-155; b) L. F. Silvian, J. M. Wang, T. A. Steitz, *Science* **1999**, *285*, 1074-1077.
- [74] R. Fukunaga, S. Yokoyama, *J. Molecular Biol.* **2006**, *359*, 901-912.
- [75] H. Jakubowski, *Acta Biochim. Pol.* **2011**, *58*, 149-163.
- [76] L. Pauling, *The Probability of Errors in the Process of Synthesis of Protein Molecules*, Birkhauser, **1957**.
- [77] A. R. Fersht, M. M. Kaethner, *Biochemistry-US* **1976**, *15*, 3342-3346.
- [78] M. J. Wilson, D. L. Hatfield, *Biochim. Biophys. Acta* **1984**, *781*, 205-215.
- [79] T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel, *Annu. Rev. Biochem.* **2004**, *73*, 147-176.
- [80] S. Lepthien, L. Merkel, N. Budisa, *Angew. Chem., Int. Ed.* **2010**, *49*, 5446-5450.
- [81] S.-J. Oh, K.-H. Lee, H.-C. Kim, C. Catherine, H. Yun, D.-M. Kim, *Biotechnol. Bioproc. E.* **2014**, *19*, 426-432.
- [82] G. A. Rosenthal, *Amino Acids* **2001**, *21*, 319-330.
- [83] K. Katagiri, K. Tori, Y. Kimura, T. Yoshida, T. Nagasaki, H. Minato, *J. Med. Chem.* **1967**, *10*, 1149-1154.
- [84] T. Kohno, D. Kohda, M. Haruki, S. Yokoyama, T. Miyazawa, *J. Biol. Chem.* **1990**, *265*, 6931-6935.
- [85] a) L. Fowden, *Biochem. J.* **1956**, *64*, 323-332; b) E. Rubenstein, T. McLaughlin, R. C. Winant, A. Sanchez, M. Eckart, K. M. Krasinska, A. Chien, *Phytochemistry* **2009**, *70*, 100-104.
- [86] a) M. Rabinovitz, J. M. Fisher, *Biochim. Biophys. Acta* **1964**, *91*, 313-322; b) A. L. Goldberg, *P. Nat. Acad. Sci. USA* **1972**, *69*, 422-426; c) W. F. Prouty, *J. Virol.* **1975**, *16*, 1090-1093; d) Y. Klemes, J. D. Etlinger, A. L. Goldberg, *J. Biol. Chem.* **1981**, *256*, 8436-8444.
- [87] S. Lepthien, M. G. Hoesl, L. Merkel, N. Budisa, *P. Nat. Acad. Sci. USA* **2008**, *105*, 16095-16100.
- [88] a) L. A. Galli-Taliadoros, J. D. Sedgwick, S. A. Wood, H. Korner, *J. Immunol. Methods* **1995**, *181*, 1-15; b) C. W. Song, J. Lee, S. Y. Lee, *Biotechnology Journal* **2015**, *10*, 56-68.
- [89] T. Michon, F. Barbot, D. Tirrell, *Plant Biopolymer Science: Food and Non-Food Applications* **2002**, 63-72.
- [90] a) M. L. Mock, T. Michon, J. C. M. van Hest, D. A. Tirrell, *ChemBiochem* **2006**, *7*, 83-87; b) J. A. Van Deventer, J. D. Fisk, D. A. Tirrell, *ChemBiochem* **2011**, *12*, 700-702; c) Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado, D. A. Tirrell, *Angew. Chem., Int. Ed.* **2001**, *40*, 1494-1496.
- [91] P. Wang, Y. Tang, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, *125*, 6900-6906.

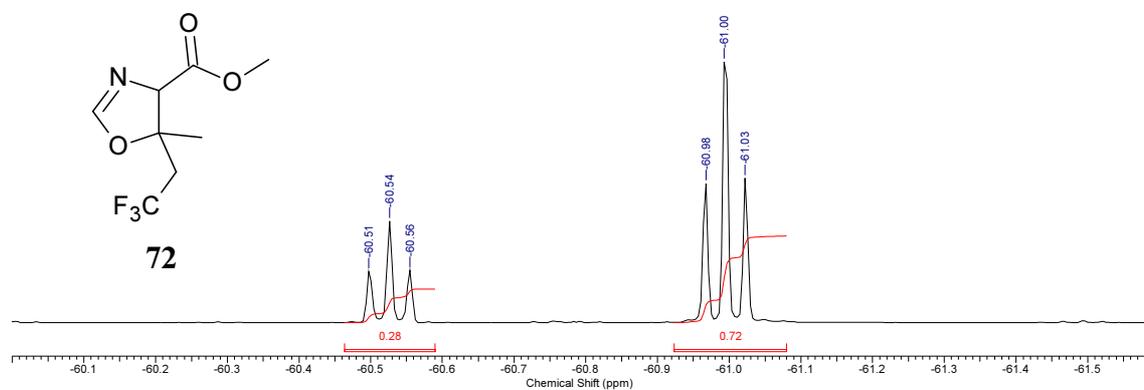
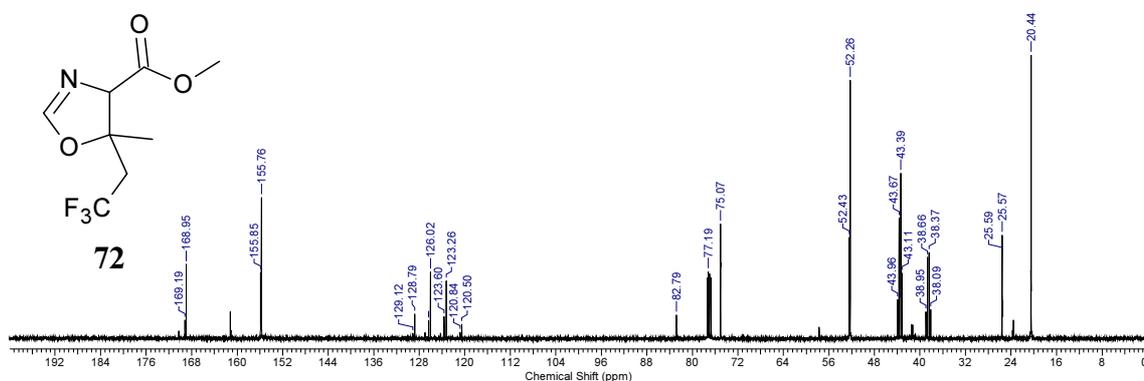
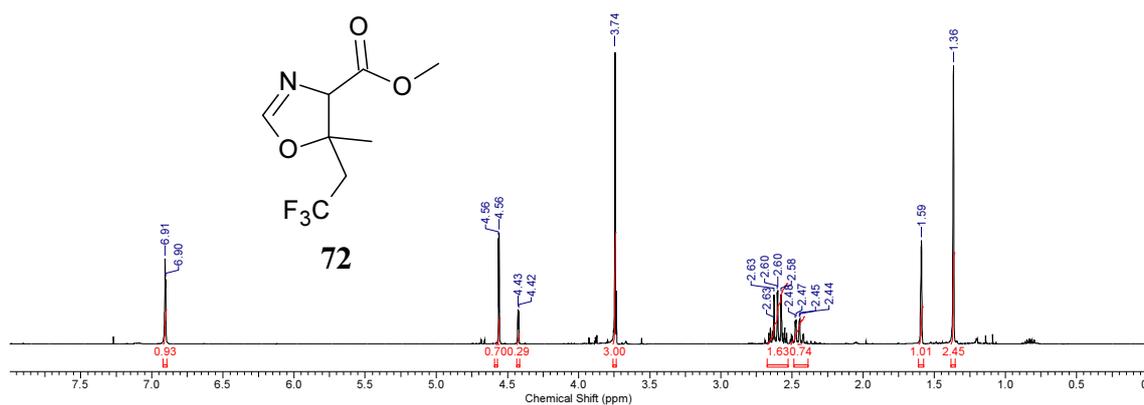
- [92] a) P. Kast, H. Hennecke, *J. Mol. Biol.* **1991**, *222*, 99-124; b) P. Kast, *Chembiochem* **2011**, *12*, 2395-2398.
- [93] a) A. J. Link, M. L. Mock, D. A. Tirrell, *Curr. Opin. Biotechnol.* **2003**, *14*, 603-609; b) N. Budisa, *Angew. Chem., Int. Ed.* **2004**, *43*, 6426-6463; c) R. E. Connor, D. A. Tirrell, *Polym. Rev.* **2007**, *47*, 9-28; d) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem., Vol 79* **2010**, *79*, 413-444; e) M. G. Hoesl, N. Budisa, *Angew. Chem., Int. Ed.* **2011**, *50*, 2896-2902; f) P. O'Donoghue, J. Ling, Y.-S. Wang, D. Soell, *Nat. Chem. Biol.* **2013**, *9*, 594-598.
- [94] a) J. A. Johnson, Y. Y. Lu, J. A. Van Deventer, D. A. Tirrell, *Curr. Opin. Chem. Biol.* **2010**, *14*, 774-780; b) K. Lang, J. W. Chin, *ACS Chem. Biol.* **2014**, *9*, 16-20; c) K. Lang, J. W. Chin, *Chem. Rev.* **2014**, *114*, 4764-4806.
- [95] a) S. Ye, A. A. Berger, D. Petzold, O. Reimann, B. Matt, B. Kokschi, *Beilstein J. Org. Chem.* **2010**, *6*; b) T. G. Heckler, L. H. Chang, Y. Zama, T. Naka, M. S. Chorghade, S. M. Hecht, *Biochemistry-US* **1984**, *23*, 1468-1473; c) M. Duca, S. Chen, S. M. Hecht, *Methods* **2008**, *44*, 87-99.
- [96] a) A. R. Fersht, C. Dingwall, *Biochemistry-US* **1979**, *18*, 1238-1244; b) J. F. Chen, N. N. Guo, T. Li, E. D. Wang, Y. L. Wang, *Biochemistry-US* **2000**, *39*, 6726-6731; c) A. R. Fersht, *Biochemistry-US* **1977**, *16*, 1025-1030.
- [97] S. A. Fraser, C. J. Easton, *Aust. J. Chem.* **2015**, *68*, 9-12.
- [98] M. Salwiczek, E. K. Nyakatura, U. I. M. Gerling, S. Ye, B. Kokschi, *Chem. Soc. Rev.* **2012**, *41*, 2135-2171.
- [99] P. Wang, A. Fichera, K. Kumar, D. A. Tirrell, *Angew. Chem., Int. Ed.* **2004**, *43*, 3664-3666.
- [100] Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado, D. A. Tirrell, *Angew. Chem., Int. Ed.* **2001**, *40*, 1494-1496.
- [101] Y. Tang, D. A. Tirrell, *J. Am. Chem. Soc.* **2001**, *123*, 11089-11090.
- [102] D. J. Stigers, Z. I. Watts, J. E. Hennessy, H.-K. Kim, R. Martini, M. C. Taylor, K. Ozawa, J. W. Keillor, N. E. Dixon, C. J. Easton, *Chem. Commun.* **2011**, *47*, 1839-1841.
- [103] S. Narayanan, M. R. S. Iyengar, P. L. Ganju, S. Rengaraju, T. Shomura, T. Tsuruoka, S. Inouye, T. Niida, *J. Antibiot.* **1980**, *33*, 1249-1255.
- [104] a) D. Datta, N. Vaidehi, D. Q. Zhang, W. A. Goddard, *Protein Sci.* **2004**, *13*, 2693-2705; b) S. P. Nadarajan, S. Mathew, K. Deepankumar, H. Yun, *J. Mol. Graph. Model.* **2013**, *39*, 79-86.
- [105] S. Doublie, *Method Enzymol.* **1997**, *276*, 523-530.
- [106] K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *P. Nat. Acad. Sci. USA* **2002**, *99*, 19-24.
- [107] a) K. L. Kiick, R. Weberskirch, D. A. Tirrell, *FEBS Lett.* **2001**, *502*, 25-30; b) K. L. Kiick, R. Weberskirch, D. A. Tirrell, *FEBS Lett.* **2001**, *505*, 465-465.
- [108] J. C. M. van Hest, K. L. Kiick, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, *122*, 1282-1288.
- [109] B. Wiltschi, W. Wenger, S. Nehring, N. Budisa, *Yeast* **2008**, *25*, 775-786.
- [110] M. Suchanek, A. Radzikowska, C. Thiele, *Nat. Methods* **2005**, *2*, 261-267.
- [111] N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann, R. Huber, *Eur. J. Biochem.* **1995**, *230*, 788-796.
- [112] M. Levine, H. Tarver, *J. Biol. Chem.* **1951**, *192*, 835-850.
- [113] H. Jakubowski, *J. Biol. Chem.* **2000**, *275*, 21813-21816.
- [114] C. Wolschner, A. Giese, H. A. Kretzschmar, R. Huber, L. Moroder, N. Budisa, *P. Nat. Acad. Sci. USA* **2009**, *106*, 7756-7761.
- [115] M. D. Vaughan, P. Cleve, V. Robinson, H. S. Duewel, J. F. Honek, *J. Am. Chem. Soc.* **1999**, *121*, 8475-8478.
- [116] a) H. Duewel, E. Daub, V. Robinson, J. F. Honek, *Biochemistry* **1997**, *36*, 3404-3416; b) N. Budisa, O. Pipitone, I. Siwanowicz, M. Rubini, P. P. Pal, T. A. Holak, M. L. Gelmi, *Chem. Biodivers.* **2004**, *1*, 1465-1475.

- [117] L. Serre, G. Verdon, T. Choinowski, N. Hervouet, J. L. Risler, C. Zelwer, *J. Mol. Biol.* **2001**, *306*, 863-876.
- [118] a) H. Jakubowski, *Cell. Mol. Life Sci.* **2004**, *61*, 470-487; b) H. Jakubowski, *Phosphorus Sulfur* **2013**, *188*, 384-395.
- [119] Z. I. Watts, C. J. Easton, *J. Am. Chem. Soc.* **2009**, *131*, 11323-11325.
- [120] I. L. C. Hernandez, M. J. L. Godinho, A. Magalhaes, A. B. Schefer, A. G. Ferreira, R. G. S. Berlinck, *J. Nat. Prod.* **2000**, *63*, 664-665.
- [121] a) P. Hemstrom, K. Irgum, *J. Sep. Sci.* **2006**, *29*, 1784-1821; b) B. Buszewski, S. Noga, *Anal. Bioanal. Chem.* **2012**, *402*, 231-247.
- [122] R. A. K. Rao, M. Ajmal, B. A. Siddiqui, S. Ahmad, *Environ. Monit. Assess.* **1999**, *54*, 289-299.
- [123] I. N. Arthur, J. E. Hennessy, D. Padmakshan, D. J. Stigers, S. Lesturgez, S. A. Fraser, M. Liutkus, G. Otting, J. G. Oakeshott, C. J. Easton, *Chem.-Eur. J.* **2013**, *19*, 6824-6830.
- [124] F. H. Vaillancourt, E. Yeh, D. A. Vosburg, S. E. O'Connor, C. T. Walsh, *Nature* **2005**, *436*, 1191-1194.
- [125] B. D. Dangel, J. A. Johnson, D. Sames, *J. Am. Chem. Soc.* **2001**, *123*, 8149-8150.
- [126] C. Benhaim, L. Bouchard, G. Pelletier, J. Sellstedt, L. Kristofova, S. Daigneault, *Org. Lett.* **2010**, *12*, 2008-2011.
- [127] P. Schadewaldt, A. Bodner-Leidecker, H. W. Hammen, U. Wendel, *Pediatr. Res.* **2000**, *47*, 271-277.
- [128] C. M. Marson, U. Grabowska, A. Fallah, T. Walsgrove, D. S. Eggleston, P. W. Baures, *J. Org. Chem.* **1994**, *59*, 291-296.
- [129] N. Muller, *J. Fluorine Chem.* **1987**, *36*, 163-170.
- [130] L. Panella, A. M. Aleixandre, G. J. Kruidhof, J. Robertus, B. L. Feringa, J. G. de Vries, A. J. Minnaard, *J. Org. Chem.* **2006**, *71*, 2026-2036.
- [131] D. D. Young, J. Torres-Kolbus, A. Deiters, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5478-5480.
- [132] D. K. Black, S. R. Landor, A. N. Patel, P. F. Whiter, *J. Chem. Soc. C* **1967**, 2260-2262.
- [133] H. Hasegawa, S. Arai, Y. Shinohara, S. Baba, *J. Chem. Soc. Perk. T. 1* **1993**, 489-494.
- [134] L. Zhang, G. Gellerstedt, *Magn. Reson. Chem.* **2007**, *45*, 37-45.
- [135] a) X. M. Zhu, R. R. Schmidt, *Tetrahedron Lett.* **2003**, *44*, 6063-6067; b) X. M. Zhu, R. R. Schmidt, *Chem.-Eur. J.* **2004**, *10*, 875-887; c) K. Pachamuthu, X. M. Zhu, R. R. Schmidt, *J. Org. Chem.* **2005**, *70*, 3720-3723.
- [136] a) L. Yu, WO 9746248 A1, **1997**; b) P. Alewood, M. Muttenthaler, Z. Dekan, WO 2011120071 A1, **2011**; c) A. D. de Araujo, M. Mobli, G. F. King, P. F. Alewood, *Angew. Chem., Int. Ed.* **2012**, *51*, 10298-10302; d) A. D. de Araujo, M. Mobli, J. Castro, A. M. Harrington, I. Vetter, Z. Dekan, M. Muttenthaler, J. Wan, R. J. Lewis, G. F. King, S. M. Brierley, P. F. Alewood, *Nat. Commun.* **2014**, *5*:3165.
- [137] a) X.-H. Chen, Z. Xiang, Y. S. Hu, V. K. Lacey, H. Cang, L. Wang, *Acs Chem. Biol.* **2014**, *9*, 1956-1961; b) Z. Xiang, V. K. Lacey, H. Ren, J. Xu, D. J. Burban, P. A. Jennings, L. Wang, *Angew. Chem., Int. Ed.* **2014**, *53*, 2190-2193.
- [138] J. P. Tam, Y. A. Lu, C. F. Liu, J. Shao, *P. Nat. Acad. Sci. USA* **1995**, *92*, 12485-12489.
- [139] H. Teramoto, K. Kojima, *Macromol. Biosci.* **2015**, *15*, 719-727.
- [140] K. Kawamura, T. Yamada, K. Kurihara, T. Tamada, R. Kuroki, I. Tanaka, H. Takahashi, N. Niimura, *Acta Crystallogr. D* **2011**, *67*, 140-148.
- [141] P. Soumillion, J. Fastrez, *Protein Eng.* **1998**, *11*, 213-217.
- [142] F. Bergmann, M. Dieckmann, P. Berg, *J. Biol. Chem.* **1961**, *236*, 1735-1740.
- [143] C. B. Marks, M. Vasser, P. Ng, W. Henzel, S. Anderson, *J. Biol. Chem.* **1986**, *261*, 7115-7118.
- [144] a) M. M. Botvinik, S. M. Avaeva, E. A. Mistryukov, *Zh. Obshch. Khim.* **1954**, *24*, 2084-2091; b) I. Photaki, V. Bardakos, *Chem. Commun.* **1966**, 818-818; c) M. Wilchek, C. Zioudrou, Patchornik A., *J. Org. Chem.* **1966**, *31*, 2865-2867; d) A. Srinivasan, R. W.

- Stephenson, R. K. Olsen, *J. Org. Chem.* **1977**, *42*, 2253-2256; e) G. Baschang, F. M. Dietrich, R. Gisler, A. Hartmann, J. Stanek, L. Tarcsay, EP5682A1, **1979**; f) I. Photaki, S. Caranikas, I. Samouilidis, L. Zervas, *J. Chem. Soc.–Perk. T. 1* **1980**, 1965-1970; g) M. Takasaki, K. Harada, *Chem. Lett.* **1984**, 1745-1746; h) J. Burton, US 4455303 A, **1984**; i) E. A. Hagen, T. Bergan, A. J. Aasen, *Acta Chem. Scand. B* **1984**, *38*, 5-14; j) R. F. Maes, EP 142387 A1, **1985**; k) K. S. Cheung, W. Boisvert, S. A. Lerner, M. Johnston, *J. Med. Chem.* **1986**, *29*, 2060-2068; l) M. Takasaki, K. Harada, *J. Chem. Soc. Chem. Comm.* **1987**, 571-573; m) D. Crich, J. W. Davies, *Tetrahedron* **1989**, *45*, 5641-5654; n) D. O. Choi, H. Kohn, *Tetrahedron Lett.* **1995**, *36*, 7011-7014; o) A. S. Murkin, M. E. Tanner, *J. Org. Chem.* **2002**, *67*, 8389-8394; p) D. Gupta, R. S. Yedidi, S. Varghese, L. C. Kovari, P. M. Woster, *J. Med. Chem.* **2010**, *53*, 4234-4247; q) R. S. Yedidi, Z. Liu, Y. Wang, J. S. Brunzelle, I. A. Kovari, P. M. Woster, L. C. Kovari, D. Gupta, *Bioche Bioph Res Co* **2012**, *421*, 413-417.
- [145] a) J. U. Bowie, J. F. Reidhaarolson, W. A. Lim, R. T. Sauer, *Science* **1990**, *247*, 1306-1310; b) S. Kamtekar, J. M. Schiffer, H. Y. Xiong, J. M. Babik, M. H. Hecht, *Science* **1993**, *262*, 1680-1685; c) I. Ladunga, R. F. Smith, *Protein Eng.* **1997**, *10*, 187-196; d) D. S. Riddle, J. V. Santiago, S. T. BrayHall, N. Doshi, V. P. Grantcharova, Q. Yi, D. Baker, *Nat. Struct. Biol.* **1997**, *4*, 805-809; e) C. E. Schafmeister, S. L. LaPorte, L. J. W. Miercke, R. M. Stroud, *Nat. Struct. Biol.* **1997**, *4*, 1039-1046; f) L. R. Murphy, A. Wallqvist, R. M. Levy, *Protein Eng.* **2000**, *13*, 149-152; g) T. P. Li, K. Fan, J. Wang, W. Wang, *Protein Eng.* **2003**, *16*, 323-330; h) L. Y. Yampolsky, A. Stoltzfus, *Genetics* **2005**, *170*, 1459-1472; i) C. Etchebest, C. Benros, A. Bornot, A. C. Camproux, A. G. de Brevern, *Eur. Biophys. J.* **2007**, *36*, 1059-1069; j) G. Zhang, K. Wang, B. Zheng, M. Cheng, Y. Li, K. Liu, L. Cai, *Chem. Commun.* **2013**, *49*, 11086-11088; k) S. Hormoz, *Sci. Rep.* **2013**, *3*.
- [146] T. Krick, N. Verstraete, L. G. Alonso, D. A. Shub, D. U. Ferreira, M. Shub, I. E. Sanchez, *Mol. Biol. Evol.* **2014**, *31*, 2905-2912.
- [147] M. E. Hepperle, D. A. Campbell, D. T. Winn, J. M. Betancort, WO2009102876A1, **2009**.
- [148] N. M. Howarth, L. P. G. Wakelin, *J. Org. Chem.* **1997**, *62*, 5441-5450.
- [149] M. Garcia, A. Serra, M. Rubiralta, A. Diez, V. Segarra, E. Lozoya, H. Ryder, J. M. Palacios, *Tetrahedron-Asymmetr.* **2000**, *11*, 991-994.
- [150] V. Denniel, P. Bauchat, D. Danion, R. Danion-Bougot, *Tetrahedron Lett.* **1996**, *37*, 5111-5114.
- [151] a) K. S. C. Reid, P. F. Lindley, J. M. Thornton, *FEBS Lett* **1985**, *190*, 209-213; b) C. C. Valley, A. Cembran, J. D. Perlmutter, A. K. Lewis, N. P. Labello, J. Gao, J. N. Sachs, *J. Biol. Chem.* **2012**, *287*, 34979-34991.
- [152] C. Ducho, R. B. Hamed, E. T. Batchelar, J. L. Sorensen, B. Odell, C. J. Schofield, *Org. Biomol. Chem.* **2009**, *7*, 2770-2779.
- [153] N. K. Jana, J. G. Verkade, *Org. Lett.* **2003**, *5*, 3787-3790.
- [154] K. Yamashita, K. Inoue, K. Kinoshita, Y. Ueda, H. Murao, WO 1999/033785 A1, **1999**.
- [155] N. Valls, M. Borregan, J. Bonjoch, *Tet. Lett.* **2006**, *47*, 3701-3705.
- [156] Y. Ozaki, S. Maeda, M. Miyoshi, K. Matsumoto, *Synthesis-Stuttgart* **1979**, 216-217.
- [157] a) M. Jacob, M. L. Roumestant, P. Viallefont, J. Martinez, *Synlett* **1997**, 691-692; b) M. D. Swift, A. Sutherland, *Org. Biomol. Chem.* **2006**, *4*, 3889-3891.
- [158] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr. D* **2011**, *67*, 235-242.
- [159] P. Emsley, K. Cowtan, *Acta Crystallogr. D* **2004**, *60*, 2126-2132.
- [160] A. W. Schuttelkopf, D. M. F. van Aalten, *Acta Crystallogr. D* **2004**, *60*, 1355-1363.

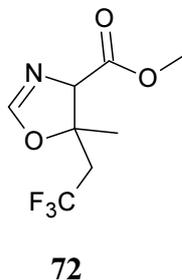
Appendix

A1. Spectroscopic characterisation of new compounds



Elemental Composition Report

Page 1



Single Mass Analysis

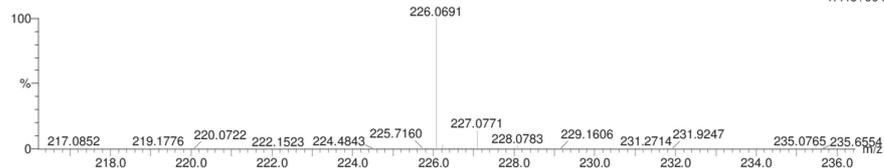
Tolerance = 3.0 PPM / DBE: min = -1.5, max = 10.0
Element prediction: Off
Number of isotope peaks used for i-FIT = 3

Monoisotopic Mass, Even Electron Ions
368 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)
Elements Used:
C: 0-50 H: 0-50 N: 0-5 O: 0-5 F: 0-3
ML 056/AJ
42580
1845 92 (4.026) Cm (91:92)

KE375

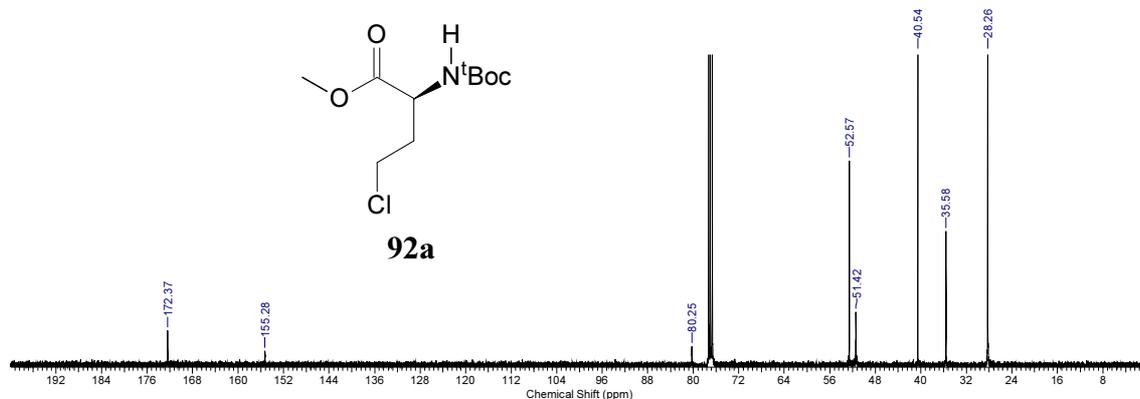
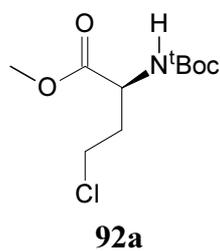
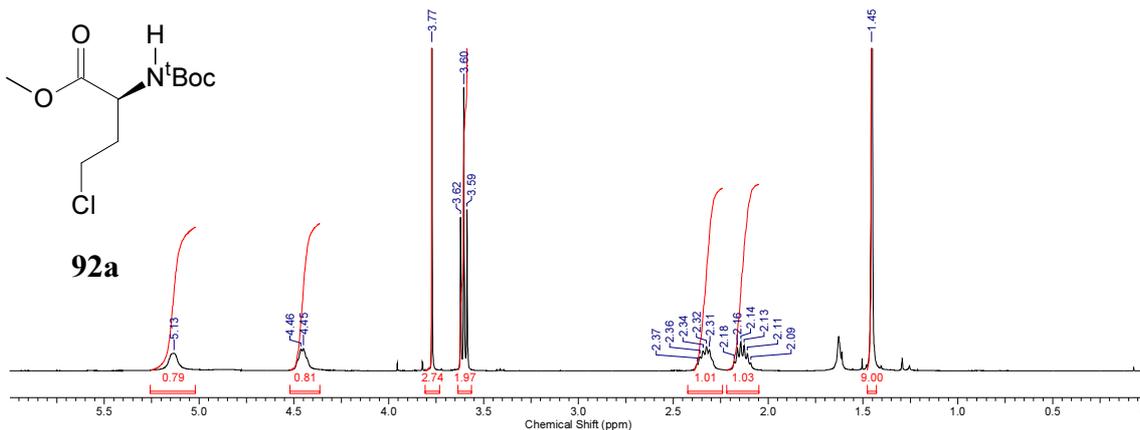
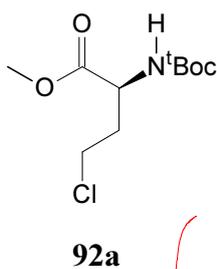
12-Oct-2015 12:05:48

1: TOF MS ES+
1.41e+004



Minimum: -1.5
Maximum: 5.0 3.0 10.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
226.0691	226.0691	0.0	0.0	2.5	113.6	C8 H11 N O3 F3



Elemental Composition Report

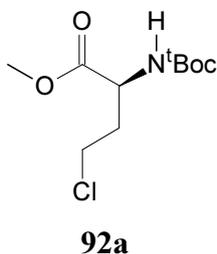
Page 1

Single Mass Analysis

Tolerance = 3.0 PPM / DBE: min = -1.5, max = 20.0

Element prediction: Off

Number of isotope peaks used for i-FIT = 3



Monoisotopic Mass, Odd and Even Electron Ions
18 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-10 H: 0-20 N: 0-1 O: 4-4 23Na: 0-1 35Cl: 0-1 37Cl: 0-1

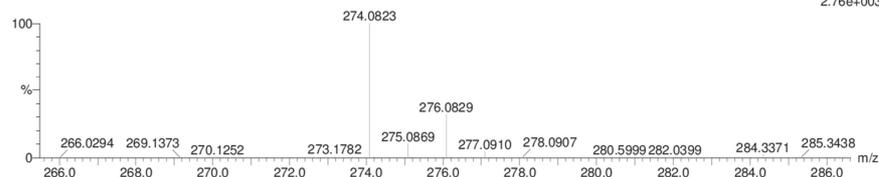
ML210EO

33726

1087.78 (3.433) Cm (78:80)

18-Jul-2014 1:7:9

1: TOF MS ES+
2.76e+003



Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
274.0823	274.0822	0.1	0.4	1.5	405.0	C10 H18 N O4 23Na 35Cl

Elemental Composition Report

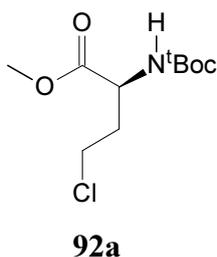
Page 1

Single Mass Analysis

Tolerance = 3.0 PPM / DBE: min = -1.5, max = 20.0

Element prediction: Off

Number of isotope peaks used for i-FIT = 3



Monoisotopic Mass, Odd and Even Electron Ions
18 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-10 H: 0-20 N: 0-1 O: 4-4 23Na: 0-1 35Cl: 0-1 37Cl: 0-1

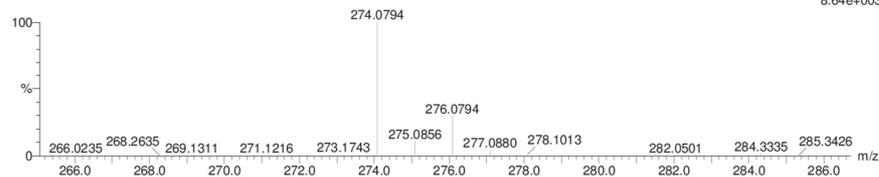
ML210EO

33726

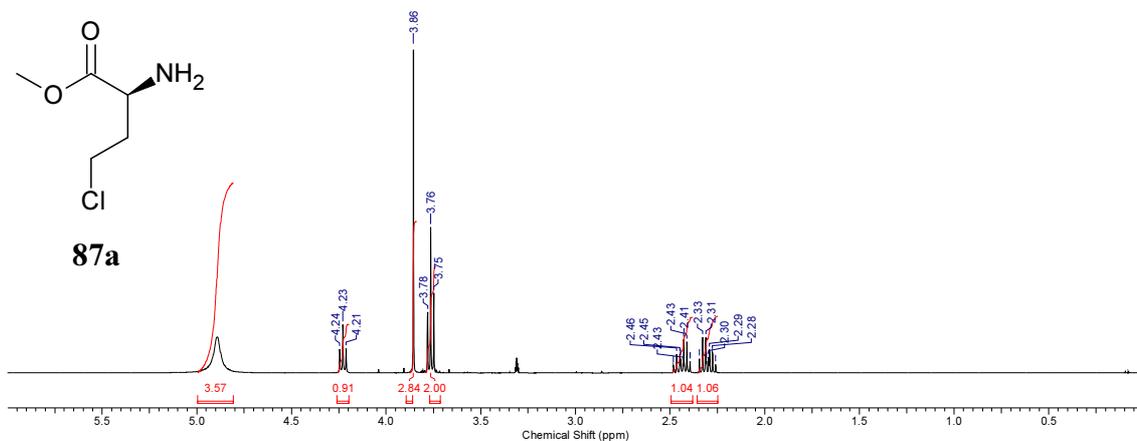
1087.76 (3.326) Cm (72:80)

18-Jul-2014 1:7:9

1: TOF MS ES+
8.64e+003

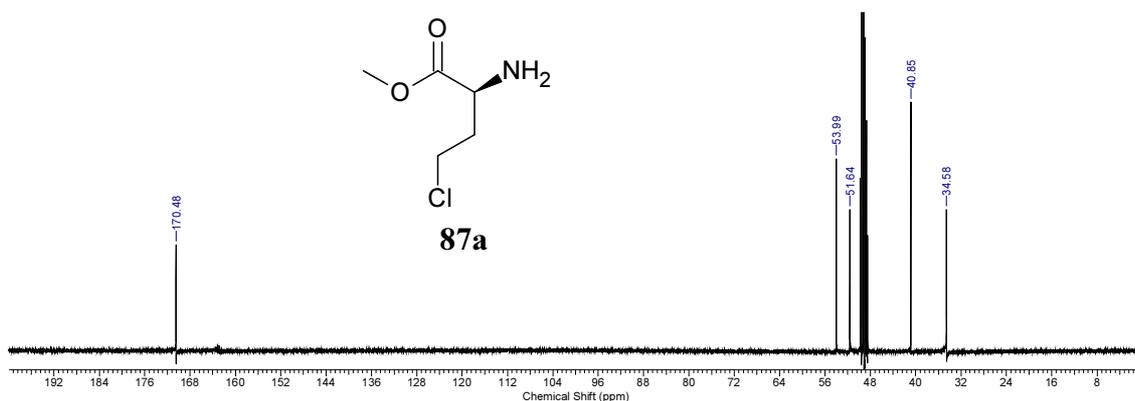


Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
276.0794	276.0793	0.1	0.4	1.5	n/a	C10 H18 N O4 23Na 37Cl



Chlorinated Amino Acids in Peptide Production

Appendix

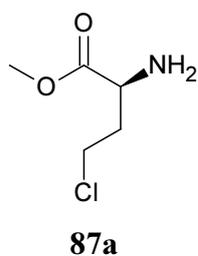


Elemental Composition Report

Page 1

Single Mass Analysis

Tolerance = 3.0 PPM / DBE: min = -1.5, max = 10.0
 Element prediction: Off
 Number of isotope peaks used for i-FIT = 3



Monoisotopic Mass, Even Electron Ions

317 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-50 H: 0-60 N: 0-5 O: 0-10 35Cl: 0-5 37Cl: 0-5

ML246a/AJ

KE375

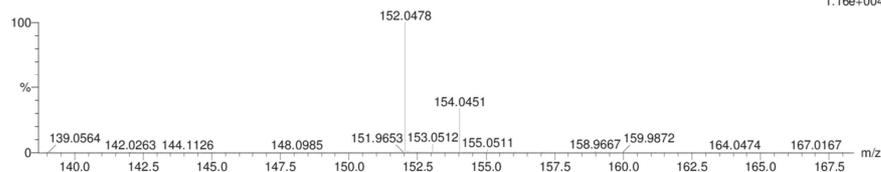
28-Jan-2015 10:19:09

36864

0143 13 (0.596) Cm (7:25)

1: TOF MS ES+

1.16e+004



Minimum:

Maximum: 5.0 3.0 -1.5

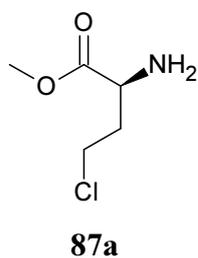
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
152.0478	152.0478	0.0	0.0	0.5	1907.8	C5 H11 N O2 35Cl

Elemental Composition Report

Page 1

Single Mass Analysis

Tolerance = 3.0 PPM / DBE: min = -1.5, max = 10.0
 Element prediction: Off
 Number of isotope peaks used for i-FIT = 3



Monoisotopic Mass, Even Electron Ions

324 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-50 H: 0-60 N: 0-5 O: 0-10 35Cl: 0-5 37Cl: 0-5

ML246a/AJ

KE375

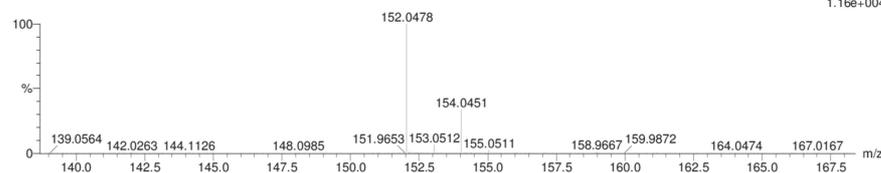
28-Jan-2015 10:19:09

36864

0143 13 (0.596) Cm (7:25)

1: TOF MS ES+

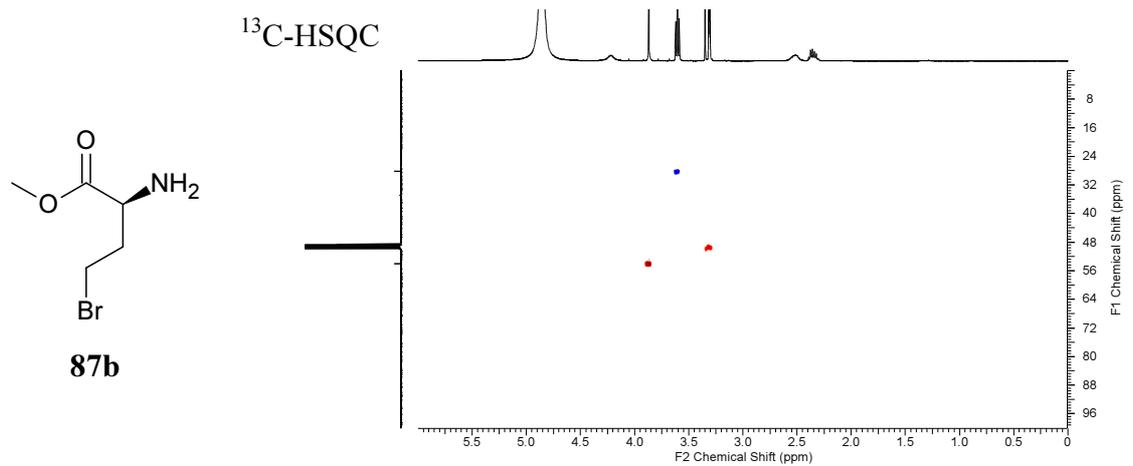
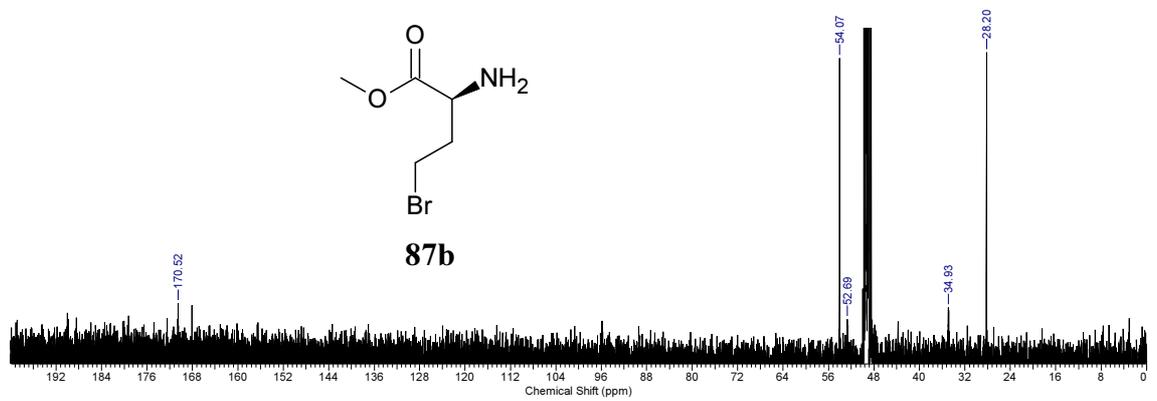
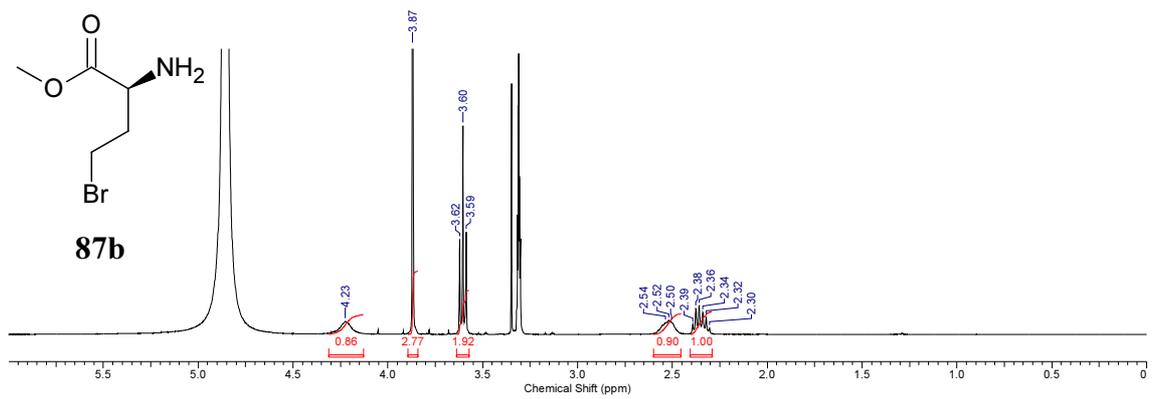
1.16e+004



Minimum:

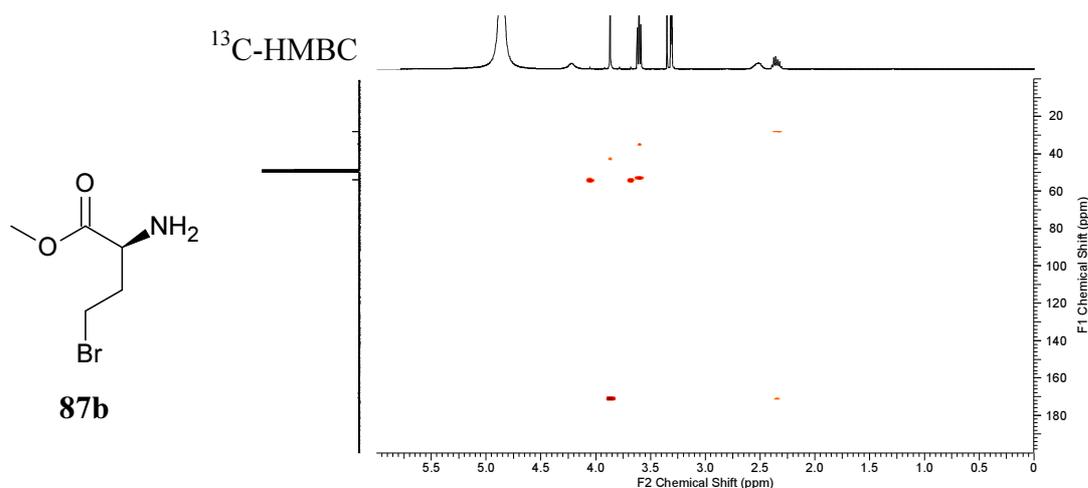
Maximum: 5.0 3.0 -1.5

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
154.0451	154.0449	0.2	1.3	0.5	n/a	C5 H11 N O2 37Cl



Chlorinated Amino Acids in Peptide Production

Appendix



Elemental Composition Report

Page 1

Single Mass Analysis

Tolerance = 3.0 PPM / DBE: min = -1.5, max = 10.0

Element prediction: Off

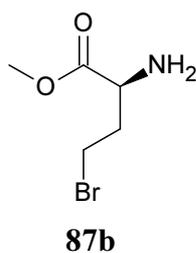
Number of isotope peaks used for i-FIT = 3

Monoisotopic Mass, Even Electron Ions

268 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-50 H: 0-60 N: 0-5 O: 0-10 79Br: 0-5 81Br: 0-5



ML242a/AJ

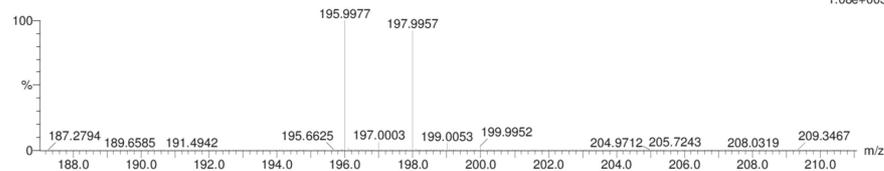
36863

0142.24 (1.048) Cm (22:24)

KE375

28-Jan-2015 10:10:38

1: TOF MS ES+
1.08e+003



Minimum: -1.5
Maximum: 5.0 3.0 10.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
195.9977	195.9973	0.4	2.0	0.5	493.7	C5 H11 N O2 79Br

Elemental Composition Report

Page 1

Single Mass Analysis

Tolerance = 3.0 PPM / DBE: min = -1.5, max = 10.0

Element prediction: Off

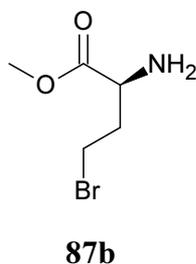
Number of isotope peaks used for i-FIT = 3

Monoisotopic Mass, Even Electron Ions

266 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-50 H: 0-60 N: 0-5 O: 0-10 79Br: 0-5 81Br: 0-5



ML242a/AJ

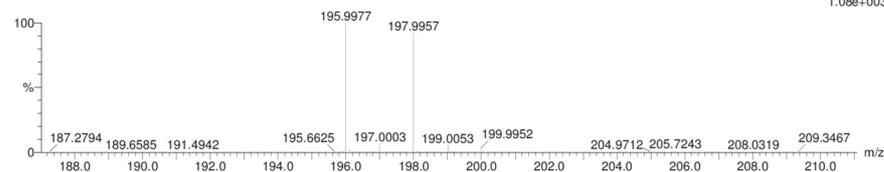
36863

0142.24 (1.048) Cm (22:24)

KE375

28-Jan-2015 10:10:38

1: TOF MS ES+
1.08e+003



Minimum: -1.5
Maximum: 5.0 3.0 10.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
197.9957	197.9953	0.4	2.0	0.5	12.1	C5 H11 N O2 81Br

A2. X-Ray data and structure validation reports

The disk attached to the back cover of this Thesis contains the pdb and mtz files of the crystal structures **C1**, **C2** and **C3** described in Chapter 4.

The following pages have the validation reports for the crystal structures **C1**, **C2** and **C3**.

Structure C1



Preliminary Full wwPDB X-ray Structure Validation Report ⓘ

Aug 10, 2015 – 04:26 AM EDT

DISCLAIMER

This is a preliminary version of the new style of wwPDB validation report.
This report is produced by the wwPDB validation pipeline
before deposition or annotation of the structure.
This is not an official wwPDB validation report and is not a proof of deposition.
This report should not be submitted to journals.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

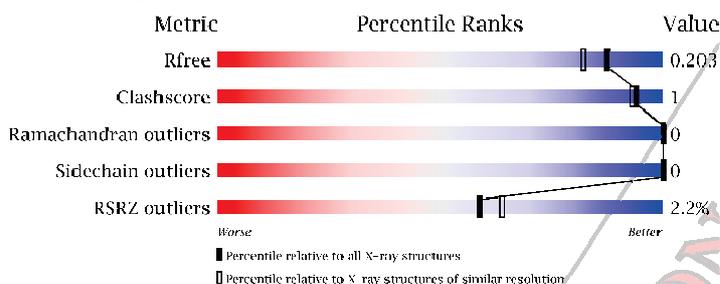
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.17 November 2013
Xtrriage (Phenix)	:	dev-1439
EDS	:	stable24195
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable24195

1 Overall quality at a glance (i)

The reported resolution of this entry is 1.70 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	2456 (1.70-1.70)
Clashscore	79885	2929 (1.70-1.70)
Ramachandran outliers	78287	2878 (1.70-1.70)
Sidechain outliers	78261	2878 (1.70-1.70)
RSRZ outliers	66119	2456 (1.70-1.70)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	E	223	
2	I	56	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	SO4	S	4	-	X
3	SO4	S	5	-	X

2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 2430 atoms, of which 0 are hydrogen and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	E	223	Total 1637	1017	279	327	14	0	2	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	I	56	Total 447	281	80	78	8	0	2	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: unknown).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
3	S	1	Total 5	4	1	0	0
3	S	1	Total 5	4	1	0	0
3	S	1	Total 5	4	1	0	0
3	S	1	Total 5	4	1	0	0
3	S	1	Total 5	4	1	0	0

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: unknown).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
4	C	1	Total 1	1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	W	320	Total 320	O 320	0	1

PRELIMINARY VALIDATION REPORT

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1:



- Molecule 2:



4 Data and refinement statistics 

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α , β , γ	74.56Å 80.60Å 124.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	67.64 – 1.70 34.09 – 1.70	Depositor EDS
% Data completeness (in resolution range)	96.3 (67.64-1.70) 96.3 (34.09-1.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.59 (at 1.70Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.154 , 0.190 0.169 , 0.203	Depositor DCC
R_{free} test set	1997 reflections (5.27%)	DCC
Wilson B-factor (Å ²)	18.8	Xtriage
Anisotropy	0.067	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.42 , 40.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 39890 reflections	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	2430	wwPDB-VP
Average B, all atoms (Å ²)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.79% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CA, SO4

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	E	1.16	0/1674	1.09	7/2269 (0.3%)
2	I	1.25	0/463	1.34	5/618 (0.8%)
All	All	1.18	0/2137	1.15	12/2887 (0.4%)

There are no bond length outliers.

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	I	53	ARG	NE-CZ-NH2	-10.10	115.25	120.30
2	I	50	ASP	CB-CG-OD1	7.84	125.36	118.30
1	E	194	ASP	CB-CG-OD1	7.75	125.28	118.30
2	I	53	ARG	NE-CZ-NH1	6.99	123.80	120.30
1	E	72	ARG	NE-CZ-NH2	-6.44	117.08	120.30
2	I	20	ARG	NE-CZ-NH1	6.13	123.36	120.30
1	E	168[A]	ASP	CB-CG-OD2	-6.09	112.82	118.30
1	E	168[B]	ASP	CB-CG-OD2	-6.09	112.82	118.30
1	E	194	ASP	CB-CG-OD2	-5.42	113.42	118.30
1	E	87	PHE	CB-CG-CD1	5.41	124.59	120.80
1	E	242	THR	CA-CB-CG2	-5.12	105.23	112.40
2	I	3	ASP	CB-CG-OD1	5.09	122.89	118.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit,

and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	1637	0	1597	6	0
2	I	447	0	432	1	0
3	S	25	0	0	1	0
4	C	1	0	0	0	0
5	W	320	0	0	0	0
All	All	2430	0	2029	6	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 1.

All (6) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:E:172:LYS:HE3	3:S:5:SO4:O2	2.15	0.46
1:E:239:ILE:O	1:E:243:ILE:HG12	2.19	0.43
1:E:243:ILE:HD13	1:E:243:ILE:HG21	1.76	0.42
1:E:35:VAL:HG13	1:E:37:TYR:CZ	2.55	0.41
1:E:62:ALA:HB1	1:E:95:VAL:HG13	2.02	0.41
1:E:200:SER:OG	2:I:15:LYS:C	2.59	0.40

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	E	223/223 (100%)	221 (99%)	2 (1%)	0	100	100
2	I	56/56 (100%)	55 (98%)	1 (2%)	0	100	100
All	All	279/279 (100%)	276 (99%)	3 (1%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	186/184 (101%)	186 (100%)	0	100	100
2	I	46/44 (104%)	46 (100%)	0	100	100
All	All	232/228 (102%)	232 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	I	24	ASN

5.3.3 RNA [i](#)

There are no RNA chains in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 1 is modelled with single atom - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected

value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	SO4	S	1	-	4,?,?	0.56	0	6,?,?	0.24	0
3	SO4	S	2	-	4,?,?	0.10	0	6,?,?	0.72	0
3	SO4	S	3	-	4,?,?	0.70	0	6,?,?	0.86	0
3	SO4	S	4	-	4,?,?	0.58	0	6,?,?	0.95	0
3	SO4	S	5	-	4,?,?	0.16	0	6,?,?	0.33	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	SO4	S	1	-	-	0/0/?/?	0/0/?/?
3	SO4	S	2	-	-	0/0/?/?	0/0/?/?
3	SO4	S	3	-	-	0/0/?/?	0/0/?/?
3	SO4	S	4	-	-	0/0/?/?	0/0/?/?
3	SO4	S	5	-	-	0/0/?/?	0/0/?/?

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	E	223/223 (100%)	-0.13	5 (2%) 59 64	13, 19, 30, 45	0
2	I	56/56 (100%)	-0.39	1 (1%) 65 71	12, 16, 28, 50	0
All	All	279/279 (100%)	-0.19	6 (2%) 59 64	12, 19, 30, 50	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	121	SER	4.1
1	E	120	ASN	3.1
1	E	122	ARG	3.0
1	E	83	GLY	2.8
2	I	26	LYS	2.3
1	E	84	ASN	2.3

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(\AA^2)	Q<0.9
3	SO4	S	4	5/?	0.11	10.92	38,40,40,45	0
3	SO4	S	5	5/?	0.17	4.47	46,49,60,61	0
3	SO4	S	2	5/?	0.08	-0.65	17,19,23,24	0
4	CA	C	1	1/?	0.04	-1.01	21,21,21,21	0
3	SO4	S	1	5/?	0.04	-1.29	16,17,18,20	0
3	SO4	S	3	5/?	0.04	-2.64	26,26,27,27	0

6.5 Other polymers [i](#)

There are no such residues in this entry.

Structure C2



Preliminary Full wwPDB X-ray Structure Validation Report ⓘ

May 5, 2015 – 01:13 AM EDT

DISCLAIMER

This is a preliminary version of the new style of wwPDB validation report.
This report is produced by the wwPDB validation pipeline
before deposition or annotation of the structure.
This is not an official wwPDB validation report and is not a proof of deposition.
This report should not be submitted to journals.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

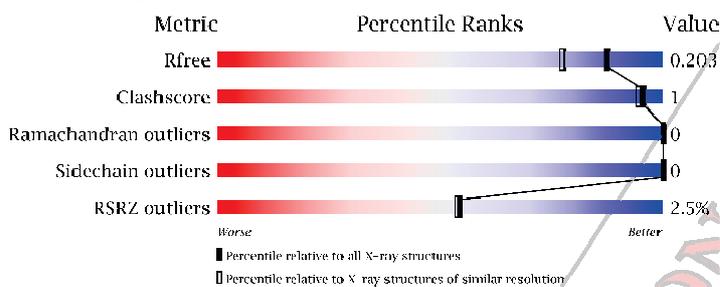
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.17 November 2013
Xtriage (Phenix) : dev-1439
EDS : stable24195
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable24195

1 Overall quality at a glance i

The reported resolution of this entry is 1.60 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	1872 (1.60-1.60)
Clashscore	79885	2199 (1.60-1.60)
Ramachandran outliers	78287	2126 (1.60-1.60)
Sidechain outliers	78261	2125 (1.60-1.60)
RSRZ outliers	66119	1872 (1.60-1.60)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	E	223	
2	I	56	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	SO4	S	4	-	X
3	SO4	S	5	-	X

2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 2427 atoms, of which 0 are hydrogen and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	E	223	1638	1018	279	327	14	0	3	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	Cl	N	O				S
2	I	56	447	280	1	80	78	8	0	2	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: unknown).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	S	1	Total	O	S	0	0
			5	4	1		
3	S	1	Total	O	S	0	0
			5	4	1		
3	S	1	Total	O	S	0	0
			5	4	1		
3	S	1	Total	O	S	0	0
			5	4	1		
3	S	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: unknown).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	1	Total	Ca	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	W	316	Total 316	O 316	0	1

PRELIMINARY VALIDATION REPORT

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1:



- Molecule 2:



4 Data and refinement statistics i

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α , β , γ	74.63Å 80.99Å 124.13Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	67.83 – 1.60 32.69 – 1.60	Depositor EDS
% Data completeness (in resolution range)	100.0 (67.83-1.60) 100.0 (32.69-1.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$	-	Xtrriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.163 , 0.193 0.176 , 0.203	Depositor DCC
R_{free} test set	2520 reflections (5.29%)	DCC
Wilson B-factor (Å ²)	(Not available)	Xtrriage
Anisotropy	(Not available)	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.42 , 44.9	EDS
Estimated twinning fraction	No twinning to report.	Xtrriage
L-test for twinning	$\langle L \rangle =$ (Not available), $\langle L^2 \rangle =$ (Not available)	Xtrriage
Outliers	(Not available)	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	2427	wwPDB-VP
Average B, all atoms (Å ²)	19.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: (Not available)

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CA, SO4, SCB

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	E	1.29	3/1678 (0.2%)	1.23	9/2274 (0.4%)
2	I	1.38	1/455 (0.2%)	1.49	4/605 (0.7%)
All	All	1.31	4/2133 (0.2%)	1.29	13/2879 (0.5%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	175	TYR	CE2-CZ	5.73	1.46	1.38
1	E	154	TYR	CE1-CZ	-5.35	1.31	1.38
2	I	21	TYR	CE2-CZ	-5.11	1.31	1.38
1	E	169	SER	CA-CB	5.03	1.60	1.52

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	I	53	ARG	NE-CZ-NH2	-12.78	113.91	120.30
1	E	72	ARG	NE-CZ-NH2	-9.71	115.45	120.30
2	I	53	ARG	NE-CZ-NH1	9.49	125.05	120.30
1	E	194	ASP	CB-CG-OD1	8.76	126.18	118.30
1	E	45	TYR	CB-CG-CD1	7.95	125.77	121.00
1	E	87	PHE	CB-CG-CD1	7.73	126.21	120.80
2	I	20	ARG	NE-CZ-NH1	6.74	123.67	120.30
2	I	50	ASP	CB-CG-OD1	6.02	123.72	118.30
1	E	72	ARG	NE-CZ-NH1	5.87	123.23	120.30
1	E	119	LEU	CB-CG-CD1	5.83	120.91	111.00
1	E	45	TYR	CZ-CE2-CD2	5.47	124.72	119.80
1	E	78	ILE	CG1-CB-CG2	5.45	123.38	111.40
1	E	184	PHE	CB-CG-CD1	5.31	124.52	120.80

There are no chirality outliers.

There are no planarity outliers.

5.2 Close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	1638	0	1603	4	0
2	I	447	0	423	2	0
3	S	25	0	0	0	0
4	C	1	0	0	0	0
5	W	316	0	0	1	0
All	All	2427	0	2026	5	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 1.

All (5) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:I:24:ASN:HD22	2:I:31[A]:GLN:NE2	2.05	0.54
1:E:130:THR:HG23	5:W:315:HOH:O	2.11	0.51
1:E:62:ALA:HB1	1:E:95:VAL:HG13	1.97	0.45
1:E:243:ILE:HD13	1:E:243:ILE:HG21	1.62	0.44
1:E:200:SER:OG	2:I:15:LYS:C	2.59	0.40

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	E	224/223 (100%)	221 (99%)	3 (1%)	0	100	100
2	I	55/56 (98%)	53 (96%)	2 (4%)	0	100	100
All	All	279/279 (100%)	274 (98%)	5 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	187/184 (102%)	187 (100%)	0	100	100
2	I	45/43 (105%)	45 (100%)	0	100	100
All	All	232/227 (102%)	232 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA chains in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 1 is modelled with single atom - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	SO4	S	1	-	4,?,?	0.60	0	6,?,?	0.33	0
3	SO4	S	2	-	4,?,?	0.12	0	6,?,?	0.69	0
3	SO4	S	3	-	4,?,?	0.87	0	6,?,?	1.12	0
3	SO4	S	4	-	4,?,?	0.68	0	6,?,?	0.96	0
3	SO4	S	5	-	4,?,?	0.15	0	6,?,?	0.58	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	SO4	S	1	-	-	0/0/?/?	0/0/?/?
3	SO4	S	2	-	-	0/0/?/?	0/0/?/?
3	SO4	S	3	-	-	0/0/?/?	0/0/?/?
3	SO4	S	4	-	-	0/0/?/?	0/0/?/?
3	SO4	S	5	-	-	0/0/?/?	0/0/?/?

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	E	223/223 (100%)	-0.21	6 (2%) 52 51	10, 17, 28, 52	0
2	I	55/56 (98%)	-0.46	1 (1%) 65 65	10, 13, 25, 40	0
All	All	278/279 (99%)	-0.26	7 (2%) 54 54	10, 16, 28, 52	0

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	121	SER	4.5
1	E	122	ARG	3.5
1	E	120	ASN	3.1
2	I	57	GLY	2.6
1	E	83	GLY	2.5
1	E	84	ASN	2.3
1	E	130	THR	2.1

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(\AA^2)	Q<0.9
3	SO4	S	4	5/?	0.11	9.52	30,35,38,39	0
3	SO4	S	5	5/?	0.23	6.57	44,45,56,56	0
3	SO4	S	3	5/?	0.08	1.96	27,27,28,32	0
3	SO4	S	1	5/?	0.05	-0.10	16,16,17,19	0
3	SO4	S	2	5/?	0.06	-0.47	15,16,20,22	0
4	CA	C	1	1/?	0.04	-1.09	19,19,19,19	0

6.5 Other polymers [i](#)

There are no such residues in this entry.

PRELIMINARY VALIDATION REPORT

Structure C3



Preliminary Full wwPDB X-ray Structure Validation Report ⓘ

Aug 10, 2015 – 02:13 AM EDT

DISCLAIMER

This is a preliminary version of the new style of wwPDB validation report.
This report is produced by the wwPDB validation pipeline
before deposition or annotation of the structure.
This is not an official wwPDB validation report and is not a proof of deposition.
This report should not be submitted to journals.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

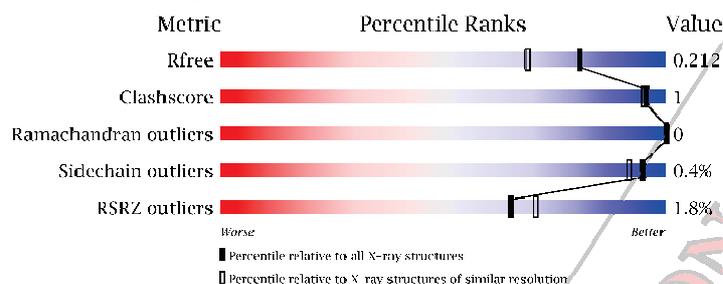
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.17 November 2013
Xtriage (Phenix) : dev-1439
EDS : stable24195
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable24195

1 Overall quality at a glance i

The reported resolution of this entry is 1.70 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	2456 (1.70-1.70)
Clashscore	79885	2929 (1.70-1.70)
Ramachandran outliers	78287	2878 (1.70-1.70)
Sidechain outliers	78261	2878 (1.70-1.70)
RSRZ outliers	66119	2456 (1.70-1.70)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	E	223	
2	I	56	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	SO4	S	4	-	X
3	SO4	S	5	-	X

2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 2449 atoms, of which 0 are hydrogen and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	E	223	1637	1017	279	327	14	0	2	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	Cl	N	O				S
2	I	56	447	280	1	80	78	8	0	2	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: unknown).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	S	1	Total	O	S	0	0
			5	4	1		
3	S	1	Total	O	S	0	0
			5	4	1		
3	S	1	Total	O	S	0	0
			5	4	1		
3	S	1	Total	O	S	0	0
			5	4	1		
3	S	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: unknown).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	1	Total	Ca	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	W	339	Total 339	O 339	0	1

PRELIMINARY VALIDATION REPORT

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1:



- Molecule 2:



4 Data and refinement statistics 

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α , β , γ	74.85Å 80.72Å 124.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	67.68 – 1.70 27.44 – 1.70	Depositor EDS
% Data completeness (in resolution range)	99.1 (67.68-1.70) 99.1 (27.44-1.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$	-	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.163 , 0.201 0.176 , 0.212	Depositor DCC
R_{free} test set	2071 reflections (5.30%)	DCC
Wilson B-factor (Å ²)	(Not available)	Xtriage
Anisotropy	(Not available)	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.42 , 45.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle =$ (Not available), $\langle L^2 \rangle =$ (Not available)	Xtriage
Outliers	(Not available)	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	2449	wwPDB-VP
Average B, all atoms (Å ²)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: (Not available)

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CA, SO4, RCB

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	E	1.02	0/1674	0.93	4/2269 (0.2%)
2	I	1.15	2/455 (0.4%)	1.21	5/605 (0.8%)
All	All	1.05	2/2129 (0.1%)	1.00	9/2874 (0.3%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	I	49	GLU	CD-OE1	5.51	1.31	1.25
2	I	22	PHE	CG-CD2	5.02	1.46	1.38

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	I	53	ARG	NE-CZ-NH2	-9.85	115.37	120.30
2	I	3	ASP	CB-CG-OD1	8.52	125.97	118.30
1	E	194	ASP	CB-CG-OD2	-8.23	110.89	118.30
1	E	194	ASP	CB-CG-OD1	7.74	125.26	118.30
2	I	20	ARG	NE-CZ-NH1	6.57	123.58	120.30
2	I	53	ARG	NE-CZ-NH1	6.24	123.42	120.30
1	E	72	ARG	NE-CZ-NH2	-5.78	117.41	120.30
2	I	3	ASP	CB-CG-OD2	-5.62	113.24	118.30
1	E	48	CYS	CA-CB-SG	5.16	123.29	114.00

There are no chirality outliers.

There are no planarity outliers.

5.2 Close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	1637	0	1597	4	0
2	I	447	0	423	2	0
3	S	25	0	0	0	0
4	C	1	0	0	0	0
5	W	339	0	0	2	0
All	All	2449	0	2020	5	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 1.

All (5) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:E:239:ILE:O	1:E:243:ILE:HG12	2.16	0.45
2:I:24:ASN:HD22	2:I:31[A]:GLN:CD	2.23	0.42
1:E:120:ASN:ND2	5:W:159:HOH:O	2.53	0.41
1:E:200:SER:OG	2:I:15:LYS:C	2.59	0.41
1:E:243:ILE:HG13	5:W:30:HOH:O	2.21	0.41

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	E	223/223 (100%)	221 (99%)	2 (1%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	I	55/56 (98%)	54 (98%)	1 (2%)	0	100	100
All	All	278/279 (100%)	275 (99%)	3 (1%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	186/184 (101%)	186 (100%)	0	100	100
2	I	45/43 (105%)	44 (98%)	1 (2%)	64	43
All	All	231/227 (102%)	230 (100%)	1 (0%)	95	92

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	I	6	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA chains in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 6 ligands modelled in this entry, 1 is modelled with single atom - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	SO4	S	1	-	4,?,?	0.26	0	6,?,?	0.45	0
3	SO4	S	2	-	4,?,?	0.22	0	6,?,?	0.91	0
3	SO4	S	3	-	4,?,?	0.32	0	6,?,?	0.76	0
3	SO4	S	4	-	4,?,?	0.19	0	6,?,?	0.97	0
3	SO4	S	5	-	4,?,?	0.10	0	6,?,?	0.23	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	SO4	S	1	-	-	0/0/?/?	0/0/?/?
3	SO4	S	2	-	-	0/0/?/?	0/0/?/?
3	SO4	S	3	-	-	0/0/?/?	0/0/?/?
3	SO4	S	4	-	-	0/0/?/?	0/0/?/?
3	SO4	S	5	-	-	0/0/?/?	0/0/?/?

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

PRELIMINARY VALIDATION REPORT

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	E	223/223 (100%)	-0.08	5 (2%) 59 64	15, 22, 34, 60	0
2	I	56/56 (100%)	-0.32	0 100 100	14, 19, 31, 51	0
All	All	279/279 (100%)	-0.13	5 (1%) 65 71	14, 22, 35, 60	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	121	SER	3.9
1	E	120	ASN	3.9
1	E	122	ARG	3.2
1	E	149	SER	2.4
1	E	130	THR	2.1

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(\AA^2)	Q<0.9
3	SO4	S	5	5/?	0.20	4.18	52,54,66,67	0
3	SO4	S	4	5/?	0.07	2.14	41,42,45,47	0
3	SO4	S	2	5/?	0.09	0.28	21,22,27,28	0
3	SO4	S	3	5/?	0.05	-0.47	26,27,27,28	0
3	SO4	S	1	5/?	0.04	-0.88	19,20,21,23	0
4	CA	C	1	1/?	0.02	-1.58	24,24,24,24	0

6.5 Other polymers [i](#)

There are no such residues in this entry.