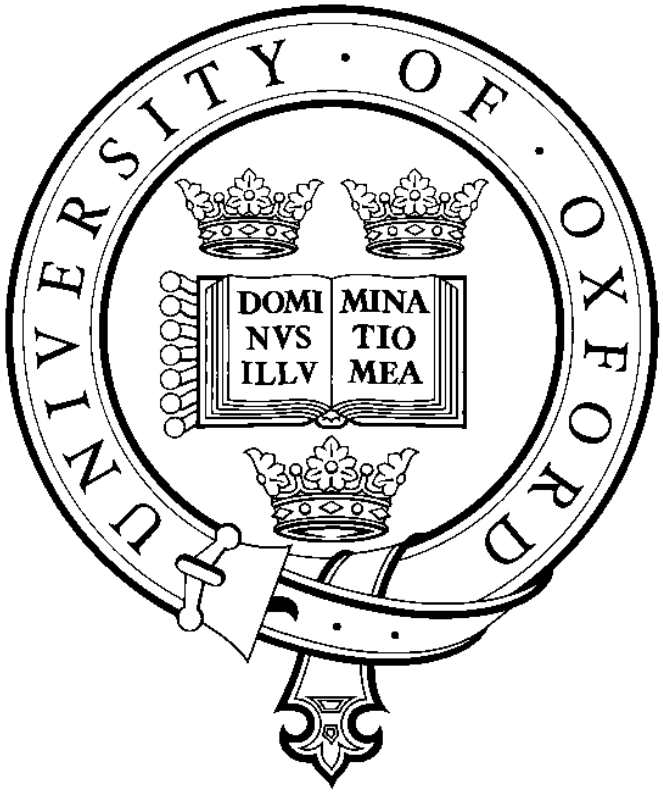
Unnatural Amino Acids as Metal-Mediated Probes of Biological Function



A thesis submitted to the Board of the Faculty of Physical Sciences in partial fulfilment of the degree of

Doctor of Philosophy

at the

University of Oxford

Bhaskar Bhushan

St. Edmund Hall, University of Oxford

Trinity 2014

Dedicated to my family and friends

Thesis Abstract

Unnatural Amino Acids as Metal-Mediated Probes of Biological Function

Bhaskar Bhushan
St. Edmund Hall
University of Oxford
Submitted for the degree of Doctor of Philosophy in Organic Chemistry
Trinity Term 2014

Conjugation reactions on proteins have been used to access various post-translational modifications, for targeted delivery of drugs, for microscopy, and in studying receptor-ligand interactions. However, the ability to modify native proteins is constrained by the reactive functionalities of naturally occurring amino acids. This has driven research into the incorporation of unnatural amino acids (UAAs) into proteins. Research in this area has been motivated both by the possibility of increasing the breadth of chemical techniques for protein modification by introducing novel 'bio-orthogonal' reactive groups via UAA incorporation, as well as generating well-defined conjugates by the site-selective incorporation of these UAAs into proteins. The objective of this thesis is to both expand the diversity of UAAs for access to new metal-mediated reactions on proteins, as well as to utilise these reactions to reveal functional information about a range of biological systems. A brief introduction into current protein conjugation and UAA incorporation methods will be made in Chapter 1.

In Chapter 2, the genetic incorporation of alkene-bearing UAAs into recombinant proteins expressed in both bacterial and mammalian systems is discussed. This technique is demonstrated to enable Ru-catalysed olefin cross-metathesis (CM) reactions on the resultant proteins. This work builds upon previously established methods to chemically incorporate CM handles onto proteins. The rational design of UAAs, as well as assays and modelling studies to screen them for recognition by the cellular incorporation machinery are discussed in detail. The expression of a range of alkene-tagged recombinant proteins, their complete characterisation, as well as the development of a more general protocol for on-protein CM is elucidated.

In Chapter 3, the utility of UAA incorporation to probe mammalian cell translation systems is examined. Incorporation of an azide-bearing UAA, in addition to heavy stable-isotope labelled amino acids is used to uncover a previously unreported system of protein synthesis in mammalian cell nuclei, along with rapid metabolic degradation of the synthesised peptides. Various orthogonal methods for the detection of this system as well as possible reasons for its conservation are discussed.

In Chapter 4, UAA incorporation and metal-mediated bioconjugation reactions are utilised in the development of a novel and generally applicable proteomics technique. This technique is used to determine quantitative changes in cell proteomes in response to external stimuli, and may be applied to systems to which traditional proteomics techniques cannot, such as *ex vivo* primary cells.

Finally, in Chapter 5, further applications of UAA incorporation are discussed. Preliminary results are reported in efforts to use UAAs in the vibrational Raman microscopic imaging of biological systems, in generating HIV vaccines, and inducing T-cell stimulation.

Table of Contents

Acknowledgement.....10

List of Publications.....11

Selection of Abbreviations.....12

Chapter 1: Introduction: Unnatural Amino Acids as Metal-mediated probes of biological function

1.1: Introduction..... 13

1.2: Modification Reactions on Proteins 14

1.3: UAA-enabled Conjugation Reactions on Proteins..... 15

1.4: Azide-alkyne cycloadditions for protein modification..... 16

1.5: Olefin Cross-Metathesis for protein modification 19

1.6: Incorporation of Unnatural Amino Acids into Proteins 23

1.6.1: 'Sense' codon reassignment 23

1.6.2: 'Non-sense' codon reassignment..... 26

1.7: Thesis objective..... 28

1.8 References:..... 29

Chapter 2: Development of the Cross-Metathesis Reaction on Proteins: From Chemical to Genetic Incorporation of Metathesis Reactive Tags

2.1 Introduction..... 36

2.2: Design of UAAs for Genetic Incorporation of CM-reactive tags on proteins 38

2.3: MetRS Activity Assays 41

2.3.1: AP4A continuous colourimetric assay 42

2.3.2: Malachite Green assay 43

2.3.3: Radiometric activity assay 46

2.4 Modelling UAA binding to MetRS 48

2.5: Incorporation trials 50

2.6: Protein function assays 57

2.6.1: Activity assay for SsβG Ahc49 P152A C344S 9Ile 57

2.6.2: Activity assay for SarZΔ(7-142) Ahc20 Ahc64: 57

2.7: CM reaction method optimisation 59

2.8: CM on Ahc-incorporated proteins 63

2.9: CM Challenges 66

2.9.1: Investigation of alternative scavenging conditions 66

2.9.2: Effect of CM conditions on protein function 69

2.10: Conclusions and Discussion 72

2.11: Methods 74

2.11.1: General Consideration 74

2.11.2: Amino acid synthesis 75

2.11.3: MetRS Activity Assays 80

2.11.4: Small Molecule CM Seac Reaction Kinetics 81

2.11.5: SBL Activity Assay 83

2.11.6: "In-gel" Tryptic Digestion-MSMS Analysis of Expressed Proteins 83

2.11.7: General Procedure for Expression of MetRS 84

2.11.8: General Procedure for Expression of Histone H3 K9C M90L C96S G102A C110A Single Ahc120 85

2.11.9: Cross Metathesis on Histone H3 K9C M90L C96S G102A C110A Ahc120 with allyl alcohol 88

2.11.10: General procedure for expression of Qβ-K16M 89

2.11.11: Cross Metathesis on Qβ-K16Ahc with allyl alcohol 92

2.11.12: General Procedure for Expression of NP276 and SsβG 92

2.11.13:Cross Metathesis on Np276 G2P M21I Ahc61 with allyl alcohol 95

2.11.14: General Procedure for Expression of IgG-Fc 97

2.11.15:Cross Metathesis on IgG-Fc Ahc32 Met208: 100

2.11.16: General Procedure for Expression of SarZΔ(7-142) 100

2.12: References 104

Chapter 3: Synthesis and Rapid Turnover of Peptides in Nucleus and Cytoplasm of Mammalian Cells

3.1: Introduction 106

3.2: Fluorescent Non-Canonical Amino Acid Tagging (FUNCAT) applied to nuclear translation 107

3.3: Pulse-chase FUNCAT experiments 110

3.4: Colocalisation of Nascent Peptides and Nascent RNA 112

3.5: Raman Microscopy 120

3.6: D-SIMS 121

3.7: NMR on HeLa cells 127

3.8: Puromycin labelling 129

3.9: RNA FISH 132

3.10: Mass spectrometry 135

3.11: Conclusion and Discussion 138

3.12: Methods 142

3.12.1: Cell culture and fixation 142

3.12.2: Fluorescent detection of Aha by CuAAC 143

3.12.3: Immuno-labeling 143

3.12.4: Fluorescence microscopy 143

3.12.5: Detecting Aha-tagged peptides biochemically 144

3.12.6: D-SIMS 144

3.12.7: RNA fluorescent *in situ* hybridization (RNA FISH) 145

3.12.8: Half-lives 147

3.12.9: Molecular cloning 147

3.12.10: Synthesis of 5-Ethynyl Uridine (EU) 149

3.12.11: Synthesis of Azidohomoalanine 153

3.12.12: Aha - EU CuAAC 156

3.12.13: Synthesis of $^{13}C_2$-EU 156

3.12.14: Synthesis of ^{15}N-Aha 158

3.12.15: CuAAC on Aha-labelled Haemagglutinin peptide 159

SPPS Synthesiser parameters: 159

Protocol for manual cleavage: 159

Protocol for reverse-phase purification of Aha-haemagglutinin-peptide: 160

3.13: References 161

Chapter 4:Quantitative Non-Canonical Amino acid Tagging

4.1: Introduction 164

4.2: Stable Isotope Labelling by Amino acids in Cell culture 164

4.3: Bioorthogonal Non-Canonical Amino acid Tagging 166

4.4: Quantitative Non-Canonical Amino Acid Tagging Work-flow 167

4.5: Metabolic impact of Azidohomoalanine 169

4.6: Application of QuaNCAT to primary T-cells 172

4.6.1: Statistical analysis of MS data 173

4.7: Results 177

4.8: Conclusion and Discussion 180

4.9: Methods 181

4.9.1: General Consideration 181

4.9.2: Synthesis of Azidohomoalanine 182

4.9.3: Synthesis of tris-triazole ligand 4.2 182

4.9.4: SILAC and azidohomoalanine labeling 183

4.9.5: Protein extraction, reduction and alkylation. 183

4.9.6: Copper-catalyzed azide-alkyne cycloaddition of protein samples with alkynyl-biotin 184

4.9.7: Pull-down and enzymatic digest of biotinylated proteins. 184

4.9.8: Mass spectrometry and data analysis. 185

4.9.9: Biosynthetic labelling by ^{35}S-Cysteine pulse and pulse-chase 186

4.10: References 187

Chapter 5: Outlook

5.1: Introduction 189

5.2: Unnatural amino acids as contrast-forming agents in Confocal Raman Microscopy 189

5.3: Raman data acquisition and imaging of *E. coli* B834(DE3) cells 191

5.4: Raman microscopy outlook 194

5.5 Towards a new HIV vaccine by chemical cross-linking of viral envelope glycoprotein gp140: 195

5.6: CuAAC on gp140 199

5.7: Mass spectrometric analysis of cross-linked gp140 200

5.8: gp140 Chemical Cross-Linking Outlook 203

5.9: T-cell receptor cross-linking 205

5.10: T-Cell Receptor Chemical Cross-Linking Outlook 208

5.11: Methods 209

5.11.1: General Procedure for Expression of NP276 proteins..... 209

5.11.2: General procedure for gp120 / gp140 expression 210

5.11.3: Procedure for Cross-linking Jurkat T-cell receptor..... 214

5.11.4: Procedure for Western Blot..... 214

5.11.5: Creating Liq. N2 stocks for HEK 293T cells..... 215

5.11.6: QIAGEN Gigaprep Kit Protocol 216

5.11.7: Synthesis of Trimeric PEG linker 217

5.11.8: Synthesis of Alkynyl biotin..... 218

5.12: References..... 219

Appendix to Chapter 4.....224

Acknowledgement

I would like to thank my family for their constant support during my time at Oxford, without whom none of this work would have been possible. I would like to thank them for inspiring me to become a scientist, encouraging me to follow my interests, and for keeping me curious and inquisitive throughout my life.

I would also like to thank the Rhodes Trust and the EPA Cephalosporin fund for helping to fund my DPhil studies at Oxford. I would especially like to thank the Registrar of the Rhodes Trust, Mary Eaton, Dr. Chris Blackwell at the Department of Organic Chemistry, Sr Silvana Dallanegra and Sr Sue Acheson for providing excellent support for students at Oxford, for which I am incredibly grateful.

I would like to thank Dr. Yuya Lin, Dr. Balakumar Vijayakrishnan, Dr. Lukas Lercher, Peter Crane, and Andrew Bissette for proof-reading this thesis. I would like to thank them again, along with Sonia De Munari, Dr. Macarena Sanchez-Navarro, Dr. Sylvain Royer, Phin Chooi, Hua Wang, Dr. Chris Spicer, Dr. Smita Gunnoo, Dr. Thomas Parsons, Dr. Sabyasachi Baboo, Dr. Vincent Geoghegan, and Amy Webb for their ceaseless support, as well as for providing useful scientific discussions during my DPhil study. I would also like to thank the support staff at the Department of Organic Chemistry, Exeter College, St Edmund Hall, and The Rhodes House.

I would like to thank Prof. Peter Cook, Prof. Oreste Acuto, and Dr. Guillaume Stewart-Jones for their advice during our collaborative projects, and for helping me to develop my skills as an experimental scientist.

I would like to thank all of my friends at Oxford for making my time here memorable, including (but not limited to) Kerrie Thornhill, Timothy Cheng, Kamal Wood, Andrew Bissette, Caroline Huang, Nithum Thain, Sonia De Munari, Krittika Bhattacharjee, Dhvani Mehta, Rakesh Sharma, Anisha Sharma, Arghya Sengupta, Alexis Brown, Alex West, Aparajita Kashyap, the members of ‘The Peer Revue’, members of the BGD group, and my housemates at 11 Norham Gardens.

Finally, I would like to thank my supervisor, Prof. Ben Davis, for the fantastic opportunity to carry out research in his lab at the Department of Organic Chemistry, University of Oxford. I would like to thank him for his advice, his guidance, and for helping to keep my passion for research burning.

List of publications

Sections of this thesis have been featured in the following publications:

1) QuaNCAT: quantitating proteome dynamics in primary cells. Howden, A.J.; Geoghegan, V.; Katsch, K.; Efstathiou, G.; Bhushan, B.; Boutureira, O.; Thomas, B.; Trudgian, D. C.; Kessler, B. M.; Dieterich, D. C.; Davis, B. G.; Acuto, O. *Nat. Methods* **2013**, 10 (4), 343-6.

2) Rapid cross-metathesis for reversible protein modifications via chemical access to Se-allyl-selenocysteine in proteins. Lin, Y. A.; Boutureira, O.; Lercher, L.; Bhushan, B.; Paton, R. S.; Davis, B. G. *J. Am. Chem. Soc.* **2013**, 135 (33), 12156-9

3) Most Human Proteins Made in Both Nucleus and Cytoplasm Turn Over within Minutes. Baboo, S.; Bhushan, B.; Jiang, H.; Grovenor, C. R. M.; Pierre, P.; Davis, B. G.; Cook, P. R. *PLoS One* **2014**, 9 (6), e99346

Manuscript in Preparation:

1) Genetic Incorporation of Olefin Cross-Metathesis Reaction Tags for Protein Modification. Bhushan, B.; Lin, Y. A.; Bilyard, M.; Hudson K. L.; Lercher, L.; Davis, B. G.

Selection of Abbreviations

Aha	=	azidohomoalanine
Ahc	=	*S*-allylhomocysteine
Ahs	=	*Se*-allylhomoselenocysteine
aq.	=	aqueous
Boc	=	*tert*-butoxycarbonyl
BONCAT	=	Bio-Orthogonal Non-Canonical Amino acid Tagging
^{o}C	=	degrees celcius
CM	=	cross-metathesis
cm^{-1}	=	centimetre inverse / wavenumber
CuAAC	=	copper catalysed azide-alkyne cycloaddition reaction
DCM	=	dichloromethane
DMEM	=	Dulbecco's Modified Eagle Medium
DMF	=	N,N -dimethyl formamide
DMSO	=	dimethylsulfoxide
DTT	=	dithiothreitol
ESI	=	Electrospray ionisation
HPLC	=	High Performance Liquid Chromatography
Hz	=	hertz
IPTG	=	Isopropyl β-D-1-thiogalactopyranoside
LB	=	Luria Broth
LC-MS	=	liquid chromatography mass spectrometry
Lit.	=	literature reference
M	=	molar
MALDI	=	matrix-assisted laser desorption / ionisation
MetRS	=	methionyl-tRNA synthetase
m, µ, n prefix	=	milli-, micro-, nano-
mRNA	=	messenger RNA
NMR	=	Nuclear Magnetic Resonance
OD	=	optical density
PBS	=	phosphate buffer saline
PTM	=	post-translational modification
QuaNCAT	=	Quantitative Non-Canonical Amino acid Tagging
R.b.	=	Round-bottom flask
RPM	=	revolutions per minute
R.T.	=	room temperature
s, min, h	=	seconds, minutes, hours
Sac	=	*S*-allylcysteine
sat.	=	saturated
SDS-PAGE	=	Sodium dodecylsulphate poly-acrylamide gel electrophoresis
Seac	=	*Se*-allylselenocystine
SILAC	=	Stable Isotope Labelling by Amino acids in Cell culture
SOC	=	Super Optimal broth with Catabolite repression
t-BuOH	=	*tert*-butanol
TFA	=	Trifluoroacetic acid
THF	=	Tetrahydrofuran
TLC	=	Thin Layer chromatography
tRNA	=	transfer-RNA

Chapter 1

Introduction: Unnatural Amino Acids as Metal-mediated probes of biological function

1.1: Introduction

Proteins are biomacromolecules which perform a vast range of functions in biological systems, including providing mechanical structure, catalysis of metabolic reactions, cell signalling, and transport. Proteins predominantly comprise 20 naturally occurring α-amino acids, which differ in their side chain lengths and chemical properties. The sequence and composition of its constituent amino acids are major factors in dictating a protein's structure and function. However, most proteins are subsequently further modified chemically or enzymatically via diverse processes, known collectively as post-translational modification (PTM), which may significantly modify their structure and function.[1]

PTMs are intimately involved in a number of cellular processes. The phosphorylation of serine, threonine and tyrosine residues is often used by biological systems as switches to regulate protein activity, cellular signalling, and metabolic regulation.[2-5] The acetylation of lysine residues on histone proteins is regulated by histone acetyltransferases (HATs) and histone deacetylases (HDACs), and plays a vital role in gene transcription regulation by modifying DNA binding.[6,7] Acetylation of non-histone proteins[8] also has been implicated in regulating the activity and stability of proteins spanning a vast range of cellular function including transcription factors,[9] cellular transport proteins,[10] cytoskeletal proteins,[11] DNA repair proteins,[12] and chaperones.[13] Other important PTMs include glycosylation[14-16] and sumoylation[17] which modify protein stability, activity, folding, localisation, and immunogenicities. Protein lipidation is implicated in the membrane localisation of proteins,[18] while ubiquitination targets proteins for degradation.[19]

The significance of PTMs is underscored by the fact that their disruption, and mutation of PTM sites, is implicated in a number of diseases[1,20] including cancer,[21,22] cardiovascular disorders,[23] liver disease,[24] and Alzheimer's disease.[25,26] This has led to several PTMs being considered as molecular targets for drug therapies.[27] However, the study of PTMs is

complicated by several factors. PTMs which serve as regulatory markers are often reversible in biological systems, in addition to existing at low levels, thus hindering the isolation of homogenous protein constructs.[28] This is further complicated when the protein target has multiple possible modification sites.

In order to access homogenous PTM constructs, the chemical modification of proteins may be employed to conjugate the requisite modifications onto protein scaffolds. Techniques for the protein modification, however, have usually been limited not only by the range of side-chain functionalities in constituent amino acids, but also by the need of the reactions to be compatible in aqueous media, and be minimally perturbative to protein structures as well.[29]

1.2: Modification Reactions on Proteins

Traditionally, conjugation reactions on proteins have utilised the reactivity of naturally occurring amino acids such as lysine and cysteine. Techniques included the use of nucleophiles,[30] reductive aminations,[31-34] native chemical ligation (NCL) reactions,[35-40] disulfide formation,[41,42] thiol-ene,[43,44] and thiol-yne coupling reactions.[45,46] In addition to accessing PTMs, bioconjugation reactions have been instrumental in developing strategies for targeted delivery of drugs and imaging agents,[47] single molecule fluorescence spectroscopy,[48] and in studying receptor-ligand interactions.[49] However, a common drawback of all of these techniques include the fact that lysine and cysteine (and more generally, amine and thiol functional groups) are highly biologically abundant. Conjugation techniques which rely upon functionalisation of these groups then are not site-specific, and both off-site and multi-site reactivity are inevitable in complex biological systems.

Alternative strategies to bioconjugation involve the use of unnatural amino acids (UAAs), which may be site-selectively incorporated into proteins, and then be used as exclusive targets for modification reaction in a 'tag-and-modify' strategy.[50] A more thorough description of the structures, uses and methods of incorporation of various UAAs will be provided in Section **1.6**.

1.3: UAA-enabled Conjugation Reactions on Proteins

Bearing side-chains with unique reactive groups, the UAA-incorporation enables the use of so-called "bio-orthogonal" chemistry, i.e, reactions which specifically target these unnatural reactive groups without perturbing or cross-reacting with biologically abundant molecules.[51,52] UAA-incorporation has enabled the use of a larger variety of conjugation reactions on proteins including Staudinger ligations[53,54] and their traceless variants,[55,56] Inverse electron-demand Diels-Alder cycloadditions,[57,58] photo-induced 1,3-dipolar cycloadditions,[59,60] Suzuki-Miyaura cross-coupling,[61,62] hydrazone and oxime formations,[63,64] Michael additions,[65,66] azide-alkyne cycloaddition reactions,[67,68] and olefin cross-metathesis[69,70] (Fig. 1.1).

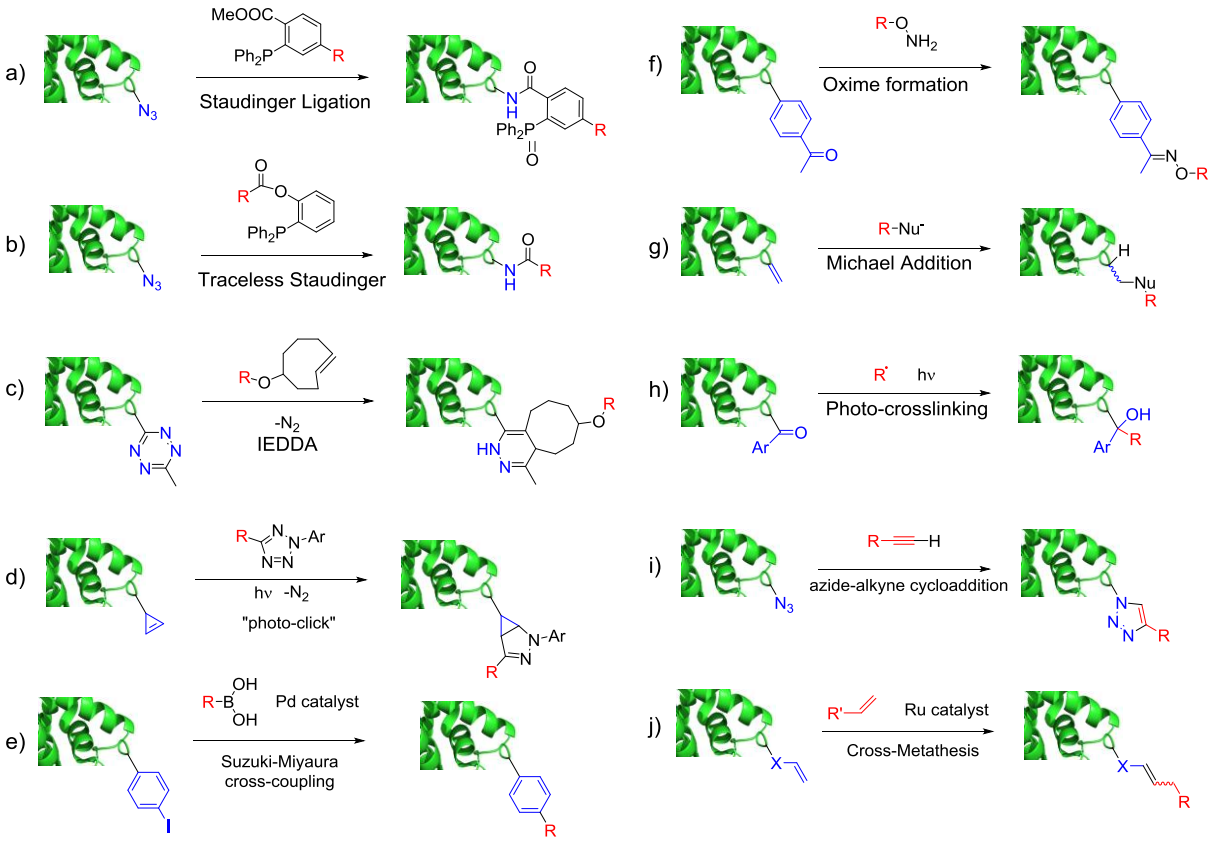


Fig. 1.1: Representative UAA-enabled reactions on proteins. a) Staudinger ligation, b) Traceless Staudinger, c) Inverse electron-demand Diels-Alder (IEDDA), d) photo-induced dipolar cycloaddition, e) suzuki-Miyaura cross-coupling, f) oxime formation, g) Michael addition, h) photo-crosslinking, i) azide-alkyne cycloaddition, j) olefin cross metathesis

Of these UAA-enabled protein modification techniques, the application of azide-alkyne cycloadditions and olefin cross-metathesis to various biological systems will be discussed in later chapters. The development of these reactions for protein modification is discussed next.

1.4: Azide-alkyne cycloadditions for protein modification

In 2002, the use of copper (I) to catalyse the fusion of azides and alkynes to generate 1, 4 - disubstituted 1, 2, 3-triazoles was first reported.[71,72] Although analogous to the thermally activated Huisgen 1,3-dipolar azide-alkyne cycloaddition,[73] the copper (I) catalysed azide-alkyne cycloaddition (CuAAC) reaction was found to proceed via a metallacycle intermediate (Fig. 1.2), and was found to occur about seven magnitudes faster even at room temperature.

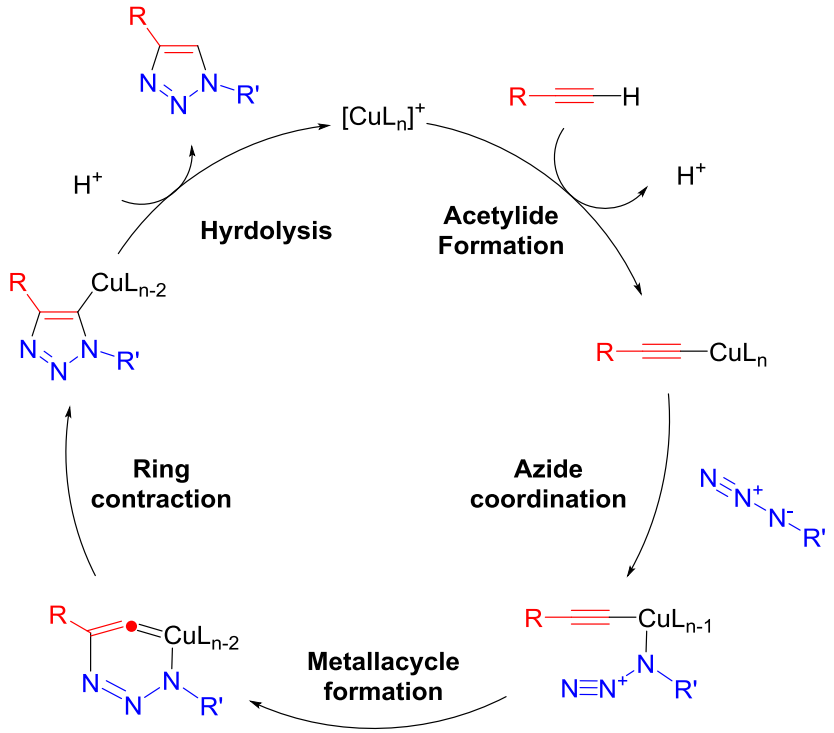


Fig. 1.2: Proposed catalytic cycle for the copper (I) catalysed azide-alkyne cycloaddition.[71]

The reaction was also found to be tolerant of the presence of both water and oxygen, stimulating interest in its application in biology. As both azide and alkyne functional groups are generally absent in biological systems, application of CuAAC would first involve labelling the system with one reactive partner, followed by conjugation with the other. The reaction was first applied to labelling the cowpea mosaic virus capsid. The protein scaffold was first functionalised with either alkynes or azides at lysine or cysteine residues, followed by CuAAC-mediated labelling with fluorophores containing the respective reaction partner.[74] However, the reaction was non site-specifc due to lack of specificity at the primary functionalisation step on lysine residues. Furthermore, the reaction was found to require both an organic co-solvent as well as an additional reducing agent to convert Cu (II) to Cu (I) *in situ*.

Development of azide and alkyne bearing UAAs such as *O*-propargyltyrosine (**1.1**),[75] *p*-azidophenylalanine (**1.2**),[75] L-azidohomoalanine (Aha, **1.3**),[76] and L-homopropargylglycine (Hpg, **1.4**)[77] allowed for the site-selective labelling of proteins with CuAAC-reactive groups. Further optimisation of CuAAC reaction conditions led to the use of copper (I) bromide and triazole based ligands, used to prevent the disproportionation of Cu (I) in water. The resultant reaction was found to allow for efficient conjugations, thereby abolishing the need for additional reducing agents,[78] which might otherwise affect protein structure by disrupting disulfide bonds.[79] Non-reducing CuAAC reactions also displayed improved conversions and higher protein recovery.[80] CuAAC has since become a widely used technique in bioconjugation, being applied to a range of systems such as generating glycoprotein mimics,[79,81] protein PEGylation,[82] identifying newly synthesised proteins,[83] labelling cell surfaces,[78,84,85] and in fluorescent microscopy.[86]

However, the exposure of live cells to the copper (I) catalyst has been reported to be harmful to cells, causing arrest of replication in *E. coli*.[84] A potential cause may be due to generation of reactive oxygen species (ROS) by copper (I), causing cellular damage.[87] However, with the development of newer triazole based ligands such as THPTA (**1.5**),[88] BTTES (**1.6**),[89] and BTTPS (**1.7**),[90] the toxic effects of copper (I) have been found to be dramatically reduced, and reaction rates accelerated (Fig. 1.3).

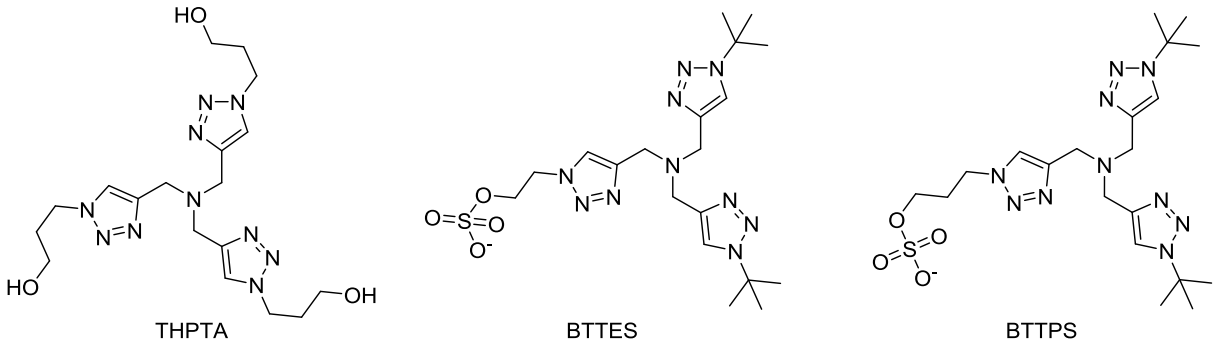


Fig. 1.3: Triazole ligands for CuAAC. THPTA: tris(3-hydroxypropyltriazolylmethyl)amine, BTTES: 2-(4-(bis(tert-butytriazolylmethyl)aminomethyl)triazolyl))ethanesulfonic acid, BTTPS: 3-(4-(bis(tert-butytriazolylmethyl)aminomethyl)triazolyl))propanesulfonic acid

An alternate strategy to circumvent the toxic effects of copper (I) involves the use of strained cyclooctyne species (**1.7**),[91] that undergo spontaneous azide-alkyne cycloadditions to form less strained triazoles. The use of such species was first reported in bioconjugation reactions with azide-labelled glycoproteins expressed on CHO cell surfaces, with no observable toxic

effects.[92] Following this, a series of strained cyclooctynes were designed to enhance the rate of the strain-promoted azide-alkyne cycloaddition (SPAAC) reaction. Modified cyclooctyne scaffolds include difluorinated cyclooctyne (**1.8**),[93] dibenzocyclooctyne (**1.9**),[94] biarylazacyclooctyne (**1.10**),[95] and cyclopropyl fused cyclooctyne (**1.11**)[96] (Fig. 1.4). These have successfully been used in a range of systems, such as in labelling azidohomoalanine-incorporated proteins in Rat-1 fibroblasts,[97] monitoring trafficking of toxic proteins in live cells,[98] and labelling membrane glycans in zebrafish embryos.[99,100]

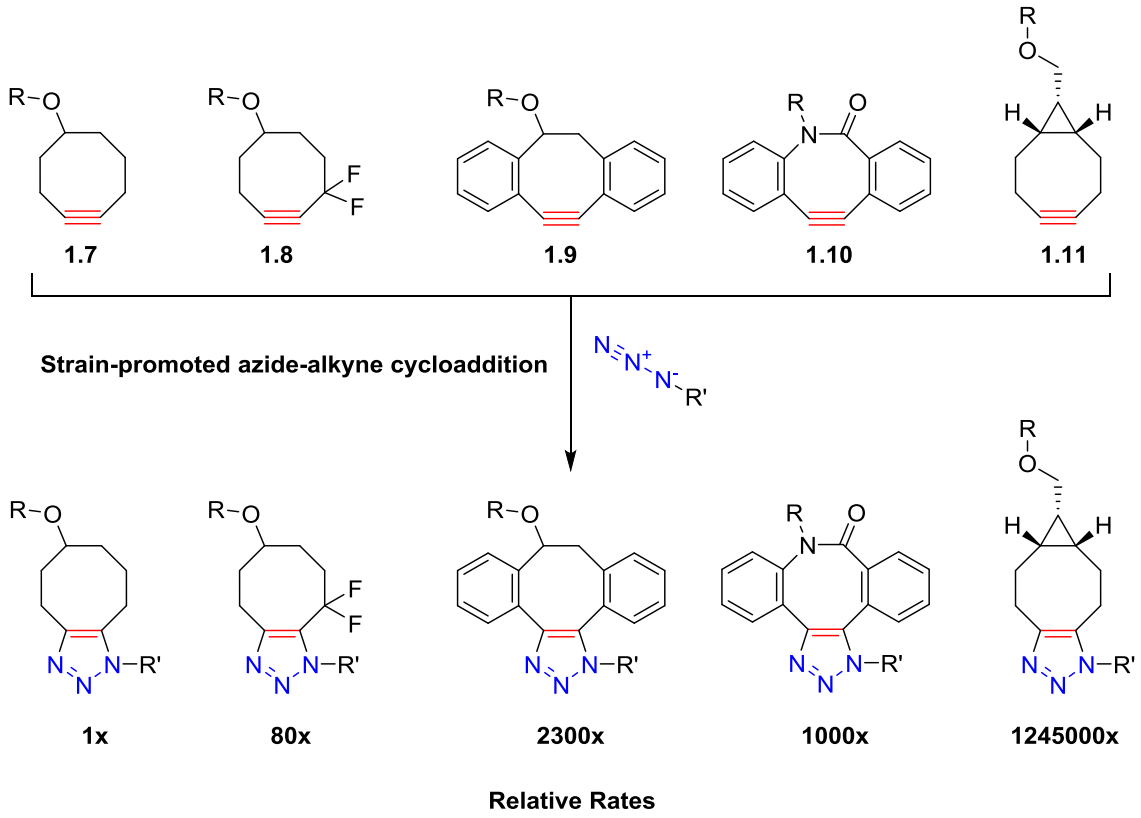


Fig. 1.4: Strained cyclooctynes for SPAAC, and their relative reaction rates.

1.5: Olefin Cross-Metathesis for protein modification

Olefin metathesis is a versatile C-C bond forming reaction (Fig. 1.5).

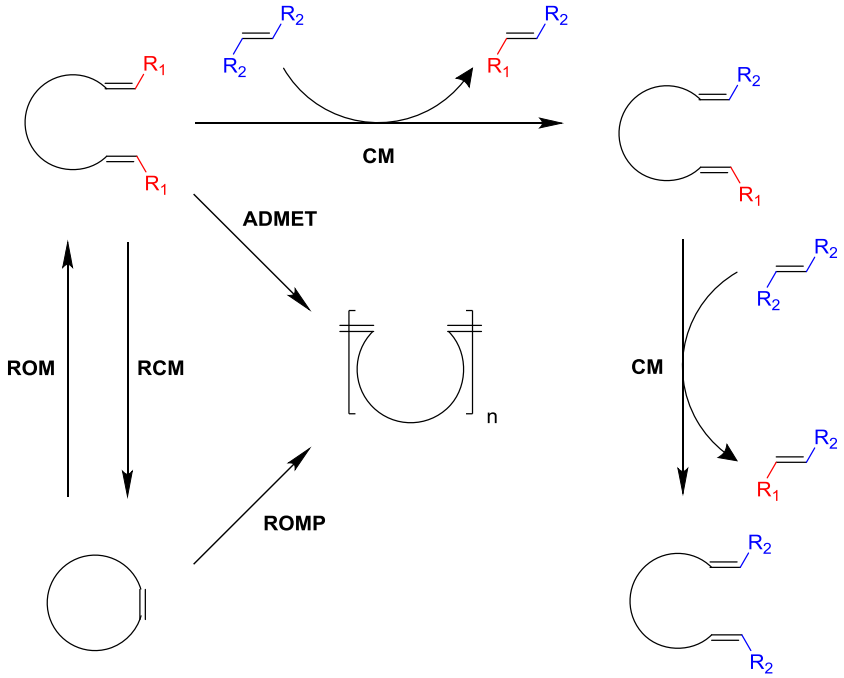
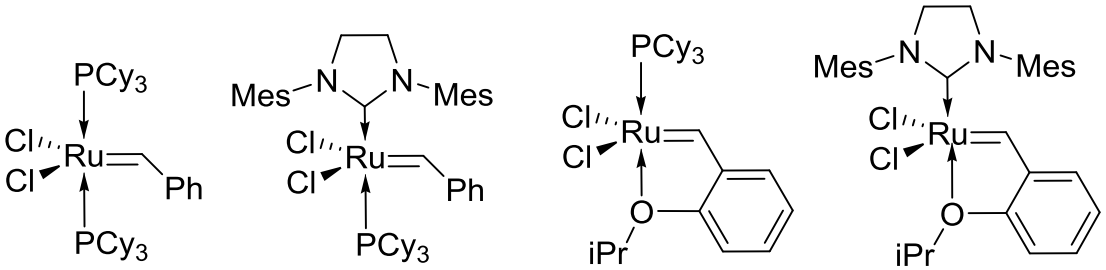


Fig. 1.5: Possible olefin metathesis reactions on a representative diene. CM: cross metathesis, RCM: ring-closing metathesis, ROM: ring-opening metathesis, ADMET: acyclic diene metathesis, ROMP: ring-opening metathesis polymerisation.

Additionally, CM is compatible with conditions for protein modification owing to the high air-stability, chemoselectivity, as well as tolerance to various functional groups such as amides, alcohols and carboxylic acids, exhibited by ruthenium-based metathesis complexes.[101]

One of the first applications of CM on peptides involved the dimerisation of cyclic peptides functionalised with L-homoallylglycine (Hag) using a ruthenium catalyst.[102] Using the first-generation Grubbs catalyst (**1.12**), CM with simple alkenes like styrene has also been demonstrated on the homologous series of alkene containing UAAs dehydroalanine (Dha, **1.13**), vinylglycine (**1.14**), allylglycine (**1.15**), and homoallylglycine (Hag, **1.16**). This study revealed that which Dha was CM-unreactive, reactivity increased upon increasing the UAA side chain length.[103,104] However, all CM products were consistently contaminated with significant amounts of self-metathesis side products. Later efforts to optimise reaction conditions and catalyst structures to favour CM products over self-metathesis led to the use of the second generation Grubbs catalyst (**1.17**).[105] Studies on CM-mediated glycopeptide

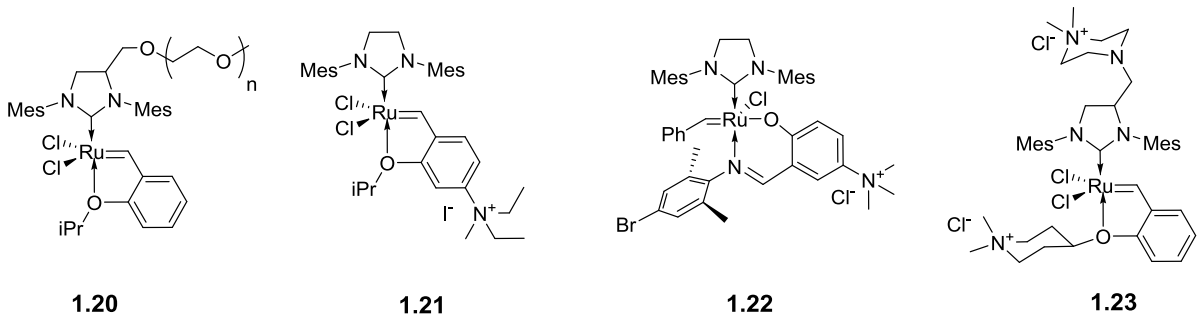
synthesis revealed that with the use of the second generation Hoveyda-Grubbs catalyst **(1.19)**, unprotected alkene functionalised sugars were found to be tolerated in the reaction.[101] This served to motivate the use of CM to generate analogues of natural glycosylation PTMs on proteins.



1.12 **1.17** **1.18** **1.19**

Fig. 1.6: Ruthenium based catalysts for olefin metathesis. **1.12**: Grubbs Catalyst first generation, **1.17**: Grubbs Catalyst second generation, **1.18**: Hoveyda-Grubbs Catalyst first generation, **1.19**: Hoveyda-Grubbs catalyst second generation.

However in all of these cases, the reaction necessitated the use of organic solvents to solubilise both the catalyst as well as peptide substrates, as well as the use of either dry solvents or inert atmospheres. In order to apply CM in aqueous systems, two strategies have been applied. First, a range of water soluble metathesis catalysts have been developed based on imparting solubility through functionalisation of traditional metathesis catalyst scaffolds. Olefin metathesis has since been demonstrated on water-soluble alkenes using ruthenium based catalysts which have been rendered water soluble either though PEGylation (**1.20**),[106] or through the introduction of one (**1.21** and **1.22**),[107,108] or two quarternary ammonium groups (**1.23**, Fig. 1.7).[109]



1.20 **1.21** **1.22** **1.23**

Fig. 1.7: Water soluble metathesis catalysts

Alternatively, traditional commercially available metathesis catalysts (Fig. 6) have been applied to aqueous CM using organic co-solvents such as DMF, methanol,[40] dimethoxyethane, or acetone[110,111]. Alternatively, the use of surfactants such as Polyoxyethanyl α-tocopheryl sebacate (PTS) allows for the co-solubility of traditional metathesis catalysts as well as peptide substrates, allowing for CM in water.[112] Further attempts were made to develop CM into a viable reaction for the functionalisation of proteins, using the organic co-solvent strategy.

Lin and co-workers screened a range of alkene-containing UAAs for aqueous CM reactivity using 30% *t*-butanol as a co-solvent, with **1.19** as a catalyst. Of all of the amino acids investigated, only the allyl sulfide containing UAAs *S*-allylcysteine (Sac, **1.24**) and *S*-allylhomocysteine (Ahc, **1.25**) were found to be CM-reactive.[113] Following on from the observation that allylic hydroxyl groups induce acceleration of olefin metathesis,[114-116] a systematic study of CM reactivity was carried out on benzylic *N*-, *O*-, and *S*- allyl species. Though all were found to be CM reactive, reaction on *S*-allyl species was most efficient,[113] where the allylic chalcogen heteroatom provides crucial coordination to the metal catalyst centre (Fig. 1.8).[69,113]

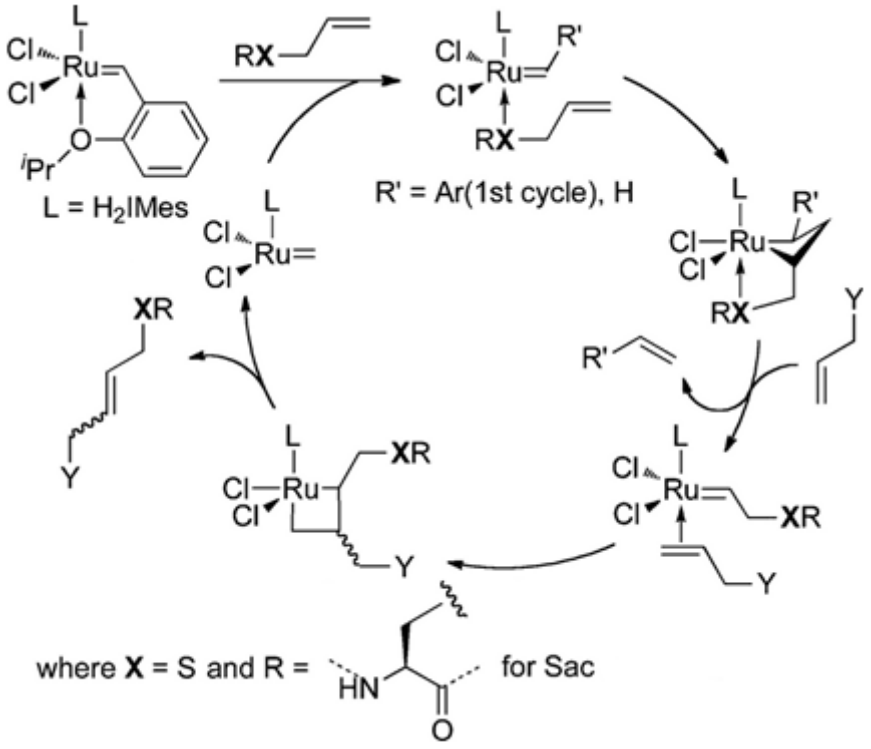


Fig. 1.8: Chalcogen-assisted CM mechanism with the Hoveyda-Grubbs 2nd Gen Catalyst

Metathesis catalyst degradation has been reported to occur both in protic solvents and also with the use of allylic alcohols as CM partners.[114,117,118] However, a key contributing factor to the superior reactivity of allyl sulfide substrates may be that CM reaction rates far exceeded rates of that catalyst decomposition, effectively out-competing the degradation reaction in water.

Dehydroalanine (Dha, **1.13**) is an effective Michael-type acceptor for a range of nucleophiles. Several routes for the chemical installation of Dha onto proteins has been reported, including oxidative elimination from incorporated phenylselenocysteine residues,[119] sulfonate elimination from native serines,[120,121] or using various reagents from native cysteine residues.[65,66,122] The resulting Dha may then be utilised to generate thioether-functionalized proteins via the conjugate addition of the corresponding thiol nucleophile. Using this method, the chemical installation of *S*-allylcysteine (Sac, **1.24**) has been demonstrated in a single-cysteine mutant of subtilisin from *Bacillus lentus* (SBL-S156C) via 1,4-addition of allyl thiol to Dha.[123]

CM was carried out on chemically incorporated Sac-labelled SBL with catalyst **1.19**, allyl alcohol as a CM partner, and 30% *t*-butanol as an organic co-solvent. Although the protein was observed to be unreactive, possibly due to unproductive coordination of the Ru-catalyst with proximal Lewis basic residues, the reactivity was restored upon the addition of $MgCl_2$ in the reaction solution.[113] The latter served as a milder, biocompatible alternative to the previously reported sacrificial Lewis base $Ti(O^iPr)_4$, which has been reported to disrupt unproductive chelation in metathesis.[124,125] This allowed for modification of SBL with allyl-functionalised sugars and PEG as well as allyl alcohol. Using these conditions, subsequent reports have outlined ring closing metathesis (though not CM) on Sac-incorporated peptides, [126] and *O*-crotylserine (**1.26**) incorporated proteins.[127] It has recently been reported that the replacement of sulfur by selenium offers a significant CM rate enhancement, as well as broadening the scope of reaction partners that can be used.[70]

1.6: Incorporation of Unnatural Amino Acids into Proteins

As discussed in section **1.3**, the ability to modify native proteins is constrained by the reactive functionalities of naturally occurring amino acids, and is often restricted to a much smaller subset than that. This has driven research into the incorporation of UAAs into cellular proteins, in order to both expand the reaction space by introducing novel reactive groups as well as reduce off-site reactivity by site-selective protein modification. Additionally, the UAAs may also function as spectroscopic and imaging probes, for installation mimics of natural PTMs or even impart novel functionalities to enzymes.[128,129] Apart from chemical methods for site-specific installation of UAAs on purified recombinant proteins, efforts have been directed towards utilising the cellular metabolic machinery to incorporate UAAs co-translationally for broader *in vivo* applications.

This has been attempted broadly in two ways. The first method involves 'sense' codon reassignment, i.e. reassigning a natural amino acid's codon for UAA incorporation by utilising auxotrophic cell lines. This codon reassignment exploits the flexibility of native aminoacyl-tRNA synthetases (aaRSs) in accepting UAAs instead of their natural substrates for tRNA loading upstream of protein synthesis.[130] The second method involves reassignment of 'non-sense' codons such as the amber stop codon (TAG) by additionally encoding an orthogonal aminoacyl tRNA synthetase/tRNA pair (aaRS/tRNA). This encoded pair is specific for the UAA of choice and complementary to the codon to be reassigned. The pair is so designed as to not cross react with the existing cellular translational machinery, but is recognised by ribosomes to allow UAA incorporation. These two methods will be described in detail in the following sections.

1.6.1: 'Sense' codon reassignment

The recognition of amino acids and their analogues by the corresponding aminoacyl-tRNA synthetases (aaRSs) is a major factor in determining their incorporation efficiency *in vivo*.[77,131,132] Therefore, in order to effectively replace a naturally existing AA in a protein sequence with a UAA, native tRNA must be aminoacylated with the UAA by the corresponding aaRS. Additionally, even if the aaRS is flexible enough to recognise the UAA and catalyse the aminoacylation reaction, UAA-aminoacylation needs to successfully out-compete native AA-aminoacylation. This may be accomplished by utilising cell lines

auxotrophic for the native AA, with a UAA for incorporation supplemented in the media. Auxotrophic cell strains do not possess *de novo* biosynthetic mechanisms for the amino acid and instead would require its supplementation in growth media.[29] For example, the *E. coli* strain B834(DE3) is auxotrophic for methionine.[133]

For efficient incorporation, the protein is usually placed under tight induction control, e.g., using a genetically encoded *lac* repressor in DE3 cells. Cells are then typically allowed to deplete internal stocks of the natural amino acid and replace it with the UAA prior to induction of protein expression. Additionally, UAA incorporation efficiency may be enhanced by over-expressing the corresponding aaRS in the same cells.[134] Mutation of the aaRS itself to allow for broader substrate specificity has also been reported to improve yields of UAA-incorporated proteins.[135,136]

One of the most extensively utilised amino acid for UAA incorporation through sense codon reassignment is methionine.[76,77,134,137] As described in section **1.4**, azide and alkyne containing amino acids L-azidohomoalanine (Aha, **1.3**),[76] and L-homopropargylglycine (Hpg **1.4**)[77] have both been successfully incorporated into proteins expressed in Met-auxotrophic cells, enabling CuAAC on recombinant proteins. Aha has also been used to enable Staudinger ligations on proteins as well.[76,138] Aha replacement has also been demonstrated in a range of mammalian cell lines, underscoring its utility in diverse biological systems as a general Met analogue.[139,140] Furthermore, the replacement of Met with Aha has been found to be largely non-toxic, and non-perturbative to protein structure and function.[141-144]

UAA incorporation has also been reported via the reassignment of 'sense' codons for other amino acids such as proline,[145] leucine,[146] isoleucine,[147] valine,[148,149] histidine,[150] tyrosine,[151] tryptophan,[152] and phenylalanine[153] in corresponding auxotrophic cell lines (Fig. 1.9).

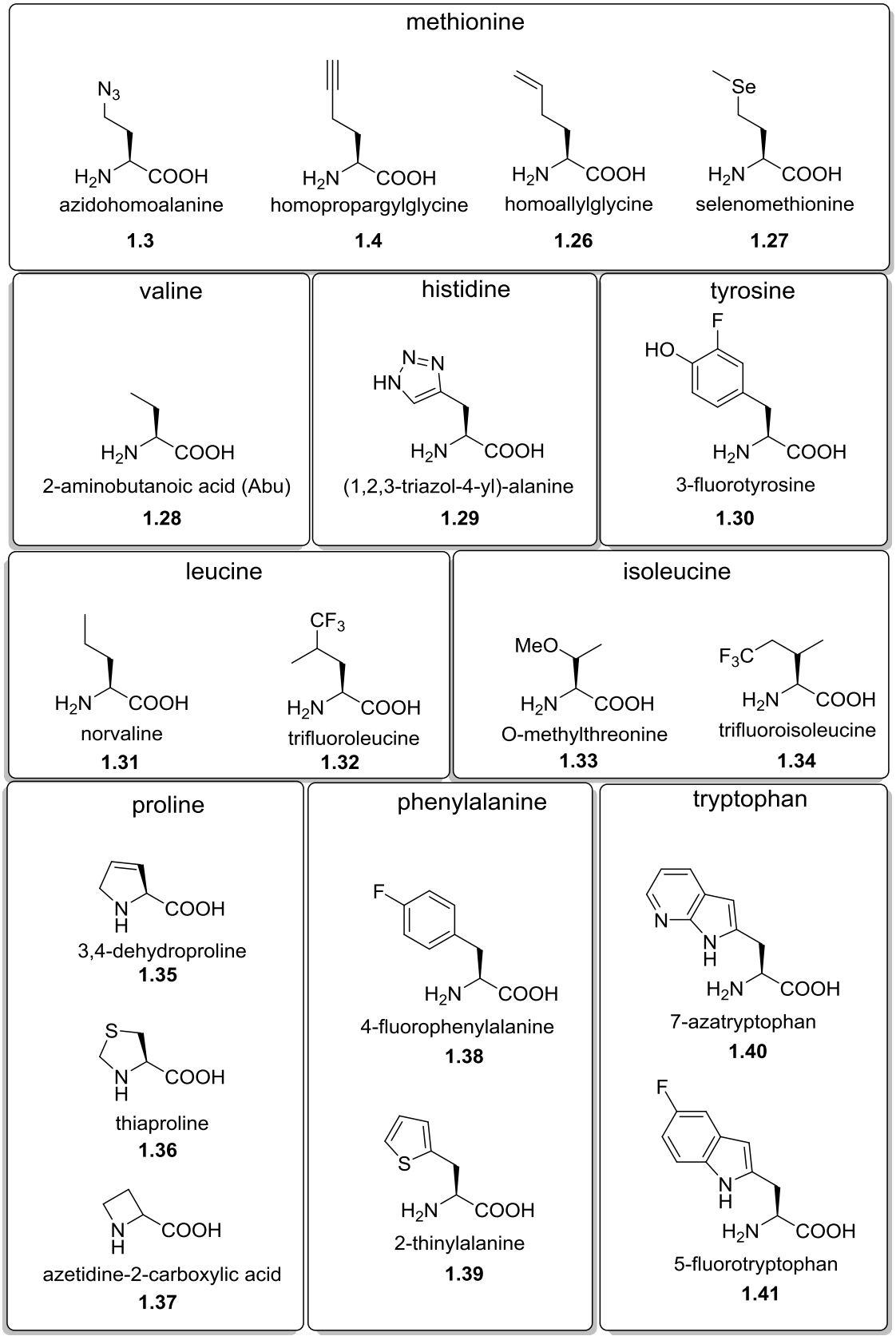


Fig. 1.9: Examples of UAAs incorporated by sense codon reassignment, grouped by replaced AAs.

1.6.2: 'Non-sense' codon reassignment

A key step during protein biosynthesis is the binding of aminoacyl-tRNA to mRNA on the ribosomal A site prior to peptide chain elongation. This binding is mediated by matching the A site codon with the complementary anticodon on the tRNA. Although 64 possible codons exist, only 61 complementary anticodon-bearing tRNAs are present in biological systems. The three remaining codons instead are utilised for termination of translation,[154] and are named 'amber' (U/TAG), 'ochre' (U/TAA) and 'opal' (U/TGA). Non-sense codon reassignment (or 'stop codon suppression') utilises these codons for unnatural amino acid incorporation, particularly the amber stop codon, which is the rarest of the three in *E. coli*.[155] This was first demonstrated in a cell-free expression system using chemically pre-aminoacylated engineered tRNA possessing an anticodon complementary to the amber non-sense codon.[156] However, due to this chemical aminoacylation, this method was not generally applicable in biological systems.

In this case however, unlike sense-codon reassignment, the substrate flexibility of native aaRSs cannot be exploited for UAA incorporation as no natural AA is being replaced by the UAA. Therefore, both the stop codon-complementary 'suppressor' tRNA and a mutant aaRS must be genetically encoded so that the aminoacylation reaction may occur *in vivo*, upstream of incorporation. In order to obtain highly site specific incorporation, the mutant aaRS must be carefully chosen such that it only aminoacylates the suppressor tRNA with the UAA for incorporation, and does not cross-react with endogenous natural AAs or non-suppressor tRNAs.[157] The observed cross-reactivity of glutaminyl-tRNA synthetase with a genetically encoded amber codon-suppressor tRNA in *S. cerevisiae* first allowed for the genetic incorporation of Gln by non-sense codon reassignment.[158] This served as a particularly insightful example of the cross-species orthogonality of aaRS-tRNA pairs, leading to the application of such systems toward UAA incorporation.

This was first successfully demonstrated by the incorporation of the UAA p-methoxyphenylalanine by non-sense codon reassignment in *E. coli* using the *Methanococcus jannaschii* (*M. jann*) TyrRS with amber codon-suppressor tRNA.[157] Subsequently, several hundred UAAs have been incorporated by non-sense codon reassignment bacteria,[159] mammalian cells,[160] and even multicellular organisms such as *C. elegans*,[161,162] and *Drosophila*.[163]

Since mutagenesis and directed evolution of aaRSs may be carried out to improve affinities for various UAAs,[164-166] incorporation of a much more diverse variety of UAAs is possible by non-sense codon reassignment as compared to sense-codon reassignment. Examples of functionalities enabled by non-sense codon reassignment include azide,[167] and alkyne functionalised UAAs for staudinger ligations and CuAAC,[159,168] as well as UAAs functionalised with reactive handles for IEDDA[57] and 'photo-click'-based protein modification reactions.[169] The incorporation of aryl halide-functionalised UAAs has also been reported,[170] as has their use in enabling Pd-catalysed Suzuki-Miyaura cross-coupling on proteins[62] and cells.[61] Aryl boronic acid functionalised UAAs have also been successfully incorporated and used for Suzuki cross-coupling.[171] In addition to accessing PTMs through cross-coupling reactions, UAAs acting as native PTMs and PTM mimics have also been incorporated directly through non-sense codon reassignment.[172] Spectroscopically active UAAs have also been successfully incorporated, such as fluorinated UAAs for ^{19}F NMR,[173,174] nitrile functionalised UAAs for IR and FRET,[175,176] fluorescent napthalene-[177], coumarine-,[178] and dansyl-functionalised[179] UAAs (Fig. 1.10).

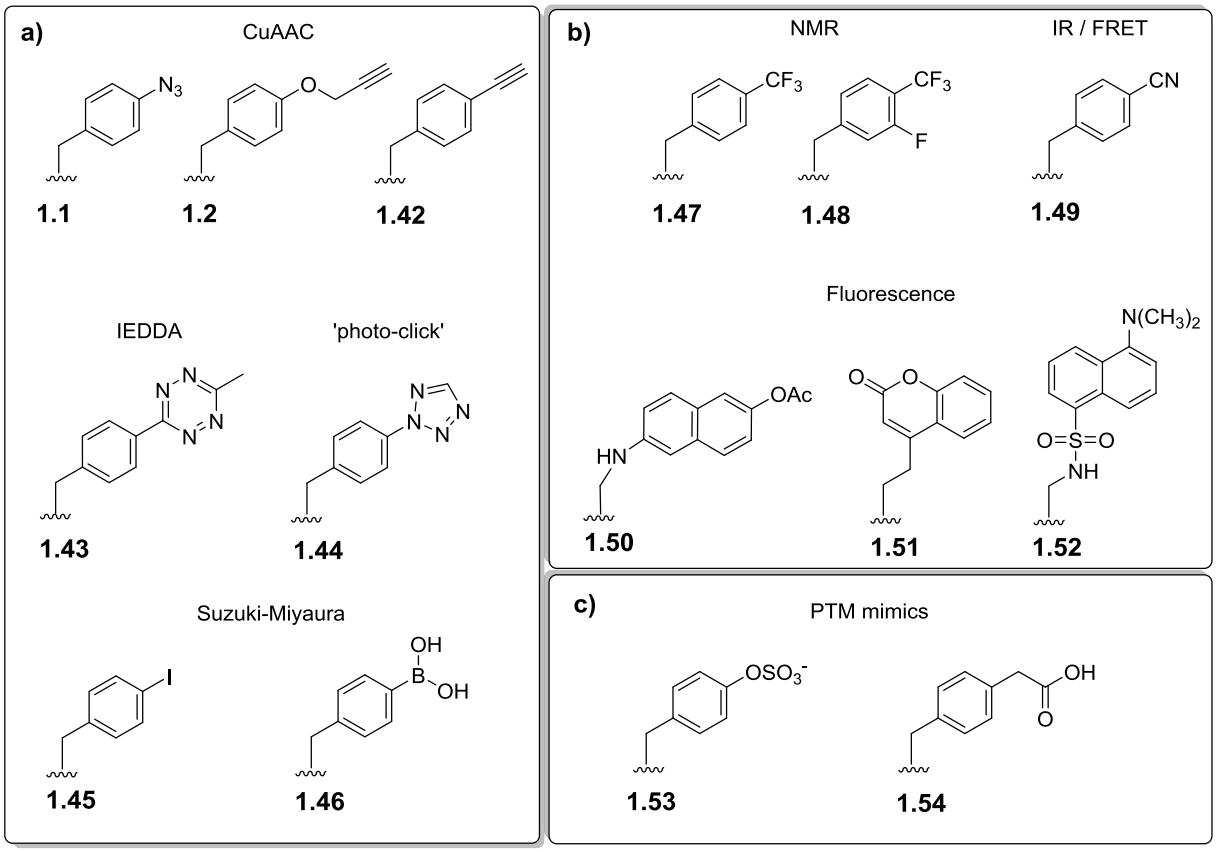


Fig. 1.10: UAAs incorporated through non-sense codon reassignment, grouped by function: a) cross-coupling, b) spectroscopically active, c) direct PTMs and PTM mimics

In addition to the amber stop codon, the ochre[180] and opal[181] stop codons may be reassigned for UAA incorporation as well, allowing for dual non-sense codon reassignment for the simultaneous incorporation of different UAAs.[182] This has been successfully demonstrated in the incorporation of UAAs functioning as FRET pairs to study protein conformation.[183,184] The reassignment of all 3 non-sense codons cannot be done simultaneously, however, as one must still be utilised to terminate protein translation. Initial work in our group also suggests that non-sense codon reassignment is compatible with sense-codon reassignment as well, allowing for the simultaneous incorporation of UAAs in auxotrophic cells (Crane et. al, unpublished results).

For the incorporation of > 2 UAAs by non-sense codon reassignment, the use of quadruplet codons have been investigated, which would allow for the reassignment of up to 256 new codons. Quadruplet codon reassignment has been successfully demonstrated,[181,185] though poor dual-incorporation efficiency stimulated interest in the development of orthogonal quadruplet-codon based translation systems in cells.[186] This system utilises orthogonal mRNA carrying an modified Shine-Dalgarno sequence for initiation of translation which is not read by native ribosomes, but is recognised and bound by ribosomes containing genetically encoded rRNA. Ribosome mutation at the A site then allows for efficient accommodation of quadruplet codons.[187] This system has been used for the site-specific dual incorporation of UAAs for intramolecular cross-linking[188] or dual labelling.[189]

1.7: Thesis objective

The use of UAAs as biological probes will be expanded upon in the following chapters. In Chapter 2, the genetic incorporation of alkene-bearing UAAs via sense-codon reassignment in bacterial and mammalian proteins to enable Ru-catalysed olefin cross-metathesis (CM) reactions will be discussed. This work builds upon previously established methods to chemically incorporate CM handles onto proteins. In this chapter, the rational design of UAAs, as well as assays to screen for recognition by cellular incorporation machinery will be discussed in detail. The expression of a range of alkene-tagged recombinant proteins, their complete characterisation, as well as the development of a more general protocol for on-protein CM will be elucidated.

In chapter 3, the utility of UAA incorporation to probe mammalian cell translation systems will be discussed. Incorporation of an azide-bearing UAA has been used to discover a

remarkable system of protein synthesis in mammalian cell nuclei at short incorporation times (~ 5 s), along with rapid metabolic degradation of the synthesised peptides. In addition to the metal-mediated protein labelling techniques used to discover this system, this chapter describes several orthogonal UAA-based labelling and analysis techniques used to confirm these observations. Possible reasons both for the existence of such a system as well as its previous lack of detection will be put forth.

In Chapter 4, the use of UAA incorporation and bioconjugation reactions will be described in the development of a novel and generally applicable proteomics technique. This technique may be used to determine quantitative changes in cell proteomes in response to external stimuli, and may be applied to systems to which traditional proteomics techniques cannot be easily applied, such as primary cells.

Finally, in Chapter 5, further applications of UAA incorporation will be discussed. Preliminary results will be reported in efforts to use UAAs in the vibrational Raman microscopic imaging of biological systems, in generating HIV vaccines, and inducing T-cell stimulation.

1.8 References:

(1) Xin, F.; Radivojac, P. *Bioinformatics* **2012**, *28*, 2905.
(2) Johnson, L. N.; Lewis, R. J. *Chem Rev* **2001**, *101*, 2209.
(3) Cohen, P. *Nat Cell Biol* **2002**, *4*, E127.
(4) Chooi, K. P.; Galan, S. R.; Raj, R.; McCullagh, J.; Mohammed, S.; Jones, L. H.; Davis, B. G. *J Am Chem Soc* **2014**, *136*, 1698.
(5) Schieven, G. L. *Curr Top Med Chem* **2009**, *9*, 1038.
(6) Grunstein, M. *Nature* **1997**, *389*, 349.
(7) Jenuwein, T.; Allis, C. D. *Science* **2001**, *293*, 1074.
(8) Glozak, M. A.; Sengupta, N.; Zhang, X.; Seto, E. *Gene* **2005**, *363*, 15.
(9) Luo, J.; Li, M.; Tang, Y.; Laszkowska, M.; Roeder, R. G.; Gu, W. *Proc Natl Acad Sci U S A* **2004**, *101*, 2259.
(10) Bannister, A. J.; Miska, E. A.; Görlich, D.; Kouzarides, T. *Curr Biol* **2000**, *10*, 467.
(11) Serrador, J. M.; Cabrero, J. R.; Sancho, D.; Mittelbrunn, M.; Urzainqui, A.; Sánchez-Madrid, F. *Immunity* **2004**, *20*, 417.
(12) Cohen, H. Y.; Lavu, S.; Bitterman, K. J.; Hekking, B.; Imahiyerobo, T. A.; Miller, C.; Frye, R.; Ploegh, H.; Kessler, B. M.; Sinclair, D. A. *Mol Cell* **2004**, *13*, 627.
(13) Bali, P.; Pranpat, M.; Bradner, J.; Balasis, M.; Fiskus, W.; Guo, F.; Rocha, K.; Kumaraswamy, S.; Boyapalle, S.; Atadja, P.; Seto, E.; Bhalla, K. *J Biol Chem* **2005**, *280*, 26729.
(14) Moremen, K. W.; Tiemeyer, M.; Nairn, A. V. *Nat Rev Mol Cell Biol* **2012**, *13*, 448.
(15) Nairn, A. V.; Aoki, K.; dela Rosa, M.; Porterfield, M.; Lim, J. M.; Kulik, M.; Pierce, J. M.; Wells, L.; Dalton, S.; Tiemeyer, M.; Moremen, K. W. *J Biol Chem* **2012**, *287*, 37835.

(16) Kwong, P. D.; Doyle, M. L.; Casper, D. J.; Cicala, C.; Leavitt, S. A.; Majeed, S.; Steenbeke, T. D.; Venturi, M.; Chaiken, I.; Fung, M.; Katinger, H.; Parren, P. W.; Robinson, J.; Van Ryk, D.; Wang, L.; Burton, D. R.; Freire, E.; Wyatt, R.; Sodroski, J.; Hendrickson, W. A.; Arthos, J. *Nature* **2002**, *420*, 678.
(17) Geiss-Friedlander, R.; Melchior, F. *Nat Rev Mol Cell Biol* **2007**, *8*, 947.
(18) Nadolski, M. J.; Linder, M. E. *FEBS J* **2007**, *274*, 5202.
(19) Lecker, S. H.; Goldberg, A. L.; Mitch, W. E. *J Am Soc Nephrol* **2006**, *17*, 1807.
(20) Li, S.; Iakoucheva, L. M.; Mooney, S. D.; Radivojac, P. *Pac Symp Biocomput* **2010**, 337.
(21) Radivojac, P.; Baenziger, P. H.; Kann, M. G.; Mort, M. E.; Hahn, M. W.; Mooney, S. D. *Bioinformatics* **2008**, *24*, i241.
(22) Krueger, K. E.; Srivastava, S. *Mol Cell Proteomics* **2006**, *5*, 1799.
(23) Van Eyk, J. E. *Circ Res* **2011**, *108*, 490.
(24) Newton, B. W.; Russell, W. K.; Russell, D. H.; Ramaiah, S. K.; Jayaraman, A. *J Proteome Res* **2009**, *8*, 1663.
(25) Gong, C. X.; Liu, F.; Grundke-Iqbal, I.; Iqbal, K. *J Neural Transm* **2005**, *112*, 813.
(26) Russell, C. L.; Koncarevic, S.; Ward, M. A. *J Alzheimers Dis* **2014**, *41*, 345.
(27) Karve, T. M.; Cheema, A. K. *J Amino Acids* **2011**, *2011*, 207691.
(28) Davis, B. G. *Science* **2004**, *303*, 480.
(29) de Graaf, A. J.; Kooijman, M.; Hennink, W. E.; Mastrobattista, E. *Bioconjug Chem* **2009**, *20*, 1281.
(30) Kalkhof, S.; Sinz, A. *Anal Bioanal Chem* **2008**, *392*, 305.
(31) Jentoft, N.; Dearborn, D. G. *J Biol Chem* **1979**, *254*, 4359.
(32) Jentoft, N.; Dearborn, D. G. *Anal Biochem* **1980**, *106*, 186.
(33) Gildersleeve, J. C.; Oyelaran, O.; Simpson, J. T.; Allred, B. *Bioconjug Chem* **2008**, *19*, 1485.
(34) McFarland, J. M.; Francis, M. B. *J Am Chem Soc* **2005**, *127*, 13490.
(35) Schnölzer, M.; Kent, S. B. *Science* **1992**, *256*, 221.
(36) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776.
(37) Kent, S. B. *Chem Soc Rev* **2009**, *38*, 338.
(38) Kochendoerfer, G. G.; Chen, S. Y.; Mao, F.; Cressman, S.; Traviglia, S.; Shao, H.; Hunter, C. L.; Low, D. W.; Cagle, E. N.; Carnevali, M.; Gueriguian, V.; Keogh, P. J.; Porter, H.; Stratton, S. M.; Wiedeke, M. C.; Wilken, J.; Tang, J.; Levy, J. J.; Miranda, L. P.; Crnogorac, M. M.; Kalbag, S.; Botti, P.; Schindler-Horvat, J.; Savatski, L.; Adamson, J. W.; Kung, A.; Kent, S. B.; Bradburne, J. A. *Science* **2003**, *299*, 884.
(39) Bang, D.; Kent, S. B. *Angew Chem Int Ed Engl* **2004**, *43*, 2534.
(40) Torbeev, V. Y.; Kent, S. B. *Angew Chem Int Ed Engl* **2007**, *46*, 1667.
(41) Bernardes, G. J.; Gamblin, D. P.; Davis, B. G. *Angew Chem Int Ed Engl* **2006**, *45*, 4007.
(42) Gamblin, D. P.; Garnier, P.; van Kasteren, S.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G. *Angew Chem Int Ed Engl* **2004**, *43*, 828.
(43) Li, Q. F.; Yang, Y.; Maleckis, A.; Otting, G.; Su, X. C. *Chem Commun (Camb)* **2012**, *48*, 2704.
(44) Jones, M. W.; Mantovani, G.; Ryan, S. M.; Wang, X.; Brayden, D. J.; Haddleton, D. M. *Chem Commun (Camb)* **2009**, 5272.
(45) Lo Conte, M.; Staderini, S.; Marra, A.; Sanchez-Navarro, M.; Davis, B. G.; Dondoni, A. *Chem Commun (Camb)* **2011**, *47*, 11086.
(46) Minozzi, M.; Monesi, A.; Nanni, D.; Spagnolo, P.; Marchetti, N.; Massi, A. *J Org Chem* **2011**, *76*, 450.
(47) Stephanopoulos, N.; Tong, G. J.; Hsiao, S. C.; Francis, M. B. *ACS Nano* **2010**, *4*, 6014.
(48) Kapanidis, A. N.; Weiss, S. *The Journal of Chemical Physics* **2002**, *117*, 10953.
(49) Rup, B.; O'Hara, D. *AAPS J* **2007**, *9*, E148.

(50) Chalker, J. M.; Bernardes, G. J.; Davis, B. G. *Acc Chem Res* **2011**, *44*, 730.
(51) Sletten, E. M.; Bertozzi, C. R. *Angew Chem Int Ed Engl* **2009**, *48*, 6974.
(52) Prescher, J. A.; Bertozzi, C. R. *Nat Chem Biol* **2005**, *1*, 13.
(53) Lemieux, G. A.; De Graffenried, C. L.; Bertozzi, C. R. *J Am Chem Soc* **2003**, *125*, 4708.
(54) Naganathan, S.; Ye, S.; Sakmar, T. P.; Huber, T. *Biochemistry* **2013**, *52*, 1028.
(55) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. *Org Lett* **2000**, *2*, 1939.
(56) Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007.
(57) Seitchik, J. L.; Peeler, J. C.; Taylor, M. T.; Blackman, M. L.; Rhoads, T. W.; Cooley, R. B.; Refakis, C.; Fox, J. M.; Mehl, R. A. *J Am Chem Soc* **2012**, *134*, 2898.
(58) Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W. *J Am Chem Soc* **2012**, *134*, 10317.
(59) Song, W.; Wang, Y.; Qu, J.; Lin, Q. *J Am Chem Soc* **2008**, *130*, 9654.
(60) Yu, Z.; Pan, Y.; Wang, Z.; Wang, J.; Lin, Q. *Angew Chem Int Ed Engl* **2012**, *51*, 10600.
(61) Spicer, C. D.; Davis, B. G. *Chem Commun (Camb)* **2013**, *49*, 2747.
(62) Spicer, C. D.; Davis, B. G. *Chemical Communications* **2011**, *47*, 1698.
(63) Liu, H.; Wang, L.; Brock, A.; Wong, C. H.; Schultz, P. G. *J Am Chem Soc* **2003**, *125*, 1702.
(64) Kim, C. H.; Axup, J. Y.; Dubrovska, A.; Kazane, S. A.; Hutchins, B. A.; Wold, E. D.; Smider, V. V.; Schultz, P. G. *J Am Chem Soc* **2012**, *134*, 9918.
(65) Chalker, J. M.; Lercher, L.; Rose, N. R.; Schofield, C. J.; Davis, B. G. *Angewandte Chemie International Edition* **2012**, *51*, 1835.
(66) Bernardes, G. J.; Chalker, J. M.; Errey, J. C.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130*, 5052.
(67) Liu, J.; Xu, Y.; Stoleru, D.; Salic, A. *Proc Natl Acad Sci U S A* **2012**, *109*, 413.
(68) Tom Dieck, S.; Müller, A.; Nehring, A.; Hinz, F. I.; Bartnik, I.; Schuman, E. M.; Dieterich, D. C. *Curr Protoc Cell Biol* **2012**, *Chapter 7*, Unit7.11.
(69) Lin, Y. A.; Chalker, J. M.; Davis, B. G. *J. Am. Chem. Soc.* **2010**, *132*, 16805.
(70) Lin, Y. A.; Boutureira, O.; Lercher, L.; Bhushan, B.; Paton, R. S.; Davis, B. G. *J. Am. Chem. Soc.* **2013**, *135*, 12156.
(71) Himo, F.; Lovell, T.; Hilgraf, R.; Rostovtsev, V. V.; Noodleman, L.; Sharpless, K. B.; Fokin, V. V. *J Am Chem Soc* **2005**, *127*, 210.
(72) Tornøe, C. W.; Christensen, C.; Meldal, M. *J Org Chem* **2002**, *67*, 3057.
(73) Gothelf, K. V.; Jørgensen, K. A. *Chem Rev* **1998**, *98*, 863.
(74) Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J Am Chem Soc* **2003**, *125*, 3192.
(75) Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G. *J Am Chem Soc* **2003**, *125*, 11782.
(76) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 19.
(77) van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 1282.
(78) Link, A. J.; Vink, M. K. S.; Tirrell, D. A. *Nat. Protocols* **2007**, *2*, 1879.
(79) van Kasteren, S. I.; Kramer, H. B.; Jensen, H. H.; Campbell, S. J.; Kirkpatrick, J.; Oldham, N. J.; Anthony, D. C.; Davis, B. G. *Nature* **2007**, *446*, 1105.
(80) van Kasteren, S. I.; Kramer, H. B.; Gamblin, D. P.; Davis, B. G. *Nat Protoc* **2007**, *2*, 3185.
(81) Fernandez-Gonzalez, M.; Boutureira, O.; Bernardes, G. J. L.; Chalker, J. M.; Young, M. A.; Errey, J. C.; Davis, B. G. *Chemical Science* **2010**, *1*, 709.
(82) Deiters, A.; Cropp, T. A.; Summerer, D.; Mukherji, M.; Schultz, P. G. *Bioorg Med Chem Lett* **2004**, *14*, 5743.

(83) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proceedings of the National Academy of Sciences* **2006**, *103*, 9482.
(84) Link, A. J.; Tirrell, D. A. *J Am Chem Soc* **2003**, *125*, 11164.
(85) Beatty, K. E.; Xie, F.; Wang, Q.; Tirrell, D. A. *J Am Chem Soc* **2005**, *127*, 14150.
(86) Dieterich, D. C.; Hodas, J. J.; Gouzer, G.; Shadrin, I. Y.; Ngo, J. T.; Triller, A.; Tirrell, D. A.; Schuman, E. M. *Nat Neurosci* **2010**, *13*, 897.
(87) Kennedy, D. C.; McKay, C. S.; Legault, M. C.; Danielson, D. C.; Blake, J. A.; Pegoraro, A. F.; Stolow, A.; Mester, Z.; Pezacki, J. P. *J Am Chem Soc* **2011**, *133*, 17993.
(88) Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. *Bioconjug Chem* **2010**, *21*, 1912.
(89) Soriano Del Amo, D.; Wang, W.; Jiang, H.; Besanceney, C.; Yan, A. C.; Levy, M.; Liu, Y.; Marlow, F. L.; Wu, P. *J Am Chem Soc* **2010**, *132*, 16893.
(90) Wang, W.; Hong, S.; Tran, A.; Jiang, H.; Triano, R.; Liu, Y.; Chen, X.; Wu, P. *Chem Asian J* **2011**, *6*, 2796.
(91) Wittig, G.; Krebs, A. *Chemische Berichte* **1961**, *94*, 3260.
(92) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J Am Chem Soc* **2004**, *126*, 15046.
(93) Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. *Proc Natl Acad Sci U S A* **2007**, *104*, 16793.
(94) Ning, X.; Guo, J.; Wolfert, M. A.; Boons, G. J. *Angew Chem Int Ed Engl* **2008**, *47*, 2253.
(95) Jewett, J. C.; Sletten, E. M.; Bertozzi, C. R. *J Am Chem Soc* **2010**, *132*, 3688.
(96) Dommerholt, J.; Schmidt, S.; Temming, R.; Hendriks, L. J.; Rutjes, F. P.; van Hest, J. C.; Lefeber, D. J.; Friedl, P.; van Delft, F. L. *Angew Chem Int Ed Engl* **2010**, *49*, 9422.
(97) Beatty, K. E.; Fisk, J. D.; Smart, B. P.; Lu, Y. Y.; Szychowski, J.; Hangauer, M. J.; Baskin, J. M.; Bertozzi, C. R.; Tirrell, D. A. *Chembiochem* **2010**, *11*, 2092.
(98) Zheng, S.; Zhang, G.; Li, J.; Chen, P. R. *Angew Chem Int Ed Engl* **2014**, *53*, 6449.
(99) Baskin, J. M.; Dehnert, K. W.; Laughlin, S. T.; Amacher, S. L.; Bertozzi, C. R. *Proc Natl Acad Sci U S A* **2010**, *107*, 10360.
(100) Laughlin, S. T.; Baskin, J. M.; Amacher, S. L.; Bertozzi, C. R. *Science* **2008**, *320*, 664.
(101) Brüchner, P.; Koch, D.; Voigtmann, U.; Blechert, S. *Synthetic Communications* **2007**, *37*, 2757.
(102) Clark, T. D.; Ghadiri, M. R. *J. Am. Chem. Soc.* **1995**, *117*, 12364.
(103) E. Gibson, S.; C. Gibson, V.; P. Keen, S. *Chemical Communications* **1997**, 1107.
(104) C. G. Biagini, S.; E. Gibson, S.; P. Keen, S. *Journal of the Chemical Society, Perkin Transactions 1* **1998**, 2485.
(105) McGarvey, G. J.; Benedum, T. E.; Schmidtmann, F. W. *Org Lett* **2002**, *4*, 3591.
(106) Hong, S. H.; Grubbs, R. H. *J. Am. Chem. Soc.* **2006**, *128*, 3508.
(107) Gułajski, Ł.; Michrowska, A.; Narożnik, J.; Kaczmarska, Z.; Rupnicki, L.; Grela, K. *ChemSusChem* **2008**, *1*, 103.
(108) Binder, J. B.; Guzei, I. A.; Raines, R. T. *Adv Synth Catal* **2007**, *349*, 395.
(109) Skowerski, K.; Szczepaniak, G.; Wierzbicka, C.; Gulajski, L.; Bieniek, M.; Grela, K. *Catalysis Science & Technology* **2012**, *2*, 2424.
(110) Connon, S. J.; Rivard, M.; Zaja, M.; Blechert, S. *Advanced Synthesis & Catalysis* **2003**, *345*, 572.
(111) Binder, J. B.; Blank, J. J.; Raines, R. T. *Org Lett* **2007**, *9*, 4885.
(112) Lipshutz, B. H.; Aguinaldo, G. T.; Ghorai, S.; Voigtritter, K. *Org Lett* **2008**, *10*, 1325.
(113) Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. J.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130*, 9642.
(114) Hoye, T. R.; Zhao, H. *Org Lett* **1999**, *1*, 1123.
(115) Nicolaou, K. C.; Leung, G. Y.; Dethe, D. H.; Guduru, R.; Sun, Y. P.; Lim, C. S.; Chen, D. Y. *J Am Chem Soc* **2008**, *130*, 10019.

(116) Imahori, T.; Ojima, H.; Yoshimura, Y.; Takahata, H. *Chemistry* **2008**, *14*, 10762.
(117) Dinger, M. B.; Mol, J. C. *Organometallics* **2003**, *22*, 1089.
(118) Werner, H.; Grünwald, C.; Stüer, W.; Wolf, J. *Organometallics* **2003**, *22*, 1558.
(119) Wang, J.; Schiller, S. M.; Schultz, P. G. *Angew Chem Int Ed Engl* **2007**, *46*, 6849.
(120) Weiner, H.; White, W. N.; Hoare, D. G.; Koshland, D. E. *J Am Chem Soc* **1966**, *88*, 3851.
(121) Strumeyer, D. H.; White, W. N.; Koshland, D. E. *Proc Natl Acad Sci U S A* **1963**, *50*, 931.
(122) Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernandez-Gonzalez, M.; Bernardes, G. J. L.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G. *Chemical Science* **2011**, *2*, 1666.
(123) Chalker, J. M.; Lin, Y. A.; Boutureira, O.; Davis, B. G. *Chem Commun* **2009**, 3714.
(124) Fürstner, A.; Langemann, K. *J. Am. Chem. Soc.* **1997**, *119*, 9130.
(125) Yang, Q.; Xiao, W. J.; Yu, Z. *Org Lett* **2005**, *7*, 871.
(126) Cochrane, S. A.; Huang, Z.; Vederas, J. C. *Org Biomol Chem* **2013**, *11*, 630.
(127) Ai, H. W.; Shen, W.; Brustad, E.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2010**, *49*, 935.
(128) Liu, C. C.; Schultz, P. G. *Annu Rev Biochem* **2010**, *79*, 413.
(129) Young, T. S.; Schultz, P. G. *J Biol Chem* **2010**, *285*, 11039.
(130) Budisa, N. *Angew Chem Int Ed Engl* **2004**, *43*, 6426.
(131) Hartman, M. C.; Josephson, K.; Lin, C. W.; Szostak, J. W. *PLoS One* **2007**, *2*, e972.
(132) Datta, D.; Vaidehi, N.; Zhang, D.; Goddard, W. A. *Protein Sci.* **2004**, *13*, 2693.
(133) Wood, W. B. *J Mol Biol* **1966**, *16*, 118.
(134) Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Lett* **2001**, *502*, 25.
(135) Ibba, M.; Hennecke, H. *FEBS Lett* **1995**, *364*, 272.
(136) Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A. *FEBS Lett* **2000**, *467*, 37.
(137) van Hest, J. C. M.; Tirrell, D. A. *FEBS Letters* **1998**, *428*, 68.
(138) Nepomniaschiy, N.; Grimminger, V.; Cohen, A.; DiGiovanni, S.; Lashuel, H. A.; Brik, A. *Org Lett* **2008**, *10*, 5243.
(139) Baboo, S.; Bhushan, B.; Jiang, H.; Grovenor, C. R. M.; Pierre, P.; Davis, B. G.; Cook, P. R. *PLoS ONE* **2014**, *9*, e99346.
(140) Howden, A. J.; Geoghegan, V.; Katsch, K.; Efstathiou, G.; Bhushan, B.; Boutureira, O.; Thomas, B.; Trudgian, D. C.; Kessler, B. M.; Dieterich, D. C.; Davis, B. G.; Acuto, O. *Nature Methods* **2013**, *10*, 343.
(141) Banerjee, P. S.; Ostapchuk, P.; Hearing, P.; Carrico, I. S. *J Virol* **2011**, *85*, 7546.
(142) Taskent-Sezgin, H.; Chung, J.; Banerjee, P. S.; Nagarajan, S.; Dyer, R. B.; Carrico, I.; Raleigh, D. P. *Angew Chem Int Ed Engl* **2010**, *49*, 7473.
(143) Merkel, L.; Beckmann, H. S.; Wittmann, V.; Budisa, N. *Chembiochem* **2008**, *9*, 1220.
(144) Ma, Y.; Biava, H.; Contestabile, R.; Budisa, N.; di Salvo, M. L. *Molecules* **2014**, *19*, 1004.
(145) Budisa, N.; Minks, C.; Medrano, F. J.; Lutz, J.; Huber, R.; Moroder, L. *Proc Natl Acad Sci U S A* **1998**, *95*, 455.
(146) Tang, Y.; Tirrell, D. A. *Biochemistry* **2002**, *41*, 10635.
(147) Wang, P.; Tang, Y.; Tirrell, D. A. *J Am Chem Soc* **2003**, *125*, 6900.
(148) Döring, V.; Mootz, H. D.; Nangle, L. A.; Hendrickson, T. L.; de Crécy-Lagard, V.; Schimmel, P.; Marlière, P. *Science* **2001**, *292*, 501.
(149) Porter, T. H.; Smith, S. C.; Shive, W. *Arch Biochem Biophys* **1977**, *179*, 266.
(150) Ikeda, Y.; Kawahara, S.; Taki, M.; Kuno, A.; Hasegawa, T.; Taira, K. *Protein Eng* **2003**, *16*, 699.
(151) Lu, P.; Jarema, M.; Mosser, K.; Daniel, W. E. *Proc Natl Acad Sci U S A* **1976**, *73*, 3471.

(152) Budisa, N.; Alefelder, S.; Bae, J. H.; Golbik, R.; Minks, C.; Huber, R.; Moroder, L. *Protein Sci* **2001**, *10*, 1281.
(153) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471.
(154) Alberts, B.; 5th ed.; Garland Science,: New York ; Abingdon, 2008.
(155) Alff-Steinberger, C.; Epstein, R. *J Theor Biol* **1994**, *168*, 461.
(156) Mendel, D.; Cornish, V. W.; Schultz, P. G. *Annu Rev Biophys Biomol Struct* **1995**, *24*, 435.
(157) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498.
(158) Liu, D. R.; Schultz, P. G. *Proc Natl Acad Sci U S A* **1999**, *96*, 4780.
(159) Miyake-Stoner, S. J.; Miller, A. M.; Hammill, J. T.; Peeler, J. C.; Hess, K. R.; Mehl, R. A.; Brewer, S. H. *Biochemistry* **2009**, *48*, 5953.
(160) Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G. *Angew Chem Int Ed Engl* **2009**, *48*, 4052.
(161) Greiss, S.; Chin, J. W. *J Am Chem Soc* **2011**, *133*, 14196.
(162) Parrish, A. R.; She, X.; Xiang, Z.; Coin, I.; Shen, Z.; Briggs, S. P.; Dillin, A.; Wang, L. *ACS Chem Biol* **2012**, *7*, 1292.
(163) Bianco, A.; Townsley, F. M.; Greiss, S.; Lang, K.; Chin, J. W. *Nat Chem Biol* **2012**, *8*, 748.
(164) Peters, F. B.; Brock, A.; Wang, J.; Schultz, P. G. *Chem Biol* **2009**, *16*, 148.
(165) Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. *Nat Struct Biol* **2003**, *10*, 425.
(166) Turner, J. M.; Graziano, J.; Spraggon, G.; Schultz, P. G. *Proc Natl Acad Sci U S A* **2006**, *103*, 6483.
(167) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J Am Chem Soc* **2002**, *124*, 9026.
(168) Deiters, A.; Schultz, P. G. *Bioorg Med Chem Lett* **2005**, *15*, 1521.
(169) Wang, J.; Zhang, W.; Song, W.; Wang, Y.; Yu, Z.; Li, J.; Wu, M.; Wang, L.; Zang, J.; Lin, Q. *J Am Chem Soc* **2010**, *132*, 14812.
(170) Xie, J.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G. *Nat Biotechnol* **2004**, *22*, 1297.
(171) Brustad, E.; Bushey, M. L.; Lee, J. W.; Groff, D.; Liu, W.; Schultz, P. G. *Angew Chem Int Ed Engl* **2008**, *47*, 8220.
(172) Liu, C. C.; Schultz, P. G. *Nat Biotechnol* **2006**, *24*, 1436.
(173) Miyake-Stoner, S. J.; Refakis, C. A.; Hammill, J. T.; Lusic, H.; Hazen, J. L.; Deiters, A.; Mehl, R. A. *Biochemistry* **2010**, *49*, 1667.
(174) Jackson, J. C.; Hammill, J. T.; Mehl, R. A. *J Am Chem Soc* **2007**, *129*, 1160.
(175) Taskent-Sezgin, H.; Chung, J.; Patsalo, V.; Miyake-Stoner, S. J.; Miller, A. M.; Brewer, S. H.; Mehl, R. A.; Green, D. F.; Raleigh, D. P.; Carrico, I. *Biochemistry* **2009**, *48*, 9040.
(176) Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G. *J Am Chem Soc* **2006**, *128*, 13984.
(177) Chatterjee, A.; Guo, J.; Lee, H. S.; Schultz, P. G. *J Am Chem Soc* **2013**, *135*, 12540.
(178) Wang, J.; Xie, J.; Schultz, P. G. *J Am Chem Soc* **2006**, *128*, 8738.
(179) Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. *Proc Natl Acad Sci U S A* **2006**, *103*, 9785.
(180) Köhrer, C.; Xie, L.; Kellerer, S.; Varshney, U.; RajBhandary, U. L. *Proc Natl Acad Sci U S A* **2001**, *98*, 14310.
(181) Anderson, J. C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 9598.
(182) Köhrer, C.; Yoo, J. H.; Bennett, M.; Schaack, J.; RajBhandary, U. L. *Chem Biol* **2003**, *10*, 1095.

(183) Kim, J.; Seo, M. H.; Lee, S.; Cho, K.; Yang, A.; Woo, K.; Kim, H. S.; Park, H. S. *Anal Chem* **2013**, *85*, 1468.
(184) Wu, B.; Wang, Z.; Huang, Y.; Liu, W. R. *Chembiochem* **2012**, *13*, 1405.
(185) Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. *Proc Natl Acad Sci U S A* **2004**, *101*, 7566.
(186) Rackham, O.; Chin, J. W. *Nat Chem Biol* **2005**, *1*, 159.
(187) Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat Biotechnol* **2007**, *25*, 770.
(188) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, *464*, 441.
(189) Sachdeva, A.; Wang, K.; Elliott, T.; Chin, J. W. *J Am Chem Soc* **2014**, *136*, 7785.

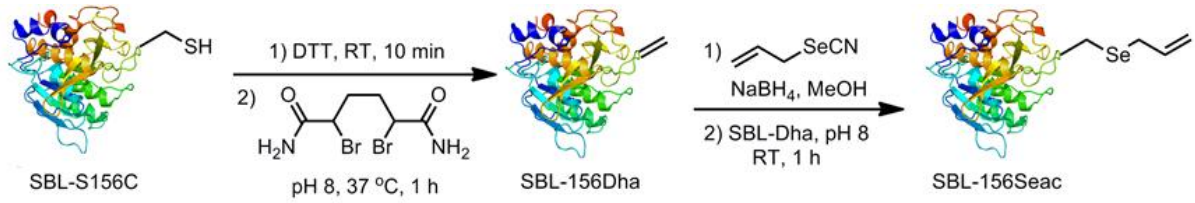
Chapter 2

Development of the Cross-Metathesis Reaction on Proteins: From Chemical to Genetic Incorporation of Metathesis Reactive Tags

2.1 Introduction

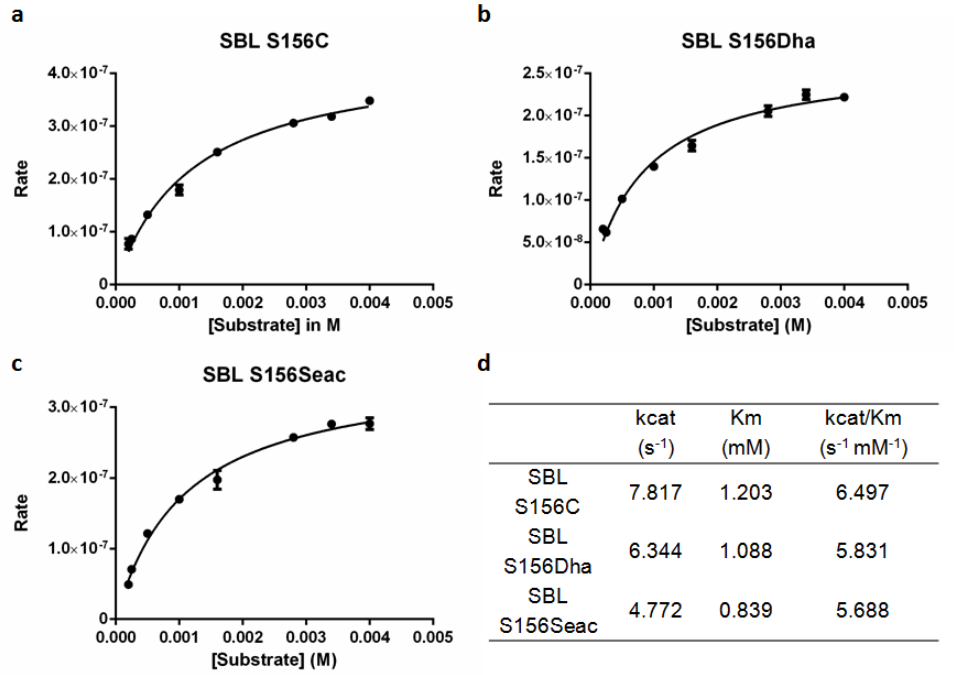
Olefin cross-metathesis (CM) is a viable reaction for the modification of alkene-containing proteins. The alkene functional group can be installed into proteins by incorporation of some olefin containing UAAs.[1-4] Though aliphatic alkene-containing UAAs such as homoallylglycine (Hag, **1.16**)[4] are both self-metathesis and cross-metathesis reactive as monomeric protected amino acids in organic solvents[5] and can be metabolically incorporated into proteins, they are completely unreactive in olefin cross-metathesis (CM) reactions in aqueous media.[6] The chemically installed UAA *S*-allylcysteine (Sac, **2.1**) enables CM on proteins, where the allylic chalcogen heteroatom provides crucial coordination to the metal catalyst centre (See Chapter 1, Section **1.5**).[6]

The chemical installation of *Se*-allylselenocysteine (Seac, **2.2**) onto proteins was carried out in collaboration with Dr. Yuya A. Lin (YAL), Dr. Omar Boutureira and Lukas Lercher. Seac installation was found to enhance the rate of CM, and allowed for an increased breadth of metathesis partners compared to Sac, such as allyl acetamides.[7,8] Seac was found to be CM-reactive with allyl alcohol as a model metathesis partner (Section **2.11.4**). Seac-incorporated proteins were prepared by the conversion of the single cysteine mutant of the protein subtilisin from *Bacillus lentus* (SBL S156C) to its corresponding dehydroalanine variant (SBL-156Dha) by a previously reported Bis-alkylation - elimination protocol.[9] This was converted to SBL-156Seac by the addition of allyl selenolate, generated in situ by the reaction of allyl selenocyanate with stoichiometric amounts of sodium borohydride in degassed methanol (Scheme 2.1).



Scheme 2.1: Two-step synthesis of SBL-156Seac from SBL-S156C

In order to confirm that the conversion of SBL S156C to the corresponding Dha and Seac analogues did not cause an appreciable loss of enzymatic activity, the enzyme kinetic parameters k_{cat} and K_m were determined with Succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide as a substrate (Fig. 2.1, Section **2.11.5** for full protocol).[10]



	kcat (s^{-1})	Km (mM)	kcat/Km (s^{-1} mM^{-1})
SBL S156C	7.817	1.203	6.497
SBL S156Dha	6.344	1.088	5.831
SBL S156Seac	4.772	0.839	5.688

Fig. 2.1: Michaelis-menten plots of the hydrolysis of Suc-AAPF-pNA by (a) SBL S156C , (b) SBL S156Dha, and (c) SBL S156Seac. Panel (d) summarises the kinetics parameters of these reactions.

The results were seen to show that conversion of residue Cys 156 to Dha, and then to Seac did not lead to an appreciable loss in enzymatic activity. With the substrate in hand, CM reactions were carried out by (YAL) with allyl alcohol as a model CM-reactive partner using previously reported CM conditions.[11] Briefly, a solution of the Hoveyda-Grubbs 2nd generation catalyst **(1.19)** in t-BuOH (43 µL of a solution of 0.6 mg in 287 µL t-BuOH) was added to a solution of SBL-156Seac (10.4 µM, 100 µL of 0.4 mg/mL solution) in 50 mM sodium phosphate buffer (pH 5) with $MgCl_2 \cdot 6H_2O$ (0.7 µL of a 1 mg/µL solution in water,

3.7 μmol). The Reaction was shaken at RT for 1 minute before adding allyl alcohol (0.5 μL, 7.5 μmol). Reaction was mixed on a vortex immediately after addition and was shaken at RT. At various time points, the reaction was passed through a PD SpinTrap G-25 (GE Healthcare) to remove small molecules, and the resulting solution analyzed by LC–MS.

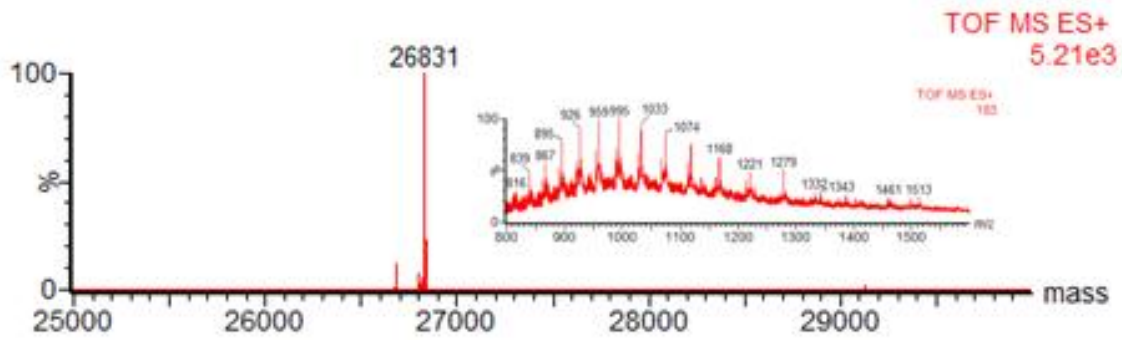


Fig. 2.2. ESI–MS spectrum of CM between SBL-156Seac and allyl alcohol. Product mass calculated = 26832 Da, Observed = 26831 Da

The reaction showed >95% conversion to the CM product within 15 minutes. CM was also successfully demonstrated by YAL with *O*-allyl-2-acetamido-2-deoxy-*β*-D-glucopyranoside as well as allyl acetamide, previously reported to be CM-unreactive with Sac-incorporated proteins.[7,8]

However, this reaction protocol was subsequently discovered to not be generally applicable, as proteins were discovered to precipitate in the presence of 30% t-BuOH. Efforts towards designing a general CM protocol for on-protein reactions will be described in section **2.7**. Furthermore, in order to expand the applicability of CM to complex biological systems, efforts were directed towards designing genetically incorporating metathesis reactive tags into proteins. These will be discussed next.

2.2: Design of UAAs for Genetic Incorporation of CM-reactive tags on proteins

Schultz and co-workers have reported the incorporation of *O*-crotyl serine(**1.26**) in yeast via amber stop codon suppression, and demonstrated its reactivity in on-protein intramolecular ring closing metathesis (RCM),[12] however CM with genetically incorporated residues has yet to be demonstrated. The alternative 'sense' codon reassignment method was used for UAA incorporation, utilising methionine auxotrophic cell lines. Here, the flexibility of methionyl-

tRNA synthetase (MetRS) was probed in order to metabolically label proteins with allyl chalcogen containing UAAs. [13]

Fersht and Dingwall first demonstrated that L-ethionine may be mischarged by the MetRS without being metabolically edited.[14] In contrast, L-homocysteine, a natural competitor of methionine, is transformed to the corresponding thiolactone to avoid being misread in translation.[15] Side chain length and heteroatom position therefore may play a critical role in metabolic recognition. We explored key aspects of the tolerance of MetRS toward positioning of the key side-chain heteroatoms required for CM with concomitant side chain length alteration (Fig. 2.3).

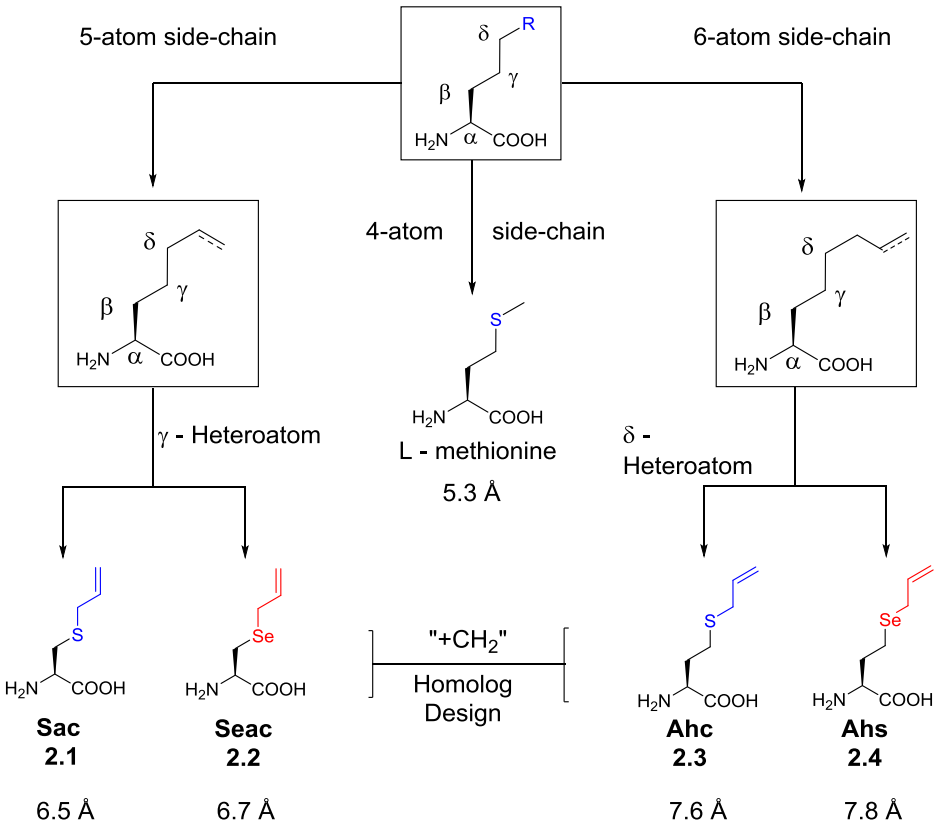
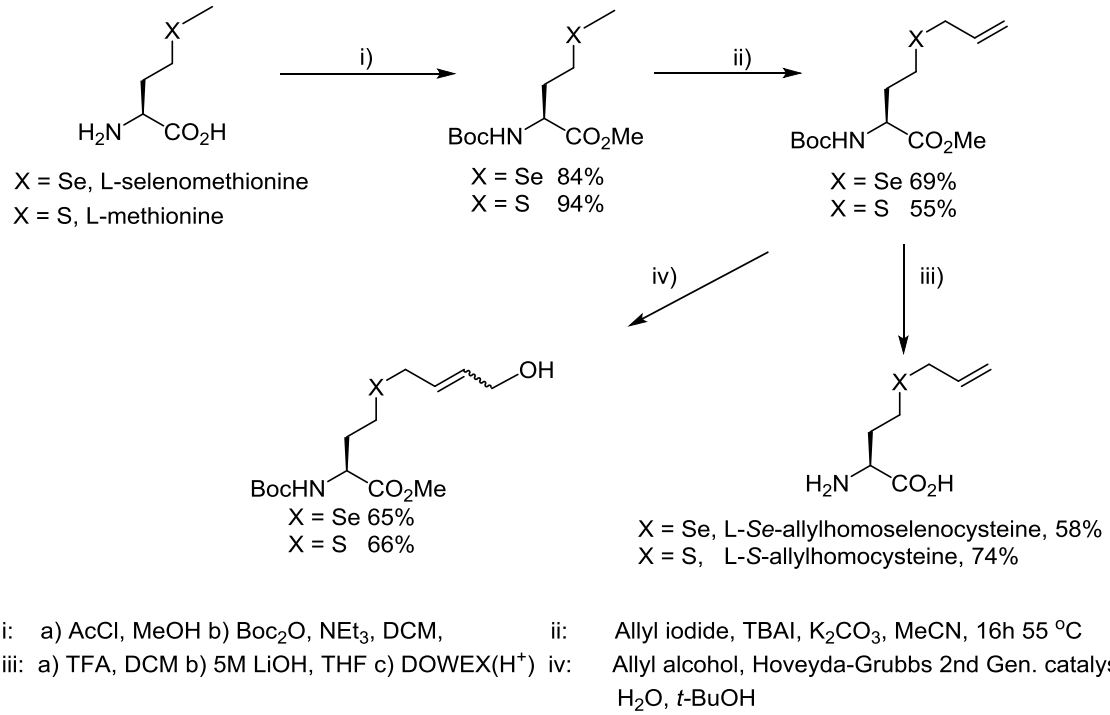


Figure 2.3. Design of the γ-heteroatom containing methionine analogues *S*-allylcysteine (Sac, **2.1**) and *Se*-allylselenocysteine (Seac, **2.2**), and the δ-heteroatom containing analogues *S*-allylhomocysteine (Ahc, **2.3**) and *Se*-allyl homoselenocysteine (Ahs, **2.4**). The through-space side chain length from Cα to the side chain terminus is displayed below each structure.

Two types of substrates were designed and synthesised, those with a γ-heteroatom (**2.1, 2.2**) and those with a δ-heteroatom (**2.3**, **2.4**). As the selenium analogue of Sac was found to have an enhanced rate of reaction in aqueous CM, the metabolic incorporation efficiencies of

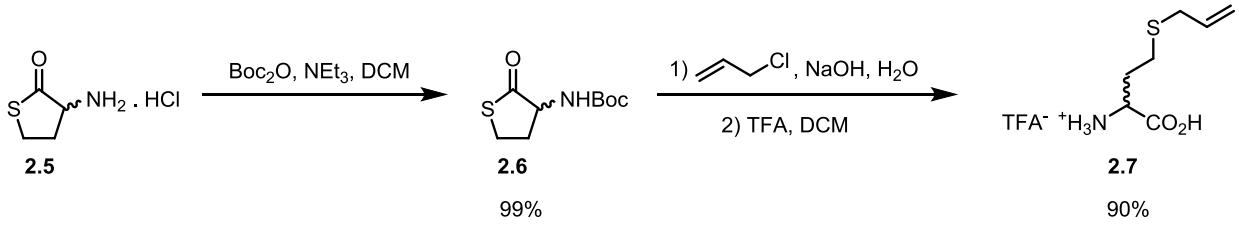
selenium analogues **2.2** and **2.4** were also considered. As none possessed a free thiol in their side chains, unlike L-homocysteine, none of the amino acids chosen here could be transformed into their corresponding chalcogen lactones.

2.3 and **2.4** were synthesised from L-methionine and L-selenomethionine respectively using a novel demethylative allylation strategy that proved direct and efficient (Scheme 2.2). CM could also be demonstrated on both (Kieran Hudson Master's Thesis).



Scheme 2.2: Synthesis of and CM reaction on *S*-allylhomocysteine (**2.3**) and Se-allylhomo-selenocysteine (**2.4**)

2.3 could also be readily synthesised from the commercially available DL-homocysteine thiolactone as a racemate (Scheme 2.3).



Scheme 2.3: Synthesis of racemic *S*-allylhomocysteine

In order to determine the incorporation efficiency of the synthesised Met analogues, *E. coli* methionyl-tRNA synthetase (MetRS) was expressed in BL21(DE3) cells.[16] The MetRS gene synthesis was custom-ordered from GeneArt, ligated into the pET-303 vector with a C-terminal His-tag for purification via an N-terminal NsiI digestion site and a C-terminal XhoI digestion site. The plasmid was received, amplified in XL-10 Gold cells, and transformed into BL21 (DE3) cells for expression (Fig. 2.4).

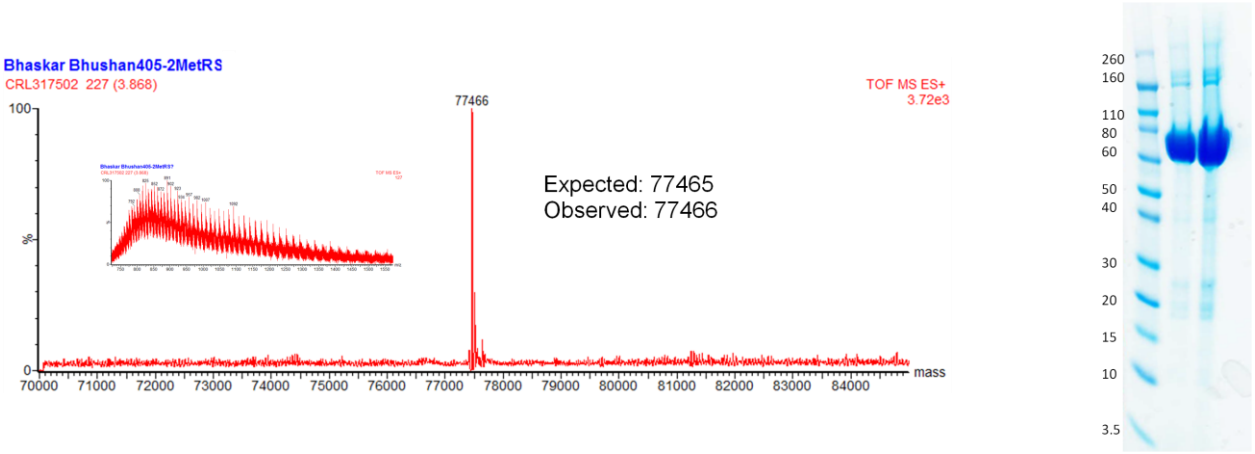
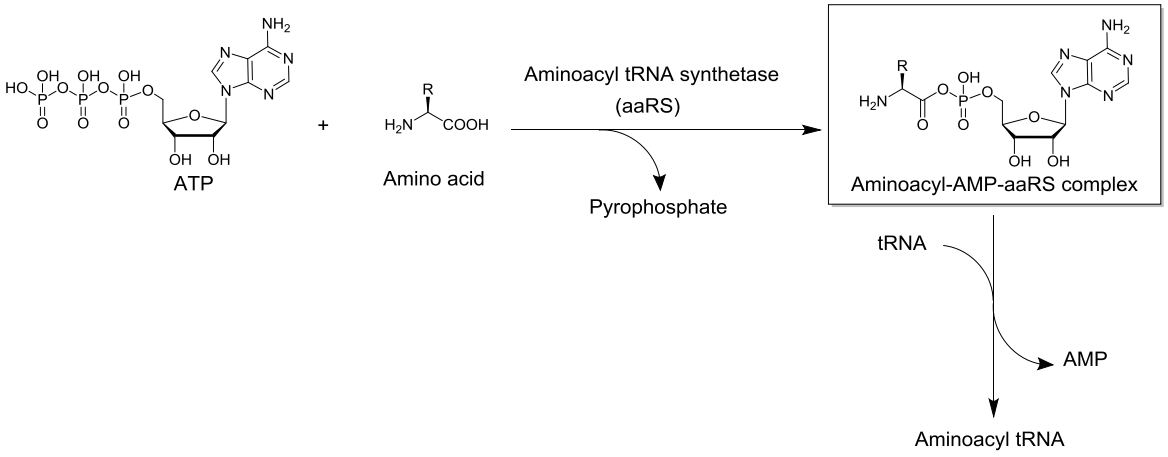


Fig. 2.4: LC-MS spectrum and SDS-PAGE of MetRS

2.3: MetRS Activity Assays

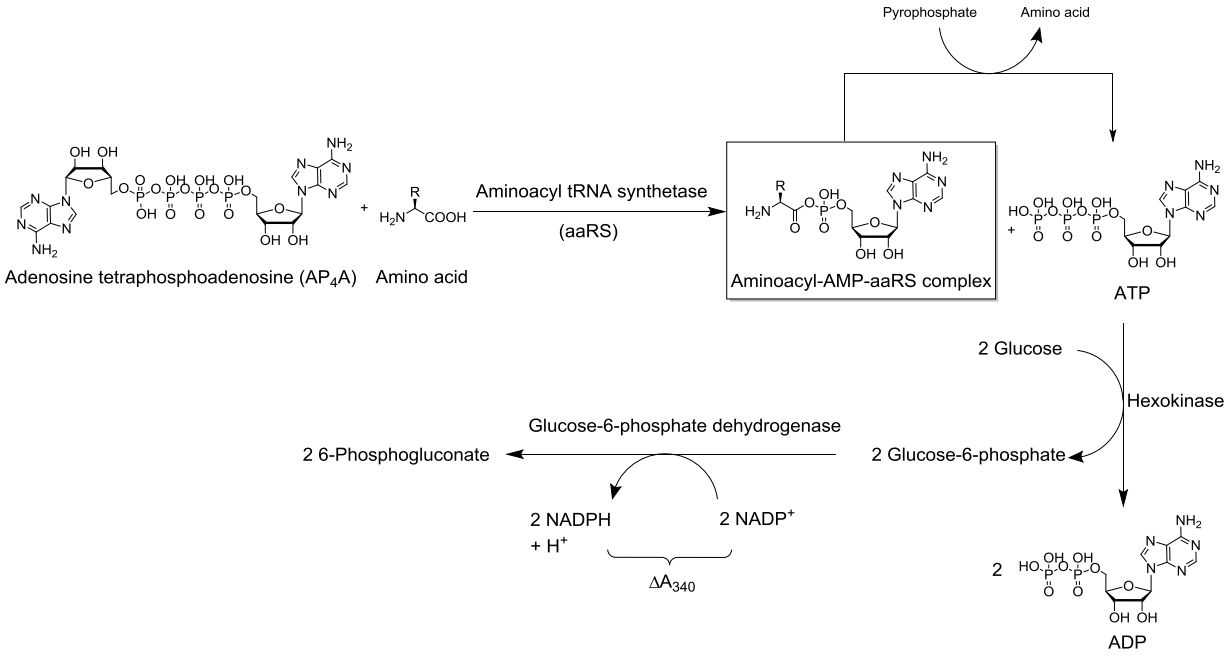
The incorporation efficiency of amino acids and their analogues *in vivo* is controlled to a significant extent by the rate of their activation by the corresponding aminoacyl-tRNA synthetases (aaRS). [4,17,18] Upstream of tRNA aminoacylation, the aaRS first forms an aminoacyl-AMP-aaRS complex upon reacting with ATP and their cognate amino acid substrates. This is then followed by aminoacylation of tRNA with the release of AMP (Scheme 2.4).



Scheme 2.4: Mechanism of aminoacyl tRNA synthetases

2.3.1: AP_4A continuous colourimetric assay

This assay utilises Adenosine tetraphosphoadenosine (AP_4A),[19] which is cleaved by aaRS in the presence of their cognate amino acid substrate to generate ATP and the aminoacyl-AMP-aaRS complex. In the presence of excess pyrophosphate, turnover of this complex occurs to generate another molecule of ATP. Using a coupled enzyme system, hexokinase and glucose-6-phosphate dehydrogenase may then be used to utilise the generated ATP to first phosphorylate glucose to glucose-6-phosphate, then oxidise it to 6-phosphogluconate with the reduction of NADP to NADPH, resulting in an increase in absorbance at 340 nm (A_{340}, Scheme 2.5).



Scheme 2.5: Coupled enzyme assay for determining amino acid recognition by corresponding aminoacyl tRNA synthetase

A preliminary assay was carried out to test the activity of the expressed MetRS towards Methionine at different concentrations (Fig. 2.5, Section **2.11.3** for full protocol):

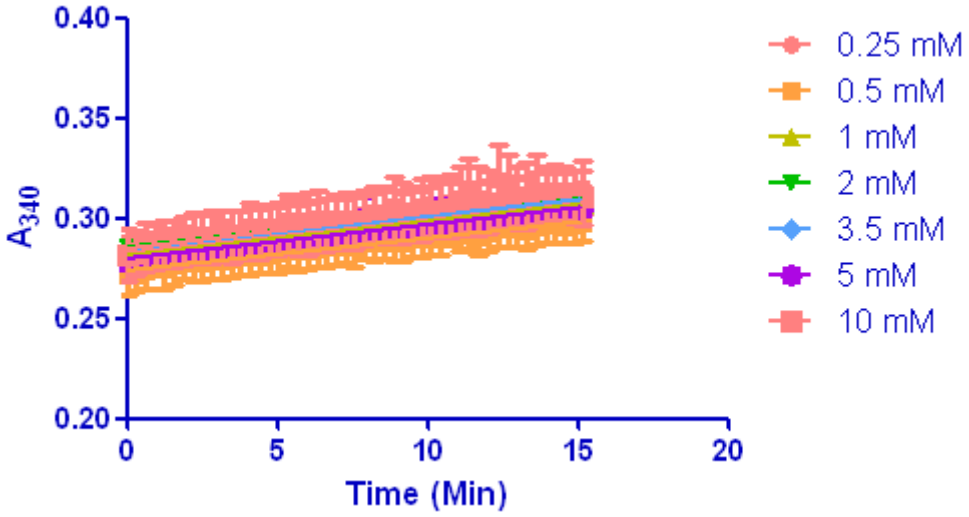


Fig. 2.5: A_{340} v/s time at different [Met]

However, no correlation between [Met] and A_{340} was seen. When different concentrations of ATP were used as a positive control for the coupled assay, there was no correlation seen between [ATP] and ΔA_{340}, suggesting that the hexokinase and/or glucose-6-phosphate dehydrogenase were inactive.

2.3.2: Malachite Green assay

This assay utilises the consumption of ATP and release of pyrophosphate (PPi) by aaRS (Scheme 2.4). The released pyrophosphate may be hydrolysed further to inorganic phosphate by pyrophosphatase (PPiase), and the phosphate release quantified at various time points with Malachite green and ammonium molybdate, leading to an increase in absorbance at 650 nm (A_{650}).[20] In order to show activity of the enzyme, MetRS was buffer exchanged into aminoacylation buffer (Section **2.11.3**), and incubated at 37 ºC with *E. coli* total tRNA, PPiase, ATP, and 1 mM DTT with 1 mM L-Methionine. In the control samples, Methionine and tRNA were absent (due to limiting amounts of the tRNA). After 30 min, reactions were quenched with 100 mM EDTA (50 uL), and developed for 30 min at r.t. with the addition of 10 uL Malachite green solution (containing ammonium molybdate). Absorbance was measured at 650 nm. A strong absorbance at 650 nm was detected in the presence of Met and tRNA, but not in the absence, along with the visual formation of a coloured malachite green-phosphomolybdate complex.

In order to obtain kinetics parameters, the assay was repeated with different concentrations of Met between 0.1 mM and 10 mM, with samples aliquoted every 15 minutes into EDTA to arrest the reaction, then treated with malachite green and ammonium molybdate. However, at all concentrations of Met, the reaction contents were observed to precipitate, leading to inaccurate readings of A_{620}.

In order to identify the problem , an exhaustive factorial analysis of the buffer components was carried out. Wells were prepared with different combinations of the buffer components (Table 2.1), each treated with 15 uL malachite green reagent, and A_{650} measured (Fig. 2.6). No MetRS or methionine was added in any sample.

Well	No.	100 uM ATP	1 mg/mL tRNAs	2 u/mL PPiase	1 mM DTT
A1	1	✓	✓	✓	✓
B1	2	✓	✓	✓	✖
C1	3	✓	✓	✖	✓
D1	4	✓	✓	✖	✖
E1	5	✓	✖	✓	✓
F1	6	✓	✖	✓	✖
G1	7	✓	✖	✖	✓
H1	8	✓	✖	✖	✖
A2	9	✖	✓	✓	✓
B2	10	✖	✓	✓	✖
C2	11	✖	✓	✖	✓
D2	12	✖	✓	✖	✖
E2	13	✖	✖	✓	✓
F2	14	✖	✖	✓	✖
G2	15	✖	✖	✖	✓
H2	16	✖	✖	✖	✖

Table 2.1: Factorial combinations used to determine source of precipitation in malachite green assay

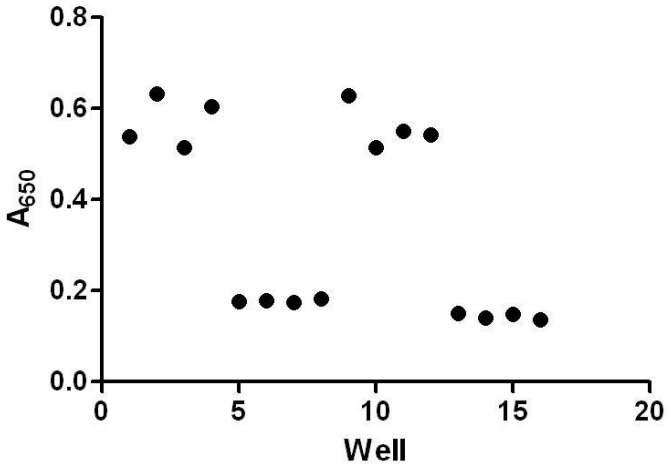


Fig. 2.6: Absorbance at 650 nm for samples as described in Table 2.1

The results showed that all wells containing tRNA at the working concentration of 1 mg/mL, regardless of the presence / absence of other buffer components, showed a high A_{650} in the absence of MetRS or AA.

A 'calibration curve' of tRNA concentrations was made, confirming that the tRNAs were interfering with the detection of phosphate upon treatment with malachite green (Fig. 2.7).

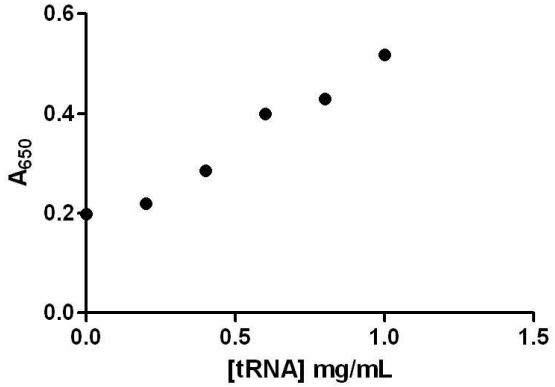


Fig. 2.7: Absorbance at 650 nm for varying concentrations of tRNA.

In order to identify the cause of this interference, tRNAs were treated with a control reagent of conc. H_2SO_4 + ammonium molybdate + Tween-20, i.e. the malachite green working reagent without the malachite green ('empty' reagent), in addition to control samples treated with the complete malachite green reagent, or with HEPES buffer alone. A_{650} values were immediately measured (Fig. 2.8).

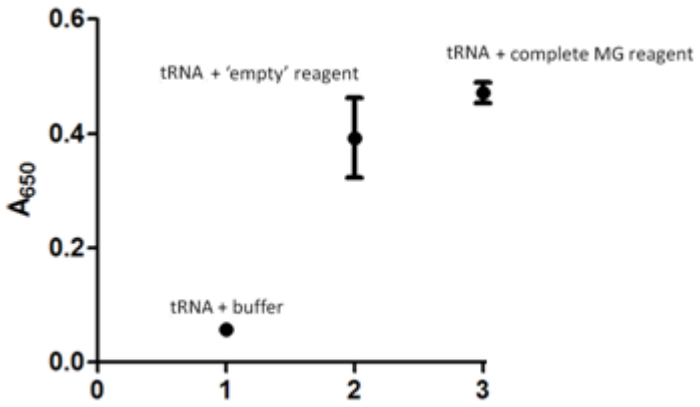


Fig. 2.8: Absorbance at 650 nm for tRNA treated with either 1) buffer alone, 2) Buffer in the presence of H_2SO_4, ammonium molybdate and Tween-20, or 3) with complete malachite green reagent.

The results indicated that the tRNA solution itself was forming a precipitate upon acidification in the presence of sulfuric acid. The acidic pH (~pH 2.0) was necessary to quench the aminoacylation reaction, as well as to keep malachite green itself in solution, as the reagent is nearly insoluble at pH 7. However, this seems to be precipitating the tRNA, thus interfering with the colourimetric assay itself.

2.3.3: Radiometric activity assay

MetRS reacts with ATP and a substrate AA to form a MetRS-AA-adenylate complex. In the presence of excess radioactive ^{32}P-pyrophosphate, the reverse reaction is catalysed to generate radioactive ^{32}P-ATP, the concentration of which increases with time.[4,21] Acidified activated charcoal may then be used to quench the reaction and adsorb the ^{32}P ATP, but not ^{32}P-pyrophosphate. The progress of the reaction may then be monitored by measuring the amount of radiation adsorbed onto charcoal at various time points.

To determine which Met analogues were recognised by MetRS, the enzyme was buffer exchanged into reaction buffer (See Methods Section), to which ^{32}P-PPi (in the form of sodium pyrophosphate with a specific activity of 0.5 TBq/mol) was added to a final concentration of 2 mM, along with 5 mM of either Met, Sac, Seac, Ahc, or Ahs. After 20 min, 20 uL from each reaction was quenched over acidified charcoal (Methods Section), which was washed twice and then transferred to a scintillation vial for scintillation counting (Fig. 2.9).

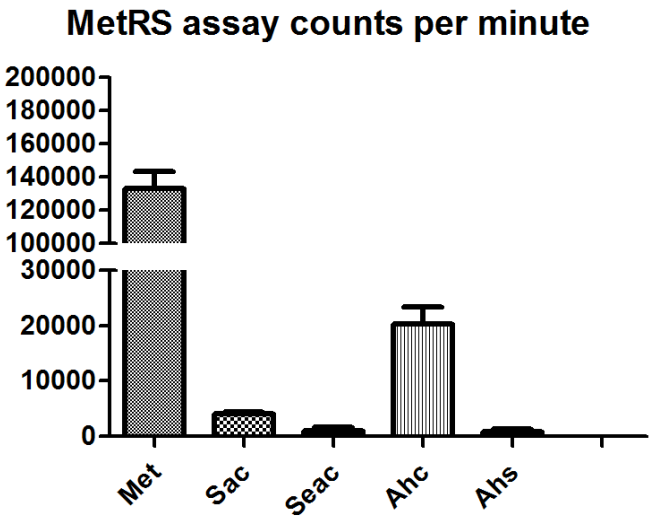


Fig. 2.9: Scintillation counts per minute from reactions with 5 mM Met, Sac, Seac, Ahc and Ahs.

Radioactive ATP generated was highest for the natural substrate, Met, as expected. Amount generated by Met is ~ 6.5x more than the next highest AA (Ahc), ~ 45x more than Sac, with Seac and Ahs being the lowest.

Next, the kinetics of MetRS with Met and Ahc were measured. The reactions were set up as above, and at various time points, 20 uL of the reactions were quenched as above. However, inconsistent data was observed due to losses in charcoal upon repeated washing by centrifugation and subsequent aspiration with a pipette (Fig. 2.10).

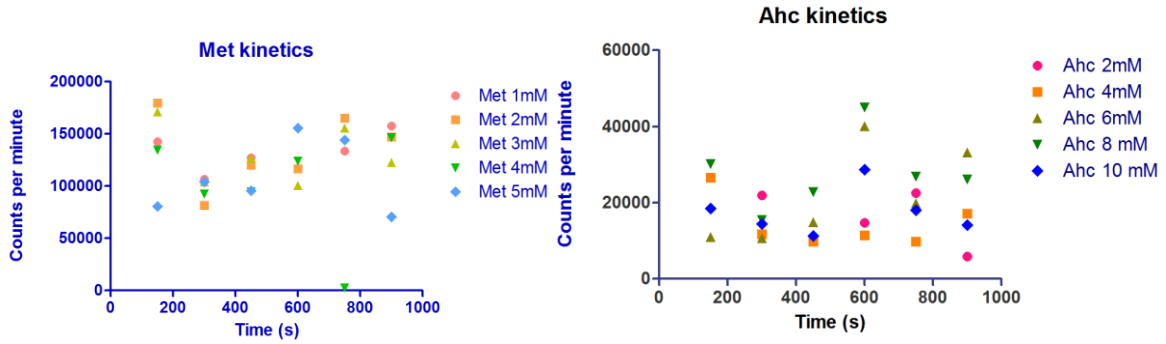


Fig. 2.10: Scintillation counts per minute v/s time from reactions with 5 mM Met, and Ahc.

To remedy this, an alternative method of washing the charcoal was devised (Methods Section). A filter-tip was modified to allow for loss-less washing of charcoal by filtration rather than centrifugation. Consistent data was now obtained, allowing for the determination of Michaelis-Menten kinetics parameters for Met and Ahc (Fig. 2.11) using Prism V6.02 (GraphPad Software Inc.)

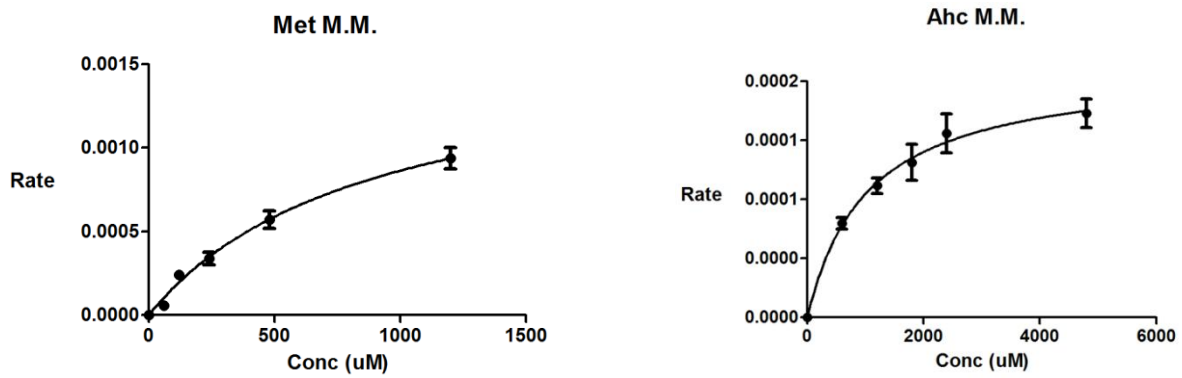


Fig. 2.11: Michaelis-Menten curves for MetRS with Met and Ahc as substrates.

The kinetics parameters were determined to be:

Met $k_{cat}/K_m = 0.041\ s^{-1}\ \mu M^{-1}$, $k_{cat} = 35.30\ s^{-1}$ $K_m = 878.3\ \mu M$

Ahc $k_{cat}/K_m = 0.0045\ s^{-1}\ \mu M^{-1}$, $k_{cat} = 4.62\ s^{-1}$ $K_m = 1033\ \mu M$

Michaelis-Menten kinetics parameters k_{cat} and K_m were determined for Met and Ahc. Comparison of the k_{cat}/K_m parameters for Met and Ahc shows that the activation by MetRS of the former is 9-fold more effective than of the latter.

2.4 Modelling UAA binding to MetRS

The docking of Met analogues Sac, Seac, Ahc and Ahs, as well as their corresponding adenylates into the active site of MetRS was carried out by gradient descent minimisation of the energy of conformationally randomised ligand structures using Autodock Tools Ver. 1.5.6 and Autodock Vina, which use a custom force field based on AMBER.[22] The enzyme scaffold used for docking was that of Met-bound MetRS.[23] All analogues were found to occupy the binding site of the native methionine (Fig. 2.12), and were calculated to have similar binding energies of -17.2 ± 0.4 kJ / mol.

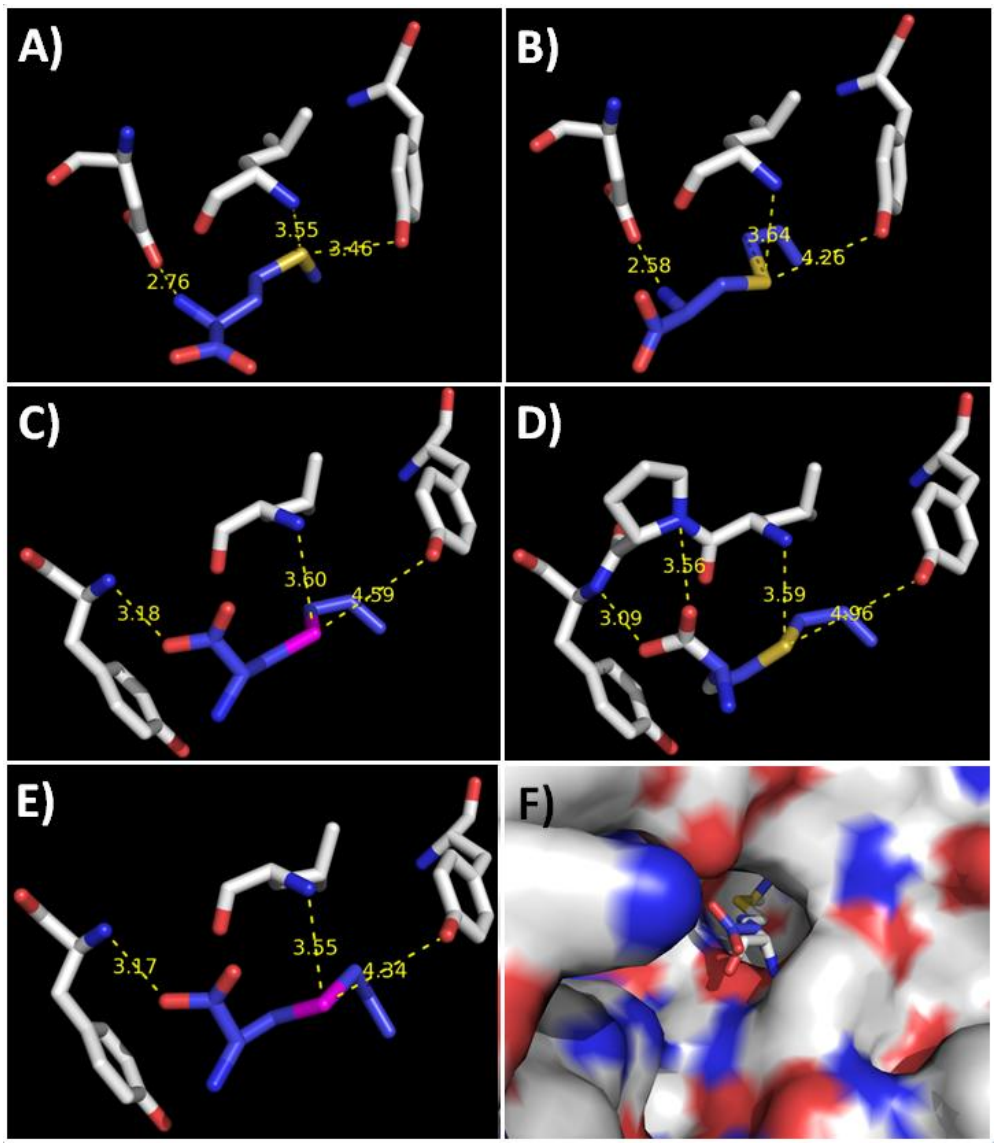


Figure 2.12. Interactions of docked amino acids with MetRS for A) Methionine, B) Sac, C) Seac, D) Ahc and E) Ahs. In panel F), the docked Met (blue backbone) is overlaid on the crystal structure Met (white backbone) with the MetRS surface shown.

Consistent with previous studies,[18] it was found that major contributors to the binding energy in MetRS include electrostatic interactions of the backbone nitrogen with the side chain of Asp 52, as well as hydrogen bonds made by the side chain heteroatom with the side chain OH of Tyr 260 and the backbone amide of Leu 13 next to the hydrophobic pocket where the terminal methyl group is situated (Fig. 2.12). Sac and Seac were seen to maintain the same contacts. However, due to the increased side-chain length, the heteroatom was seen to be pushed out of the hydrophobic pocket, and lay ~ 0.5 Å further away from the oxygen of Tyr 260 (Fig. 2.12-B and C). In Seac, Ahc, and Ahs the contact with Asp 52 was seen to be abolished, and a new contact was formed between the backbone carboxyl group and the

backbone amide of Tyr 15 (Fig. 2.12-C, D and E). Ahc was additionally stabilized by a contact between the backbone carboxyl group of 3 and the backbone amide of Pro 14 in addition to Tyr 15.

In general, it was observed that increasing the side-chain length causes decreased stabilization by Tyr 260 for all analogues as compared to Met. However, Ahc makes an additional contact with Pro 14, potentially stabilizing it further than Sac, Seac, and Ahs.

However, the flexibility of MetRS may cause it to adopt subtly but significantly different active site conformations when actually bound to 1, 2, 3, and 4 than in the Met-bound crystal structure used here, perhaps leading to a difference in energies of bound enzyme-substrate complexes.

From our models and from previous binding studies,[18,23] it was reasoned that the position of the bulky heteroatom may be another key factor in determining incorporation efficiency. The γ-heteroatom in Sac might be unfavourably positioned for subsequent aminoacylation even though the amino acid side-chain lengths are similar to that of L-ethionine. On the other hand, while the side chains of Ahc and Ahs are 2.0 – 2.5 Å longer than that of L-ethionine, both have a heteroatom at the same position as the natural substrates of MetRS.

2.5: Incorporation trials

Incorporation of each Met analogue was then investigated in the nucleosome constituent protein histone H3.[24,25] The gene for Histone H3 with a single Met at position 120 was ordered from GeneArt. The sequence, contained a 5' NdeI and a 3' BamHI restriction site. The gene was delivered in the pMA-T vector[26], from where it was subcloned into pET-15b to generate a construct containing a GSSHHHHHH- leader sequence to allow for easy purification via Ni-NTA agarose. The His-tagged H3 protein was expressed from B834(DE3) cells using the Met-auxotrophic B834(DE3) *E. coli* strain (Method section). Briefly, cells were grown up in LB media supplemented with antibiotics to an OD_{600} of 0.6 to 0.8, washed thrice with met-free Selenomet base media, resuspended in Selenomet base media supplemented with Selenomet nutrient mix, 250 mg/L UAA, and antibiotics, incubated at 30

ºC for 30 min, then at 37 ºC for 30 min, and expression induced for 16 h with 1 mM IPTG at 30 ºC.

Initial trials were carried out with Ahc, the UAA which was demonstrated to have the greatest k_{cat} and K_m with MetRS among the Met analogues. The protein was expressed in inclusion bodies, from where it was extracted, purified by Ni-NTA agarose resin, and analysed by LCMS (Fig. 2.13, Table 2.2). However, the expressed protein was observed to be heterogeneous, where a large proportion of the protein was seen to have been post-translationally modified by gluconoylation and phosphogluconoylation, which N-terminally His-tagged proteins are susceptible to.[27]

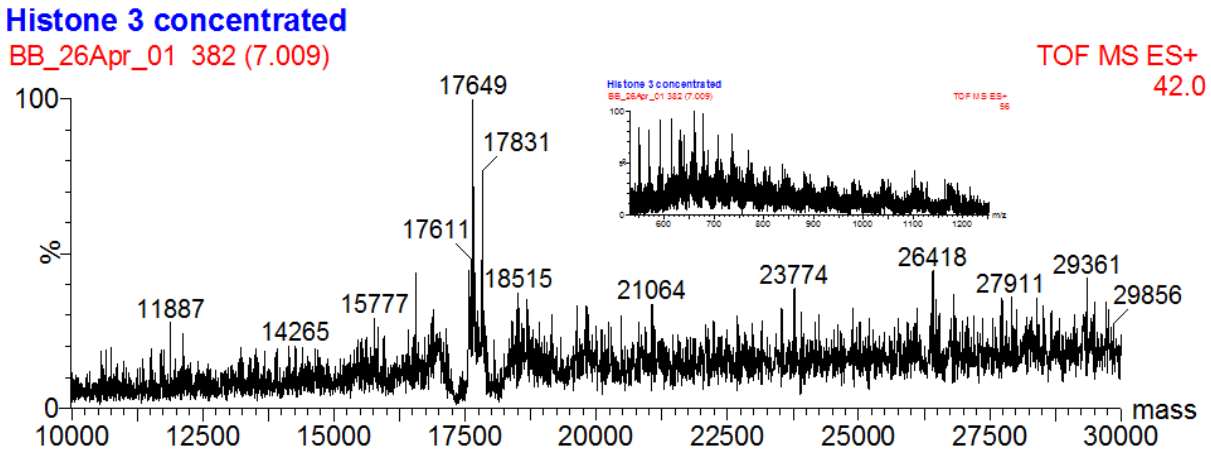


Fig. 2.13: LCMS analysis of His-tagged H3 single Ahc

Modification	Mass shift	Mass expected	Mass seen
None (only Ahc)	+0	17392	-
Gluconoylation	+178	17570	17567
Phosphogluconoylation	+258	17650	17649
Gluconoylation + phosphogluconoylation	+436	17828	17831

Table 2.2: Modifications of His-tagged H3 single Ahc observed by LCMS

As a result, it was decided to attempt the expression of single Ahc H3 without a His-tag. The gene for Histone H3 K9C M90L C96S G102A C110A Single M120 in pET3-d vector was obtained as a gift from the Chris Schofield group. The protein was instead purified by size-exclusion FPLC (See Methods section). Upon repeating expression, the major protein species observed was found to be the Met-incorporated H3, with little (~20%) Ahc incorporation (Fig. 2.14).

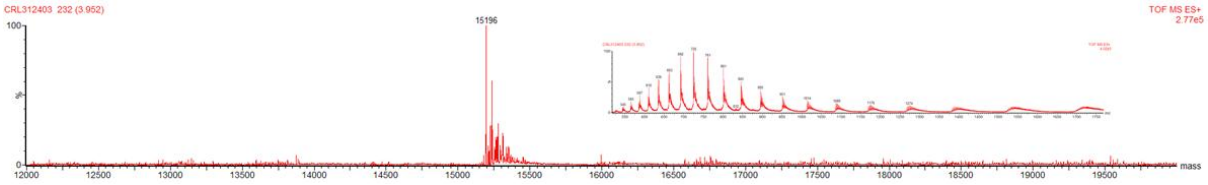


Fig. 2.14: LCMS analysis of H3 K9C M90L C96S G102A C110A expressed in B834(DE3) in a pET3-d vector. Calculated Mass for Met incorporation: 15196 Da, Ahc incorporation: 15222

The expression was repeated in B834 (DE3) pLysS cells,[28] which suppress endogenous protein expression in the presence of chloramphenicol to allow for double selection of transformant colonies with ampicillin and chloramphenicol. During the subsequent expression, extra washing steps were carried out with met-free media, and incubation times prior to induction were doubled. However, after expression of protein, extraction from inclusion bodies, and size-exclusion FPLC purification, predominantly Met-incorporation was seen. Ahc incorporation was approximately the same amount (15-20%) as in previous expression trial. It was then reasoned that the pET3-d vector, which does not encode the lactose inhibitor LacI,[29] might be the cause of leaky expression of Met-containing H3 during the growth phase of the cells prior to induction. An expression trial on 100 mL scale was carried out to test whether the cells were expressing H3 during their growth phase. The effect of 1% w/v glucose during the growth phase was also investigated (Fig. 2.15).[30]

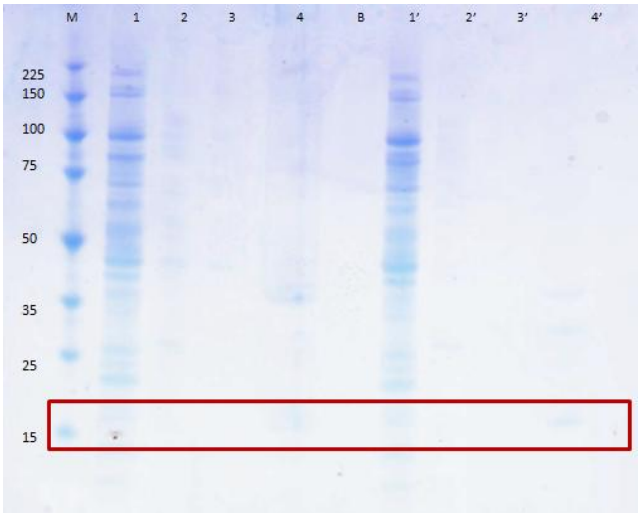


Fig. 2.15: SDS-PAGE of trials of leaky expression in H3. M: Marker, 1 & 1': Supernatant of inclusion body 1st wash with TW buffer; 2 & 2': Supernatant of inclusion body 2nd wash with TW buffer; 3 & 3': Supernatant of inclusion body wash with Histone Wash Buffer; 4 & 4': Supernatant of inclusion body extraction with unfolding buffer. B: Blank lane

1', 2', 3', 4' => With 1% glucose

1, 2, 3, 4 => Without 1% glucose

In both cases, a band at ~15 KDa was observed in the inclusion body unfolded extract, suggesting that expression was indeed leaky in this system.

The gene from this vector was excised with NcoI and BlpI, and ligated into LacI-encoding vector pET-28a. Choosing these restriction enzymes allowed for the excision of the His-tag form the vector backbone, allowing for the expression of His-tag free H3. The vector was transformed into B834(DE3) cells for expression, which was carried out in Selenomet metdia supplemented with 500 mg/L DL-Ahc (Fig. 2.16).

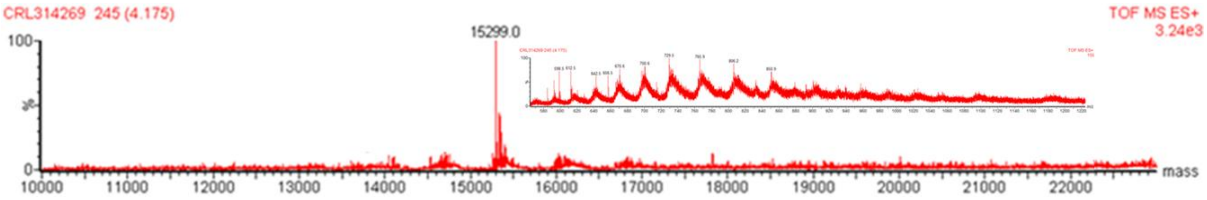


Fig. 2.16: LC-MS spectrum of Ahc-incorporated H3. Expected mass: 15222 Da, Observed: 15299 Da (2-mercaptoethanol adduct).

15299 Da peak corresponds to the disulfide-bridged complex of H3 K9C M90L C96S G102A C110A Ahc 120 with 2-mercaptoethanol (~15222 + 78 -2), which was present in the FPLC running buffer. Upon treatment of the protein with DTT for 30 min, the mass of H3 Ahc 120 with a free cysteine was seen (Fig. 2.17). Incorporation levels of Ahc were found to be > 95% as determined by LC-MS.

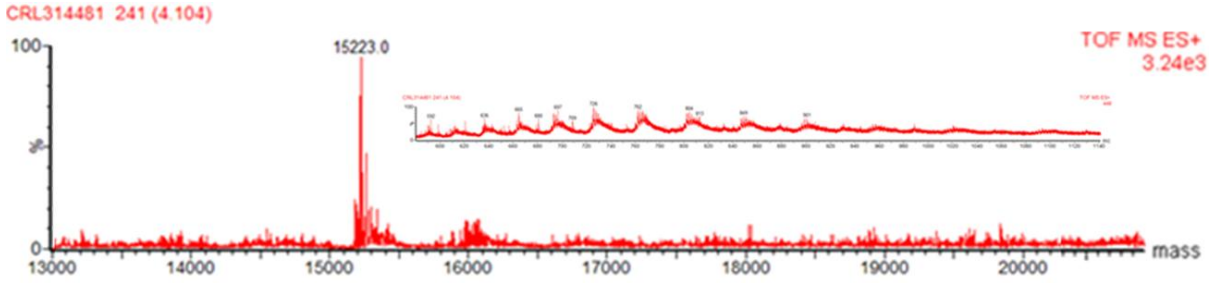


Fig. 2.17: LC-MS spectrum of Ahc-incorporated H3. Expected mass: 15222 Da, Observed: 15223 Da

Expression of H3 was attempted in selenomet media supplemented with either Sac, Seac, or Ahs. Incorporation levels of these analogues were found to be < 5% (Fig. 2.18).

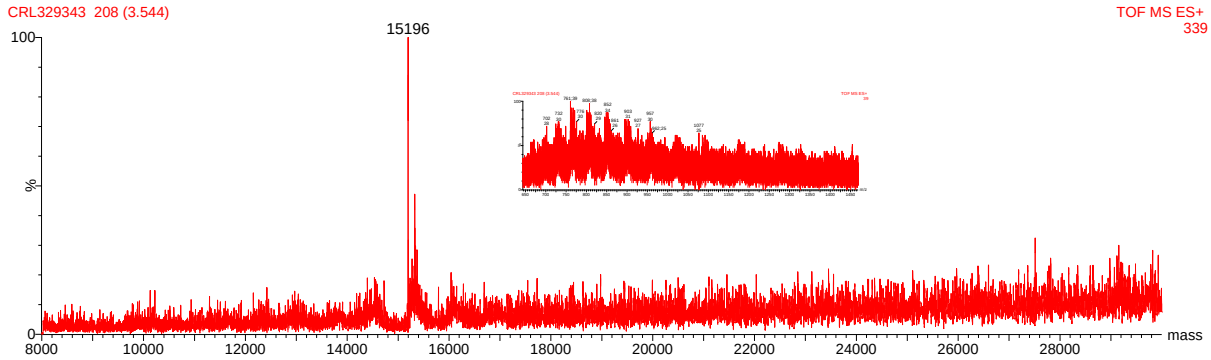


Fig. 2.18: LC-MS characterisation of H3 with attempted Sac incorporation. Expected mass for Met incorporation: 15196 Da. Observed: 15196 Da. Similar spectra were observed for Seac and Ahs incorporation.

Furthermore, in selenomet media supplemented with either of the selenium-containing amino acids Seac and Ahs, a red precipitate was seen to form after 16h of incubation at 37 ºC. This precipitate was not observed in selenomet media supplemented with either Met, Sac, or Ahc. This may be due to metabolic degradation of selenium-containing amino acids, as this red colour was not develop in aqueous solutions of Seac or Ahs incubated at 37 ºC for longer

periods of time (> 5 days). Samples of the media supernatant as well as Methanol-, acetonitrile-, DMSO- and DMF-extracted cell lysate supernatants were analysed by MS in order to identify for selenium-containing degradation species in the samples via selenium's unique isotopic distribution. However, no selenium-containing species could be detected in either intact or HCD-fragmented modes. This implied that the selenium-containing compounds were not ionising well in either positive or negative ion modes during MS, suggesting that the coloured precipitate may comprise non-ionising aggregates of elemental selenium.

The results from Met-RS screening and the H3 protein incorporation experiments suggested that of the analogues under investigation, Ahc would be the most suitable for genetic incorporation. Proteins of varying structure and function were chosen. In addition to histone H3, monomeric single-site incorporation of Ahc was successfully carried out in the pentapeptide repeat protein Np276 (Fig. 2.18),[31] and in the TIM-barrel β-glycosidase SsβG (Fig. 2.19).[32]

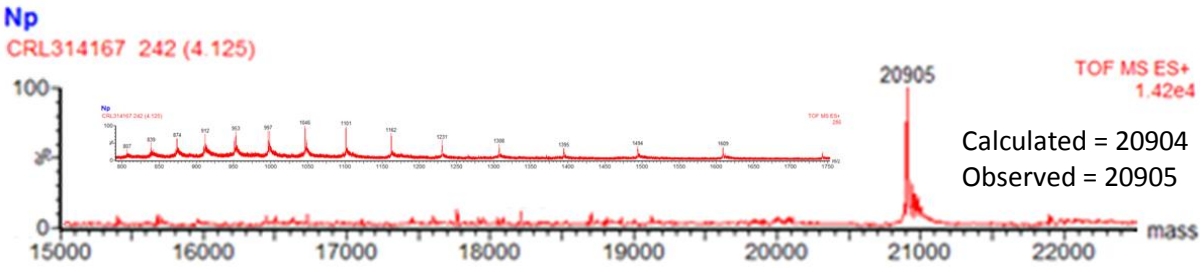


Fig. 2.18: LC-MS characterisation of Np276 G2P M21I Ahc61

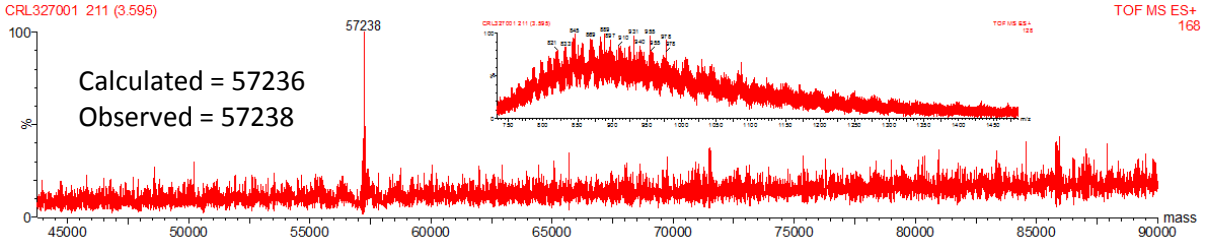


Fig. 2.19: LC-MS characterisation of SsβG Ahc49 P152A C344S 9Ile

Excellent levels (>95%) of multi-site genetic incorporation of Ahc were also demonstrated in monomeric targets such as the DNA binding protein SarZ (Fig. 2.20),[33] as well as in the multimeric bacteriophage coat protein Qβ (Fig. 2.21).[34]

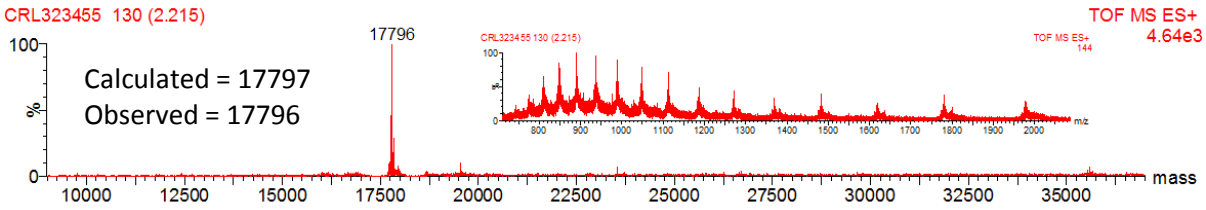


Fig. 2.20: LC-MS characterisation of SarZΔ(7-142) Ahc20 Ahc64

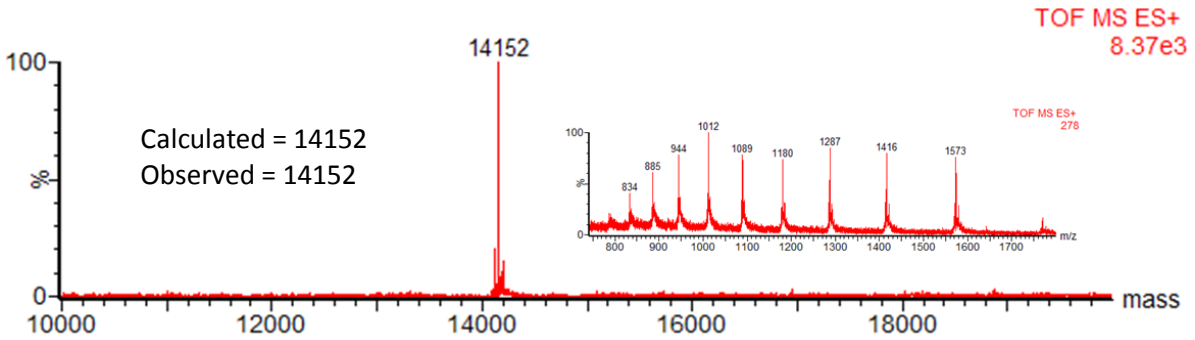


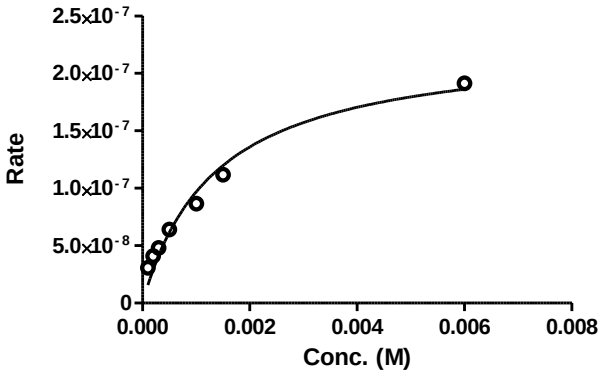
Fig. 2.21: LC-MS characterisation of Qβ K16Ahc

In addition to LCMS analysis, all proteins were characterised by MS-MS analysis, circular dichroism and gel electrophoresis. (Section **2.11.8** to **2.11.14**) Ahc-incorporated proteins were then investigated for the retention of function after replacement of Met with Ahc.

2.6: Protein function assays

2.6.1: Activity assay for SsβG Ahc49 P152A C344S 9Ile

Enzyme activity was assayed by monitoring the enzymatic liberation of *p*-nitrophenol from *p*-nitrophenyl-β-D-galactopyranoside.[35] Reactions were monitored for an increase in absorbance at 405 nm, with substrate concentrations ranging from 0.2 to 6.0 mM in 200 µL 50 mM sodium phosphate pH 6.5 containing 10 µL of 0.05 mg/mL enzyme stock, and run at 45 °C.



k_{cat} = 5.248 s^{-1}

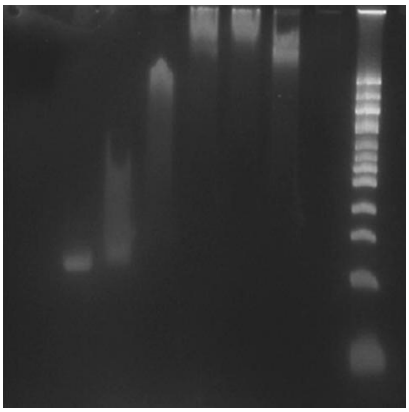
K_m = 1.237 mM

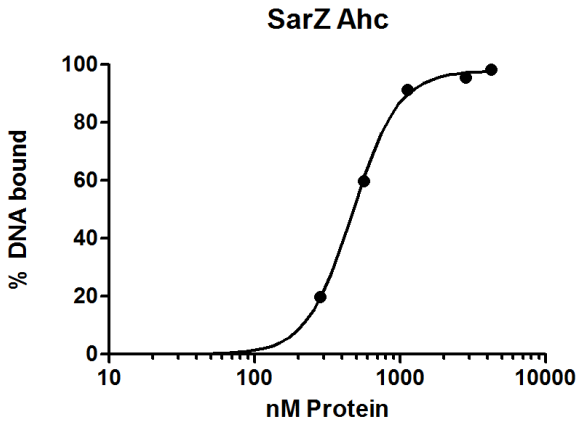
k_{cat}/K_m = 4.242 s^{-1} mM^{-1}

Kinetics parameters k_{cat} and K_m were found to be of the order of the wild-type enzyme.[35]

2.6.2: Activity assay for SarZΔ(7-142) Ahc20 Ahc64:

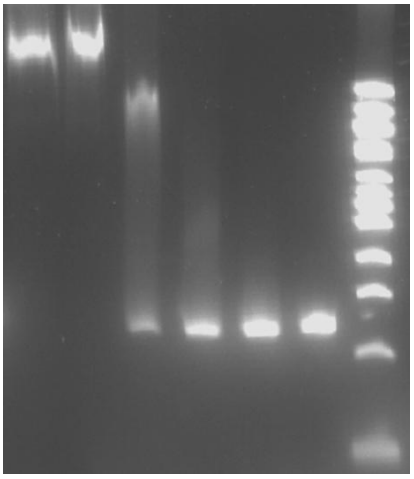
The activity of SarZΔ(7-142) Ahc20 Ahc64 was assayed by an electrophoretic mobility shift assay (EMSA) of the promoter region of the *hla* gene, which encodes α-hemolysin in *Staphylococcus aureus*.[33] The *hla* promoter gene was ordered from GeneArt and amplified by PCR. The binding of SarZ was carried out with 50 µg of purified DNA per well, run in 0.5x TBE buffer at 100 V for 90 min on a 6% Tris-glycine gel at 4 ºC in 0.5X TBE buffer. The gel was then incubated with SyBr Gold DNA stain prior to visualisation.

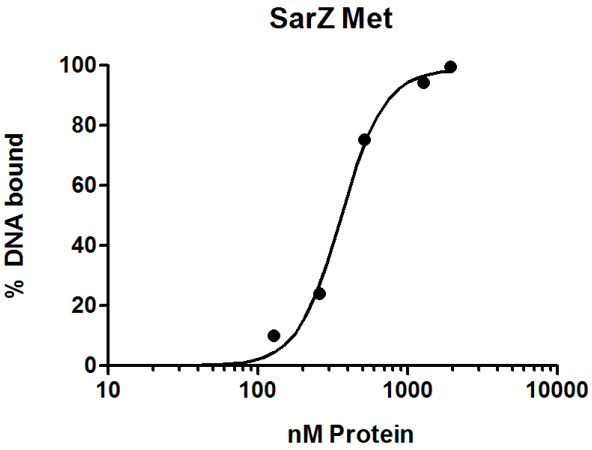




Sar Z-Ahc K_d = 469 nM

The K_d value was found to be 10x higher than the reported value for the wild-type enzyme.[33] The assay was then repeated with the Met-incorporated enzyme, as expressed in B834(DE3) cells.





Sar Z-Met K_d = 358 nM

Relative to WT Met-incorporated SarZ expressed in the same cell line, the K_d of the Ahc-incorporated protein was found to be of the same order.

Together with CD data, these results indicate that Ahc-incorporated proteins retain both global secondary and tertiary structure as well as function.

Pleasingly, the partial incorporation of Ahc could also be demonstrated into the Fc region of IgG expressed in HEK 293T cells, indicating that Ahc is a general analogue for Met, rather than being specific only for *E. coli.* The expressed protein was partially deglycosylated of N-linked glycans by treatment with PNGase-F prior to analysis by LCMS (Fig. 2.22). The single-site incorporation of Ahc was confirmed by MSMS (Section **2.11.14**).

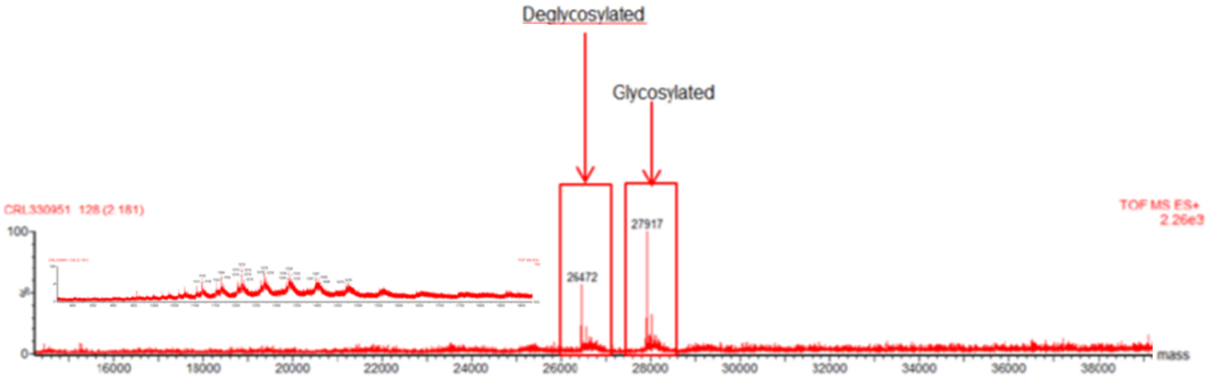


Fig. 2.22: LC-MS analysis of Ahc-incorporated IgG-Fc xpressed in HEK 293-T cells. Masses observed in deglycosylated peak cluster: 26446 (Met incorporated, 19%), 26472 (Single Ahc incorporation, 81%), no dual Ahc-incorporation.

Ahc incorporation was also attempted in engineered single-Met mutants of the biotin-binding protein traptavidin,[36] hen egg white lysozyme,[37] the *Arabidopsis thaliana* fructose-Bisphosphate aldolase AT3G52930,[38] and the Staphylococcus aureus transcription regulators MgrA[39] and SarA.[40,41] However, further work was not attempted on these proteins due to poor expression levels (< 0.5 mg/L) when expressed in B834(DE3), B834(DE3) pLysS as well as BL21-CodonPlus®(DE3)-RIL-X cell lines.

2.7: CM reaction method optimisation

YAL had previously discovered that Qβ functionalised with Ahc was seen to precipitate in solution upon the addition of 10,000 eq. of $MgCl_2$ and the Hoveyda-Grubbs' 2nd generation catalyst (HG-II, **1.19**) in 30% *t*-BuOH, conditions optimised for CM on SBL.[11] Initial reactions with Np276 Ahc-61, carried out under the same conditions, also showed heavy precipitation, and no protein was observed in the LC-MS spectrum of the reaction supernatant.

Previous work in our group has shown that $MgCl_2$ is necessary for the cross-metathesis reaction to take place, presumably by preventing the unfavourable coordination of protein carboxylate groups to the Hoveyda-Grubbs Second Generation (HG-II) catalyst.[11] It was therefore decided to screen for the minimum amount of $MgCl_2$ needed for the reaction to occur, without 'salting out' the protein when the solution of HG-II in t-BuOH was added. For this experiment, the serine protease subtilisin *Bacillus lentus* (SBL) with a chemically incorporated[42] S-allylcysteine was chosen, as it was not previously seen to precipitate under these reaction conditions.

SBL contains 5 Asp, 5 Glu, and 1 C-terminus carboxylate. 5 eq. of $MgCl_2$ per acidic group were used (i.e. 55 eq. in total) to attempt to prevent this unfavourable coordination of SBL to HG-II, with both positive and negative controls (Fig. 2.23).

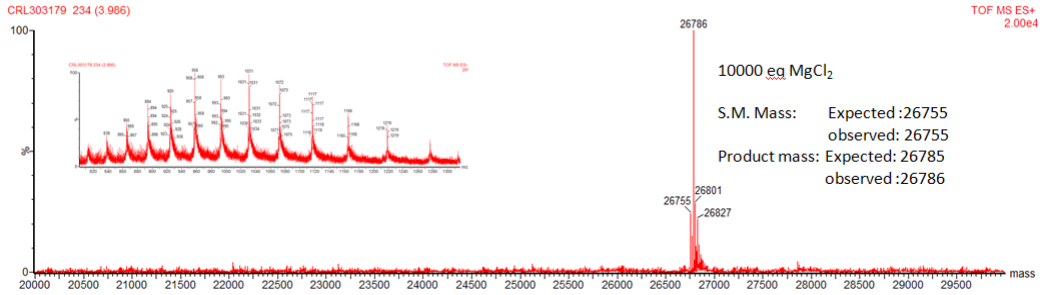


Fig. 2.23a: LCMS of SBL Sac156 after CM with allyl alcohol, 10000 eq. $MgCl_2$(+ve control)

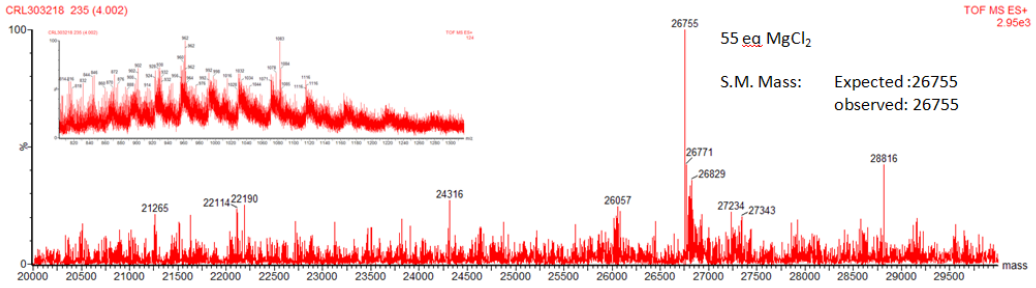


Fig.2.23b: LCMS of SBL Sac156 after CM with allyl alcohol, 55 eq. $MgCl_2$

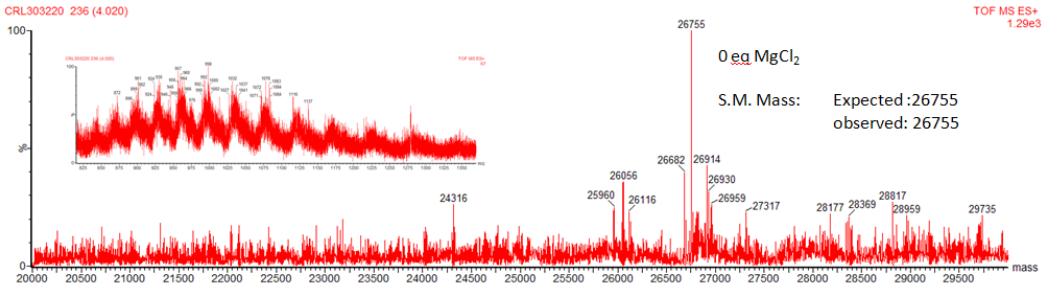


Fig. 2.23c: LCMS of SBL Sac156 after CM with allyl alcohol, 0 eq. $MgCl_2$ (-ve control).

From these results, it was observed that in the absence of $MgCl_2$, no conversion by CM was seen to occur, and only the starting material's mass could be seen in the LC-MS spectrum of the reaction. 55 eq. $MgCl_2$ was found to not be sufficient for the cross-metathesis reaction to proceed, though at 10,000 eq., as before, conversion to the CM product was observed. Experiments were repeated with 3000, 1500 and 500 eq. $MgCl_2$, but again only the starting material could be observed. At 5000 eq., though protein signal could be observed for SBL, Np276 was still seen to precipitate.

In order to prevent the precipitation of proteins, in the presence of 30% t-BuOH and 5000 eq. $MgCl_2$, mild denaturing conditions were applied such as 2.5% of the cationic detergent Cetyltrimethylammonium bromide (CTAB),[43] *n*-Dodecyl β-D-maltoside,[44] or the anionic detergent SDS (lower concentrations of detergent did not prevent the precipitation of Np276). However, in all cases the protein solution remained coloured even after desalting and dialysis, indicating the presence of (tightly and perhaps coordinately-bound) ruthenium species. A complex signal was observed in the LC-MS spectrum indicating the presence of metal adducts. It was then decided to use stronger denaturing conditions, such as either 8M urea or 5M guanidine hydrochloride, in 50 mM sodium phosphate.

In addition, a factorial analysis was carried out of the CM reaction in order to determine and exclude reaction conditions which led to the non-productive but tight association of ruthenium complexes with substrate proteins which prevented their detection. For this, non-functionalised myoglobin was used as a substrate. The factors under consideration were:

1) Denaturant used:
   Either 8 M urea or 5 M guanidine hydrochloride in 50 mM sodium phosphate pH 8.0

2) Post-reaction desalting carried out by PD Spintrap:
   Desalting carried out before LC-MS, or reaction crude injected directly to LC-MS

3) Effect of catalyst addition:
   200 eq. HG-II in 30% *t*-BuOH, or 30% *t*-BuOH only

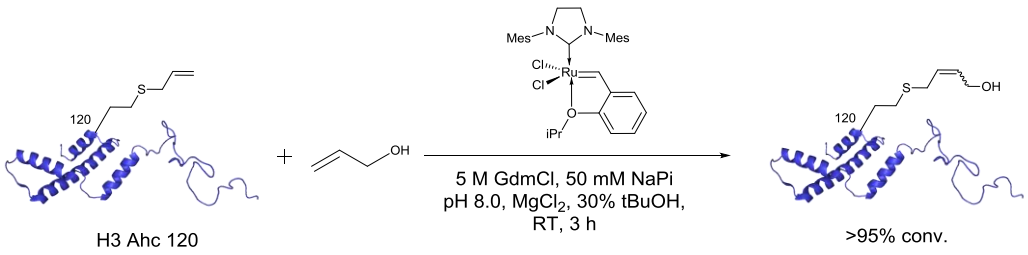
4) Addition of 3-mercaptopropanoic acid (3MPA),[45] a reported chemical scavenger for palladium species (whose role was to be investigated as a possible Ruthenium scavenger):
   400 eq. scavenger added (2 eq. relative to catalyst), or no scavenger added

   Total conditions = 2^4 = 16 reactions

A signal from myoglobin was only observed in the LC-MS of solutions where no HG-II catalyst had been added, implicating it as the cause of loss of protein signal. This catalyst is likely to be co-ordinately bound to the protein, as it was not eliminated by a PD-spintrap column. Furthermore, these results suggest that a combination of 30% t-BuOH in 8 M Urea or 5M guanidine hydrochloride, together with 5000 eq. $MgCl_2$ does not interfere with the LCMS analysis of proteins. However, 3-mercaptopropanoic acid was found to not be an efficient scavenger for HG-II at a 2:1 eq. ratio. To determine whether 3-mercaptopropanoic acid acted as a scavenger for HG-II at all, a further 75 eq. (relative to catalyst) were added to all samples where HG-II had been added, and the tubes incubated at r.t. for 30 min. . Upon analysis of the samples by LCMS, a clear protein signal was now observed where there had been none earlier. These reaction conditions were then applied to Ahc-incorporated proteins.

2.8: CM on Ahc-incorporated proteins

With Ahc-containing proteins in hand, their reactivity toward CM was tested. Allyl alcohol was used as a model metathesis partner.



Using conditions optimised in section **2.7**, CM was carried out on histone H3 Ahc120. Briefly, 60 µg of the lyophilised protein was reconstituted in denaturing buffer (5 M guanidinium chloride-50 mM sodium phosphate pH 8.0) to which were added 5000 eq. $MgCl_2.6H_2O$, 100 eq. HG-II in 30% *t*BuOH solution, and 5000 eq. allyl alcohol. The reaction was run at RT for 3 hours with shaking. 3-mercaptopropionic acid (75 eq. w.r.t. catalyst) was used as a metal scavenger for ruthenium species and the mixture was shaken for further 30 minutes at RT. The protein sample was then purified from small molecules by Vivaspin® (10kDa MWCO) followed by a PD-Spintrap column (Full experimental procedures in section **2.11.9**). LCMS analysis of the protein sample revealed > 95% conversion to the desired CM product (Fig. 2.24).

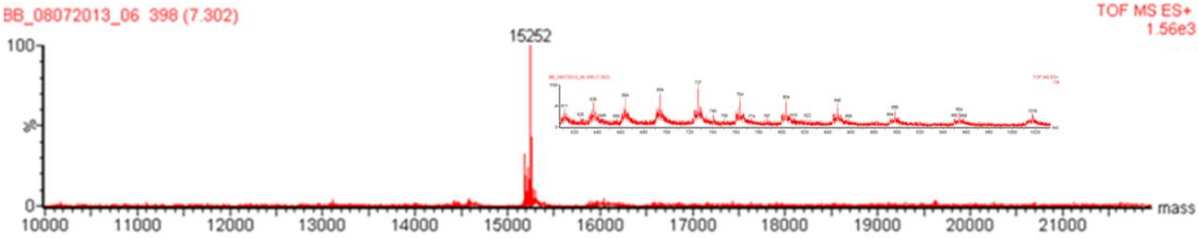
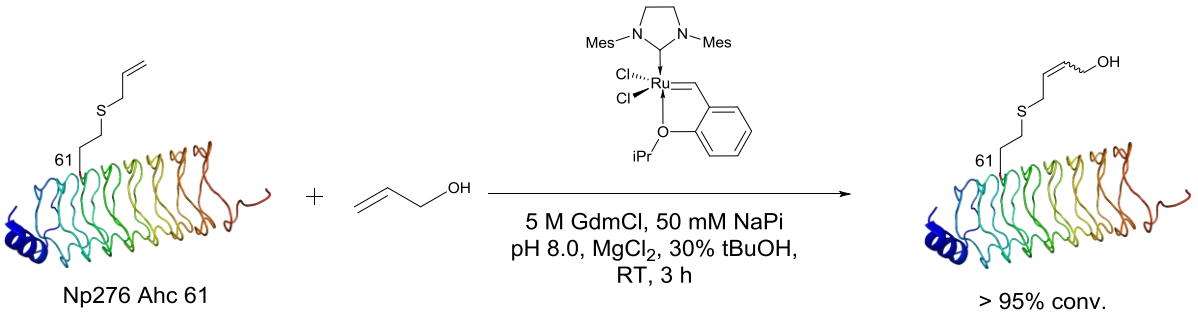


Fig. 2.24: LCMS analysis of CM reaction of histone H3 Ahc120 with allyl alcohol. Product calculated mass = 15252 Da, found = 15252 Da



Successful CM was also demonstrated using the same reaction and scavenging conditions on Np276 (Fig 2.25):

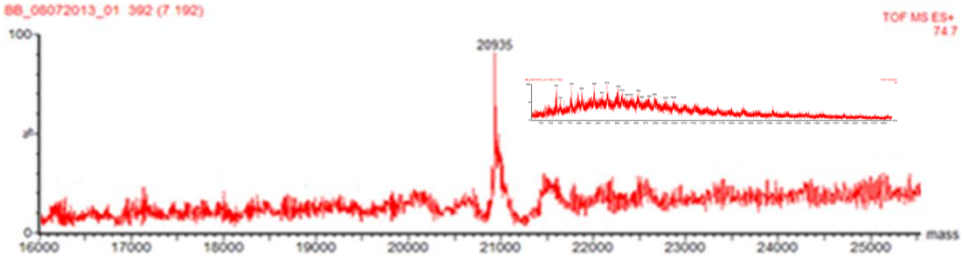
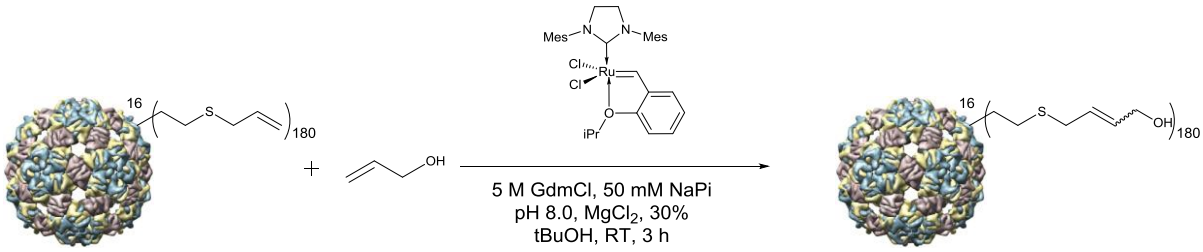


Fig. 2.25: LCMS analysis of CM reaction of Np276 Ahc61 with allyl alcohol. Product calculated mass = 20935 Da, found = 20935 Da



Pleasingly, CM was also demonstrated on the intact Qβ virus-like particle containing 180 reaction sites per particle, with 82% conversion over 3 h. The protein sample was treated with TCEP before analysis by LCMS (Fig. 2.26)

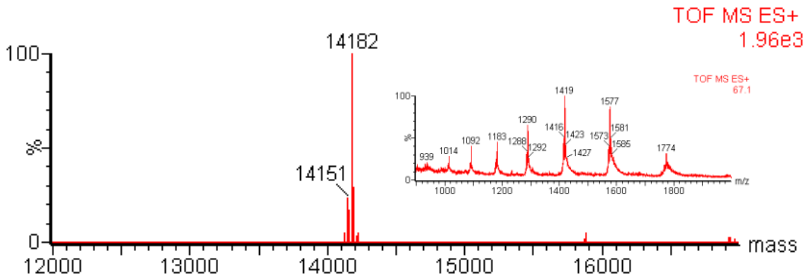
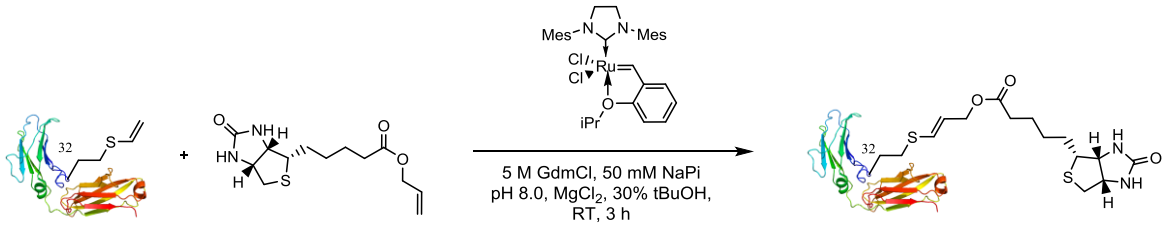
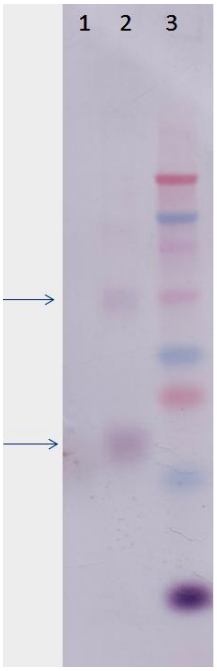


Fig. 2.26: LCMS analysis of CM reaction of Qβ Ahc16 with allyl alcohol. Stating material mass = 14152 Da, found =14151 Da. Product calculated mass = 14182 Da, found = 14182 Da



CM was carried out on glycosylated IgG-Fc Ahc32 Met208 with alkene-functionalised biotin, and functionalisation was confirmed by performing an anti-biotin Western Blot with alkaline-phosphatase conjugated Extravidin, visualised with BCIP/NBT (Fig 2.27):



Lane 3: Marker

Lane 2: Biotin-functionalised IgG-Fc Dimer (top) and monomer (bottom)

Lane 1: IgG-Fc mixed with an equivalent amount of biotin prior to SDS-PAGE

Fig. 2.27: Anti-biotin Western Blot

These conditions were then subsequently successfully applied to H3 proteins with chemically installed Seac by YAL, to yield a chemically 're-writeable' moiety for CM-based protein modification.[8]

2.9: CM Challenges

2.9.1: Investigation of alternative scavenging conditions

CM was attempted on SsβG Ahc49 P152A C344S 9Ile with allyl alcohol as a metathesis partner. After purification of the reaction mixture from small molecules by Vivaspin® (10kDa MWCO) followed by a PD-Spintrap column, only signal from metal adducts was observed in the ESI-MS. Similar adducts were observed from CM reactions of SarZ Ahc20 Ahc64 with both allyl alcohol as well as allylamine as CM partners.

A factorial analysis for scavenging Ru from Met-incorporated SsβG was carried out. Factors chosen were:

1. Scavenger used: DTT or 3-mercaptopropanoic acid (used previously)

2. Eq. of scavenger (relative to Ru): 3 eq. (used by Spicer and co workers for Palladium scavenging,[45] or 75 eq. (used previously)

3. 'Desalting' method: 3 h Dialysis (20 kDa MWCO) or Vivaspin + PD-spintrap

4. Scavenger incubation time: 15 min or 60 min

Total experiments = $2^4 = 16$

16 samples of 50 ug of SsβG-Met43 in 5M GdmCl and 50 mM NaPi were treated with 10000 eq $MgCl_2$ and 120 eq of H.G. (II) catalyst in t-BuOH to give a final concentration of 30% tBuOH. Samples were incubated at r.t. for 5 min, and then treated with the scavengers for either 15 or 60 minutes. After desalting, samples were analysed by LC-MS.

Of these, the following combination of factors showed effective scavenging of the Ru-catalyst from SsβG. The conditions used were:
1) DTT, 75 equivalents, desalting by dialysis, and 15 min incubation time (Fig. 2.28)
2) DTT, 75 equivalents, desalting by Vivaspin + PD spintrap, and 15 min incubation time (Fig. 2.29)

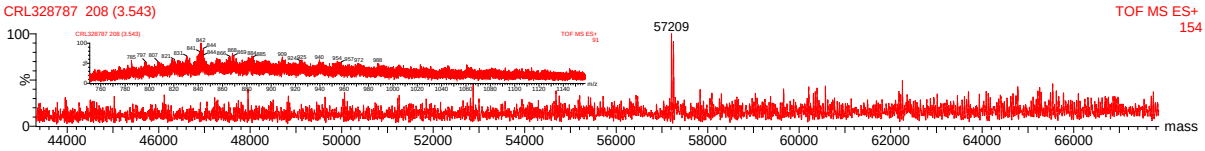


Fig. 2.28: ESI-MS spectrum for SsβG-Met43 scavenged of Ru-catalyst by DTT (75 eq.) for 15 min, followed by desalting by dialysis. Expected mass: 57209 Da, Observed: 57209 Da

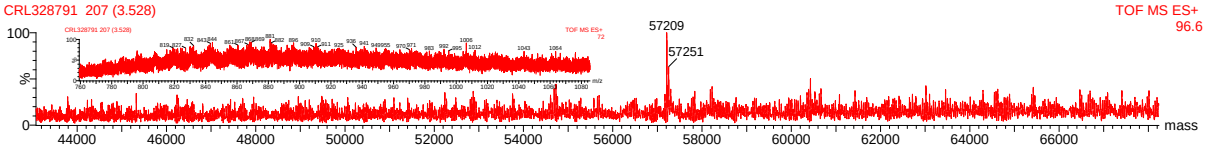


Fig. 2.29: ESI-MS spectrum for SsβG-Met43 scavenged of Ru-catalyst by DTT (75 eq.) for 15 min, followed by desalting by vivaspin and PD-SPintrap. Expected mass: 57209 Da, Observed: 57209 Da

This showed that, for SsβG-Met43 at least, DTT is preferable to 3-mercaptopropanoic acid as a ruthenium scavenger, and that 75 equivalents w.r.t catalyst were better than 15. The desalting method didn't seem to be very important, though incubation time was.

These scavenging conditions were then applied to the CM reaction on SsβG-Ahc43. However, strong signals from small molecule adducts were seen to persist, and a protein ion series was not seen to form. This indicated that the coordination of the Ru-catalyst to SsβG-Ahc43 was stronger than the corresponding Met analogue, and that effective scavenging under identical conditions was not possible for the former. CM reactions were then carried out on SsβG-Ahc43 systematically. The following factors were investigated (Table 2.3):

1. Equivalents of catalyst used: 120 eq (used for H3, Qβ and Np276) or 10 eq (reducing equivalents may improve protein recovery after desalting, though may reduce reaction conversion)

2. Scavenger used: DTT or 3-mercaptopropanoic acid (protein from CM reaction, after treatment with both scavengers, was visible at the correct position in SDS-PAGE)

3. Equivalents of $MgCl_2$ used: 10000 eq (used for H3, QB and Np276) or 100 eq. (reducing $MgCl_2$ content may prevent protein crashing out during buffer exchange, though may lead to Ru-adduct formation)

4. Duration of dialysis: 2 h (according to manufacturer's website, dialysis should be complete between 15 min to 2 h) or 16 h

5. Additional dialysis step: ± additional dialysis into 20 mM Tris from 5 M GdmCl. Lower buffer salt content improves protein ionisation for SsβG-Ahc starting material, may do the same for reaction product as well.

At all stages, aliquots were run on an SDS-PAGE, which confirmed presence of the protein in all samples. No protein degradation was observed.

Expt. No.	Eq. Catalyst	Eq.$MgCl_2$ used	Scavenger	2h dialysis against GdmCl	16h dialysis against GdmCl	6h dialysis against 20mM Tris
1	100	100	DTT	Metal adducts seen	Metal adducts seen	Metal adducts seen
2	100	100	3-MPA	Metal adducts seen	Metal adducts seen	Metal adducts seen
3	10	100	DTT	Metal adducts seen	Metal adducts seen	Metal adducts seen
4	10	100	3-MPA	Metal adducts seen	Metal adducts seen	Metal adducts seen
5	100	100	DTT	Metal adducts seen	Metal adducts seen	Metal adducts seen
6	100	100	3-MPA	Metal adducts seen	Metal adducts seen	Metal adducts seen
7	10	100	DTT	Metal adducts seen	Metal adducts seen	Metal adducts seen
8	10	100	3-MPA	Metal adducts seen	Metal adducts seen	Metal adducts seen

Table 2.3: Experimental conditions for factorial study of CM between SsβG-Ahc43 and allyl alcohol. Similar results were observed for 10,000 eq. of $MgCl_2$ as well.

In all cases, strong metal adduct signals were seen, and no protein ion series were seen to form. All samples were run on MALDI, but no protein signal was seen.

2.9.2: Effect of CM conditions on protein function

CM on proteins could potentially be used to modulate their activity. For example, the ability of SarZ to bind negatively charged DNA could be increased by introducing additional positively charged groups on its surface through CM with allylamine. In order to test this hypothesis, CM between SarZ Ahc20 Ahc64 and allylamine was carried out for 3 hours in non-denaturing buffer (50 mM sodium phosphate pH 8.0 without guanidine hydrochloride or urea). The reaction crude was dialysed against 20 mM Tris 100 mM NaCl without scavenging, and the protein used in a DNA binding EMSA (Fig. 2.30):

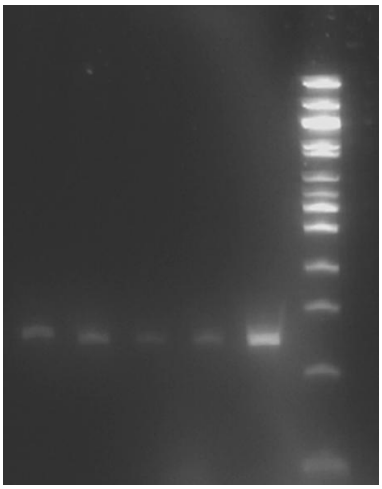


Fig. 2.30: EMSA for SarZ - allylamine CM reaction crude with hla promoter DNA

It was found that the protein no longer bound the *hla* promoter DNA at the same concentration, possibly either due to denaturation during the reaction in the presence of 30% t-BuOH, or due to persistence of ruthenium species on the protein surface. The mobility shift assay was repeated after scavenging the reaction with 3-mercaptopropanoic acid, and dialyzing against 20 mM Tris, 100 mM NaCl (Fig. 2.31, decreasing concentration of SarZ from left to right).

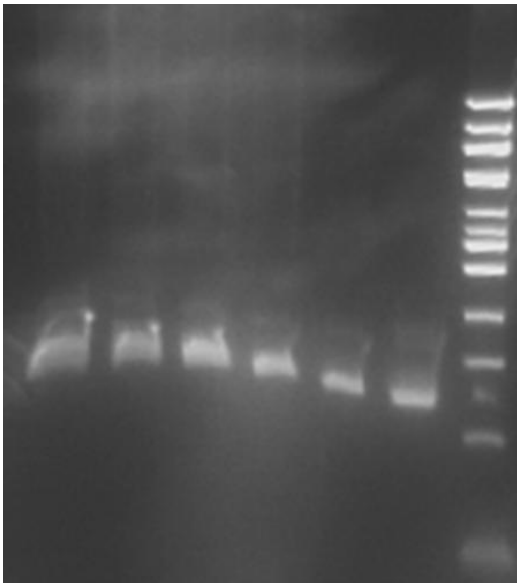


Fig. 2.31: EMSA for SarZ - allylamine CM reaction after scavenging with hla promoter DNA

Some DNA binding by SarZ was seen, though significantly less than before the CM reaction. In order to study the effect of *t*-BuOH, SarZ was CM-reacted with allylamine for 3h, or treated with only 30% *t*-BuOH for 1h, then dialysed into 50 mM phosphate buffer. The CM-reacted protein was treated with 3-mercaptopropanoic acid to scavenge bound Ru-species from the protein, and dialysed into 50 mM phosphate buffer. Both samples were incubated with the hla promoter DNA and a mobility shift assay carried out (Fig. 2.32).

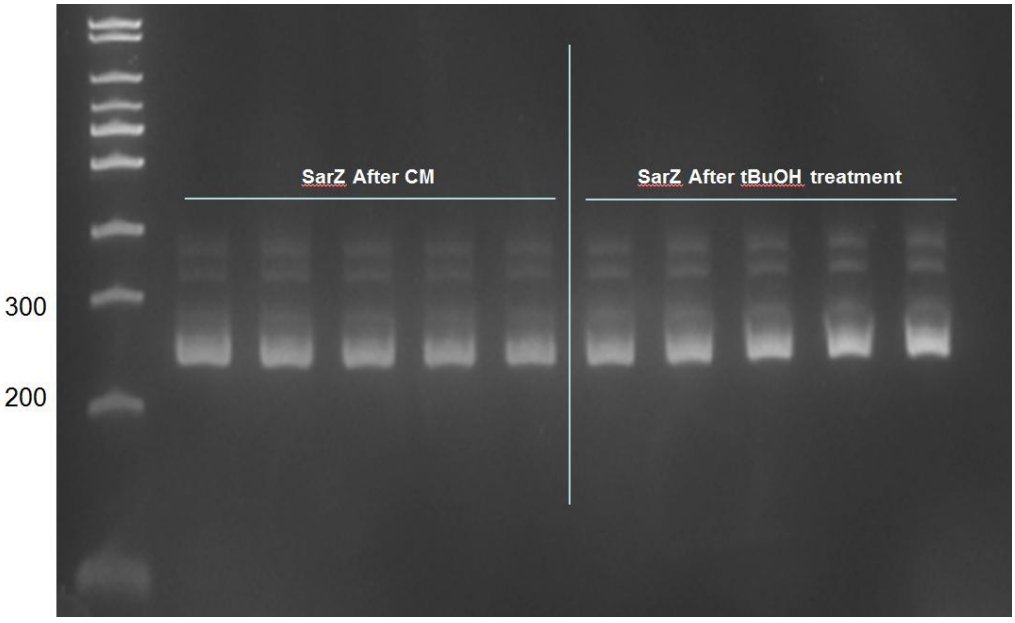


Fig. 2.32: EMSA for SarZ - allylamine CM reaction after scavenging, or SarZ treated with 30% *t*-BuOH for 1h only, with hla promoter DNA

DNA binding was studied with SarZ treated with 30% t-BuOH for varying times (Fig. 2.34), using identical concentrations of DNA substrate as used in section **2.6.2**, and a protein concentration 50% higher than the concentration at which complete binding was observed:

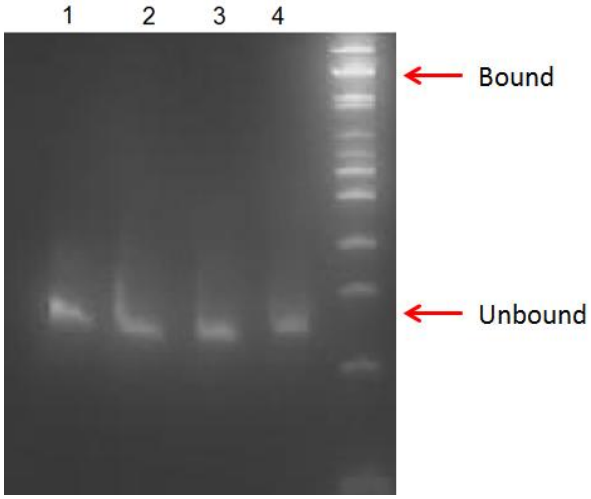


Fig. 2.34: Electrophoretic mobility shift assay for denatured SarZ with *hla* promoter DNA. 1) SarZ immediately treated with 30% t-BuOH, 2) SarZ treated with 30% t-BuOH for 1 h, 3) SarZ treated with 30% t-BuOH for 2 h, 4) SarZ dialysed back into 50 mM sodium phosphate.

Upon contact with t-BuOH, the DNA binding of SarZ was seen to be significantly and immediately reduced, and was not observed to recover upon dialysing into non-denaturing sodium phosphate buffer.

From crystal structure on PDB,[46] the secondary structure of SarZ was seen to be 56.5% alpha helix, 17.8% beta, 25.8% coil. Analysis of the CD spectrum of SarZ Ahc (Section **2.11.16**) with the online service Dichroweb,[47,48] showed that the Ahc-incorporated protein's secondary structure was 56.4% alpha, 19.3% beta, 24.3% coil, in good agreement with the Met-incorporated protein's crystal structure.

After 2h exposure to 30% t-BuOH, the secondary structure of SarZ was seen to change to 48.3% alpha, 33.8% beta, 17.9% coil. After 16h dialysis back into non-denaturing 50 mM sodium phosphate pH 8.0, the native secondary structure was not seen to be recovered, and instead was found to be 63.3% alpha and 36.7% beta. This suggests that, unlike SBL,[49] the activity of enzymes in general may not be retained when exposed to 30% *t*-BuOH during CM.

2.10: Conclusions and Discussion

The development of the chemically installed UAAs *S*-allylcysteine (Sac) and *Se*-allylselenocysteine (Seac) has provided a great impetus toward the use of CM in protein modifications. The installation of these UAAs allowed for the fascile access to post-translational modifications and bioconjugations in recombinant proteins, with minimal perturbation of the protein scaffolds. This has served to drive the search for general genetically incorporated CM tags for potential use in native biological systems, instead of recombinant, purified proteins alone. A study of reports on the CM-reactivities of chemically and genetically incorporated alkene-functionalised tags indicated that CM-reactive allyl chalcogen containing UAAs (e.g., Sac and Seac) were not reported to be readily genetically incorporated in *E. coli*. Conversely, readily incorporated alkene-containing UAAs (e.g., Hag) were not reactive in aqueous CM, a consequence of the 'carbon-only' side-chain. Based on this, four allyl chalcogen containing UAAs, *S*-allylcysteine (Sac), *Se*-allylselenocysteine (Seac), *S*-allyl homocysteine (Ahc), and *Se*-allyl homoselenocysteine (Ahs), were investigated as potential Met surrogates for the genetic incorporation of metathesis reactive tags.

In this chapter, it was demonstrated that, of the UAAs under consideration, the previously unexplored amino acid *S*-allyl homocysteine (Ahc) - though not Sac, Seac, or Ahs - is an effective methionine surrogate, and is recognised by the cognate *E. coli* Methionyl-tRNA synthetase MetRS. Comparison of the Michaelis-Menten kinetics parameters (k_{cat}/K_m ratios) for this enzyme showed that enzymatic efficiency for Met is 9-fold greater than for Ahc. Nevertheless, efficient Ahc incorporation into proteins was successfully demonstrated by the translational apparatus of Met-auxotrophic *E. coli* (> 95%). This now allows genetic control of olefin cross metathesis in proteins using sense (Met) codon reassignment. Pleasingly, incorporation wasalso demonstrated in mammalian HEK 293T cells (~80% single-site), indicating the use of Ahc as a general cross-species Met surrogate.

The resulting Ahc-containing proteins were found to retain their secondary structure as determined by CD, as well as their function, as demonstrated in the case of the enzyme SsβG and the DNA binding protein SarZ. Of these, Ahc- incorporated H3, Np276 and Qβ proteins were found to be reactive in CM reactions as well, using allyl alcohol as a model CM partner. Though a general reaction and metal-scavenging protocol still bears further investigating for

carrying out CM on SsβG and SarZ., nevertheless, these reactions serve as a first step toward a more general use of olefin cross-metathesis in chemical biology.

The CM catalyst used in this study was the Hoveyda-Grubbs 2nd generation catalyst, which is compatible with aqueous CM, though the presence of 30% *t*-BuOH was necesssary to keep the catalyst soluble. To prevent protein aggregation in the presence of the organic co-solvent, (unlike the previously reported CM reactions on chemically incorporated Sac and Seac tags on SBL) the reaction necessitated the use of strong denaturing conditions (5M guanidine hydrochloride or Urea). A CD-based protein folding and activity analysis revealed that both the native structure as well as the DNA binding ability of SarZ was lost upon treatment with 30% *t*-BuOH, and not recovered upon dialysis into non-denaturing buffer, suggesting that a more thorough refolding protocol may be necessary, such as solid phase-assisted,[50] or chromatography-based protein refolding.[51]

Alternatively, with the continuing development of water soluble catalysts for cross-metathesis,[52,53] there is great motivation for the development of *in vivo* Ahc-enabled CM in cellular systems, thereby enabling access to a broad range of biological applications.

2.11: Methods

2.11.1: General Consideration

Proton nuclear magnetic resonance (^{1}H NMR) spectra were recorded on a Bruker AV400 (400 MHz) or a Bruker AVII500 (500 MHz) spectrometer, as indicated. Carbon nuclear magnetic resonance (^{13}C NMR) spectra were recorded on a Bruker AV400 (100 MHz) spectrometer or on a Bruker AVII500 (125 MHz) spectrometer, as indicated. All chemical shifts are quoted on the δ scale in ppm using residual solvent as the internal standard. Coupling constants (*J*) are reported in Hertz (Hz) with the following splitting abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, quin = quintet, and a = apparent.

Infrared (IR) spectra were recorded on a Bruker Tensor 27 Fourier Transform spectrophotometer using thin films. Absorption maxima (υ_{max}) are reported in wavenumbers (cm^{-1}). Low resolution mass spectra (LRMS) were recorded on a Waters Micromass LCT Premier TOF spectrometer using electrospray ionization (ESI) and high resolution mass spectra (HRMS) were recorded on a Bruker MicroTOF ESI mass spectrometer. Nominal and exact m/z values are reported in Daltons.

Thin layer chromatography (TLC) was carried out using Merck aluminium backed sheets coated with $60F_{254}$ silica gel. Visualization of the silica plates was achieved using a UV lamp (λ_{max} = 254 nm), and/or potassium permanganate (5% $KMnO_4$ in 1M NaOH with 5% potassium carbonate). Flash column chromatography was carried out using BDH PROLAB® 40-63 mm silica gel (VWR). Mobile phases are reported in % volume of more polar solvent in less polar solvent for binary systems (e.g. 20% EtOAc in petrol = 1:4 ethyl acetate:petrol) and in relative composition for ternary systems (e.g. 1:2:4 H_2O:^{i}PrOH:EtOAc)

The anhydrous methanol used was purchased from Sigma. All other solvents were used as supplied (Analytical or HPLC grade), without prior purification. Distilled water was used for chemical reactions and Milli-Q® purified water for protein manipulations. 'Petrol' refers to the fraction of light petroleum ether boiling in the range 40-60 ºC. All reactions using anhydrous conditions were performed using flame-dried apparatus under an atmosphere of argon or nitrogen. Brine refers to a saturated solution of sodium chloride. Anhydrous magnesium sulfate ($MgSO_4$) or sodium sulfate (Na_2SO_4) was used as drying agents after reaction workup, as indicated.

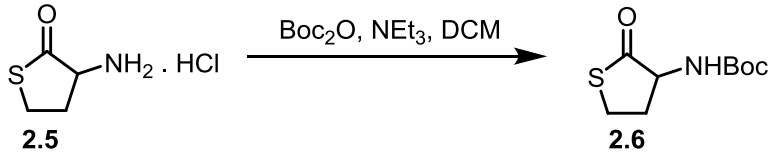
Protein Mass Spectrometry

Liquid chromatography-mass spectrometry (LC-MS) was performed on a Waters Micromass LCT Premier XE (ESI-TOF-MS) coupled to a Waters 1525 Binary HPLC pump using Chromolith FastGradient RP-18 2×50 mm column from Merck. Water (solvent A) and acetonitrile (solvent B), each containing 0.1% formic acid, were used as the mobile phase. The gradient was programmed as in shown in the table below. The electrospray source was operated with a capillary voltage of 3.2 kV and a cone voltage of 25 V for all the protein used in this report. Nitrogen was used as the nebulizer and desolvation gas at a total flow of 700 L/hr. Total mass spectra were reconstructed from the ion series using the MaxEnt algorithm preinstalled on MassLynx software (v. 4.1 from Waters) according to manufacturer's instructions.

Time(min)	Flow (mL)	%A	%B
0	0.4	95	5
1	0.4	95	5
5	0.4	0	100
8	1.0	0	100
8.1	1.0	95	5
9.5	1.0	95	5

2.11.2: Amino acid synthesis

***N*-Boc-DL-Homocysteine thiolactone (2.6)**

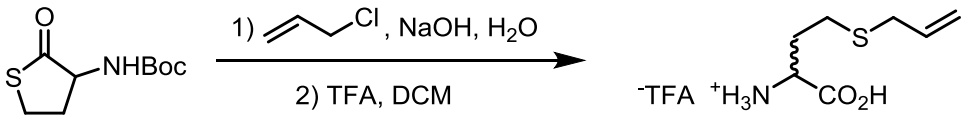


2.5 **2.6**

DL-Homocysteine thiolactone hydrochloride (3.0 g, 19.5 mmol) was added to a 100 mL round-bottomed flask with a Teflon™-coated stirrer bar and suspended in DCM (50 mL). The flask was put under argon atmosphere, stirred and cooled to 0 °C with an ice bath. Et_3N (5.0 mL, 39.0 mmol) was added portionwise. Boc_2O (5.0 g, 23.4 mmol) was added and the reaction was stirred at room temperature for 4 hours. After this time the reaction was evaporated to dryness with rotary evaporation. The resulting white solid was re-dissolved with EtOAc (150 mL), diluted with H_2O (300 mL) and transferred to a separatory funnel. The organic layer was isolated and the aqueous layer was further extracted with EtOAc (150 mL).

The combined EtOAc was washed with brine (200 mL), dried ($MgSO_4$) and then filtered. The solvent was removed under vacuum to give the desired compound as a white solid (4.2 g, 99%). NMR shows the desire product was pure enough without further purification. Spectroscopic data was identical to that previously reported.[54] m.p. = 130-133 °C. IR (υ_{max}, solid): 3350, 2978, 1677, 1519, 1368, 1300, 1239, 1157, 1053, 972, 863, 779, 612. ^{1}H NMR (400 MHz, $CDCl_3$): δ_H = 1.45 (9H, s, Boc), 1.98 (1H, qd, *J* = 12.2, 7.5, H_β), 2.72-2.94 (1H, m, H_β'), 3.23 (1H, ddd, *J* = 11.6, 7.1, 1.0, H_γ), 3.31 (1H, td, *J* = 12.1, 5.1, H_γ'), 4.12-4.45 (1H, m, H_α), 5.01 (1H, br. s., NH). ^{13}C NMR (100 MHz, $CDCl_3$): δ_C = 27.2 (C_γ), 28.3 (Boc), 32.1 (H_β), 60.5 (H_α), 80.4 (Boc, 4°), 155.5, 174.1 (2 × C=O). LRMS m/z (ESI+): Found 240.04 $[M+Na]^+$; $C_9H_{15}NO_3SNa$ requires 240.07.

_S_-Allyl-DL-Homocysteine TFA salt (2.7)

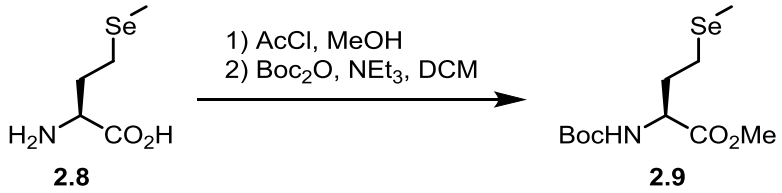


2.6 **2.7**

N-Boc-DL-Homocysteine thiolactone (2.0 g, 9.20 mmol) was added to a 100 mL round-bottomed flask and suspended in H_2O (40 mL). The suspension was stirred at room temperature and NaOH (1.5 g, 36.8 mmol) was added as a solid. After NaOH was completely dissolved, allyl chloride (4.0 mL, 46.0 mmol) was added. The emulsion was stirred vigorously at room temperature for 3 hours. After this time, TLC (50% EOAc in petrol with 0.5% AcOH) revealed disappearance of starting material (R_f = 0.6) and formation of the allylated product (R_f = 0.3). Reaction was acidified by adding 1M HCl until pH < 5 and then transferred to a separatory funnel. The product was extracted from the aqueous layer with EtOAc (2 × 100 mL). The organic layer was washed with brine (100 mL), dried with $MgSO_4$, filtered and then concentrated under reduced pressure to yield a light yellow oil. This was re-dissolved in DCM (30 mL) and TFA (6 mL) was then added. The reaction was stirred at room temperature for 2 hours after which TLC (50% EOAc in petrol with 0.5% AcOH) revealed complete consumption of starting material. The reaction was evaporated to dryness under reduced pressure to give a dark oil. The dark residue was diluted with a little bit of H_2O and cooled to 0 °C before adding NH_4OH (25% aq) dropwise until basic (pH > 8, pH paper). During this process a white, very insoluble solid crashed out. Purification by column

chromatography (1:2:7 25% NH_4OH (25% aq):MeOH:^{i}PrOH) gave the desired amino acid TFA salt as a white powder (2.4 g, 90%). m.p. = 225-226. IR (υ_{max}, solid): 2920, 1657, 1578, 1412, 1339, 1192, 1127, 988, 916, 840, 801, 724. ^{1}H NMR (400 MHz, D_2O): δ_H = 1.93-2.14 (2H, m, H_β), 2.49-2.57 (2H, m, H_γ), 3.13 (2H, d, *J* = 7.1, SCH_2), 3.74 (1H, dd, *J* = 6.9, 5.7, H_α), 5.06-5.13 (2H, m, $CH=CH_2$), 5.75 (1H, ddt, *J* = 17.1, 9.9, 7.2, 7.2, $CH=CH_2$). ^{13}C NMR (100 MHz, D_2O): δ_C = 25.7 (C_γ), 30.4 (SCH_2), 33.8 (C_β), 54.3 (C_α), 118.2 ($CH=CH_2$), 134.2 ($CH=CH_2$), 174.5 (C=O). LRMS m/z (ESI+): Found 176.07 $[M+H]^+$; $C_7H_{14}NO_2S$ requires 176.07.

N-Boc-L-Selenomethionine methyl ester (BocSeMetOMe, 2.9)

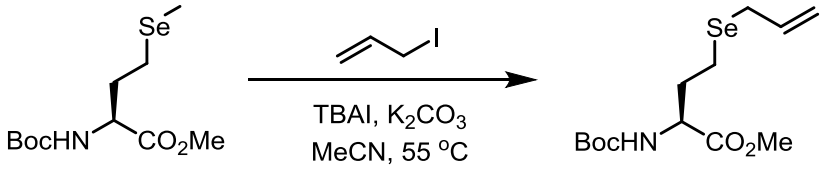


2.8 **2.9**

Anhydrous MeOH (30 mL) was added to a flame-dried 100 mL round-bottomed flask under argon with a Teflon™-coated stirrer bar. The solvent was stirred, cooled to 0 °C with an ice bath and AcCl (10 mL, 153 mmol) was added dropwise over 5 minutes. The solution was stirred for an additional 10 minutes at 0 °C, giving a concentrated solution of HCl. L-SeMet (3.0 g, 15.3 mmol) was added under a stream of argon. The ice bath was removed and the reaction stirred at room temperature for 18 hours. The solvent was then removed under reduced pressure to give the crude selenomethionine methyl ester hydrochloride as a yellow solid. This material was used immediately in the next step without purification. The crude ester was suspended in DCM (30 mL), stirred and cooled to 0 °C with an ice bath. To this, Et_3N (4.3 mL, 30.6 mmol) was added portionwise. Boc_2O (4.0 g, 18.4 mmol) was added and the reaction was stirred at room temperature for 4 hours. After this time TLC (30% EtOAc in Petrol, $KMnO_4$) showed formation of the product (R_f = 0.4). Reaction was diluted with DCM (100 mL) and transferred to a separatory funnel. The organic layer was washed with H_2O (200 mL), brine (200 mL), dried ($MgSO_4$) and then filtered. The solvent was removed under vacuum to give a clear oil. Purification by column chromatography (20% EtOAc in Petrol) yielded the titled compound as a clear oil (4.0 g, 84%). $[\alpha]_D$ = +27.5 (c = 0.77, $CHCl_3$). IR (υ_{max}, film): 3359, 2977, 1714, 1513, 1438, 1366, 1247, 1164, 1050, 1024, 779. ^{1}H NMR (400

MHz, $CDCl_3$): δ_H = 1.40 (9H, s, Boc), 1.90-2.01 (4H, m, H_β + SCH_3), 2.07-2.19 (1H, m, H_β'), 2.50 (2H, t, *J* = 7.7, H_γ), 3.71 (3H, s, CO_2CH_3), 4.27-4.42 (1H, m, H_α), 5.16 (1H, d, *J* = 7.3, NH). ^{13}C NMR (100 MHz, $CDCl_3$): δ_C = 4.1 ($SeCH_3$), 20.3 (C_γ), 28.2 (Boc), 33.2 (C_β), 52.3 (CO_2CH_3), 52.5 (C_α), 79.9 (Boc), 155.3, 172.7 (2 × C=O). HRMS m/z (ESI+): Found 334.0527 $[M+Na]^+$; $C_{11}H_{22}NO_4SeNa$ requires 334.0528.

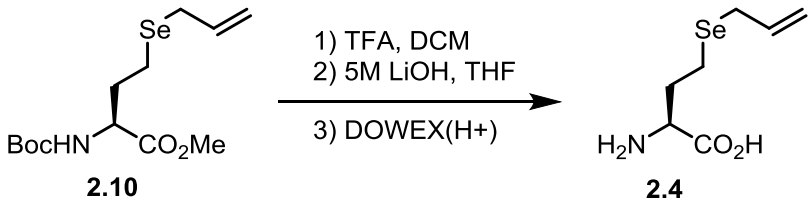
Se-allyl-N-Boc-L-Homoselenocysteine methyl ester (BocAhsOMe, 2.10) via One-pot Allylation/ Demethylation



2.9 **2.10**

BocSeMetOMe (1.2 g, 3.87 mmol) was added to a 50 mL round-bottomed flask equipped with a Teflon™-coated stirrer bar and dissolved in MeCN (15 mL) then heated to 55 °C. To this, K_2CO_3 (0.8 g, 5.80 mmol), allyl iodide (1.1 mL, 11.6 mmol) and TBAI (2.14 g, 5.80 mmol) were added subsequently. The reaction mixture was stirred for 16 hours at 55 °C and. The solvent was removed under reduced pressure to give a pale yellow oil. This was taken up with Et_2O (100 mL) and H_2O (150 mL) and transferred to a separatory funnel. The organic layer was separated and the aqueous layer was further extracted with Et_2O (2 × 75 mL). The combined organic layer was washed with brine (150 mL), before drying over $MgSO_4$, filtering and reduction under vacuum to a yellow oil. Following purification via column chromatography (10% EtOAc in Petrol) to give the product as a clear oil (0.9 g, 69%). $[\alpha]_D$ = +26.9 (c = 0.48, $CHCl_3$). IR (υ_{max}, film): 3361, 2977, 1714, 1506, 1437, 1366, 1164, 1049, 914, 757. ^{1}H NMR (400 MHz, $CDCl_3$): δ_H = 1.45 (9H, s, Boc), 1.90-2.06 (1H, m, H_β), 2.06-2.23 (1H, m, H_β'), 2.43-2.57 (2H, m, H_γ), 3.18 (2H, d, *J* = 7.6, $SeCH_2CH$=), 3.75 (3H, s, CO_2CH_3), 4.35-4.44 (1H, m, H_α), 4.97-5.09 (2H, m, $CH=CH_2$), 5.80-5.92 (1H, m, $CH_2CH=CH_2$). ^{13}C NMR (100 MHz, $CDCl_3$): δ_C = 18.0 (C_γ), 25.9 ($SeCH_2CH$=), 28.3 (Boc), 33.6 (C_β), 52.4 (CO_2CH_3), 53.6 (C_α), 116.3 ($CH=CH_2$), 134.7 ($CH_2CH=CH_2$), 154.5, 172.7 (2 × C=O). HRMS m/z (ESI+): Found 360.0671 $[M+Na]^+$; $C_{13}H_{23}NO_4SeNa$ requires 360.0685.

Se-Allyl-L-homoselenocysteine (2.4)



2.10 **2.4**

BocAhsOMe (1.83 g, 5.44 mmol) was added 100 mL round-bottomed flask equipped with a Teflon™- coated stirrer bar and then dissolved in DCM (20 mL). TFA (4 mL) was added and reaction mixture stirred at room temperature for 2 hours. After this time TLC (15% EtOAc in Petrol, ninhydrin) indicated consumption of the starting material (R_f = 0.4). The solvent was removed under reduced pressure to yield a yellow oil. The residue was re-dissolved in THF (6 mL) and cooled to 0 °C with an ice bath. LiOH (14 mL of a 5M solution in H_2O) was added dropwise and then the ice bath removed. The reaction mixture was stirred at room temperature for 2 hours. After this time the solution was diluted with H_2O (20 mL) and neutralised with DOWEX® 50WX8 (H+) until pH < 7 (pH paper). The suspension was poured onto an empty column and washed with water (100 mL, gravity flow).The flow-through was discarded and the column eluted with 5% NH_4OH (aq). The fractions containing product (identified by TLC) were combined and concentrated by rotary evaporation to give a yellow solid. This was purified by column chromatography (1:2:7 25% NH_4OH (25% aq):MeOH:^{i}PrOH) to give a white powder (700 mg, 58% yield). m.p. = 200-201 °C. IR (υ_{max}, solid): 3290, 1574, 1421, 1361, 1177, 988, 914, 861. ^{1}H NMR (400 MHz, D_2O): δ_H = 1.60-1.83 (2H, m, H_β), 2.32-2.41 (2H, m, H_γ), 3.05 (2H, d, *J* = 7.8, $SeCH_2CH$=), 3.10 (1H, dd, *J* = 7.1, 5.8, H_α), 4.83-4.95 (2H, m, $CH=CH_2$), 5.73 (1H, ddt, *J* = 17.2, 9.6, 7.8, 7.8, $CH_2CH=CH_2$). ^{13}C NMR (100 MHz, D_2O): δ_C = 19.1 (C_γ), 25.6 ($SeCH_2CH$=), 35.8 (C_β), 56.5 (C_α), 116.9 ($CH=CH_2$), 135.3 ($CH_2CH=CH_2$), 183.1 (C=O). HRMS m/z (ESI–): Found 222.0039 $[M-H]^-$; $C_7H_{12}NO_2Se$ requires 222.0039.

2.11.3: MetRS Activity Assays

1) AP_4A Protocol:

MetRS was buffer exchanged into aminoacylation buffer (50 mM HEPES pH 7.6, 10 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, 10% (v/v) DMSO, 10 mM D-glucose, 0.5 mM $NADP^+$, 1.7 $mM.min^{-1}$ hexokinase and glucose 6-phosphate dehydrogenase from baker's yeast (Sigma). In a final volume of 200 µL, containing aminoacylation buffer, 2.4 µM MetRS, 20 µM pyrophosphate and 0.25 to 10 mM Met, reactions were initiated by the addition of 0.7 mM AP_4A. Absorbance was continuously monitored at 340 nm, to detect reduction of NADP+ to NADPH, where ε_{NADPH} at 340nm = 6220 $M^{-1}cm^{-1}$

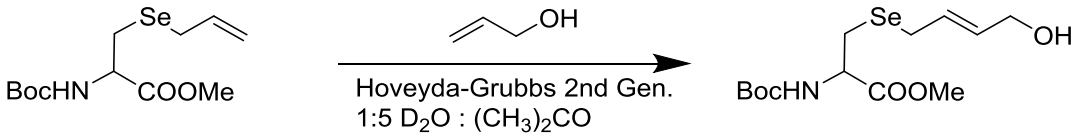
2) Malachite Green assay protocol:

MetRS was buffer exchanged into aminoacylation buffer (30 mM HEPES, 140 mM NaCl, 30 mM KCl, 40 mM $MgCl_2$), and incubated at 37 ºC (55 ug/mL enzyme conc., 100 µL volume) with 1 mg/mL *E. coli* total tRNA, 2 u/mL PPiase, 100 uM ATP, and 1 mM DTT with 0.1 mM, 0.5 mM, 1.0 mM, 2.5 mM, 5.0 mM, 7.5 mM, 10 mM, and 20 mM methionine.. At regular time intervals, 20 µL aliquots were quenched with 100 mM EDTA (80 µL), and developed for 30 min at r.t. with the addition of 10 uL Malachite green solution. Absorbance was measured at 620 nm.

3) Radioactive MetRS protocol:

To determine which UAAs were recognised by MetRS, reactions were run for 20 min in 150 uL of reaction buffer each (pH 7.6, 20 mM imidazole, 0.1 mM EDTA, 10 mM 2-mercaptoethanol, 7 mM MgCl2, 2 mM ATP, 0.1 mg/mL BSA, and 2 mM of ^{32}P-PPi (in the form of sodium pyrophosphate with a specific activity of 0.5 TBq/mol), 70 nM MetRS and 5 mM Met, Sac, Seac, Ahc, or Ahs. After 20 min, 20 uL from each reaction was quenched in 0.5 mL of a solution comprising 200 mM PPi, 7% w/v $HClO_4$, and 3% w/v activated charcoal. The charcoal was washed twice with 0.5 mL of a 10 mM PPi, 0.5% $HClO_4$ solution, and then transferred to a scintillation vial for scintillation counting. To determine Michaelis-Menten kinetics parameters, the reactions were set up as above, with aliquots quenched at regular time intervals.

2.11.4: Small Molecule CM Seac Reaction Kinetics



To a solution of *N*-Boc-Se-allylselenocysteine-methoxy ester[55] (BocSeacOMe) and Hoveyda-Grubbs 2nd generation catalyst in 1:5 D_2O: $(CD_3)_2CO$, allyl alcohol (neat) was added, and the progress of the cross-metathesis reaction was monitored by 1H NMR. NMR peak heights of the starting material relative to the residual solvent peaks were converted to concentrations at various time points. Data obtained for initial time points was fitted by a single-phase exponential decay model ($A = A_0e^{-kx}$) using Prism V6.02 (GraphPad Software Inc.) to obtain pseudo first-order rate constants at different concentrations of allyl alcohol. The observed pseudo first-order rate constants at various allyl alcohol concentrations were fitted by linear regression to determine the second order rate constant (Fig. 2.35, 2.36, 2.37).

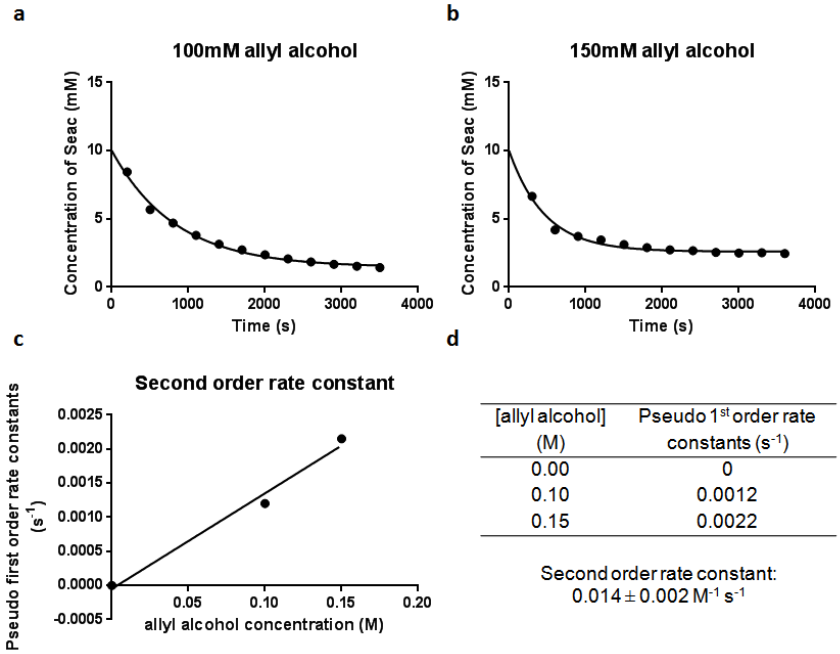
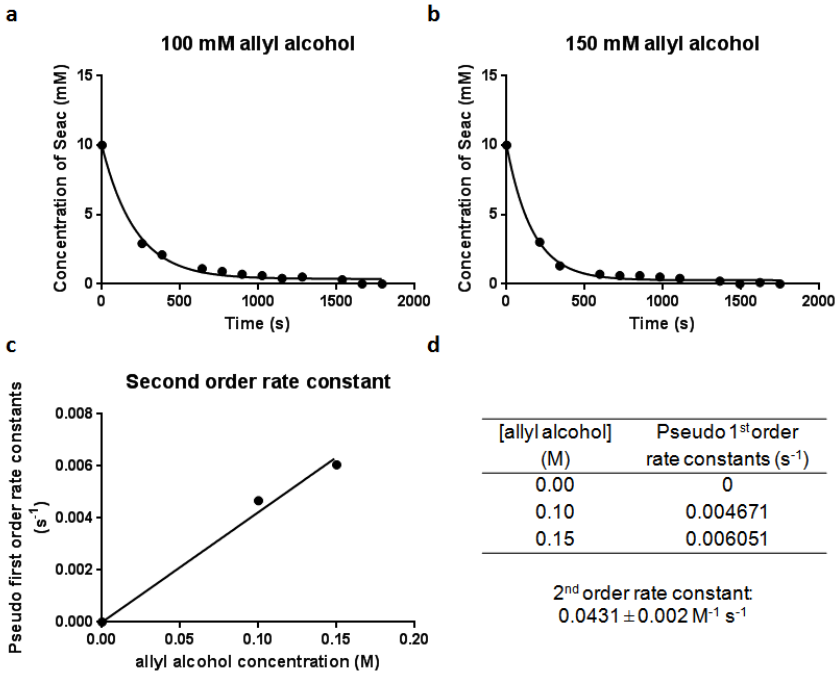


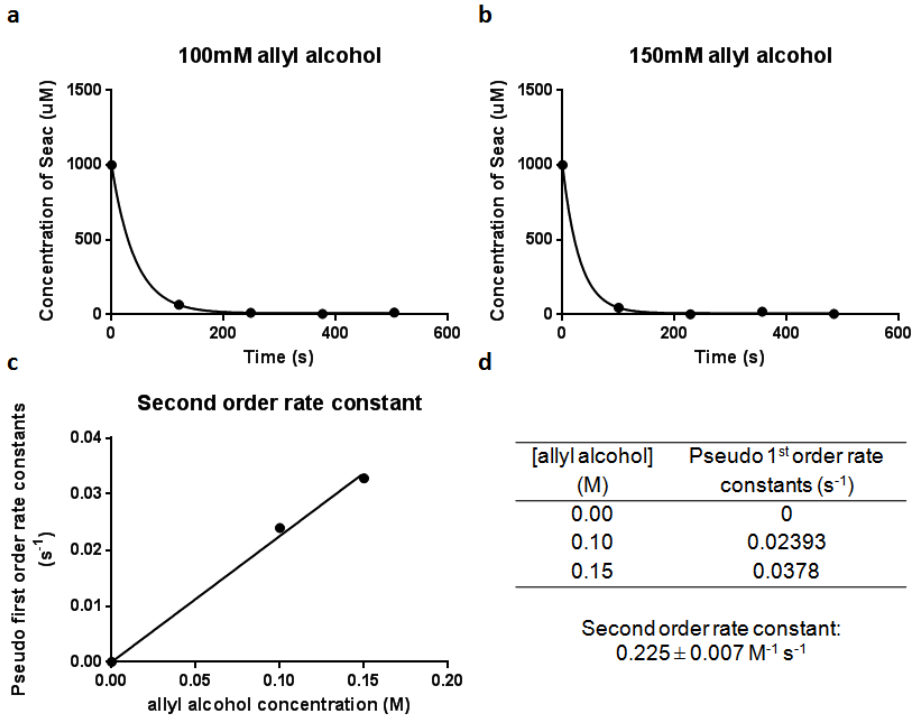
Fig 2.35: NMR signal intensity decay with 10 mM BocSeacOMe, 10 mM catalyst and allyl alcohol at (a) 100 mM and (b)150 mM. (c) Pseudo-first order rate constants v/s allyl alcohol concentration.



[allyl alcohol] (M)	Pseudo 1st order rate constants (s^{-1})
0.00	0
0.10	0.004671
0.15	0.006051

2nd order rate constant:
$0.0431 \pm 0.002\ M^{-1}\ s^{-1}$

Fig 2.36: NMR signal intensity decay with 10 mM BocSeacOMe, 50 mM catalyst and allyl alcohol at (a) 100 mM and (b) 150 mM. Panel (c) Pseudo-first order rate constants v/s allyl alcohol concentration.



[allyl alcohol] (M)	Pseudo 1st order rate constants (s^{-1})
0.00	0
0.10	0.02393
0.15	0.0378

Second order rate constant:
$0.225 \pm 0.007\ M^{-1}\ s^{-1}$

Fig 2.37: NMR signal intensity decay with 1 mM BocSeacOMe, 50 mM catalyst and allyl alcohol at (a) 100 mM and (b) 150 mM. Panel (c) Pseudo-first order rate constants v/s allyl alcohol concentration.

2.11.5: SBL Activity Assay

Stock solutions of SBL-S156C (unmodified), SBL-S156Dha, and SBL-S156Seac were prepared in 100 mM Tris with 0.005% Tween-80 (pH 8.6, 'Dilution buffer') at concentrations of 1.12 x 10^{-7} M. Dilution buffer (98 µL) was added to wells of a 96-well plate. Aliquots of the substrate N-Succinyl-Ala-Ala-Pro-Phe p-nitroanilide[56] (Suc-AAPF-pNA, 0.20 M in DMSO, Bachem) were added to each well in a final volume of 2 µL DMSO, and reactions were initiated by the addition of 100 µL of the SBL protein stocks to give 8 substrate concentrations ranging from 0.2 mM to 4.0 mM.[57] All SBL-containing were seen to turn yellow over the course of the reaction, indicating liberation of pnitroaniline (pNA) and confirming peptidase activity of all SBL samples. All protein solutions and the substrate solution alone at the same concentrations are colorless. The progress of the reaction was monitored by measuring the 410 nm (ε = 8800 M^{-1} cm^{-1}) absorbance of the liberated p-nitroaniline. Data was collected every 18 seconds, and fitted by nonlinear regression to the Michaelis-Menten model using Prism V 6.02 (GraphPad Software, Inc) to obtain the kinetic parameters k_{cat} and K_m.

2.11.6: "In-gel" Tryptic Digestion-MSMS Analysis of Expressed Proteins

"In-gel" tryptic digestions were performed following an MS-compatible protocol. Following SDS-PAGE analysis of the protein, gel bands of the desired protein were excised from the gel, diced into 1 mm cubes and placed in Eppendorf tubes. The bands were washed with 100 µL of Solution A (40 mg ammonium bicarbonate in 10 mL Milli-Q water and 10 mL HPLC grade acetonitrile) for 30 minutes and the supernatant was discarded. The process was repeated 2 times, until the gel pieces were no longer dark blue. The gel was then washed with 100 µL acetonitrile until the gel pieces dehydrated and turned white (10 minutes). The supernatant was discarded and the samples were air-dried. 100 µL of a 10 mM DTT solution in 25 mM ammonium bicarbonate was added to the dry gel pieces and incubated at 37 °C for 30 minutes. The supernatant was discarded, and the gel pieces were washed twice with 25 mM ammonium bicarbonate solution, followed by acetonitrile until the gel pieces dehydrated and turned white (10 minutes). 100 µL of 55 mM iodoacetamide solution in 25 mM ammonium bicarbonate solution was then added to the gel pieces and incubated for 60

minutes in the dark. The supernatant was then discarded and the gel pieces were washed twice with 100 µL Solution A for 10 minutes. The pieces were then washed with 100 µL acetonitrile until the gel pieces dehydrated and turned white (10 minutes). The supernatant was discarded and the samples were air-dried. 20 µL of reconstituted Promega Sequencing Grade Modified Trypsin solution (1 µg) was then added and the digest was incubated at 37 °C for 16 h. 1 µL of formic acid was then added in order to stop the digest, and the supernatant containing the peptides was transferred to a new tube. 50 µL of extraction buffer (20 µL formic acid in 10 mL Milli-Q water and 10 mL HPLC grade acetonitrile) was then added to the gel pieces and the mixture was incubated for 30 minutes. The supernatant was pooled with the supernatant from the previous step. The combined supernatant was then concentrated to dryness by SpeedVac and then transferred to MS vials in 10 µL water. The samples were analyzed by MS/MS-MS.

2.11.7: General Procedure for Expression of MetRS

Overnight cultures of single colonies of BL21 (DE3) transformed with the MetRS gene in pET303/CT-His vector were used to inoculate L.B. media (1 L) containing ampicillin (50 µg/mL) and cells incubated at 37 °C. The optical density (OD_{600}) was monitored until a value of 0.7 was obtained (~3-4 hours) and expression was induced for 4 h with the addition of 0.1 mM IPTG. The cells were harvested by centrifugation (8 minutes, 8000 rpm) to give a pellet, which was collected in a 50 mL falcon tube and resuspended in Wash Buffer (50 mM Tris HCl , pH 7.5, 100 mM NaCl, 1 mM Na-EDTA, Protease inhibitor, 5 mM 2-mercaptoethanol, 20 mL, 10 mM imidazole), freeze-thawed twice, and then lysed by sonication. The cell debris was spun down by centrifugation (23000 RPM, JA25.50, 20 min, 4 C). The supernatant was loaded onto a His-Trap Ni-FPLC column (5 mL), washed with 10 mM imidazole buffer, and eluted with a gradient between 10 mM and 400 mM imidazole. Fractions were run on a 4-12% Bis-Tris SDS-PAGE in MOPS for 35 min at 200 V. MetRS containing fractions were pooled, syringe-filtered, and dialysed against 50 mM HEPES-NaOH (pH 7.6), 100 mM KCl, 10 mM $MgCl_2$, 7 mM β-Mercaptoethanol, then stored at -80 C in 30% glycerol

2.11.8: General Procedure for Expression of Histone H3 K9C M90L C96S G102A C110A Single Ahc120

An *E. Coli* B834(DE3) glycerol stock of Histone H3 K9C M90L C96S G102A C110A Single M120 was used to inoculate L.B. media (20 mL) containing kanamycin (50 µg/mL) and incubated for 16 hours at 37 °C. The starter culture was used to inoculate Selenomet media (1 L), supplemented with Selenomet nutrient mix, L-methionine (20 mg/L) and kanamycin (50 µg/mL), the cells incubated at 37 °C, and the optical density (OD_{600}) monitored until a value of 0.7 was obtained (~3-4 hours). The cells were then pelleted by centrifugation (8 minutes, JLA-9.1000 rotor, 8000 rpm). The pellets were washed three times with SelenoMet base media (3 × 300 mL) containing kanamycin (50 µg/mL) and then transferred to prewarmed SelenoMet media (1 L) containing kanamycin (50 µg/mL) and supplemented with Selenomet nutrient mix as well as the desired unnatural amino acid (DL-Ahc, 500 mg/L). The culture was incubated for 30 minutes at 37 °C, followed by 30 min at 30 °C. Expression was induced by addition of IPTG (1 mM) and was continued for 16 hours at 30 °C (220 rpm). The cells were harvested by centrifugation (8 minutes, 8000 rpm) to give a pellet, which was collected in a 50 mL falcon tube and resuspended in Wash Buffer (50 mM Tris HCl , pH 7.5, 100 mM NaCl, 1 mM Na-EDTA, Protease inhibitor, 5 mM 2-mercaptoethanol, 20 mL), freeze-thawed twice, and then lysed by sonication. The cell debris was spun down by centrifugation (23000 RPM, JA25.50, 20 min, 4 C).

The cell pellet was re-suspended in TW buffer (Wash buffer + 1% w/v Triton-X, 30 mL), then centrifuged again. This was repeated twice more with TW buffer, then once more with Wash buffer. The pellet was distributed amongst four 2 mL eppendorf tubes with a spatula, and incubated with 125 µL DMSO each. The pellets were vortexed, and the volume was made up to 2 mL in each tube with unfolding buffer (6M Guanidine hydrochloride, 20 mM Tris HCl pH 7.5, 10 mM 2-mercaptoethanol). The contents were incubated at r.t. with shaking for 1 h, then spun down at 15000 g. The extracted protein in the supernatant was collected, and purified by gel filtration FPLC using a HiLoad Superdex 26/60 200 Column, isocratic elution with SAU-1000 buffer (7M Urea, 20 mM Sodium Acetate, 1M NaCl, 5 mM 2-mercaptoethanol, 1mM Na-EDTA, pH 5.2).

Eluted proteins were collected in 10 mL fractions, which were analysed by SDS-PAGE (NuPAGE 4-12% Bis-Tris gel, Invitrogen, MES buffer, 200 V, 35 min). Histone H3

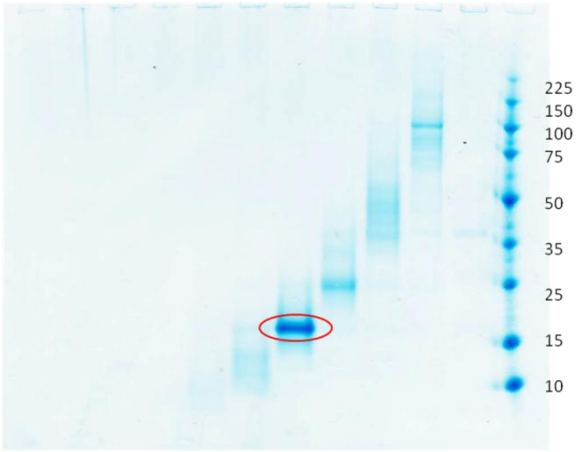
containing fractions were pooled, dialysed against water with 5 mM 2-mercaptoethanol, then lyophilised and stored at -20 °C.

Sequence Data for Histone H3 K9C M90L C96S G102A C110A Single M120:

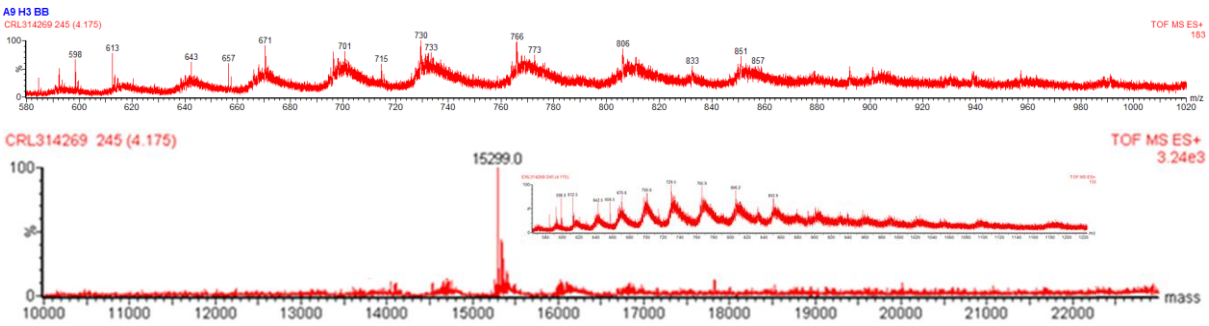
MARTKQTARCSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQ
KSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVLALQEASEAYLVALFEDTNLAAIHA
KRVTIXPKDIQLARRIRGERA

If X = Met, Calculated mass = 15196 (N-terminus Met cleaved)
If X = Ahc, Calculated mass = 15196 + 26 = 15222 (N-terminus Met cleaved)

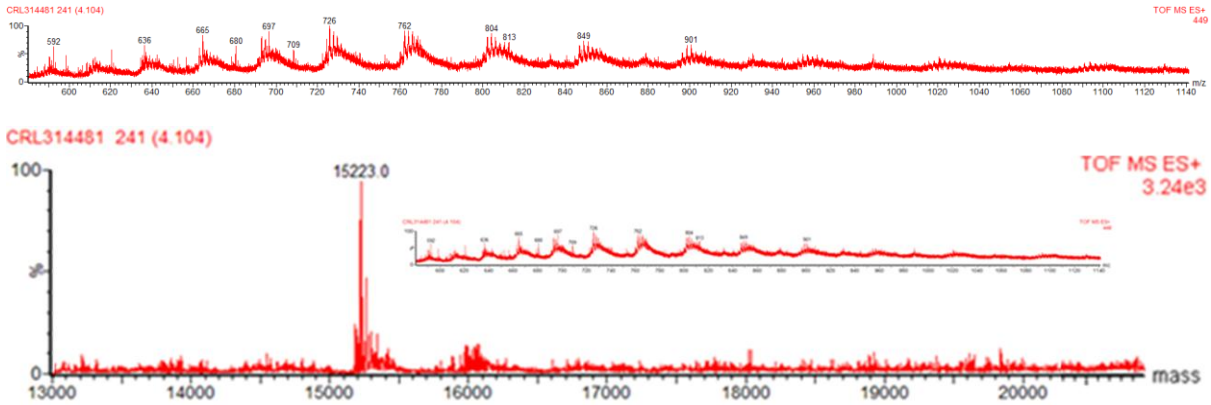
SDS-PAGE:



LCMS:



Upon treatment of the protein with DTT for 30 min, the mass of H3 Ahc 120 with a free cysteine is seen:



Expected mass: 15222 Da, Observed: 15223 Da

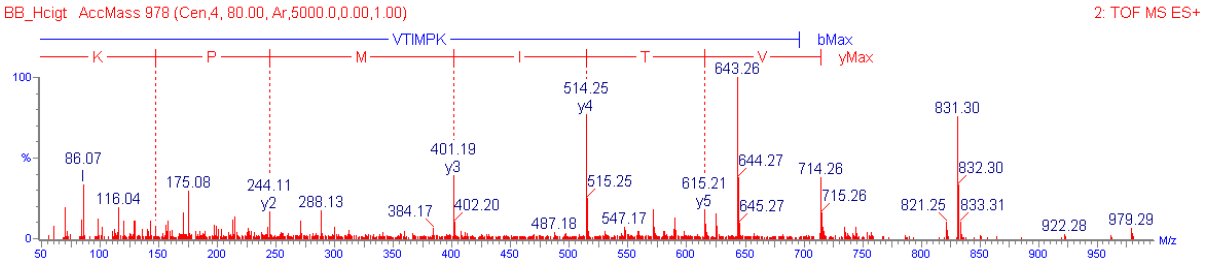
MSMS Data for Histone H3 K9C M90L C96S G102A C110A Ahc120:

ARTKQTARCSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQK STELLIRKLPFQRLVREIAQDFKTDLRFQSSAVLALQEASEAYLVALFEDTNLAAIHAK RVTIMPKDIQLARRIRGERA

The peptide containing the unnatural amino acid Ahc, denoted as “M” in this case, is highlighted in the sequence above. The MSMS data showed good coverage of the fragmented peptide and received a score of 23 on BioLynx analysis.

Observed MW: 713.4244 Precursor ion charge state: 2
M/z tolerance: 0.30 Intensity threshold: 191 (0.750%)
Modifications: Ahc (+)

a	72.08	173.13	286.21	443.27	540.33	668.42
	0.02	---	---	---	---	---
b	100.08	201.12	314.21	471.27	568.32	696.42
	---	---	---	---	---	---
	Val	**Thr**	**Ile**	**Ahc**	**Pro**	**Lys**
y	714.43	615.36	514.31	401.23	244.17	147.11
	---	0.15	0.06	0.03	0.06	0.04
z	697.40	598.33	497.28	384.20	227.14	130.08
	---	---	0.05	0.03	---	---



2.11.9: Cross Metathesis on Histone H3 K9C M90L C96S G102A C110A Ahc120 with allyl alcohol

A solution of Histone H3 K9C M90L C96S G102A C110A Ahc120 (60 µg) was reconstituted in 5 M Guanidine hydrochloride – 50 mM sodium phosphate (105 µL, pH 8.0). To this protein solution was added $MgCl_2.6H_2O$ (4 µL of a 1 mg/µL solution in water, 19.7 µmol, 5000 eq.) and Hoveyda-Grubbs 2nd generation catalyst in *t*BuOH solution (0.3 mg in 43 µL *t*BuOH, 0.47 µmol, 100 eq.). The reaction was mixed on a vortex each time after an addition. Reaction was shaken at RT for 1 minute before adding allyl alcohol (1.14 µL, 19.7 µmol, 5000 eq.). Reaction was mixed on a vortex immediately after addition and was shaken at RT for 3 hours. 3-mercaptopropionic acid (10 µL) was added as a metal scavenger for ruthenium and the mixture was shaken for further 30 minutes at RT. The protein sample was then purified from small molecules by dilution with 5M GdmCl-50mM sodium phosphate (pH 8.0, 400 µL) and then concentration by Vivaspin® (10kDa MWCO) to 100 µL. This process was repeated twice, followed by further removal of small molecules using a PD-Spintrap column.

LCMS analysis of the protein sample revealed complete conversion to the desired CM product (calculated = 15252, found = 15252, Section **2.8**).

2.11.10: General procedure for expression of Qβ-K16M

Plasmid carrying the K16M Qβ gene in p75M vector was transformed into *E. Coli* B834(DE3) competent cells and was plated on LB-agar supplemented with ampicillin (100 µg/mL) and incubated at 37 °C for 18 hours. A single colony from the plate was used to inoculate SelenoMet media (10 mL) containing ampicillin (100 µg/mL) and supplemented with L-methionine (40 µg/mL). The overnight culture were used to inoculate SelenoMet media (625 mL) containing ampicillin (100 µg/mL) and supplemented with L-methionine (40 µg/mL). The cultures were incubated (37 °C, 180 rpm) and grown to OD_{600} between 0.6 and 0.8. The cells were then pelleted by centrifugation (7,000 rpm, 8 minutes). The pellets were washed with SelenoMet media (3 × 300 mL) containing ampicillin (100 µg/mL) and then transferred to prewarmed (37 °C) SelenoMet media (625 mL) with containing ampicillin (100 µg/mL) and supplemented with amino acid **5.11** (500 mg/L, 1.7 mM) in place of methionine. The culture was incubated for 30 minutes at 37 °C, followed by 1 hour at 30 °C. Expression was induced by addition of IPTG (1 mM) and was continued for 6 hours (37 °C, 180 rpm). The cells were harvested by centrifugation (7,000 rpm, 8 minutes) and were collected and suspended in TBS buffer (20 mM Tris-HCl, 150 mM NaCl, 2 mM EDTA, pH 7.5, 30 mL) and stored at –80 °C.

The cells were thawed and sonicated (Soniprep at 10 micron amplitude, 30 seconds with 1 minute rest, 5 times, on ice) and the cell debris removed by centrifugation (20,000 rpm, 40 minutes). The clear lysate was added to an equal volume of 1:1 n-BuOH:$CHCl_3$ and mixed by inversion. The layers were separated by centrifugation (3750 rpm, swing-out, 30 minutes). The aqueous layer was isolated and the VLPs were precipitated by addition of PEG8000 (10% w/v) and NaCl (100 mM) to the aqueous layer, dissolved by vortexing and the suspension rocked at 4 °C for 30 minutes. The precipitated Qβ particles were collected by centrifugation (3750 rpm, swing-out, 30 minutes) and resuspended in 10 mL TBS buffer with the supernatant discarded. The extraction step was repeated a second time on the resuspended

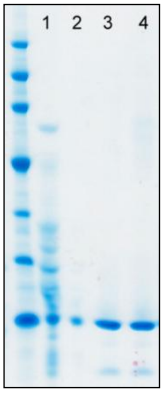
Qβ. The resulting aqueous layer was isolated and dialyzed into 50 mM sodium phosphate buffer (pH 8).

Sequence Data for Qβ-K16Ahc

AKLETVTLGNIGKDGXQTLVLNPRGVNPTNGVASLSQAGAVPALEKRVTVSVSQPSR
NRKNYKVQVKIQNPTACTANGSCDPSVTRQAYADVTFSFTQYSTDEERAFVRTELAA
LLASPLLIDAIDQLNPAY

If X = Met, Calculated mass = 14126

If X = Ahc, Calculated mass = 14126 + 26 = 14152



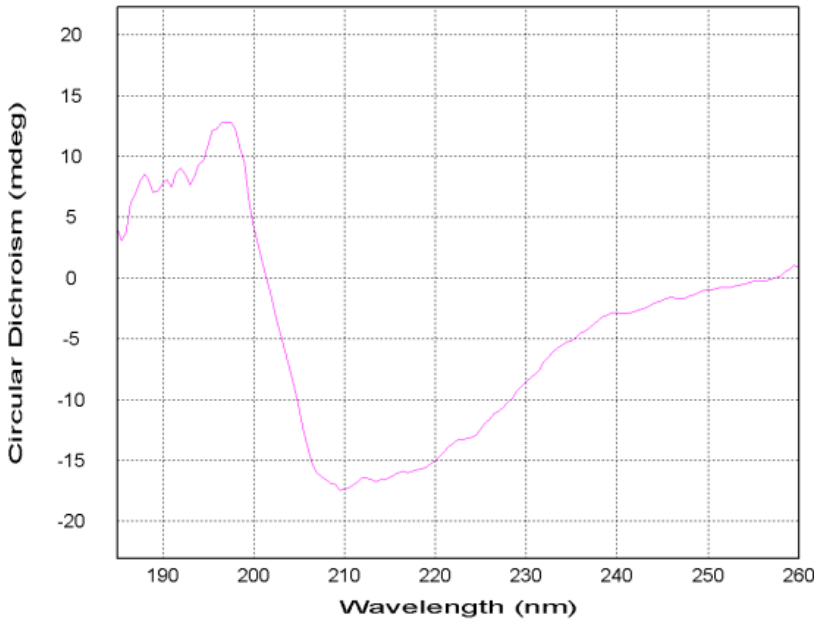
1: Qβ-Ahc aqueous layer after 1st extraction

2: Qβ-Ahc aqueous layer after 2nd extraction

3: Qβ-Ahc VLP

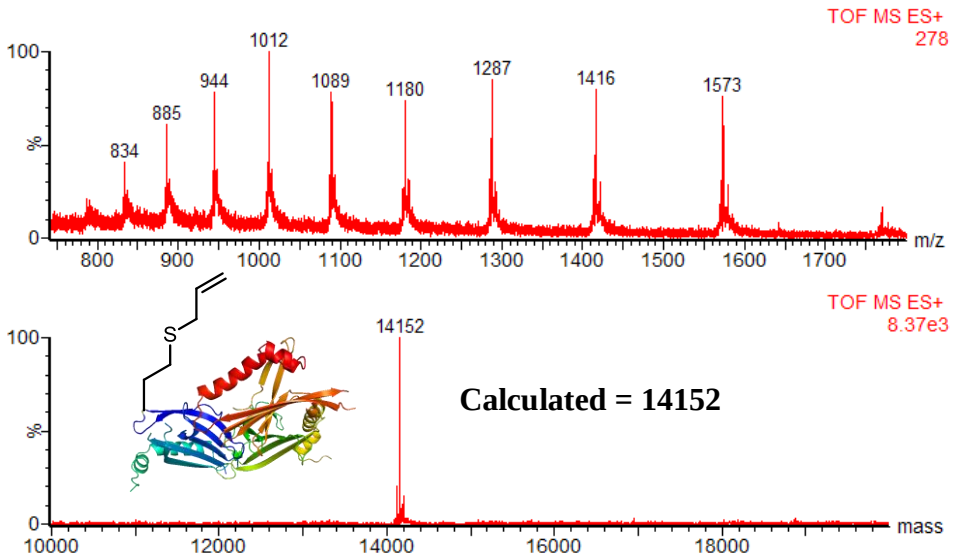
4: Qβ-Ahc VLP after dialysis

CD Spectrum:



The Qβ-VLPs were reduced by adding 2 µL of 0.5M TCEP solution to 50 µL of protein solution (c = 0.2) then small scale dialysis (30 minutes, 500 mL sodium phosphate buffer (50

mM, pH 8)) prior to analysis by LCMS. The LCMS spectra of the protein monomer are shown below.



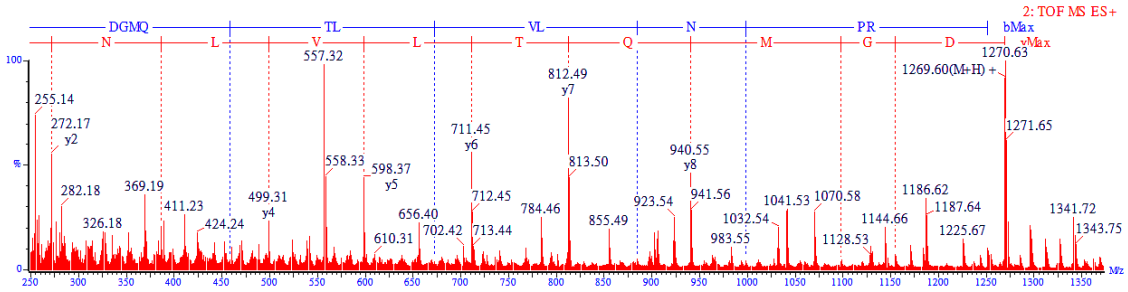
MSMS Data of Tryptic Digested Qβ-K16Ahc

AKLETVTLGNIGKDGMQTLVLNPRGVNPTNGVASLSQAGAVPALEKRVTVSVSQPS
RNRKNYKVQVKIQNPTACTANGSCDPSVTRQAYADVTFSFTQYSTDEERAFVRTELA
ALLASPLLIDAIDQLNPAY

The peptide containing the unnatural amino acid Ahc, denoted as "M" in this analysis, after tryptic digestion of Qβ-K16Ahc is highlighted in the sequence above. The MSMS data showed good coverage of the fragmented peptide and received a score of 123 on BioLynx analysis (MassLynx v4.1, from Waters).

Observed MW: 1268.6522 Precursor ion charge state: 1
M/z tolerance: 0.30 Intensity threshold: 3970 (0.750%)
Modifications: Ahc (+)

a	88.04	145.06	302.12	430.18	531.23	644.31	743.38	856.46	970.51	1067.56	1223.66
	---	---	-0.05	---	---	---	---	---	---	---	---
b	116.03	173.06	330.12	458.18	559.22	672.31	771.38	884.46	998.50	1095.56	1251.66
	---	---	---	-0.04	---	-0.03	---	-0.03	-0.02	---	---
	Asp	**Gly**	**Ahc**	**Gln**	**Thr**	**Leu**	**Val**	**Leu**	**Asn**	**Pro**	**Arg**
y	1269.67	1154.64	1097.62	940.56	812.50	711.45	598.37	499.30	386.22	272.17	175.12
	---	-0.00	0.01	0.00	0.01	-0.00	-0.00	-0.01	-0.01	0.00	---
z	1252.64	1137.61	1080.59	923.53	795.47	694.42	581.34	482.27	369.18	255.14	158.09
	---	0.02	0.01	-0.01	---	---	---	-0.01	-0.01	-0.00	---



2.11.11: Cross Metathesis on Qβ-K16Ahc with allyl alcohol

A solution of Qβ-K16Ahc (c = 2.0 mg/mL, 35 µL) in 50 mM sodium phosphate buffer (pH 8) was diluted to 700 µL with 5M guanidinium chloride (GdmCl) and concentrated back down with a Vivaspin® (10kDa MWCO) to a final volume of 100 µL (c = 0.7 mg/mL, 0.03 mM). This process was repeated twice. To this protein solution was added $MgCl_2.6H_2O$ (10 µL of a 1 mg/µL solution in water, 49 µmol, 10000 eq.) and Hoveyda-Grubbs 2nd generation catalyst in *t*BuOH solution (0.4 mg in 43 µL *t*BuOH, 0.64 µmol, 133 eq.). The reaction was mixed on a vortex each time after an addition. Reaction was shaken at RT for 1 minute before adding allyl alcohol (1.1 µL, 17 µmol, 3340 eq.). Reaction was mixed on a vortex immediately after addition and was shaken at RT for 3 hours. 3-mercaptopropionic acid (1.7 µL, 20 µmol, 4000 eq.) was added as a metal scavenger for ruthenium and the mixture was shaken for further 15 minutes at RT. The protein sample was then purified from small molecules by dilution with 5M GdmCl (400 µL) and then concentration by Vivaspin® (10kDa MWCO) to 100 µL. This process was repeated for four times. The protein sample was treated with TCEP before analysis by LCMS. LCMS analysis of the protein sample revealed 80% conversion to the desired CM product (calculated = 14182, found = 14182, Section **2.8**).

2.11.12: General Procedure for Expression of NP276 and SsβG

An *E. Coli* B834(DE3) glycerol stock transformed with either Np276 G2P M21I Single M61 in pET28a, or SsβG C344S M43 9×Ile in pET28d was used to inoculate L.B. media (20 mL) containing kanamycin (50 µg/mL) and incubated for 16 hours at 37 °C. The starter cultures were used to inoculate Selenomet media (1 L), supplemented with Selenomet nutrient mix, L-methionine (20 mg/L) and kanamycin (50 µg/mL), the cells incubated at 37 °C, and the

optical density (OD_{600}) monitored until a value of 0.7 was obtained (~3-4 hours). The cells were then pelleted by centrifugation (8 minutes, JLA-9.1000 rotor, 8000 rpm). The pellets were washed three times with SelenoMet base media (3 × 300 mL) containing kanamycin (50 µg/mL) and then transferred to prewarmed SelenoMet media (1 L) containing kanamycin (50 µg/mL) and supplemented with Selenomet nutrient mix as well as the desired unnatural amino acid (DL-Ahc, 500 mg/L). The culture was incubated for 30 minutes at 37 °C, followed by 30 min at 30 °C. Expression was induced by addition of IPTG (1 mM) and was continued for 16 hours at 30 °C (220 rpm). The cells were harvested by centrifugation (8 minutes, 8000 rpm) to give pellets, which were collected in a 50 mL falcon tube and resuspended in binding buffer (20 mM Tris-HCl, 15 mM imidazole, 0.5 M NaCl, 1% glycerol, pH 7.8) (20 mL). The cell suspension was treated with DNase (10 µL), and proteinase inhibitor tablet (1/4 tablet), and incubated at 4 °C for 2 hours with occasional votexing. The cell was then lysed by sonication (Soniprep at 10 micron amplitude, 30 seconds on with 30 s rest, 5 cycles, on ice). The cell debris was removed from the lysis mixture by centrifugation (40 minutes, JA 25.5 rotor, 20000 rpm). The protein was purified by incubation of the cleared cell lysate with HisBind Ni-NTA resin (3 mL per L of culture, Novagen) overnight at 4 °C. The resin was packed into a glass econo-column and unbound proteins washed out with binding buffer (**A**) (20 mM Tris-HCl, 15 mM imidazole, 0.5 M NaCl, 1% glycerol, pH 7.8) (15-20 CV, ~50 mL). The bound protein was eluted from the resin using elution buffer (**B**) (20 mM Tris-HCl, 500 mM imidazole, 0.5 M NaCl, 5% glycerol, pH 7.8) using a step gradient (20% B (5 mL), 40% B (5 mL), 60% B (5 mL), 80% B (5 mL), 100% B (5 mL × 4)). All eluted proteins were collected in 5 mL fractions, which were analysed by SDS-PAGE (NuPAGE 4-12% Bis-Tris gel, Invitrogen, MES buffer, 200 V, 35 min).

Fractions containing pure Np276 protein were combined subjected to ammonium sulfate precipitation (85% saturation) for 16 h at 4 °C. The precipitated proteins were pelleted by centrifugation (45 minutes, JA 25.5 rotor, 20000 RPM 4°C), desalted using a PD10 column, with 50 mM Sodium phosphate buffer, pH 8.0 as a running buffer, then analyzed by LC-MS.
Ni-NTA column eluate fractions containing SSβG proteins were combined, and concentrated down to 5 mL using a 10K MWCO Vivaspin column. The contents were loaded onto a float-a-lyzer dialysis kit for overnight dialysis against 50mM pH 8.0 Na-Pi buffer, and analysed by LC-MS.Protein solutions were flash frozen, and stored at -20 ˚C.

Sequence Data for NP276 G2P M21I M6lX

PSSHHHHHHSSGLVPRGSHIDVGKLRQLYAAGERDFSIVDLRGAVLENINLSGAILHG

AXLDEANLQQANLSRADLSGATLNGADLRGANLSKADLSDAILDNAILEGAILDEAV

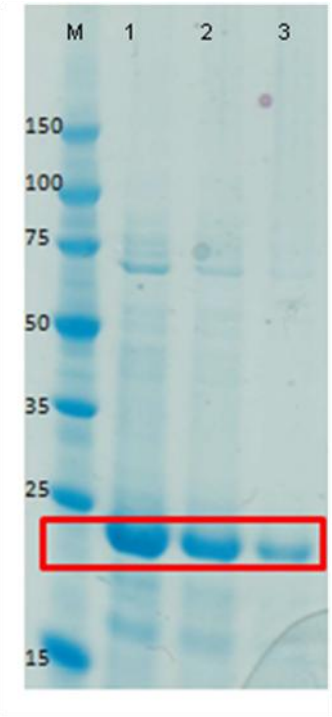
LNQANLKAANLEQAILSHANIREADLSEANLEAADLSGADLAIADLHQANLHQAALE

RANLTGANLEDANLEGTILEGGNNNLAT

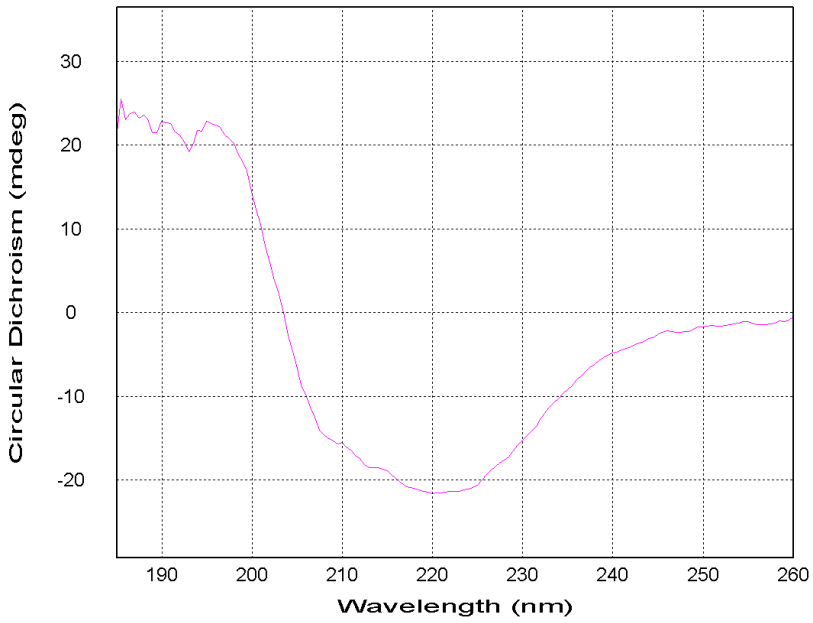
If X = Met, Calculated mass = 20878 (N-terminus Met cleaved)

If X = Ahc, Calculated mass = 20878 + 26 = 20904 (N-terminus Met cleaved)

SDS-PAGE

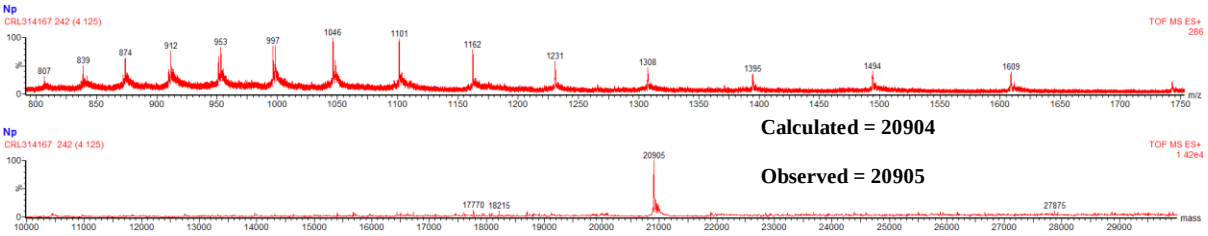


CD Spectrum:



LC-MS

NP276-Ahc61



MSMS Data for Tryptic Digested NP276 G2P M21I Ahc61

MPSSHHHHHHSSGLVPRGSHIDVGKLRQLYAAGERDFSIVDLRGAVLENINLSGAILH

GAMLDEANLQQANLSRADLSGATLNGADLRGANLSKADLSDAILDNAILEGAILDEA

VLNQANLKAANLEQAILSHANIREADLSEANLEAADLSGADLAIADLHQANLHQAA
LERANLTGANLEDANLEGTILEGGNNNLAT

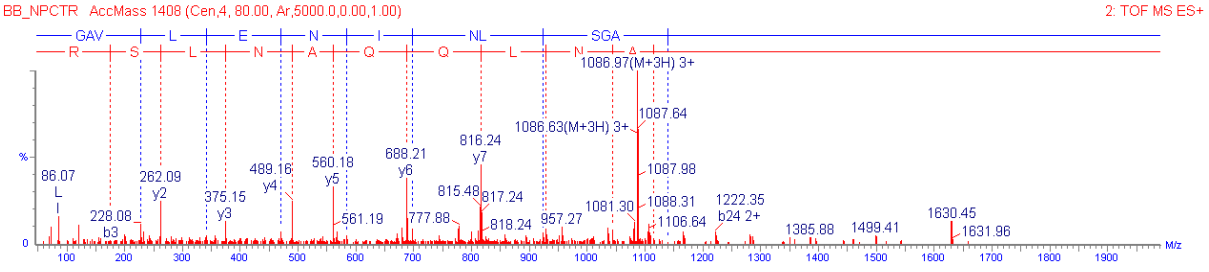
The peptide containing the unnatural amino acid Ahc, denoted as “M” in this case, is highlighted in the sequence above. The MSMS data showed good coverage of the fragmented peptide and received a score of 583 on BioLynx analysis.

Observed MW: 3257.6765 Precursor ion charge state: 3
M/z tolerance: 0.30 Intensity threshold: 140 (0.750%)
Modifications: Ahc (+)

a	30.03	101.07	200.14	313.22	442.27	556.31	669.39	783.44	896.52	983.55	1040.57	1111.61
	---	0.03	0.05	---	0.11	---	---	---	---	---	---	---
b	58.03	129.07	228.13	341.22	470.26	584.30	697.39	811.43	924.52	1011.55	1068.57	1139.61
	---	---	0.06	0.08	0.11	0.14	0.17	---	0.24	---	---	0.27
	Gly	Ala	Val	Leu	Glu	Asn	Ile	Asn	Leu	Ser	Gly	Ala
y	3258.69	3201.67	3130.64	3031.57	2918.48	2789.44	2675.40	2562.31	2448.27	2335.19	2248.16	2191.13
	---	---	---	---	---	---	---	---	---	---	---	---
z	3241.66	3184.64	3113.61	3014.54	2901.45	2772.41	2658.37	2545.28	2431.24	2318.16	2231.13	2174.10
	---	---	---	---	---	---	---	---	---	---	---	---

a	1224.70	1337.78	1474.84	1531.86	1602.90	1759.96	1873.04	1988.07	2117.11	2188.15	2302.19	2415.28
	---	---	---	---	---	---	---	---	---	---	---	---
b	1252.69	1365.77	1502.83	1559.85	1630.89	1787.95	1901.04	2016.06	2145.11	2216.14	2330.19	2443.27
	---	---	---	---	---	---	---	---	---	---	---	---
	Ile	Leu	His	Gly	Ala	Ahc	Leu	Asp	Glu	Ala	Asn	Leu
y	2120.10	2007.01	1893.93	1756.87	1699.85	1628.81	1471.75	1358.67	1243.64	1114.60	1043.56	929.52
	---	---	---	---	---	---	---	---	---	0.27	0.24	0.22
z	2103.07	1989.98	1876.90	1739.84	1682.82	1611.78	1454.72	1341.64	1226.61	1097.57	1026.53	912.49
	---	---	---	---	---	---	---	0.28	---	---	---	0.22

a	2543.33	2671.39	2742.43	2856.47	2969.56	3056.59	3212.69
	---	---	---	---	---	---	---
b	2571.33	2699.39	2770.42	2884.47	2997.55	3084.58	3240.68
	---	---	---	---	---	---	---
	Gln	Gln	Ala	Asn	Leu	Ser	Arg
y	816.43	688.37	560.32	489.28	375.24	262.15	175.12
	0.19	0.16	0.13	0.12	0.09	0.06	0.04
z	799.40	671.34	543.29	472.25	358.21	245.12	158.09
	0.18	0.16	0.14	---	---	0.10	0.04



2.11.13:Cross Metathesis on Np276 G2P M21I Ahc61 with allyl alcohol

A solution of Np276 G2P M21I Ahc61 (120 µg) in 50 mM sodium phosphate buffer (pH 8.0) was diluted to 700 µL with 5M guanidinium chloride (GdmCl) and 50 mM sodium phosphate

buffer (pH 8.0), then concentrated back down with a Vivaspin® (10kDa MWCO) to a final volume of 100 µL (c = 1.2 mg/mL, 5.74 nmol, 1 eq). This process was repeated twice. To this protein solution was added $MgCl_2.6H_2O$ (2.3 µL of a 1 mg/µL solution in water, 11.5 µmol, 2000 eq.) and Hoveyda-Grubbs 2nd generation catalyst in *t*BuOH solution (0.36 mg in 43 µL *t*BuOH, 0.57 µmol, 100 eq.). The reaction was mixed on a vortex each time after an addition. Reaction was shaken at RT for 1 minute before adding allyl alcohol (1.9 µL, 28.7 µmol, 5000 eq.). Reaction was mixed on a vortex immediately after addition and was shaken at RT for 4 hours. 3-mercaptopropionic acid (10 µL) was added as a metal scavenger for ruthenium and the mixture was shaken for further 30 minutes at RT. The protein sample was then purified from small molecules by dilution with 5M GdmCl-50mM sodium phosphate (pH 8.0, 400 µL) and then concentration by Vivaspin® (10kDa MWCO) to 100 µL. This process was repeated twice, followed by further removal of small molecules using a PD-Spintrap column. LCMS analysis of the protein sample revealed complete conversion to the desired CM product (calculated = 20934, found = 20935, Section **2.8**).

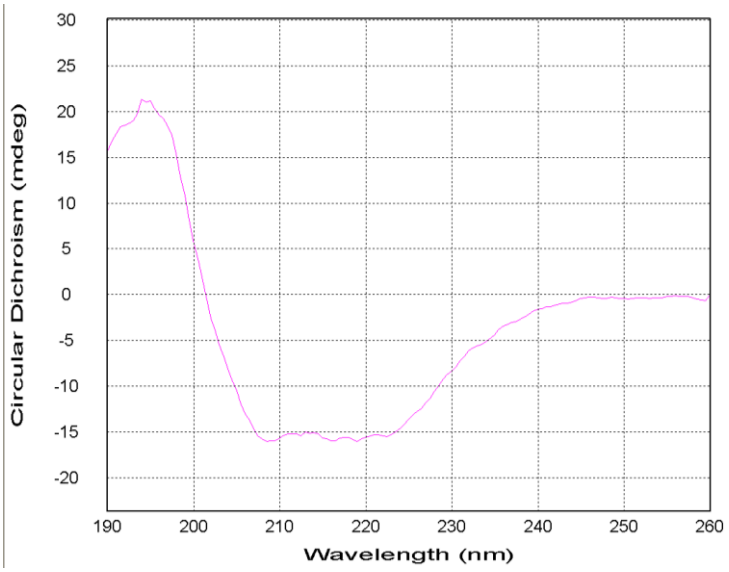
Sequence Data for SsβG M49 P152A C344S 9Ile:

GHHHHHHHSFPNSFRFGWSQAGFQSEIGTPGSEDPNTDWYKWVHDPEN**X**AAGLVSG
DLPENGPGYWGNYKTFHDNAQKIGLKIARLNVEWSRIFPNPLPRPQNFDESKQDVTE
VEINENELKRLDEYANKDALNHYREIFKDLKSRGLYFILNIYHWALPLWLHDPIRVRR
GDFTGPSGWLSTRTVYEFARFSAYIAWKFDDLVDEYSTINEPNVVGGLGYVGVKSGF
PPGYLSFELSRRAIYNIIQAHARAYDGIKSVSKKPVGIIYANSSFQPLTDKDIEAVEIAE
NDNRWWFFDAIIRGEITRGNEKIVRDDLKGRLDWIGVNYYTRTVVKRTEKGYVSLG
GYGHGSERNSVSLAGLPTSDFGWEFFPEGLYDVLTKYWNRYHLYIYVTENGIADDA
DYQRPYYLVSHVYQVHRAINSGADVRGYLHWSLADNYEWASGFSIRFGLLKVDYN
TKRLYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRH

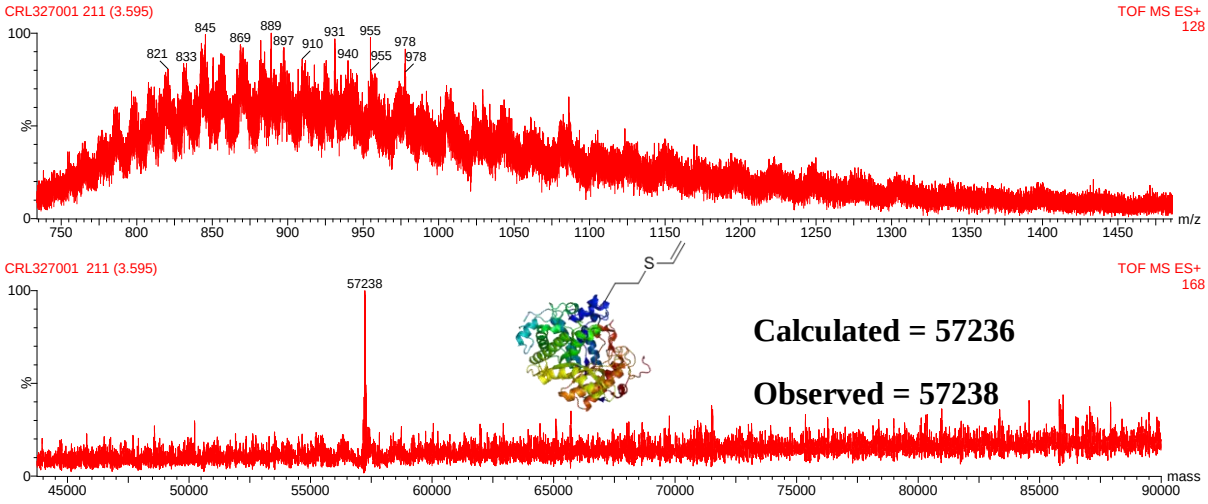
If X = Met, Calculated mass = 57210 (N-terminus Met cleaved)
If X = Ahc, Calculated mass = 57210 + 26 = 57236 (N-terminus Met cleaved)

SDS-PAGE:

Circular dichroism:



LCMS:



2.11.14: General Procedure for Expression of IgG-Fc

The Fc region of IgG was expressed in HEK 293T cells transiently transfected with pHLsec vector encoding for IgG-Fc, amplified by the Qiagen EndoFree Plasmid Giga kit. HEK293T cells were grown in two Corning HYPER*Flask* cell culture vessels in DMEM with 10% FCS and 1% penicillin-streptomycin. At 90% confluence, the cells were transfected with 2 mg of

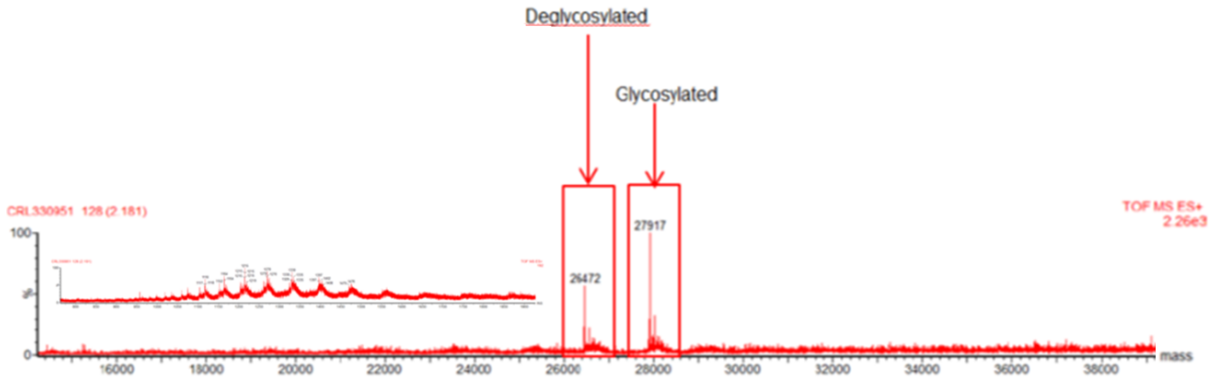
DNA per flask with linear PEI (MW 25000, Polysciences). 16 hours after transfection, the media was replaced with Met -, Cys - and Gln - free DMEM supplemented with 500 mg/L DL-Ahc, 62.6 mg/L L-cystine and 584 mg/L L-glutamine. The flasks were incubated for 4 days at 37 °C with 5% CO_2, following which the media was harvested and loose cell debris removed by centrifugation. The protein was purified by passing the supernatant through a 100 mL Talon-resin column, washing out unbound proteins with 300 mL of Buffer A (10mM Tris, 150mM NaCl, pH 8.0), and eluting the bound protein from the column with 250 mL buffer B (10mM Tris, 150mM NaCl, 250 mM imidazole, pH 8.0). The resulting protein solution was concentrated to 25 mL and dialysed into PBS. 1 mg of the protein was deglycosylated by treatment with PNGase-F for 3h at 37 °C.

Sequence data for IgG-Fc:

ETGHTCPPCPAPELLGGPSVFLFPPKPKDTLXISRTPEVTCVVVDVSHEDPEVKFNWY
VDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIE
KTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNY
KATPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVX'HEALHNHYTQKSLSLSPGG
TKHHHHHH

If X = X' = Met, Calculated mass = 26443 (no glycan)
If X = Ahc and X' = Met, Calc. mass = 26443 + 26 = 26469
If X = X' = Ahc, Calculated mass = 26469 + 26 = 26495

SDS-PAGE:



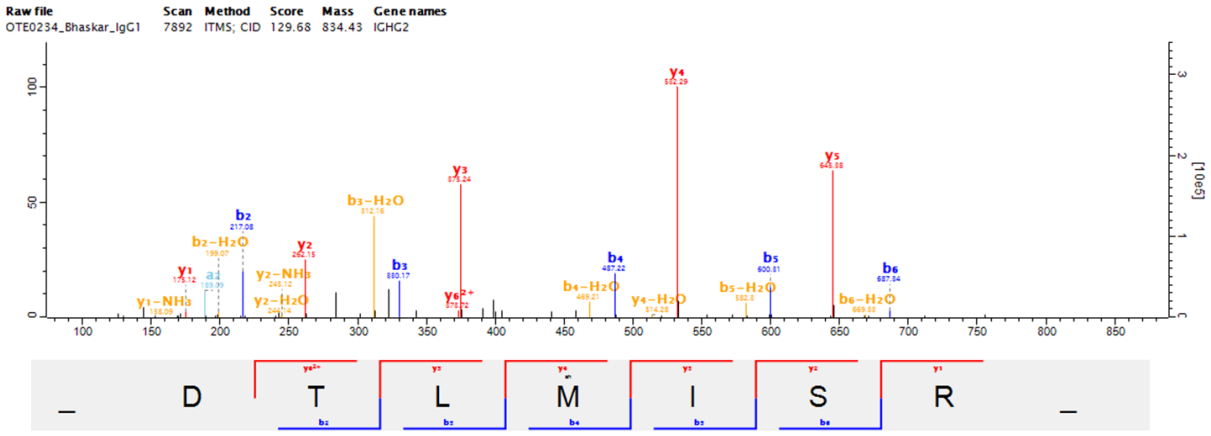
Masses observed in deglycosylated peak: 26444 (Met incorporated, 19%), 26472 (Single Ahc incorporation, 81%), no dual Ahc-incorporation

MSMS Data for IgG-Fc:

ETGHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWY
VDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIE
KTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNY
KATPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVM'HEALHNHYTQKSLSLSPGG
TKHHHHHH

The peptides containing the unnatural amino acid Ahc, denoted as M and M' in this case, are highlighted in the sequence above. The MSMS data showed good coverage of both fragmented peptides, and confirmed the incorporation of Ahc at position 32, but only of Met at position 208.

MSMS of Ahc-labelled peptide H_2NDTL-Ahc-IMSR-COOH:



2.11.15:Cross Metathesis on IgG-Fc Ahc32 Met208:

A solution of IgG-Fc Ahc32 Met208 (140 µg) was dialysed into 5 M Guanidine hydrochloride – 50 mM sodium phosphate (105 µL, pH 8.0). To this protein solution was added $MgCl_2.6H_2O$ (4 µL of a 1 mg/µL solution in water, 19.7 µmol, 4000 eq.) and Hoveyda-Grubbs 2nd generation catalyst in *t*BuOH solution (0.3 mg in 43 µL *t*BuOH, 0.47 µmol, 100 eq.). The reaction was mixed on a vortex each time after an addition. Reaction was shaken at RT for 1 minute before adding allyl biotin amide (7.5 mg, 5000 eq.). Reaction was mixed on a vortex immediately after addition and was shaken at RT for 3 hours. 3-mercaptopropionic acid (10 µL) was added as a metal scavenger for ruthenium and the mixture was shaken for further 30 minutes at RT. The protein sample was then purified from small molecules by dilution with 5M GdmCl-50mM sodium phosphate (pH 8.0, 400 µL) and then concentration by Vivaspin® (10kDa MWCO) to 100 µL. This process was repeated twice, followed by further removal of small molecules using a PD-Spintrap column. The reaction contents were run on an anti-biotin Western blot to confirm functionalization.

2.11.16: General Procedure for Expression of SarZΔ(7-142)

An *E. Coli* B834(DE3) glycerol stock containing a pET28a (Novagen) vector encoding for a truncated form of SarZ, SarZΔ(7-142) was used to inoculate L.B. media (20 mL) containing kanamycin (50 µg/mL) and incubated for 16 hours at 37 °C. The starter cultures were used to inoculate Selenomet media (1 L), supplemented with Selenomet nutrient mix, L-methionine (20 mg/L) and kanamycin (50 µg/mL), the cells incubated at 37 °C, and the optical density (OD_{600}) monitored until a value of 0.7 was obtained (~3-4 hours). The cells were then pelleted by centrifugation (8 minutes, JLA-9.1000 rotor, 8000 rpm). The pellets were washed three times with SelenoMet base media (3 × 300 mL) containing kanamycin (50 µg/mL) and then transferred to prewarmed SelenoMet media (1 L) containing kanamycin (50 µg/mL) and supplemented with Selenomet nutrient mix as well as the desired unnatural amino acid (DL-Ahc, 500 mg/L). The culture was incubated for 30 minutes at 37 °C, followed by 30 min at 30 °C. Expression was induced by addition of IPTG (1 mM) and was continued for 16 hours

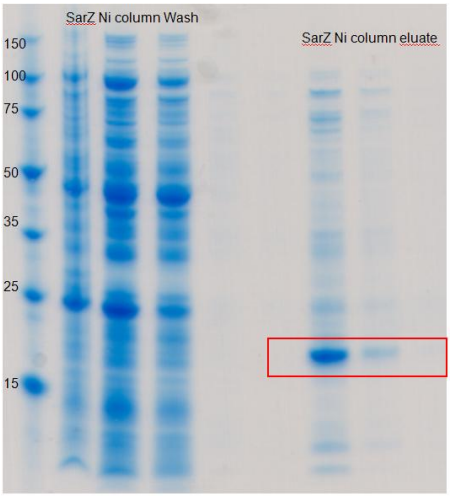
at 30 °C (220 rpm). The cells were harvested by centrifugation (8 minutes, 8000 rpm) to give pellets, which were collected in a 50 mL falcon tube and resuspended in binding buffer (20 mM Tris (pH 7.5), 300 mM NaCl, 2 mM DTT, 20 mM imidazole and 10% glycerol) (20 mL). The cell suspension was treated with DNase (10 μL), and proteinase inhibitor tablet (1/4 tablet), and incubated at 4 °C for 2 hours with occasional votexing. The cell was then lysed by sonication (Soniprep at 10 micron amplitude, 30 seconds on with 30 s rest, 5 cycles, on ice). The cell debris was removed from the lysis mixture by centrifugation (30 minutes, JA 25.5 rotor, 20000 rpm). The protein was purified by incubation of the cleared cell lysate with HisBind Ni-NTA resin (3 mL per L of culture, Novagen) overnight at 4 °C. The resin was packed into a glass econo-column and unbound proteins washed out with binding buffer (**A**) (20 mM Tris (pH 7.5), 300 mM NaCl, 2 mM DTT, 20 mM imidazole and 10% glycerol) (15-20 CV, ~50 mL). The bound protein was eluted from the resin using elution buffer (**B**) (20 mM Tris (pH 7.5), 300 mM NaCl, 2 mM DTT, 500 mM imidazole and 10% glycerol) using a step gradient (20% B (5 mL), 40% B (5 mL), 60% B (5 mL), 80% B (5 mL), 100% B (5 mL × 4)). All eluted proteins were collected in 5 mL fractions, which were analysed by SDS-PAGE (NuPAGE 4-12% Bis-Tris gel, Invitrogen, MES buffer, 200 V, 35 min).

Sequence Data for SarZΔ(7-142):

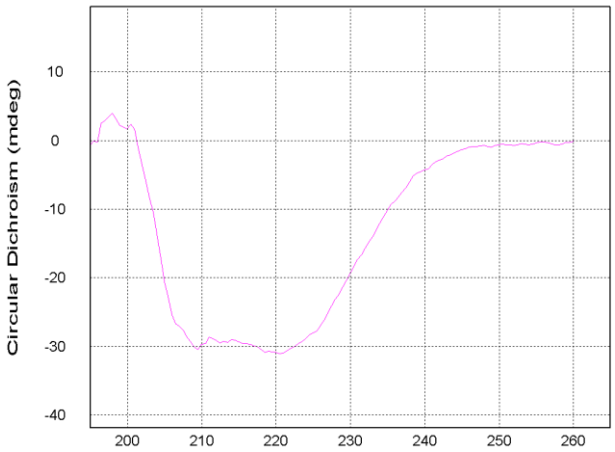
MGSSHHHHHHSSGLVPRGSHXYVENSYLSKQLCFLFYVSSKEIIKKYTNYLKEYDLT
YTGYIVLX'AIENDEKLNIKKLGERVFLDSGTLTPLLKKLEKKDYVVRTREEKDERNL
QISLTEQGKAIKSPLAEISVKVFNEFNISEREASDIINNLRNFVSKNFDYSDRK

If X = X' = Met, Calculated mass = 19495 (N-terminus Met cleaved)
If X = Met and X' = Ahc, Calc. mass = 19495 + 26 = 19521
If X = X' = Ahc, Calculated mass = 19521 + 26 = 19547

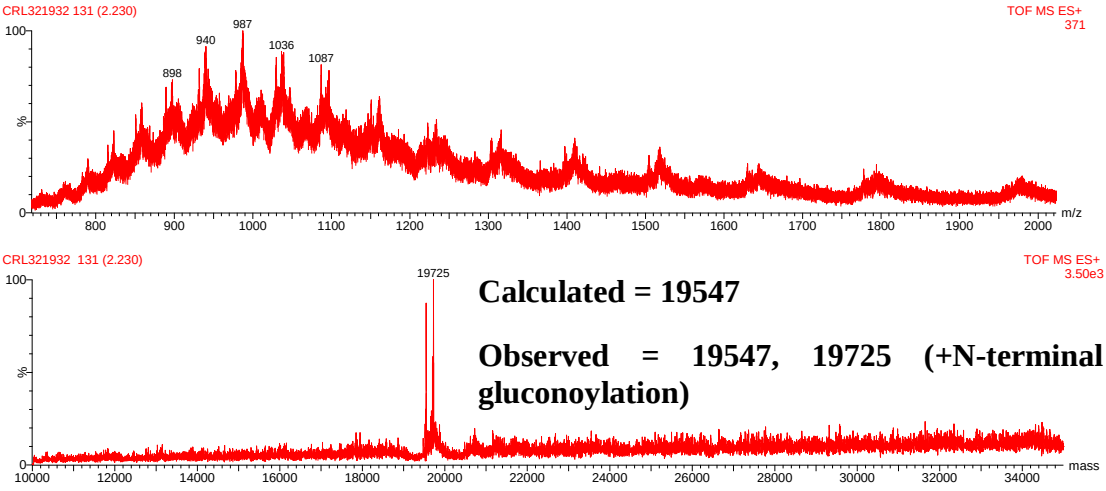
SDS-PAGE:



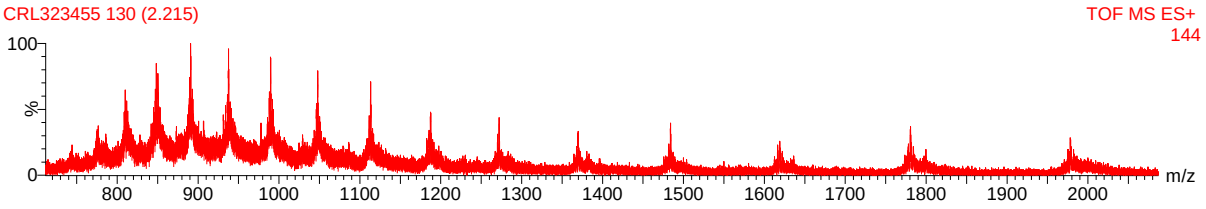
Circular dichroism:

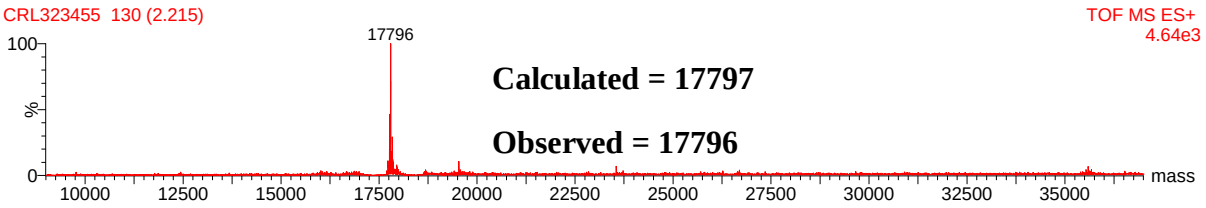


LCMS:



The protein was dialysed against thrombin cleavage buffer (20 mM Tris, 100 mM NaCl, pH 7.6), and the peptide fragment GSSHHHHHHSSGLVPR- was cleaved by agarose resin-bound thrombin from the Thrombin CleanCleave™ Kit from Sigma Aldrich. The resin was removed by centrifugation and the cleaved peptide fragment removed by dialysis of the cleaved protein against fresh cleavage buffer.





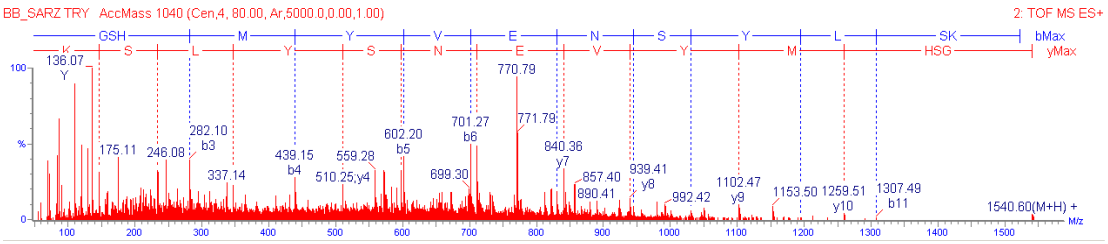
MSMS Data for Tryptic Digested SarZΔ(7-142):

MGSSHHHHHHSSGLVPRGSHMYVENSYLSKQLCFLFYVSSKEIIKKYTNYLKEYDLT
YTGYIVLMAIENDEKLNIKKLGERVFLDSGTLTPLLKKLEKKDYVVRTREEKDERNL
QISLTEQGKAIKSPLAEISVKVFNEFNISEREASDIINNLRNFVSKNFDYSDRK

The peptides containing the unnatural amino acid Ahc, denoted as "M" in this case, are highlighted in the sequence above. The MSMS data showed good coverage of both fragmented peptides, which received scores of 355 and 102 on BioLynx analysis.

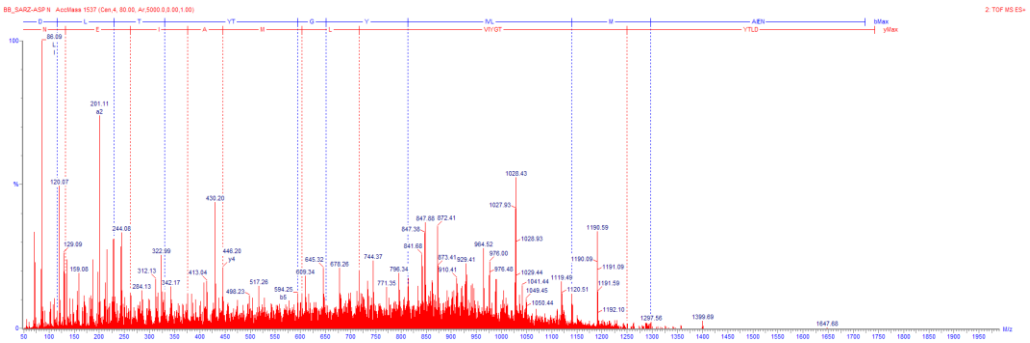
Observed MW: 1539.7021 Precursor ion charge state: 1
M/z tolerance: 0.30 Intensity threshold: 310 (0.750%)
Modifications: Ahc (+)

a	30.03	117.07	254.13	411.19	574.25	673.32	802.36	916.40	1003.44	1166.50	1279.58	1366.61	1494.71
	---	---	---	0.03	---	---	---	0.06	---	---	---	---	---
b	58.03	145.06	282.12	439.18	602.24	701.31	830.36	944.40	1031.43	1194.49	1307.58	1394.61	1522.70
	---	---	0.02	0.03	0.05	0.04	0.05	0.04	0.06	0.07	0.09	---	---
	Gly	Ser	His	Ahc	Tyr	Val	Glu	Asn	Ser	Tyr	Leu	Ser	Lys
y	1540.71	1483.69	1396.66	1259.60	1102.54	939.48	840.41	711.37	597.32	510.29	347.23	234.15	147.11
	---	---	---	0.09	0.07	0.07	0.05	0.05	0.05	0.04	0.02	0.02	0.01
z	1523.68	1466.66	1379.63	1242.57	1085.51	922.45	823.38	694.34	580.29	493.26	330.20	217.12	130.08
	---	---	---	---	---	---	---	0.04	---	0.06	---	---	---



Observed MW: 1740.8644 Precursor ion charge state: 2
M/z tolerance: 0.30 Intensity threshold: 20 (0.750%)
Modifications: S-allylhomocysteine (+)

a	88.04	201.12	302.17	465.23	566.28	623.30	786.37	899.45	998.52	1111.60	1268.66	1339.70	1452.78	1581.82	1695.87
	---	0.01	---	---	---	0.03	0.01	---	---	---	---	---	---	---	---
b	116.03	229.12	330.17	493.23	594.28	651.30	814.36	927.45	1026.51	1139.60	1296.66	1367.69	1480.78	1609.82	1723.86
	-0.03	0.02	0.02	---	0.02	0.02	0.05	---	---	0.09	0.11	---	---	---	---
	Asp	Leu	Thr	Tyr	Thr	Gly	Tyr	Ile	Val	Leu	Ahc	Ala	Ile	Glu	Asn
y	1741.87	1626.85	1513.76	1412.71	1249.65	1148.60	1091.58	928.52	815.43	716.36	603.28	446.22	375.19	262.10	133.06
	---	---	---	---	0.12	---	---	---	---	0.05	0.03	0.02	0.02	0.01	0.01
z	1724.84	1609.82	1496.73	1395.68	1232.62	1131.57	1074.55	911.49	798.40	699.33	586.25	429.19	358.16	245.07	116.03
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	-0.03



2.12: References

(1) van Hest, J. C. M.; Tirrell, D. A. *FEBS Lett.* **1998**, *428*, 68.
(2) Song, W.; Wang, Y.; Yu, Z.; Vera, C. I. R.; Qu, J.; Lin, Q. *ACS Chem. Biol.*, *5*, 875.
(3) Ngo, J. T.; Tirrell, D. A. *Acc. Chem. Res.* **2011**, *44*, 677.
(4) van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 1282.
(5) E. Gibson, S.; C. Gibson, V.; P. Keen, S. *Chem. Commun.* **1997**, 1107.
(6) Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. J.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130*, 9642.
(7) Lin, Y. A.; Chalker, J. M.; Davis, B. G. *J. Am. Chem. Soc.* **2010**, *132*, 16805.
(8) Lin, Y. A.; Boutureira, O.; Lercher, L.; Bhushan, B.; Paton, R. S.; Davis, B. G. *J. Am. Chem. Soc.* **2013**, *135*, 12156.
(9) Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernandez-Gonzalez, M.; Bernardes, G. J. L.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G. *Chemical Science* **2011**, *2*, 1666.
(10) Davis, B. G.; Lloyd, R. C.; Jones, J. B. *Bioorg Med Chem* **2000**, *8*, 1527.
(11) Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. a. J. L.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130*, 9642.
(12) Ai, H. W.; Shen, W.; Brustad, E.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2010**, *49*, 935.
(13) Budisa, N. *Angew Chem Int Ed Engl* **2004**, *43*, 6426.
(14) Fersht, A. R.; Dingwall, C. *Biochemistry* **1979**, *18*, 1250.
(15) Jakubowski, H. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 4504.
(16) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. *Nat. Biotechnol.* **2001**, *19*, 751.
(17) Hartman, M. C.; Josephson, K.; Lin, C. W.; Szostak, J. W. *PLoS One* **2007**, *2*, e972.
(18) Datta, D.; Vaidehi, N.; Zhang, D.; Goddard, W. A. *Protein Sci.* **2004**, *13*, 2693.
(19) Zamecnik, P. *Anal Biochem* **1983**, *134*, 1.
(20) Cestari, I.; Stuart, K. *Journal of Biomolecular Screening* **2013**, *18*, 490.
(21) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 19.
(22) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455.
(23) Serre, L.; Verdon, G.; Choinowski, T.; Hervouet, N.; Risler, J. L.; Zelwer, C. *J. Mol. Biol.* **2001**, *306*, 863.
(24) Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B. D.; Sun, Z.-W.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C. P.; Allis, C. D.; Jenuwein, T. *Nature* **2000**, *406*, 593.
(25) Kouzarides, T. *Cell* **2007**, *128*, 693.
(26) Nakagomi, K.; Kohwi, Y.; Dickinson, L. A.; Kohwi-Shigematsu, T. *Molecular and Cellular Biology* **1994**, *14*, 1852.
(27) Geoghegan, K. F.; Dixon, H. B. F.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.; McColl, A. S.; Kamath, A. V.; Stroh, J. G. *Analytical Biochemistry* **1999**, *267*, 169.
(28) van Heerden, A.; Browning, K. S. *Journal of Biological Chemistry* **1994**, *269*, 17454.
(29) Khan, M. A.; Miyoshi, H.; Gallie, D. R.; Goss, D. J. *Journal of Biological Chemistry* **2008**, *283*, 1340.
(30) Deutscher, J. *Current Opinion in Microbiology* **2008**, *11*, 87.
(31) Vetting, M. W.; Hegde, S. S.; Hazleton, K. Z.; Blanchard, J. S. *Protein Sci.* **2007**, *16*, 755.
(32) Aguilar, C. F.; Sanderson, I.; Moracci, M.; Ciaramella, M.; Nucci, R.; Rossi, M.; Pearl, L. H. *J. Mol. Biol.* **1997**, *271*, 789.

(33) Kaito, C.; Morishita, D.; Matsumoto, Y.; Kurokawa, K.; Sekimizu, K. *Molecular Microbiology* **2006**, *62*, 1601.
(34) Kozlovska, T. M.; Cielēns, I.; Dreilinņa, D.; Dišlers, A.; Baumanis, V.; Ose, V.; Pumpēns, P. *Gene* **1993**, *137*, 133.
(35) Hancock, S. M.; Corbett, K.; Fordham-Skelton, A. P.; Gatehouse, J. A.; Davis, B. G. *Chembiochem* **2005**, *6*, 866.
(36) Chivers, C. E.; Crozat, E.; Chu, C.; Moy, V. T.; Sherratt, D. J.; Howarth, M. *Nat Methods* **2010**, *7*, 391.
(37) Johnson, L. N.; Phillips, D. C. *Nature* **1965**, *206*, 761.
(38) Lu, W.; Tang, X.; Huo, Y.; Xu, R.; Qi, S.; Huang, J.; Zheng, C.; Wu, C. A. *Gene* **2012**, *503*, 65.
(39) Truong-Bolduc, Q. C.; Dunman, P. M.; Strahilevitz, J.; Projan, S. J.; Hooper, D. C. *J Bacteriol* **2005**, *187*, 2395.
(40) Sun, F.; Ding, Y.; Ji, Q.; Liang, Z.; Deng, X.; Wong, C. C.; Yi, C.; Zhang, L.; Xie, S.; Alvarez, S.; Hicks, L. M.; Luo, C.; Jiang, H.; Lan, L.; He, C. *Proc Natl Acad Sci U S A* **2012**, *109*, 15461.
(41) Cheung, A. L.; Nishina, K. A.; Trotonda, M. P.; Tamber, S. *Int J Biochem Cell Biol* **2008**, *40*, 355.
(42) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130*, 5052.
(43) Simpson, R. J. *Cold Spring Harbor protocols* **2010**, *2010*, pdb.prot5412.
(44) Calcutta, A.; Jessen, C. M.; Behrens, M. A.; Oliveira, C. L.; Renart, M. L.; González-Ros, J. M.; Otzen, D. E.; Pedersen, J. S.; Malmendal, A.; Nielsen, N. C. *Biochim Biophys Acta* **2012**, *1818*, 2290.
(45) Spicer, C. D.; Davis, B. G. *Chem Commun* **2011**, *47*, 1698.
(46) Poor, C. B.; Chen, P. R.; Duguid, E.; Rice, P. A.; He, C. *J Biol Chem* **2009**, *284*, 23517.
(47) Whitmore, L.; Wallace, B. A. *Nucleic Acids Res* **2004**, *32*, W668.
(48) Whitmore, L.; Wallace, B. A. *Biopolymers* **2008**, *89*, 392.
(49) Chalker, J. M.; Lin, Y. A.; Boutureira, O.; Davis, B. G. *Chem Commun* **2009**, 3714.
(50) Esmaeili, M. A.; Yazdanparast, R. *Protein J* **2008**, *27*, 334.
(51) Zhang, L.; Zhang, Q.; Wang, C. *Biomed Chromatogr* **2013**, *27*, 365.
(52) Hong, S. H.; Grubbs, R. H. *J. Am. Chem. Soc.* **2006**, *128*, 3508.
(53) Skowerski, K.; Szczepaniak, G.; Wierzbicka, C.; Gulajski, L.; Bieniek, M.; Grela, K. *Catal. Sci. Technol.* **2012**, *2*, 2424.
(54) Llewellyn, D. B.; Wahhab, A. **2009**, *50*, 3939.
(55) Lin, Y. A.; Chalker, J. M.; Davis, B. G. *J. Am. Chem. Soc.*, *132*, 16805.
(56) DelMar, E. G.; Largman, C.; Brodrick, J. W.; Geokas, M. C. *Analytical Biochemistry* **1979**, *99*, 316.
(57) Davis, B. G.; Shang, X.; DeSantis, G.; Bott, R. R.; Jones, J. B. *Bioorganic & Medicinal Chemistry* **1999**, *7*, 2293.

Chapter 3

Synthesis and Rapid Turnover of Peptides in Nucleus and Cytoplasm of Mammalian Cells

This project was carried out in collaboration with Sabyasachi Baboo and Prof. Peter Cook, Dunn School of Pathology, and Haibo Jiang and Prof. Chris Grovenor, Department of Materials, University of Oxford, and sections have been published.[1] All Aha- and heavy AA cell labelling experiments, as well as CuAAC reactions were carried out by both BB and SB. Puromycin labelling, RNA-FISH experiments, and microscopy were carried out by SB. SIMS images collected by HJ. Unless otherwise stated, all other experiments were performed by BB.

3.1: Introduction

The primary defining characteristic of eukaryotic cells in the presence of a nuclear membrane. It is widely assumed that RNA transcription occurs in the nucleus, and subsequent protein synthesis (translation) is restricted to the cytoplasm. The presence of introns in eukaryotic mRNA seems to indicate why this should be so, as prior to cytoplasmic export, mRNA transcripts are spliced to remove intronic sequences. Instead, if RNA transcription and protein translation were coupled and occurred simultaneously in the nucleus itself, then intronic sequences with their several stop codons would generate peptides as well. This would seemingly lead to the expenditure of cellular resources in expressing either unnecessary truncated peptide fragments at best, and potentially toxic peptides at worst. However, there is some evidence that is consistent with a model where a fraction of protein translation occurs within nuclei of eukaryotic cells.

First, nuclei contain the necessary components for protein translation, such as ribosomes,[2] initiation factors,[3] aminoacylated tRNAs,[4,5] and modest levels of translation factors that might support a limited amount of protein synthesis if concentrated at specific regions of translation within nuclei.[6] Second, isolated nuclei can aminoacylate tRNAs and incorporate radiolabeled amino acids into proteins,[7] although contaminating cytoplasmic machinery on the outer nuclear membrane could be responsible for such synthesis.[8]

The phenomenon of nonsense-mediated decay (NMD) provides additional evidence to support nuclear translation. NMD is a quality control mechanism that checks newly made mRNA transcripts.[9] A fraction of incorrectly spliced transcripts contain premature termination codons (PTCs) encoding truncated proteins. Such PTC-bearing transcripts are degraded by NMD.[10] A significant fraction of NMD occurs in the nucleus.[11-14] Only active cytoplasmic ribosomes are known to be able to recognise termination codons, but these could not trigger degradation of a mis-spliced transcript while it is still in the nucleus. Nuclear NMD could be explained if active nuclear ribosomes recognised PTCs, thereby allowing nuclear translation to occur.

Using pulses as short as 5 s, in this work we show that three independent labelling and visualisation techniques involving UAAs, puromycin, and amino acids tagged with 'heavy' ^{15}N and ^{13}C demonstrate incorporation into both nucleus and cytoplasm in a process sensitive to translational inhibitors, and rapidly turn over.

3.2: Fluorescent Non-Canonical Amino Acid Tagging (FUNCAT) applied to nuclear translation

FUNCAT is a technique employed to visualise newly synthesised proteins in cells.[15] Cells are pulse-labelled with unnatural amino acids such as the azide-containing Met-surrogate L-azidohomoalanine (Aha)[16] or the alkyne-bearing Met-surrogate L-homopropargylglycine (Hpg),[17] in Met-depleted media. Membranes of labelled cells may then be permeabilised with a detergent, and the Aha- or Hpg- incorporated proteins reacted with a corresponding alkyne- or azide-linked fluorophore via the [3+2] copper-catalysed azide-alkyne cycloaddition (CuAAC) reaction (see Section 3.12.2 for full protocol).

This protocol was applied in order to visualise the distribution of newly synthesised proteins in HeLa cells (Fig. 3.1). After CuAAC with alkyne-linked Alexa 555, a clear signal could be observed at 555 nm from both nuclear and cytoplasmic regions of HeLa cells labelled with Aha for 2 min. This signal was seen to decrease upon pre-treatment of the cells with 100 μg/mL of anisomycin, a translation inhibitor.[18] Although nuclear regions seemed brightest, quantitative integration shows more fluorescently labelled species in the larger area of the cytoplasm.

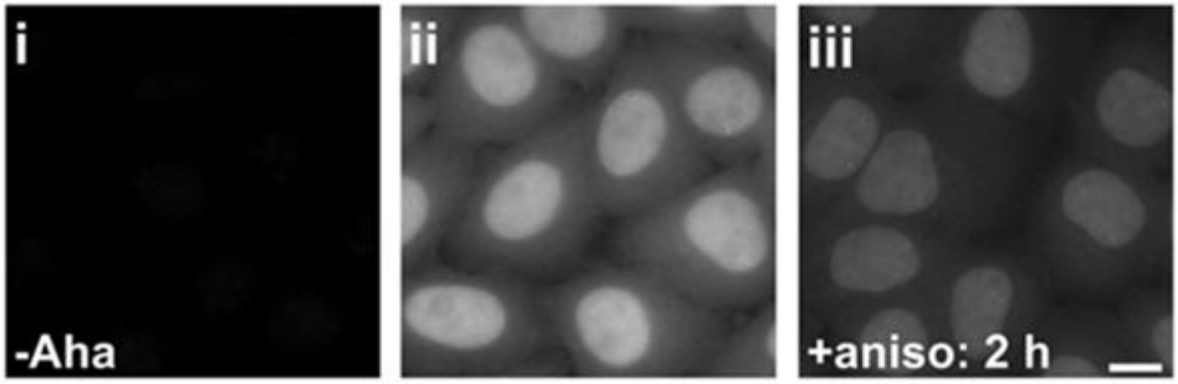


Fig. 3.1: Fluorescence at 555 nm from HeLa cells labelled with Aha for 2 min after CuAAC with Alkyne-linked Alexa-555. i) Met incorporation only (negative control), ii) Aha-labelled cells, iii) Aha-labelled cells pre-incubated with the translation inhibitor anisomycin. Bar: 10 µm

Eukaryotic ribosomes synthesise proteins at a rate of 3 to 5 amino acids per second.[19] In order to distinguish between signal from nuclear incorporation of Aha into nascent nuclear proteins from cytoplasmic incorporation into mature proteins followed by nuclear import, a short 5-second Aha incorporation pulse was used. A cytoplasmic as well as nuclear signal was observed for this short pulse time as well, and was seen to be sensitive to pre-incubation with anisomycin for 15 min. (Fig. 3.2).

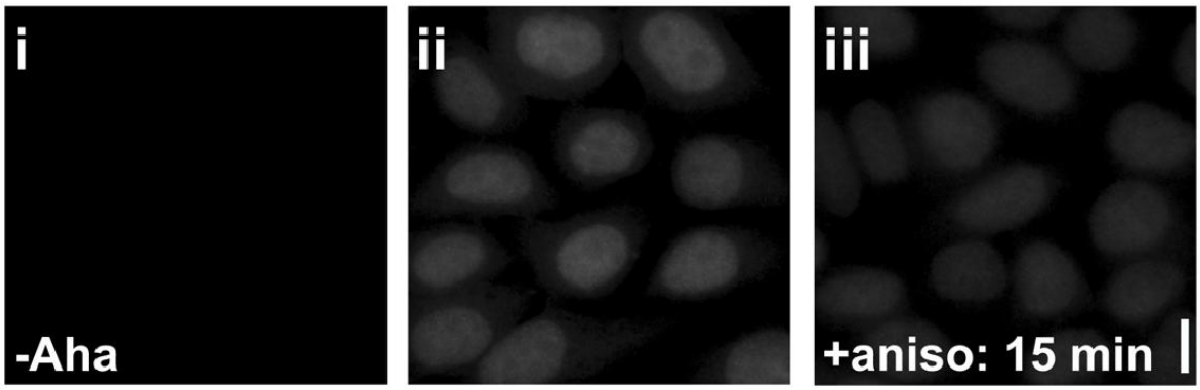


Fig 3.2: Fluorescence at 555 nm from HeLa cells labelled with Aha for 5 seconds after CuAAC with Alkyne-linked Alexa-555: i) Met incorporation only (negative control), ii) Aha-labelled cells, iii) Aha-labelled cells pre-incubated with the translation inhibitor anisomycin. Bar: 10 µm

Assuming an expression rate of 5 amino acids per second, a 5 second pulse allows for the polymerisation of 25 peptide residues. Assuming an average mature protein size of 400 peptide residues, only 6.25% (i.e., 25/400) of all proteins will be fully synthesised during this pulse. Further factoring in the abundance of methionine (and therefore of Aha) in proteins of ~ 2%, a further 60% of these fully synthesised proteins will not contain a single site for Aha

incorporation and therefore will not contribute to the observed fluorescent signal. Therefore, only 2.5% of the observed signal can be accounted for by mature labelled proteins, and the majority of the nuclear signal must arise from peptides synthesised there.

In order to distinguish between Aha incorporated into proteins and Aha bound to methionyl tRNA in an aminoacyl-tRNA complex, HeLa cells were also treated with 5U RiboShredder (Epicentre), an RNase cocktail, in PBS after fixation and immediately prior to permeabilisation and CuAAC. If a major fraction of nuclear Aha was in fact bound to tRNA, then this would be washed out during the permeabilisation and subsequent washing steps prior to CuAAC, leading to a reduction in fluorescence signal. Instead, a slight increase in the relative fluorescent intensity in the nucleus and cytoplasm were seen, to 189 ± 26% and 139 ± 22% respectively relative to RNase-untreated cells. A possible explanation may be that the RNA formerly blocked the access of the CuAAC reagents to some buried Aha-incorporated peptides, but could not do so after degradation by RiboShredder.

Importantly, this fluorescent signal was also seen to be sensitive to translation inhibitors as well as proteasomal inhibitors. As shown above, pre-treatment of the cells with anisomycin,[18] led to a decrease in the fluorescent signal. Additionally, pre-treatment with 100 µg/ml of the proteasome inhibitor MG132,[20] led to an increase in signal from a 2 minute pulse (Fig. 3.3).

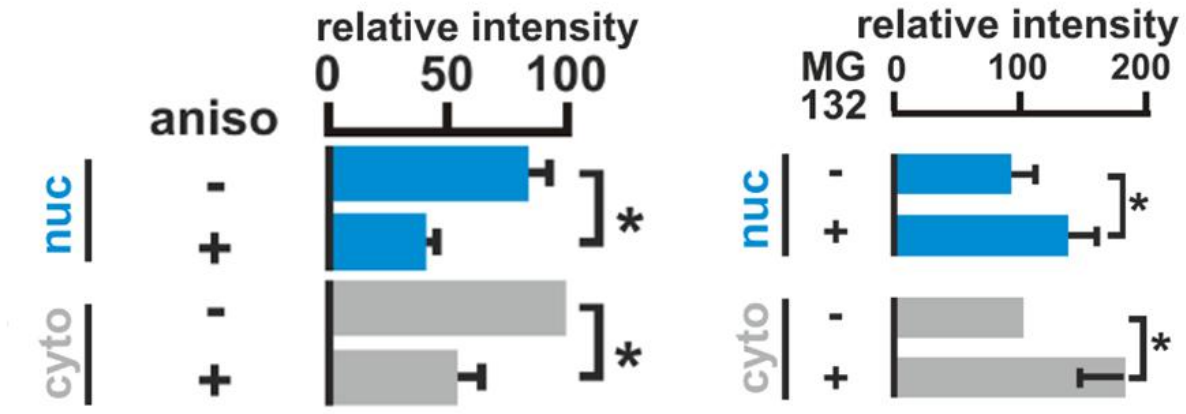


Fig. 3.3: Quantification of fluorescent intensities from cytoplasmic and nuclear region of HeLa cells in the absence and presence of either anisomycin (left) or MG132 (right). * P < 0.0001 (Student's two-tailed t test, n = 20 cells) Intensities normalised relative to the intensity in the cytoplasm in cells untreated with anisomycin or MG132.

These results suggest that Aha is in fact incorporated into proteins during the 2 minute pulse, as inhibiting protein expression by anisomycin leads to a decrease in fluorescent signal, while inhibiting protein degradation by MG132 leads to an increase in fluorescent signal.

3.3: Pulse-chase FUNCAT experiments

In order to confirm that this incorporation was a metabolic process, a pulse-chase experiment with Met was carried out. After a 2-min Aha pulse, cells were transferred to Met-supplemented DMEM for 5 min. at either 37 ºC or 4 ºC prior to fixation, permeabilisation, and CuAAC with alkynyl Alexa-555. A dramatic reduction of the signal was observed in cells 'chased' at 37 ºC compared to cells which were only exposed to the pulse (Fig. 3.4), suggesting that the Aha incorporation was metabolic, and was being replaced by the natural substrate Met during the chase. Importantly, no signal reduction was seen in cells 'chased' at 4 ºC, again suggesting metabolic incorporation. Kinetics of the pulse-chase experiment revealed that the Aha-labelled species in both the nucleus and cytoplasm had a half-life of < 1 min.

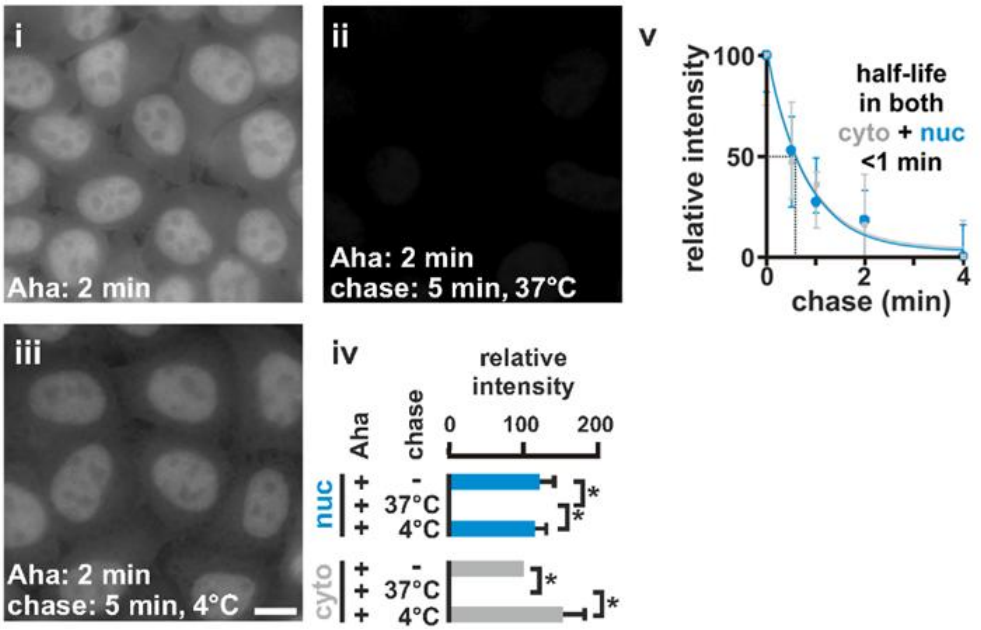


Fig 3.4: Fluorescent signal from HeLa cells during pulse-chase experiments: i) After 2 min. pulse, ii) 2 min. pulse followed by 5 min. chase at 37 ºC, iii) 2 min. pulse followed by 5 min. chase at 4 ºC iv) Total fluorescent intensity in the nucleus and cytoplasm before and after chase with Met at 37 ºC and 4 ºC after 2 min, relative to cytoplasmic intensity before the chase, v) Kinetics of pulse-chase experiment at 37 ºC Bar: 10 µm

Although these results contrast with reported half-lives of human proteins of ~20 h,[21,22] previous studies have utilised heavy-isotope pulses of several days prior to measurement in order to completely label the proteome. These observations may be reconciled if several short peptides were being synthesised and rapidly turned over without forming functional proteins or contributing to the mature proteome. Such peptides, if they exist, would be missed using longer pulses used in previous protein lifetime studies. Possible reasons for the existence of such an apparently wasteful mechanism will be discussed in Section **3.11**.

In order to determine whether these observations might be general to mammalian cells, pulse-chase experiments were repeated with primary human umbilical vein endothelial cells (HUVEC cells, Fig. 3.5). Results indicated that this phenomenon was not peculiar to HeLa cells. Increasing the pulse duration, however, allowed for sufficient incorporation to allow the fluorescent signal to survive a 5-min chase, further supporting the hypothesis that previously-employed longer pulses led to reporting of longer protein half lives

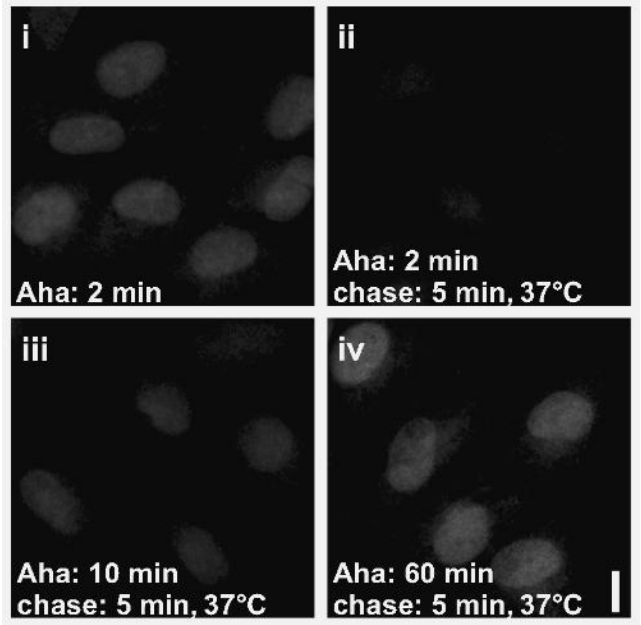


Fig. 3.5: Fluorescent signal from HUVEC cells during pulse-chase experiments: i) After 2 min. pulse, ii) 2 min. pulse followed by 5 min. chase at 37 ºC, iii) 10 min. pulse followed by 5 min. chase at 37 ºC, iv) 60 min. pulse followed by 5 min. chase at 37 ºC

These results are consistent with most newly-made Aha-tagged peptides being destroyed within minutes by the proteasome, as incubation with MG132 increases the fluorescent signal. A small fraction would then go on to contribute to the mature proteome, which is what is detected using longer pulses.

3.4: Colocalisation of Nascent Peptides and Nascent RNA

If there indeed exists a system of translation in mammalian cell nuclei, then this might be coupled to transcription. The colocalisation of nascent RNA and nascent peptide chains would serve as evidence in favour of a coupled system of transcription and nuclear translation. To label nascent RNA, HeLa cells were labelled with 5-ethynyluridine (EU, **3.1**)[17,23]. Though HeLa cells are methionine auxotrophs, they possess de novo pathways for the synthesis of uridine.[24] Therefore, labelling HeLa cells in uridine-depleted media supplemented with EU does not prevent the incorporation of uridine, which then competes with EU. In order to deplete endogenous pools of uridine monophosphate (UMP), and thus uridine, cells were incubated with 6-azauridine prior to labelling with EU. 6-azauridine is a competitive inhibitor for the conversion of orotidine monophosphate (OMP) to UMP during the *de novo* biosynthesis of uridine.[25] Using this method, a 5 minute labelling time yielded a significant fluorescent signal after CuAAC of EU with azide-linked Alexa-488 dye (Invitrogen), confirming incorporation of EU (Fig. 3.6).

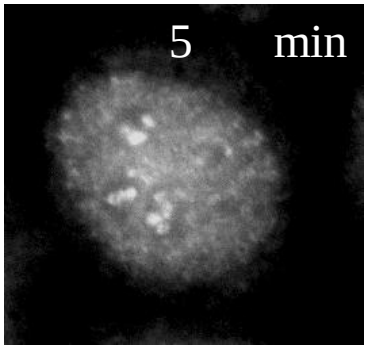
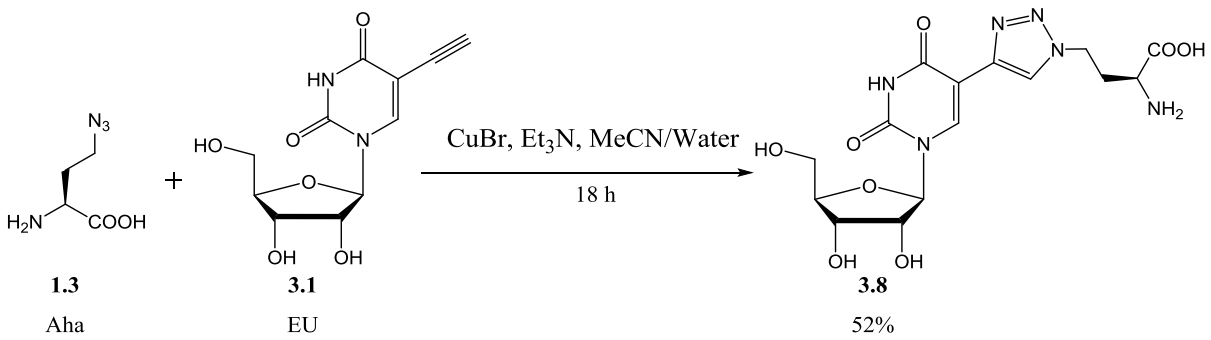


Fig 3.6: Fluorescence from EU-incorporated HeLa cells after CuAAC with azide-linked Alexa-488

In order to detect colocalisation of Aha- and EU- labelled peptides and RNA, CuAAC would be carried out. The resultant triazole species (**3.6**) was prepared as a reference (Scheme 3.1).



Scheme 3.1: Synthesis of 3.6 via CuAAC of Aha and EU.

Triazole **3.8** was seen to be fluorescent, with excitation wavelength λ_{exc} = 370 nm, and emission wavelength λ_{em} = 446 nm, thus making the triazole-forming reaction fluorogenic (Fig. 3.7). This is in agreement with the excitation and emission ranges for similar uridine-functionalised species, which have been used as fluorescent probes to study RNA folding.[26,27]

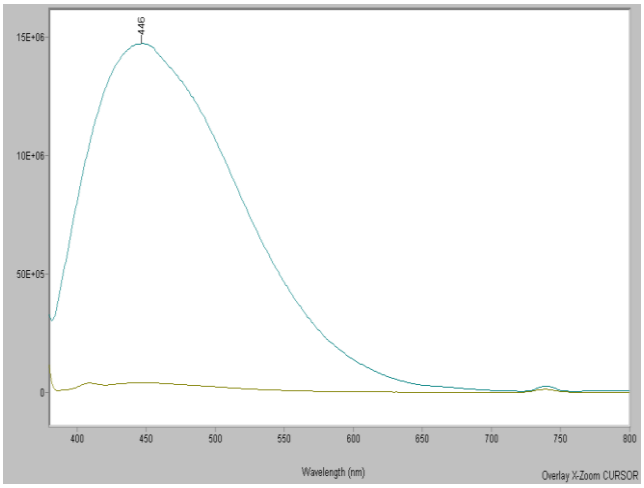


Fig. 3.7: Fluorescence emission spectrum for **3.8** (blue) in water (baseline), λ_{exc} = 370 nm.

Next, a double-incorporation experiment was carried out with Aha and EU. After labelling for 5 minutes, cells were subjected to CuAAC, giving rise to a fluorescent signal from both the nucleus and cytoplasm (Fig. 3.8). In unlabelled cells, this signal was not observed. Fluorescent signal was recorded in the DAPI channel (λ_{exc} = 358 nm, and λ_{em} = 462 for DAPI).[28]

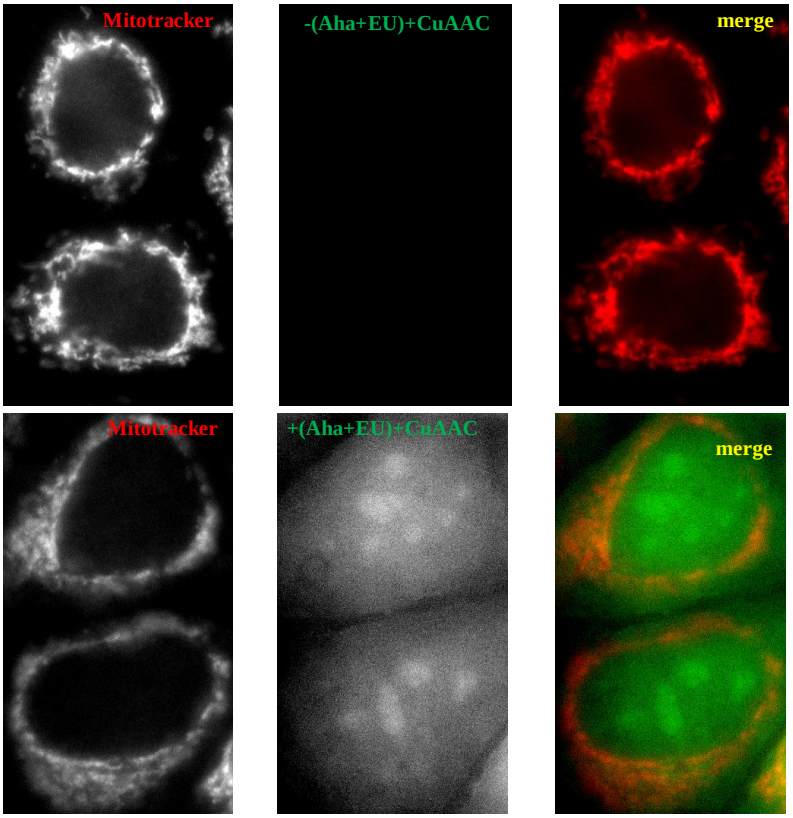


Fig. 3.8: Fluorescence obtained from HeLa cells labelled with Aha and EU (bottom row) or unlabelled HeLa cells (top row). Mitochondria stained with Mitotracker Red (λ_{exc} =579 nm, λ_{em} = 599 nm)

The formation of **3.8** in HeLa cells was also studied by two-photon fluorescence lifetime microscopy (FLIM), a technique by which a sample is imaged based on the distribution of fluorescence decay rates.[29, 30] The fluorescent lifetime of tryptophan might have been quenched upon the formation of a fluoresccent triazole species.[31] FLIM experiments were carried out at the Research Complex at Harwell with Dr Stan Botchway and Dr Tony Parker. FLIM images show that the average tryptophan fluorescent lifetime in HeLa cells (Fig. 3.9-a) which is usually 2.9 ns, is reduced to 2.1 ns in labelled cells after CuAAC (Fig. 3.9-b). This could be explained by the fact that the fluorogenic CuAAC reaction leads to the formation of an electronically active uridine-linked triazole, which may lead to the quenching of tryptophan side-chain fluorescence.

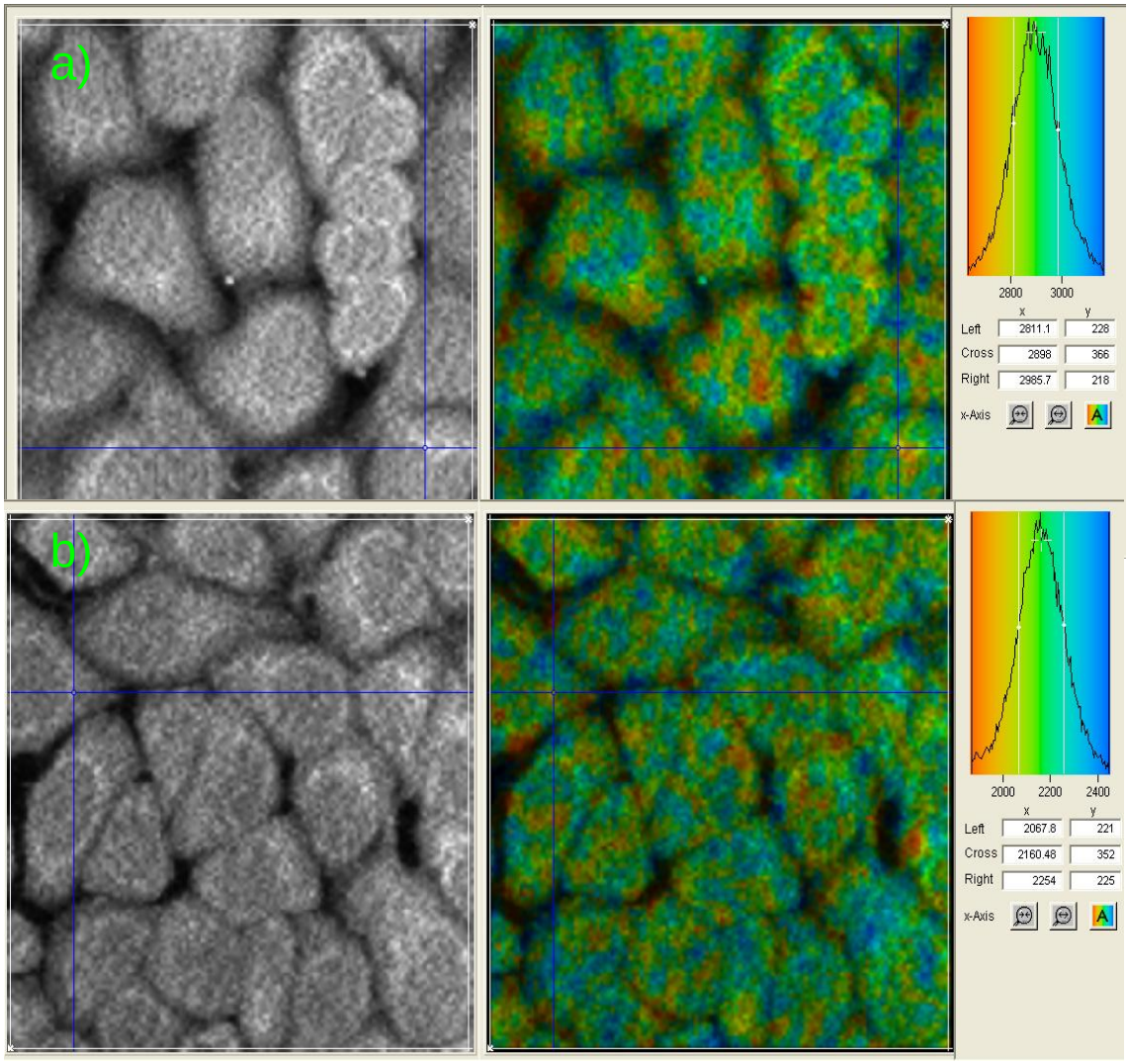


Fig. 3.9: FLIM images from HeLa cells a) control sample, average Tryptophan fluorescence lifetime 2.9 ns [right-hand distribution curve], b) post-labelling with Aha and EU, followed by CuAAC, lifetime reduced due to quenching to 2.1 ns. Excitation is at 590 nm (corresponding to 1 photon excitation at 295 nm)

However, when the cells were incubated only with the ascorbate solution provided with the Invitrogen kit, a necessary component to reduce Cu(II) to Cu(I) for CuAAC, a low intensity fluorescent signal could be observed in nucleoli (Fig. 3.10). This signal may be generated by the formation of pentosidine cross-links generated by the reaction of ascorbate with amino acid and carbohydrate-rich regions in the nucleoli via the Maillard reaction.[32,33] Although the fluorescence obtained from the Aha-EU labelled cells post-CuAAC is significantly higher than the fluorescence from cells treated only with ascorbate ($p < 0.0001$), we decided to

investigate this further. The fluorescence was not observed when cells were incubated with other reaction components alone.

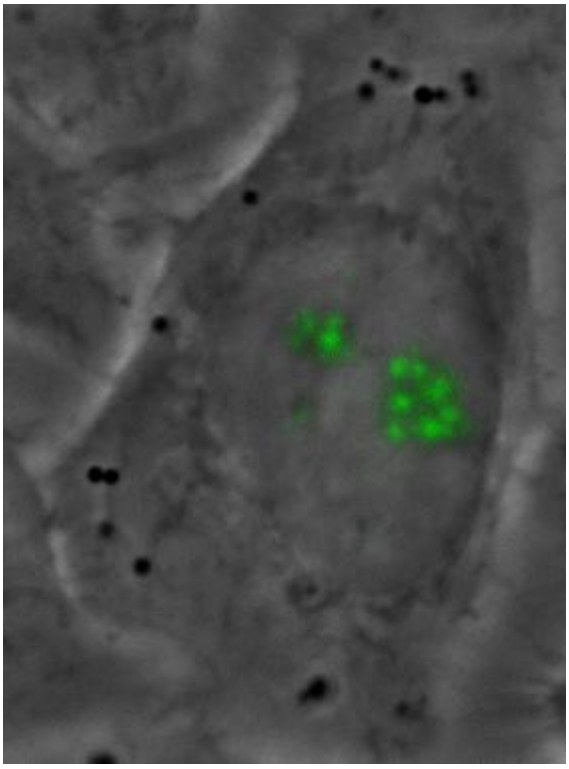


Figure 3.10: Low-intensity fluorescence from nucleoli of HeLa cells incubated with ascorbate only

Attempts were made to switch completely away from the copper(II)- and ascorbate-based Invitrogen kit to a copper(I) bromide-based labelling protocol, by pre-dissolving the copper bromide in DMSO or acetonitrile at final working concentration of 0.5% of the organic co-solvent during the reaction. However, less fluorescence increase was observed per unit time in the 555 nm channel post-labelling as compared to the Invitrogen kit, suggesting that the Alexa-555 labelling protocol via the non-reducing copper(I) bromide-based CuAAC reaction was less efficient than via the recuding Cu(II)-based CuAAC reaction . Longer labelling times would be needed to obtain significant fluorescence, which could allow for greater transport of proteins across the nuclear membrane, potentially confounding the results.

The formation of fluorescent species by ascorbate might also confound the results of the FLIM experiment on HeLa cells labelled with Aha and EU. Therefore, a comprehensive experimental analysis was carried out in order to identify the cause of the lifetime quenching of tryptophan (Table 3.1, Fig. 3.11 to 3.14).

Expt. No.	EU labelled	Aha labelled	CuAAC carried out (with ascorbate)	Tryptophan fluorescent lifetime
1	No	No	No	3.0 ns
2	No	No	Yes	2.0 ns
3	No	Yes	No	3.1 ns
4	No	Yes	Yes	2.1 ns
5	Yes	No	No	3.1 ns
6	Yes	No	Yes	2.1 ns
7	Yes	Yes	No	2.9 ns
8	Yes	Yes	Yes	2.1 ns

Table 3.1. Conditions of HeLa cells for FLIM experiments

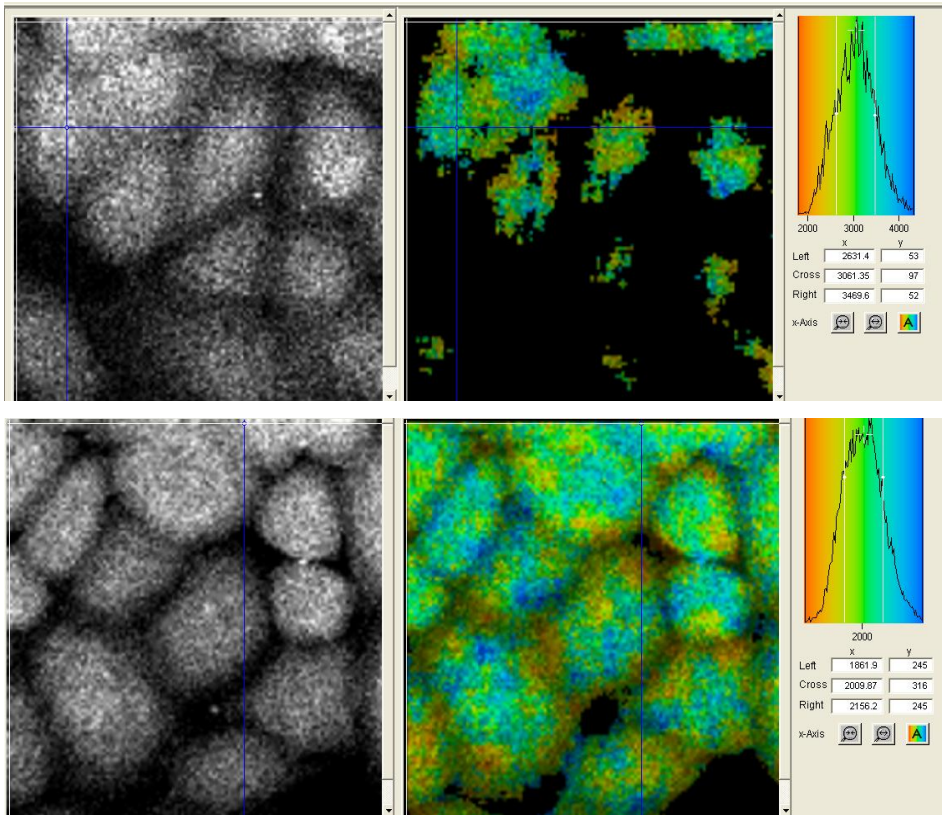


Fig. 3.11: Optical image (left panel), FLIM (centre) and lifetime distribution of tryptophan fluorescence in HeLa cells: a) (-Aha, -EU, -CuAAC), and b) (-Aha, -EU, +CuAAC).

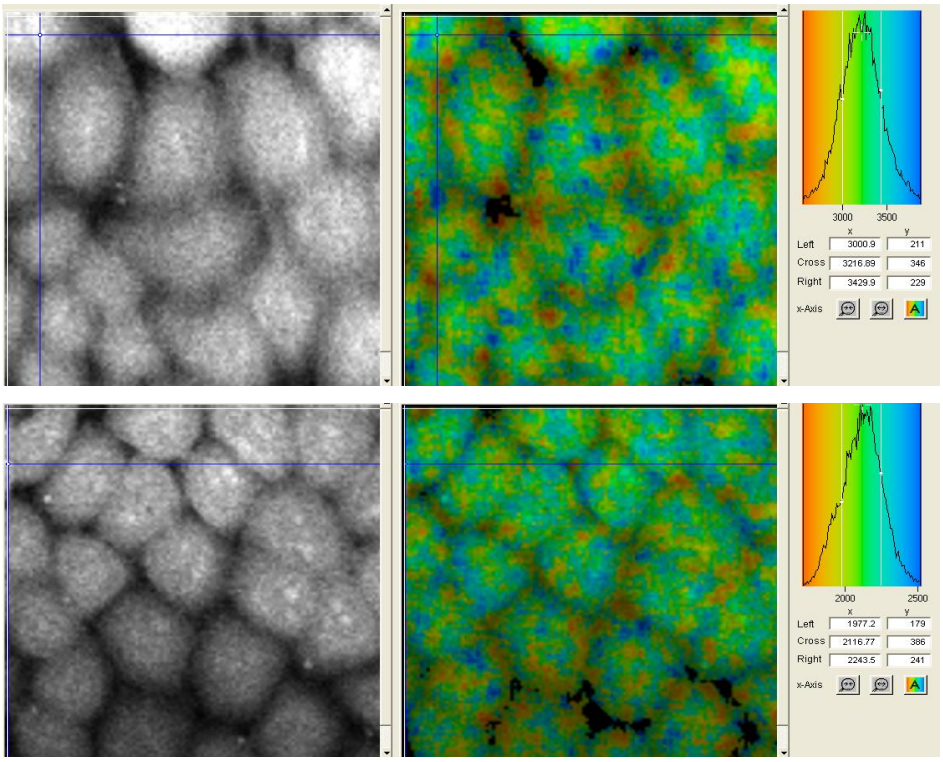


Fig. 3.12: Optical image (left panel), FLIM (centre) and lifetime distribution of tryptophan fluorescence in HeLa cells: a) (+Aha, -EU, -CuAAC), and b) (+Aha, -EU, +CuAAC).

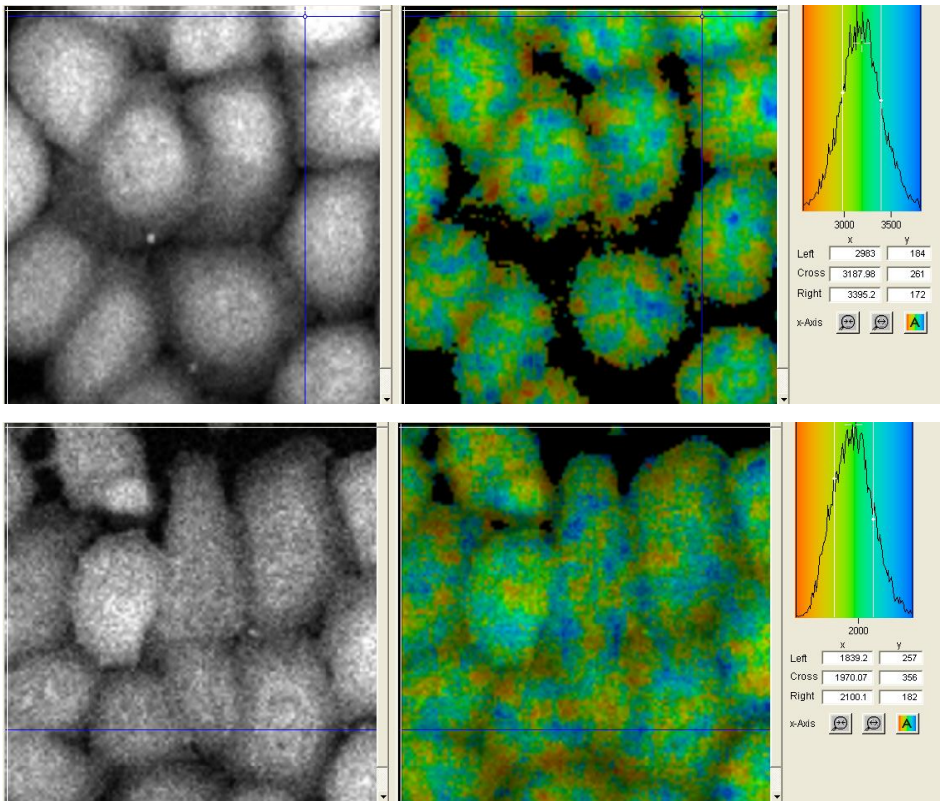


Fig. 3.13: Optical image (left panel), FLIM (centre) and lifetime distribution of tryptophan fluorescence in HeLa cells: a) (-Aha, +EU, -CuAAC), and b) (-Aha, +EU, +CuAAC)

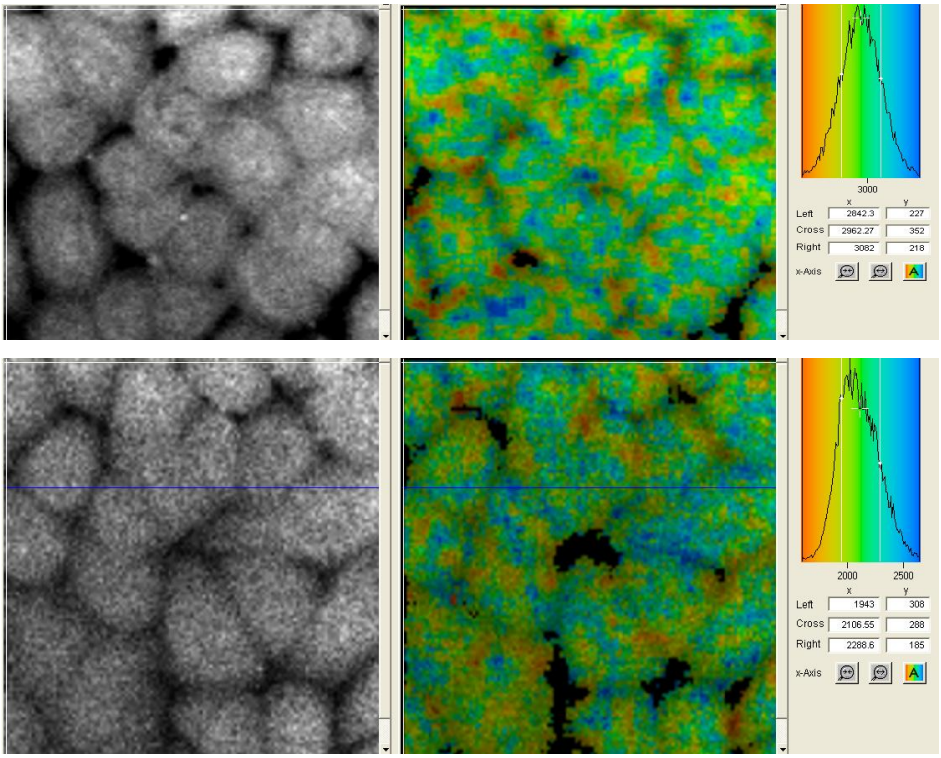


Fig. 3.14: Optical image (left panel), FLIM (centre) and lifetime distribution of tryptophan fluorescence in HeLa cells: a) (+Aha, +EU, -CuAAC), and b) (+Aha, +EU, +CuAAC)

In all the above cases, upon subjecting the fixed cells to CuAAC, the fluorescence lifetime of tryptophan was seen to decrease from 2.9 ns to 2.1 ns, regardless of whether the cells were labelled with either Aha or EU. Moreover, the decrease in fluorescence lifetime was not seen to be significantly altered when cells were incubated with both Aha and EU, indicating that the formation of the fluorescent triazole did not significantly alter the quenching of Trp fluorescence caused by ascorbate. Therefore, the in vivo formation of triazole **3.8** (and thus, the co-localisation of nascent peptides and RNA) could not be confirmed using FLIM, due to interference from ascorbate.

3.5: Raman Microscopy

Raman microscopic imaging (Chapter 5) of HeLa cells incubated with Aha and EU was also performed in order to map the intracellular distribution of these labels. The experiment was carried out at the Renishaw facility at Wotton-Under-Edge. This technique, like SIMS, would in principle eliminate the need for CuAAC to visualise these distributions, and be free from any biases of that technique. However, though a characteristic Raman stretching signal at 2100 cm^{-1} from recrystallised Aha was observed (Fig. 3.15), a significant signal could not be obtained from labelled cells using the Renishaw instrument (Fig. 3.16).

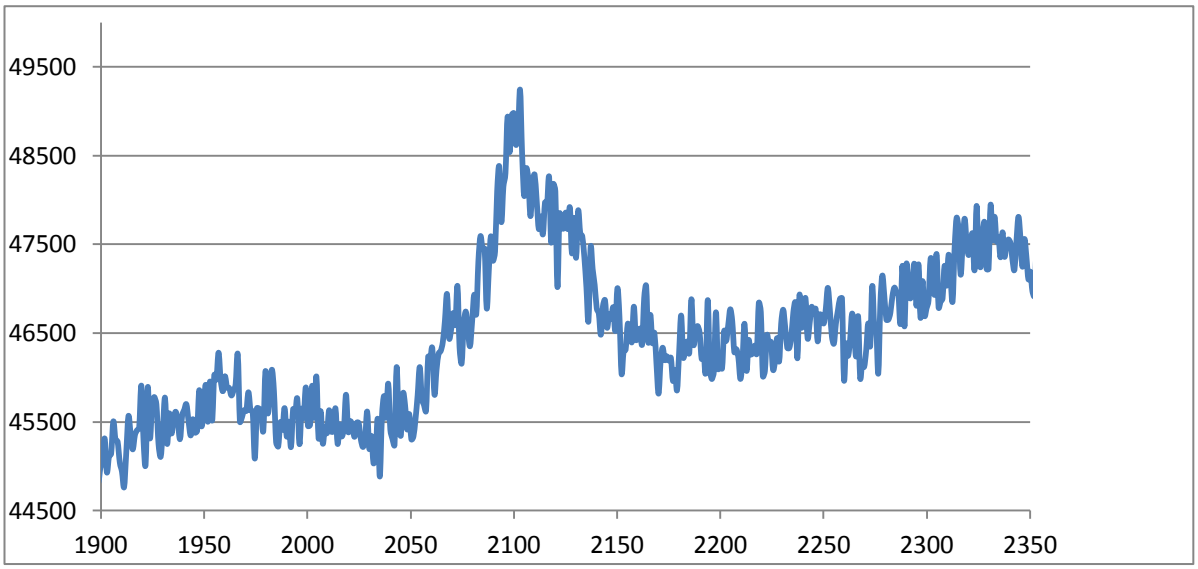


Fig. 3.15: Raman spectrum obtained from pure recrystallised Aha

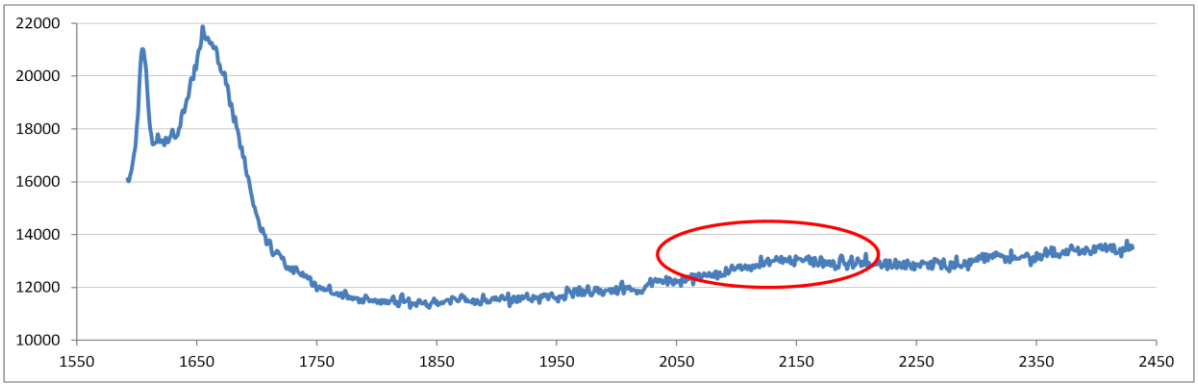


Fig. 3.16: Raman spectrum obtained from HeLa cells incubated with Aha for 5 min.

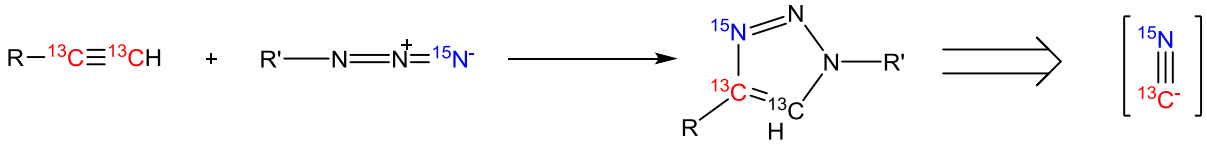
In Fig. 3.16, the highlighted region was initially thought to be a broadened azide stretching signal, but was subsequently found to be an artefact arising due to cellular fluorescence, and

found to arise from unlabelled HeLa cells as well. A higher signal could not be obtained for acquisition times up to 1 h, and a spectral map of labelled cells could not be recorded.

3.6: D-SIMS

D-SIMS (Dynamic Secondary-Ion Mass Spectrometry) is a mass-spectrometry based technique by which the surface composition of a sample may be analysed.[34,35] The technique involves raster-scanning the surface of a sample with a primary ion beam, causing the ejection of secondary ions from the surface by sputtering. The ejected secondary ions from each pixel are then analysed by MS to record m/z information, which is then combined with the spatial information of that pixel to create an image that reflects the isotopic composition of the surface. The technique has been successfully applied to biological systems, and combines high sensitivity and mass resolution ($^{13}C^{14}N^-$, 27.0064 amu, and $^{12}C^{15}N^-$, 27.0001 amu, can be distinguished), with a spatial resolution of ~100 nm in the *x* and *y* axes, and a few nanometers in the *z* axis.[36]

By specifically labelling the alkyne carbons in EU with ^{13}C, and the azide nitrogens in Aha with ^{15}N, a $^{13}C^{15}N^-$ secondary ion will be ejected from the Aha-EU complex if they have been successfully reacted by CuAAC. An increase in abundance of this heavy ion may then be used to determine the colocalisation of labelled Aha and EU (Scheme 3.2).



Scheme 3.2: Generation of $^{13}C^{15}N^-$ secondary ion by the degradation of ^{13}C and ^{15}N labelled Aha-EU complex

HeLa cells were incubated with ^{13}C-EU for 5 min. and ^{15}N-Aha for 10 seconds, fixed, in 2.5% glutaraldehyde, spun down, and embedded in epoxy resin. 1 µm sections across cell nuclei were cut for subsequent D-SIMS analysis (Fig. 3.17). SIMS images were acquired at the Begbroke Science Park.

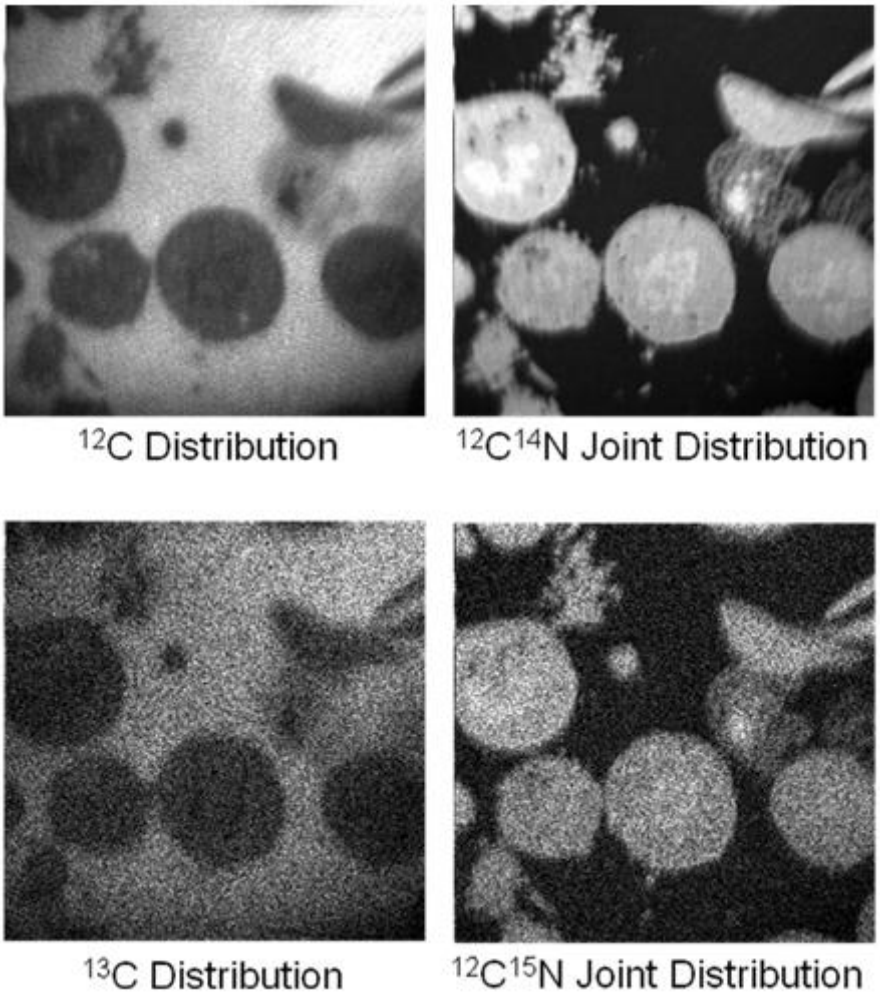


Fig. 3.17: D-SIMS maps for HeLa cell sections embedded in epoxy resin.

The signals for 'heavy' $^{12}C^{15}N^-$ and $^{13}C^-$ were seen to be lower than 'light' $^{12}C^{14}N^-$ and $^{12}C^-$, as both the labelling time with heavy isotopic species as well as the natural abundance of these heavy isotopes is low. However, a very low signal-to-noise ratio for the double labelled $^{13}C^{15}N^-$ ion was observed, making it difficult to unambiguously confirm the colocalisation of EU and Aha in these samples. The labelling on short time-scales (important to decouple nuclear translation from post-translational import) did not sufficiently label the system with heavy isotopes to fall within the detection limit of D-SIMS. Therefore, in order to maximise the amount of signal from the cell without increasing labelling times, commercially available heavy isotope-labelled SILAC[37] amino acids were used. [$^{13}C_6$, $^{15}N_4$]-L-Arginine and [$^{13}C_6$, $^{15}N_2$]-L-Lysine contain several-fold higher stable isotope labels per molecule than single ^{15}N Aha, leading to a concomitant increase in heavy isotope signal strength. Furthermore, arginine and lysine are natural amino acids, and unlike Aha, do not need to out-compete endogenous pools of amino acids during incorporation.

HeLa cells were grown for 2 min. in 'heavy' 8mM [$^{15}N_2$]Lys and 4mM [$^{15}N_4$]Arg, and the distribution of 'light' $^{12}C^{14}N^-$ ions in cell sections was first recorded from representative sections in both nuclei and cytoplasm, after area-normalisation (Fig. 3.18-i). Signal due to 'heavy' $^{12}C^{15}N^-$ ions was seen to be lower, as expected (Fig. 3.18-ii)

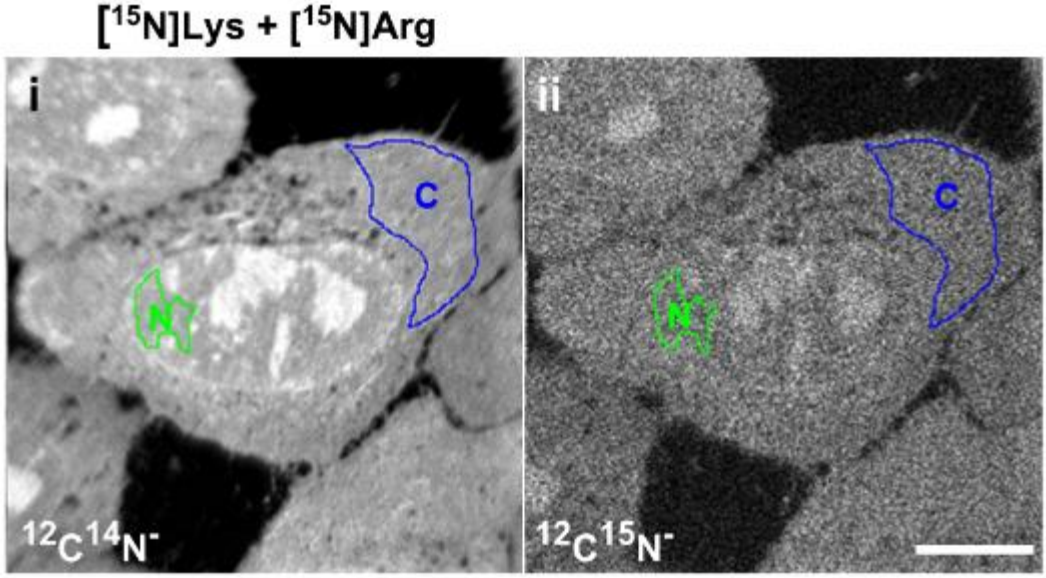


Fig. 3.18: ^{14}N and ^{15}N distribution in labelled HeLa cell sections as determined by the distribution of i) $^{12}C^{14}N^-$ and ii) $^{12}C^{15}N^-$ secondary ions by D-SIMS. Representative nuclear and cytoplasmic sections shown, Regions of interest (ROIs) were selected in the nucleus and cytoplasm that avoided nucleoli and obvious cytoplasmic vesicles in order to avoid biasing abundances.

As this signal contains contributions from naturally abundant ^{15}N in addition to the labelled amino acids, the signal from labelled cells was expressed both as a ratio relative to both heavy and light ions, as well as normalized relative to the abundance of ^{15}N in unlabelled cells (Fig. 3.19). Unlabelled control cells were seen to possess a signal ratio of (heavy) ÷ (heavy + light) which matches the theoretical natural abundance of ^{15}N of 0.0037 (row 1).[38] A 10 s pulse in heavy-amino acid containing medium increased the signal ratio ratios by ~6% in both the nucleus and cytoplasm (row 2). The signal was seen to be sensitive to the translation inhibitors as pre-treatment with anisomycin led to a reduction this increase (row 3). A 120 s pulse was seen to yield an even higher signal ratio in both compartments (row 4), which were reduced by a chase with light Lys- and Arg- containing media at 37°C (row 5) within 5 minutes, but not at 4°C (row 6). These results confirm that labelled peptides are metabolically synthesised in both nucleus and cytoplasm, and that they turn over rapidly.

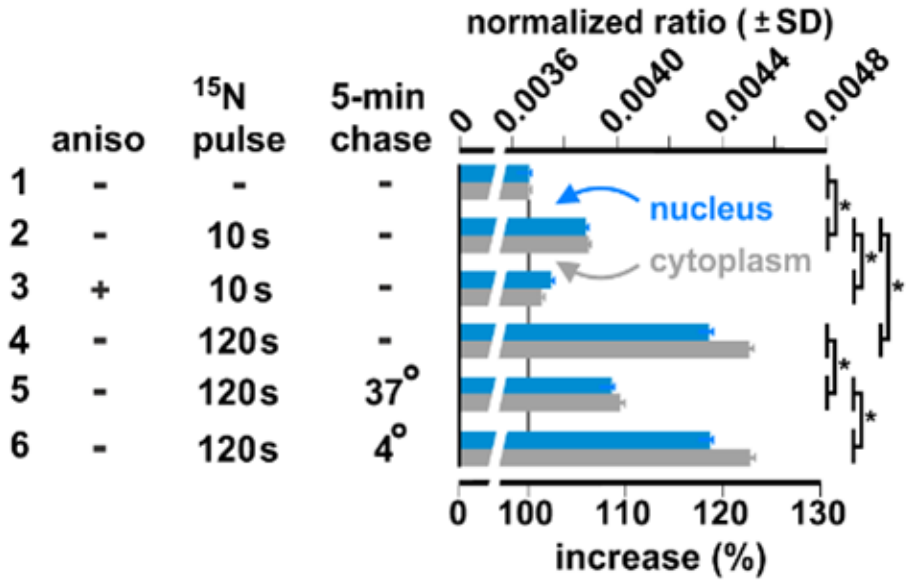


Fig. 3.19: Experimental conditions and ^{15}N abundance as determined by D-SIMS in labelled HeLa cell sections (blue: nuclear region, grey: cytoplasmic region). *: $p < 0.0001$ (Student's two-tailed *t* test; n = 19–26 cells)

The turnover rate of peptides was measured, which were expressed within a 2-min pulse labelled with either [^{15}N]Met (which is incorporated by the ribosome into both the N-terminus and internally), or [$^{15}N_2$]Lys and [$^{15}N_4$]Arg (which are not incorporated by the ribosome at the amino terminus). Most peptides labelled with [^{15}N]Met were seen to rapidly turn over in both the nucleus and cytoplasm within several tens of seconds (Fig. 3.20). Peptides tagged with [$^{15}N_2$]Lys and [$^{15}N_4$]Arg were seen to turn over more slowly.

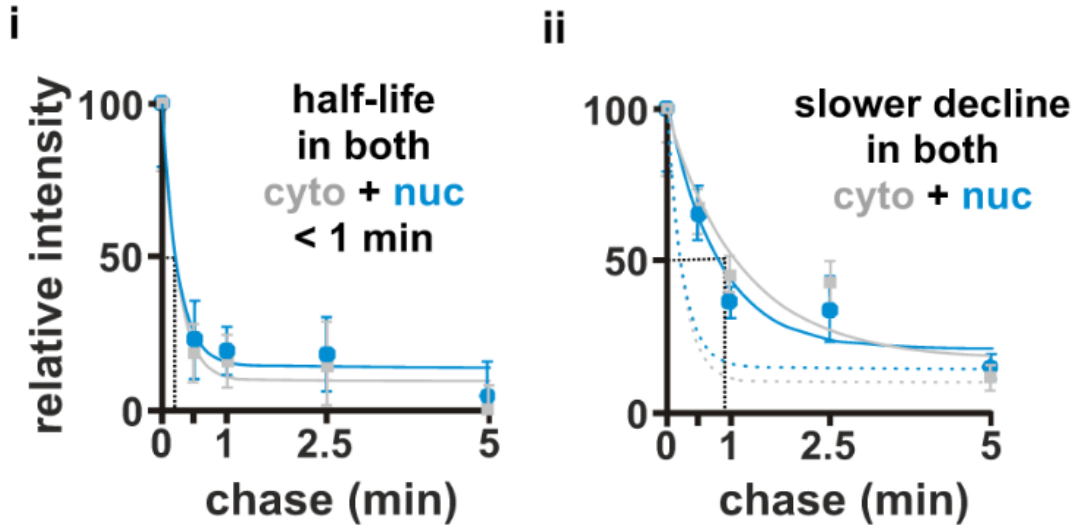


Fig. 3.20: Kinetics of ^{15}N-labelled peptide turnover in HeLa cell sections as determined by SIMS (blue: nuclear region, grey: cytoplasmic region) labelled with either i) [^{15}N] Met, or ii) [$^{15}N_2$]Lys + [$^{15}N_4$]Arg

These results are consistent with a system where translation occurs in both the nucleus and cytoplasm, and where most translating ribosomes abort soon after initiation with rapid turnover of the resulting peptides, and a minority go on to generate longer peptides to form functional proteins. By pulsing with a label like [^{15}N]Met, whose incorporation is biased toward the N-terminus of elongating peptides, most of the signal could then be attributed to short peptides that turned over quickly, synthesised by aborting ribosomes. Conversely, with labels that can only be incorporated into the middle of a peptide (i.e., [$^{15}N_2$]Lys and [$^{15}N_4$]Arg), a larger fraction is incorporated into longer peptides that turn over more slowly.

In order to confirm that the heavy isotope signal originated from peptide-bonded species rather than free amino acids, HeLa cells were starved of all amino acids for 15 min, and then incubated for 2 min. in media containing both sets of exclusively [^{13}C]-labelled amino acids and exclusively [^{15}N]-labelled amino acids . Statistically, one in four peptide bonds formed between these sets of amino acids would be expected contain a [^{13}C]-labelled carboxyl carbon covalently bonded to a [^{15}N]-labelled amide nitrogen, which would then yield $^{13}C^{15}N^-$ ions during SIMS (Fig. 3.21).[39]

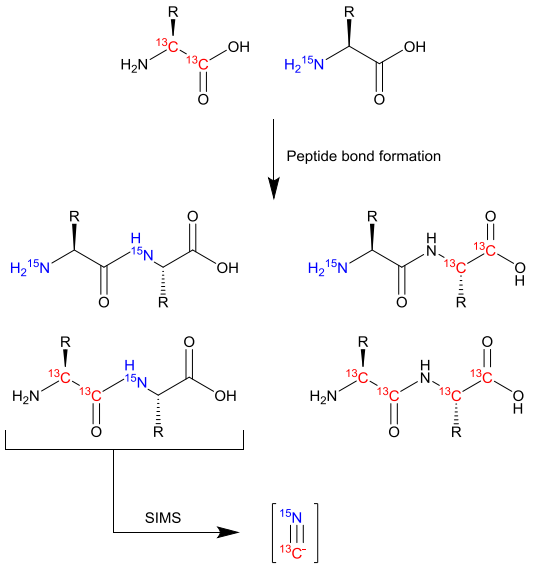


Fig. 3.21: Possible dipeptides formed from [^{13}C]-labelled amino acids and [^{15}N]-labelled amino acids, one of which would yield $^{13}C^{15}N^-$ during SIMS bombardment.

Most $^{13}C^{15}N^-$ ions would originate from the fragmentation of a ^{13}C-AA-^{15}N-AA peptide, though a minor fraction would originate from the complete fragmentation of two different molecules followed by recombination of ^{13}C from one with ^{15}N from the other.[39] Detection of

heavy-heavy ions at levels above their natural abundance should primarily reflect the presence of ^{13}C and ^{15}N atoms lying close together within the same molecule in the sample.

The distributions of $^{13}C^{15}N^-$ and $^{12}C^{14}N^-$ ions in nuclear and cytoplasmic regions were measured, and signals due to $^{13}C^{15}N^-$ were expressed as ratios relative to relative to all CN^- species. Unlabelled control cells were seen to possess a signal ratio of (heavy)÷ (light) which reflects the expected ratio given the abundance of ^{13}C and ^{15}N (Fig. 3.22 row 1). A 120 s pulse in heavy- amino acid-containing medium increased the signal ratios in both the nucleus and cytoplasm (row 2). The signal was sensitive to translation inhibitors: pre-treatment with anisomycin or cycloheximide[40] led to a reduction this increase (rows 3 and 4). The signal ratio in both compartments was reduced by a chase with light- amino acid-containing media at 37°C (row 5) within 5 minutes, but not at 4°C (row 6). These results were consistent with a significant fraction of the $^{13}C^{15}N^-$signal arising from newly made peptide-bonded species in both nucleus and cytoplasm.

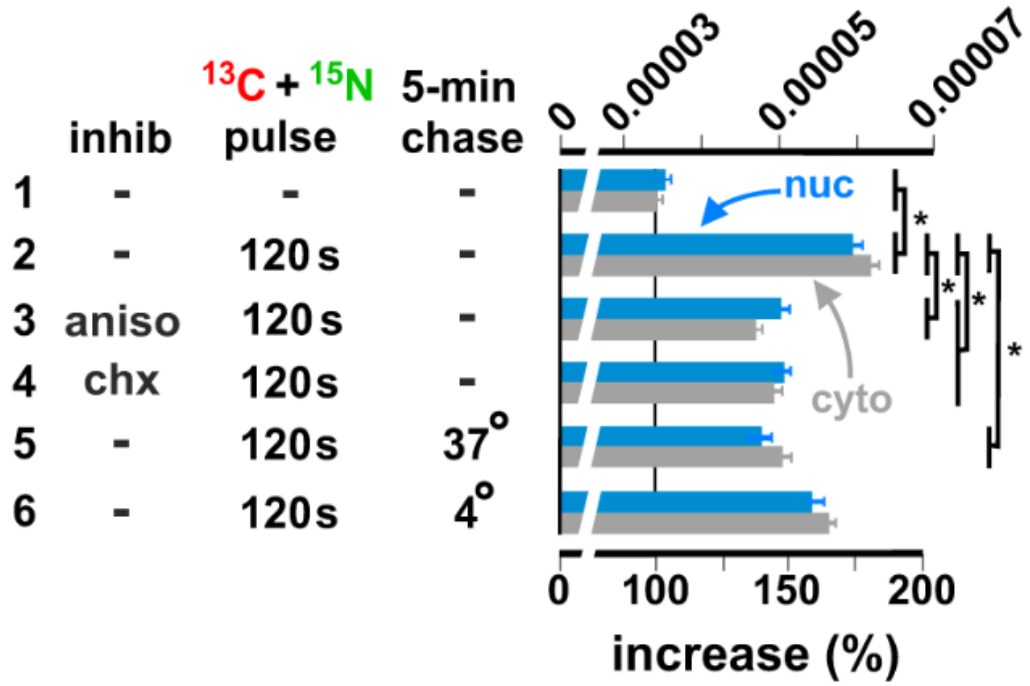


Fig. 3.22: Experimental conditions and $^{13}C^{15}N$ joint abundance in labelled HeLa cell sections as determined by D-SIMS (blue: nuclear region, grey: cytoplasmic region). *: $p < 0.0001$ (Student's two-tailed *t* test; n = 11–18 cells).

3.7: NMR on HeLa cells

In order to confirm that incorporation of amino acids was occurring in peptide bonded species, whole-cell NMR experiments were attempted on ^{13}C-amino acid labelled HeLa cells. If a cross-coupled NMR peak between the carboxyl ^{13}C of one amino acid and $^{13}C_{\alpha}$ of an adjacent amino acid was seen, that would serve as evidence in favour of peptide bond formation (Fig. 3.23). This signal was expected to comprise chemical shifts from peptide-bonded species arising from various combination of labelled amino acids in different magnetic environments inside fixed cells.

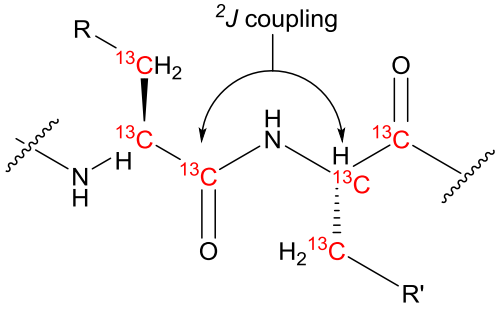


Figure 3.23: Expected ^{2}J coupling for peptide bonded ^{13}C-labelled amino acids

Cells were labelled with ^{13}C-AAs in AA-free DMEM media, with or without pre-incubation with cycloheximide. Labeling was carried out, for either 2 min. or 1 h, followed by fixation with ice-cold methanol. Cells labelled with ^{13}C-AAs for 1 h without cycloheximide were expected to have the highest signal, and were re-suspended in CD_3OD and SDS by vortexing for 15 minutes. 1-D experiments were run to see evidence of labelled molecules before 2-D coupling experiments could be run. ^{13}C NMR spectrum acquisitions were run for 2 h (2048 scans) on Brucker AVIII700 (700 MHz) and AVIII600 (600 MHz) NMR spectrometers with Dr. Tim Claridge.

On both machines, low-intensity multiplet ^{13}C signals could be seen at ~56 ppm and 16 ppm, which lie in regions where C_{α} and C_{β} were expected to be seen (Fig. 3.24). High-intensity flanking signals could be assigned to the solvent as well as SDS. No signals in the carbonyl ^{13}C region could be seen. On further inspection, however, the multiplet signal was seen to form a characteristic regular 1:3:6:7:6:3:1 septet corresponding to a -CD_3 species, rather than one arising from a mixture of different $^{13}C_{\alpha}$-$^{13}C_{\alpha}$ couplings across diferent peptide bonds or $^{13}C_{\alpha}$-$^{13}C_{\beta}$ couplings within residues. This suggests that these may be impurities in the

deuterated methanol NMR solvent, rather than signals from the cells. No signal enhancement was observed upon increasing the acquisition time to 20 h (20480 scans, Fig. 3.25).

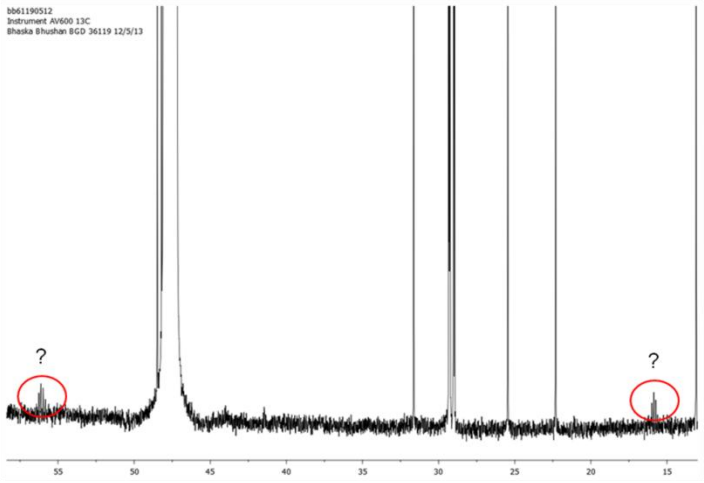


Fig 3.24: 600 MHz NMR signal for HeLa cells labelled with ^{13}C-amino acids for 1 h, 2048 scans

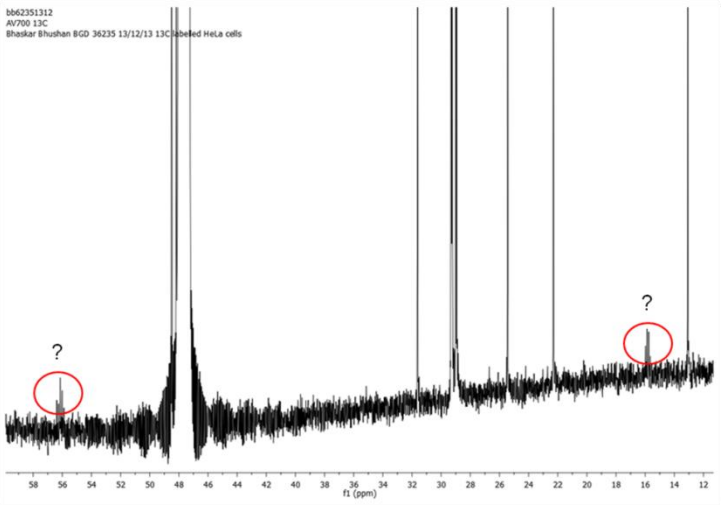


Fig 3.25: 700 MHz NMR signal for HeLa cells labelled with ^{13}C-amino acids for 1 h, 20480 scans

3.8: Puromycin labelling

Puromycin is a translation inhibitor which mimics the 3' end of tyrosyl-tRNA (Fig. 3.26).[41] It halts translation by entering the A site of ribosomes, and covalently bonding to the growing peptide chain's C-terminus.[42] This leads to the formation of nascent puromycylated peptide chains, which may be detected via immunoblotting and immunoprecipitation.[43] Alternative methods to detect puromycylated peptides include the use of fluorophore-linked puromycin,[44] or by labelling with anti-puromycin fluorescent antibodies.[45]

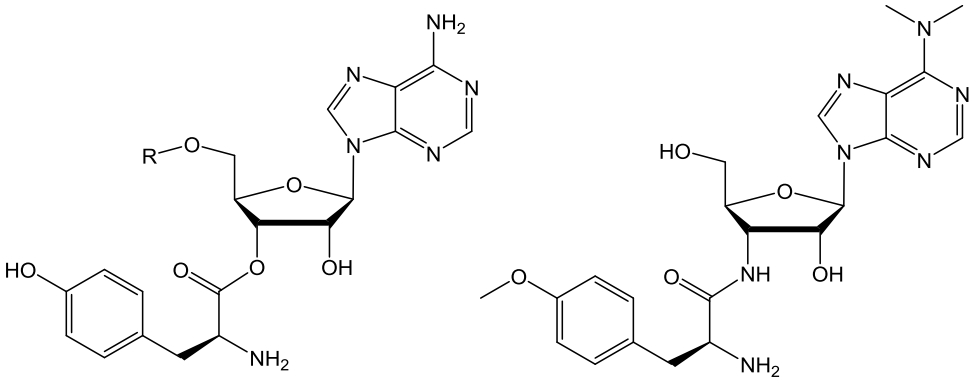


Fig. 3.26: Structures of tyrosyl tRNA terminus (left) and puromycin (right)

Nascent peptides have been reported to be generated in the nuclei of HeLa cells by labelling with puromycin followed by immunofluorescence.[46] Cells were also treated with cycloheximide, an elongation inhibitor,[40] to 'freeze' translation and to slow down dissociation of puromycylated peptides from their ribosomal sites of synthesis. These experimental conditions were modified by reducing the puromycylation time (down from 10 min. to 5 s) and fixation prior to permeabilisation and labelling with a mouse monoclonal antibody, 12D10 and Cy3-conjugated donkey anti-mouse antibody for visualisation.

After pre-treatment with cycloheximide for 15 min. and labelling with puromycin for 5 s, puromycylated species were seen in bright nuclear foci in confocal images (Fig. 3.27). Cells labelled with puromycin for 30 s displayed a more diffused signal (Fig. 3.27-C), and after 60 s, the peri-nuclear region was seen to become the brightest (Fig. 3.28-D).

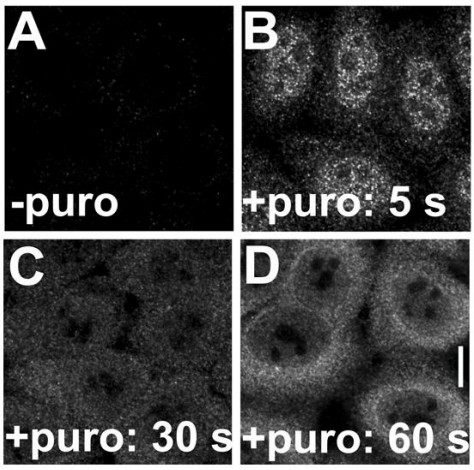


Fig. 3.27: Cy3 fluorescent signal from puromycylated species in HeLa cells.

Puromycylated peptides are known to accumulate at sites on the smooth endoplasmic reticulum (SER) after dissociation from ribosomes before passing through the Golgi apparatus to the exterior.[45-47] This may explain the results above, with puromycylated peptides migrating away from sites of synthesis (Fig. 3.27 B-D).

Additional results obtained using longer puromycin labelling times (5 min) were also consistent with this interpretation. In the absence of puromycin, background levels of fluorescence were seen (Fig. 3.28). Labelling with puromycin for 5 minutes yielded significant peri-nuclear fluorescence in addition to nuclear signal, as previously reported by David et al.[46] Importantly, this signal was seen to be sensitive to the translation inhibitor anisomycin, as pre-treatment was seen to reduce both cytoplasmic and nuclear signal.

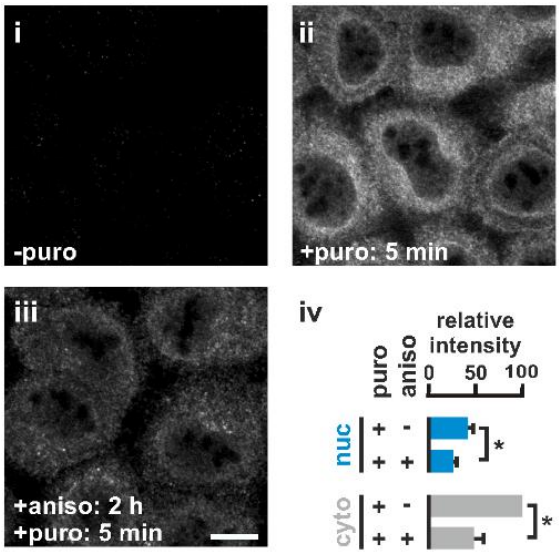


Fig. 3.28: Cy3 fluorescent signal from puromycylated species in HeLa cells labelled for 5 min: i) Unlabelled cells, ii) labelled cells, iii) labelled cells pre-incubated with anisomycin, iv) relative intensity in both compartments was seen to decrease with anisomycin treatment (*: $P < 0.0001$, Student's two-tailed t test, n = 20 cells)

Evidence in favour of a coupled system of transcription and translation in the nucleus was also obtained by pre-treatment of cells with DRB (5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole), a transcription inhibitor.[48] Inhibiting transcription was found to reduce the signal from puromycylated species in the nucleus (Fig. 3.29-iii,iv), but not the cytoplasm.As before, anisomycin treatment reduced signal in both compartments (Fig. 3.29-ii,iv).

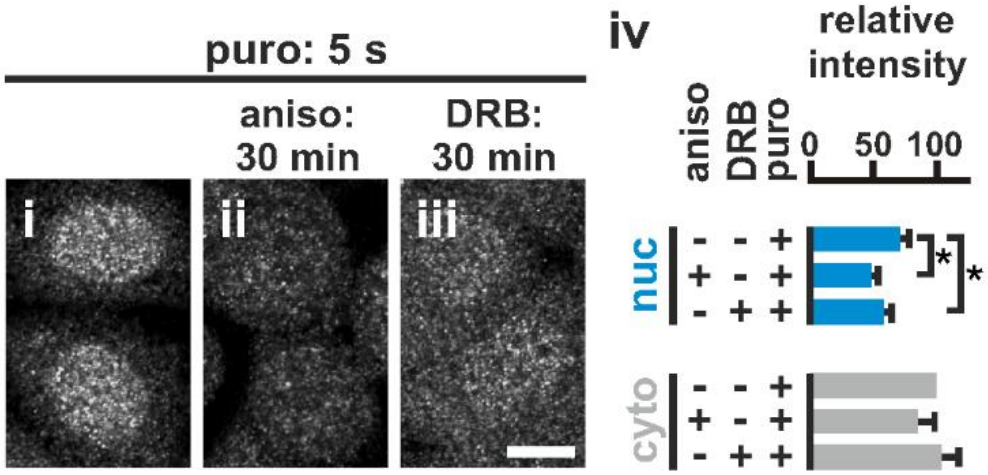


Fig. 3.29: Cy3 fluorescent signal from puromycylated species in HeLa cells labelled with puromycin for 5 s, i) without anisomycin or DRB treatment, ii) pre-treated with anisomycin, iii) pre-treated with DRB, iv) relative intensities in both compartments decrease with anisomycin treatment, but only nuclear signal is seen to decrease upon treatment with DRB.

3.9: RNA FISH

Data from three independent cell labelling techniques, i.e., Aha-incorporation, heavy AA incorporation, and puromycylation all point toward the existence of protein synthesis in the nucleus. However, peptides labelled during short pulses have half lives of less than 1 minute as determined from pulse-chase experiments, i.e., these short peptides are degraded by the proteasome quickly after they are synthesised. Attempts were then made to determine the reason for the existence of such an apparently wasteful turnover.

Prior to degradation by NMD, nascent RNA transcripts bearing PTCs might be read by active nuclear ribosomes,[7,49] with truncated peptides produced as a result. The degradation of these peptides might be reflected in the fast turnover observed in Section 3.3 and 3.6.

A test *Cd2* gene ± PTC (Fig. 3.30) was expressed using a multi-copy system; on transfection, the vector encoding *Cd2* would replicate to generate several thousands of 'mini-chromosomes' that would be co-transcribed by cellular polymerases in discrete foci.[50] *Cd2* expression is driven by a promoter normally switched off by a modified tetracycline repressor (introduced by co-transfection with a second plasmid). However, expression may be switched on by treatment with doxycycline, a member of the tetracycline group of antibiotics. Each construct encodes an SV40 *ori* (to permit replication) and a modified tetracycline promoter inducible with doxycycline. In the base vector, the Tet promoter drives expression of a protein containing an N-terminal mitochondrial signal sequence (mt), the rat *Cd2* epitope recognized by the OX34 antibody,[51,52] a shortened intron 2 of *Cd2* (with intact splice donor and acceptor sites), a ribosome pause sequence (rps) which should slow a translating ribosome, and a C-terminal EGFP. Vector 1 encodes an EGFP sequence with the segment encoding the fluorochrome deleted (nfEGFP, to free the green channel for other use). Vector 2 encodes a haemagglutinin (HA) tag inserted within the intron upstream of the first stop codon. Vector 3 also bears a PTC, 67 nucleotides upstream of the intron-exon junction, a position that should induce NMD.[53]

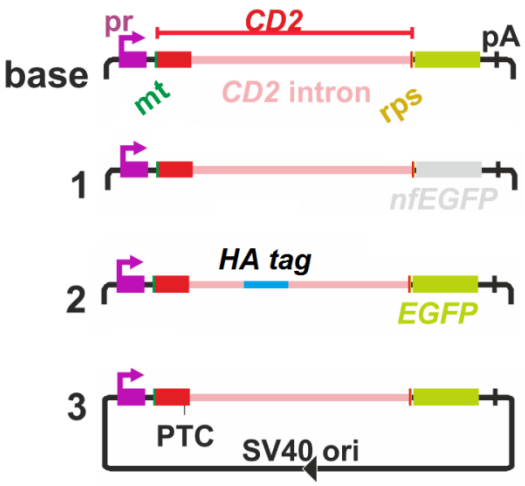


Fig. 3.30: CD2-EGFP expression constructs.

Levels of nascent *Cd2* RNA were monitored by fluorescence *in situ* hybridization (FISH) using probes targeting intronic regions of the RNA. Cos-7 cells (which encode the SV40 T antigen) were co-transfected with constructs encoding the tetracycline repressor and a second vector with the tetracycline promoter driving transcription of *Cd2* ± a PTC. After transfection, cells were treated with doxycycline (10 µM; 45 min), fixed 26 h after transfection, intronic *Cd2* RNA detected by RNA FISH, DNA stained with DAPI, and images collected using a wide-field microscope.

As expected, the (control) PTC^{-} vector was found to transcribe high levels of these intronic regions in many nucleoplasmic foci (Fig. 3.31 A, B). The PTC^{+} vector was found to yield less nuclear signal (Fig. 3.31 C-i); presumably, because the PTC-containing RNA would have been degraded by NMD. Pre-incubation with the translation inhibitor cycloheximide was found to increase signal above that seen in the control (Fig. 3.31 C-i), consistent with the inhibition of ribosomes from detecting PTC^{+} transcripts, thereby preventing their degradation by NMD. Importantly, when cells transfected with PTC^{-} constructs were treated with the splicing inhibitor spliceostatin A (SSA),[54] the fluorescent signal was seen to decrease, but treatment with anisomycin (aniso) reversed this decrease (Fig. 3.31 C-ii). This was consistent with the interpretation that spliceostatin A was expected to induce faulty splicing, generating several transcripts with PTCs (though one was not introduced through the vector). These transcripts would then be degraded by NMD, thus reducing the signal from intronic regions of C*d2*. However, if the recognition mechanism for NMD involved a translating ribosome, then

treatment with the translation inhibitor anisomycin would have inhibited downstream transcript degradation, and the signal would have been restored.[55]

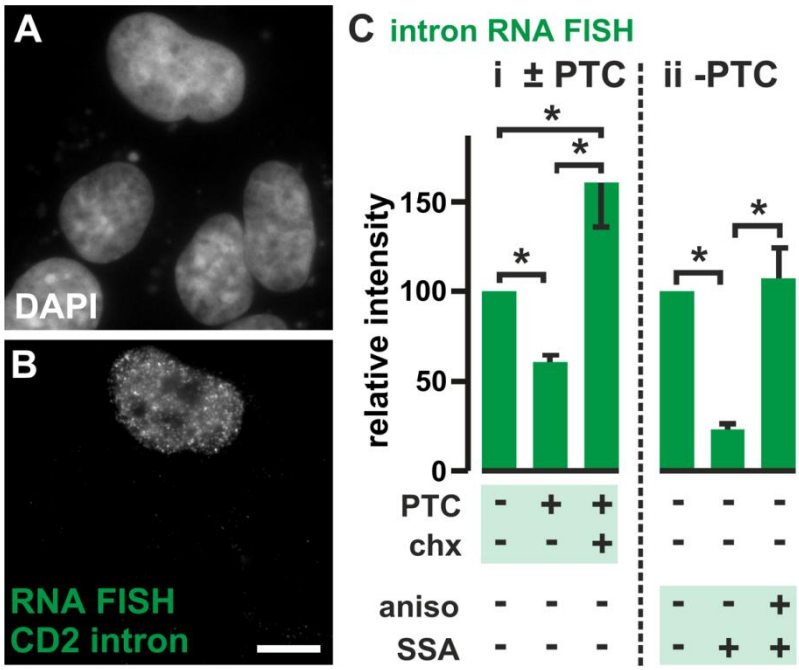


Figure 3.31. Effects of PTCs on nascent *Cd2* RNA in Cos-7 cells, A), B) Two views of one field after co-transfecting the PTC^{-} vector (only the cell at the top was transfected and expresses intronic *Cd2* RNA in nuclear foci). Bar: 10 µm. C) After subtracting background, intensities (± SD) seen in nuclei were expressed relative to the value found in untreated cells transfected with the PTC^{-} vector. *: $p < 0.0004$ (Student's two-tailed *t* test, n = 20 cells).

These results are consistent with a translating ribosome scanning nascent transcript for PTCs, and – once found – faulty transcripts and truncated peptides produced as by-products being degraded. As nascent PTC-bearing transcripts only occur in the nucleus, this is evidence in favour of nuclear translation. As the peptide turnover is significant, it follows that many nascent RNAs must encode PTCs. This interpretation is strengthened by the fact that ribosomes, NMD components, and proteasomes are all found in the nucleus (often at transcription sites) at concentrations that roughly equal those found in the cytoplasm.[7,56]

3.10: Mass spectrometry

'Vector 2' (Fig. 3.30) contained a haemagglutinin (HA) tag inserted in the intron of CD2, in-frame with previous exon and upstream of the first intronic stop codon followed by a C-terminal EGFP, i.e., Ex.(CD2) – Int.(HA-Stop) – Ex.(EGFP).

In COS-7 cells where this mRNA sequence is mis-spliced, the ribosome would read through the intronic transcript, synthesising a CD2-fused HA tag, terminating at the downstream stop codon. In correctly spliced mRNA, the ribosome would instead express eGFP, having not encountered the spliced-out HA. The HA tag was first utilised for in vivo immunofluorescence imaging to confirm expression. In COS-7 cells treated with spliceostatin A (SSA),[54] a significant HA-linked immunofluorescence signal was visible both in nuclear and cytoplasmic regions (Fig. 3.32). As both CD2 and HA are cell surface proteins,[57] this is a evidence in favour of nuclear synthesis of the nascent CD2-HA fusion protein.

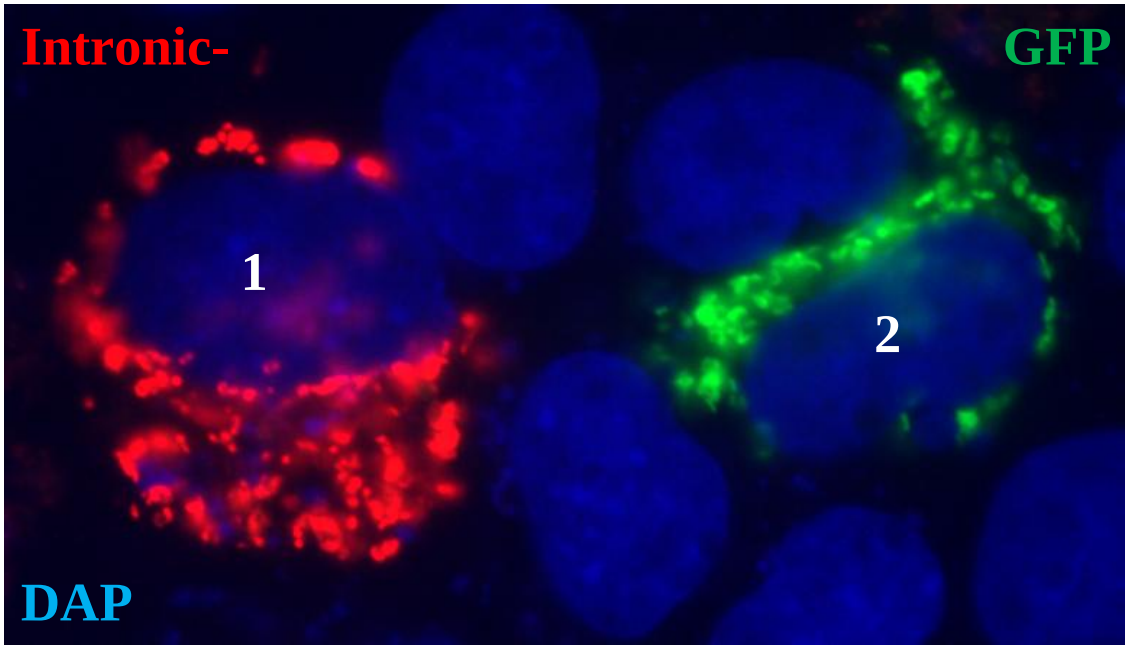


Fig. 3.32: COS-7 cells expressing CD2-intronic HA-exonic eGFP RNA constructs (48h), treated with SSA (24h). Cell 1 prominently shows mis-splicing, and the immunofluorescent signal from HA is seen to be strong in both the nucleus and cytoplasm. Cell 2 shows normal splicing, and the downstream eGFP is expressed.

In order to definitively confirm the presence of this construct in the nucleus, COS-7 cells were labelled with Aha and EU for 10s and 5 min. respectively, then subjected to CuAAC after incubating with SSA.

The full sequence of the synthesised protein was as follows:

RK(Aha)LAAVSRVLSGASQKPASRVLVASRNFANDATFLQRDSGTVWGALGHGINL
NIPNFQ(Aha)TDDIDEVRWERGSTLVAEFKRK(Aha)KPFLKSGAFEILANGDLKIKNLT
RDDSGTYNVTVYSTNGTRILNKALDLRILGKSSFLKVFHFR**TGA(Aha)YPYDVPDYA**

This sequence contains a tryptic digestion site (sites shown in green above) just before the HA peptide fragment (in blue). The CuAAC reaction product between EU and the labelled HA peptide TGA-(Aha)-YPYDVPDYA was sought by mass spectrometry from the trypsin-treated cell lysate.

The Aha-incorporated peptide was synthesised by SPPS and purified by RP-HPLC to be used as a standard for subsequent MS experiments (Fig. 3.33).

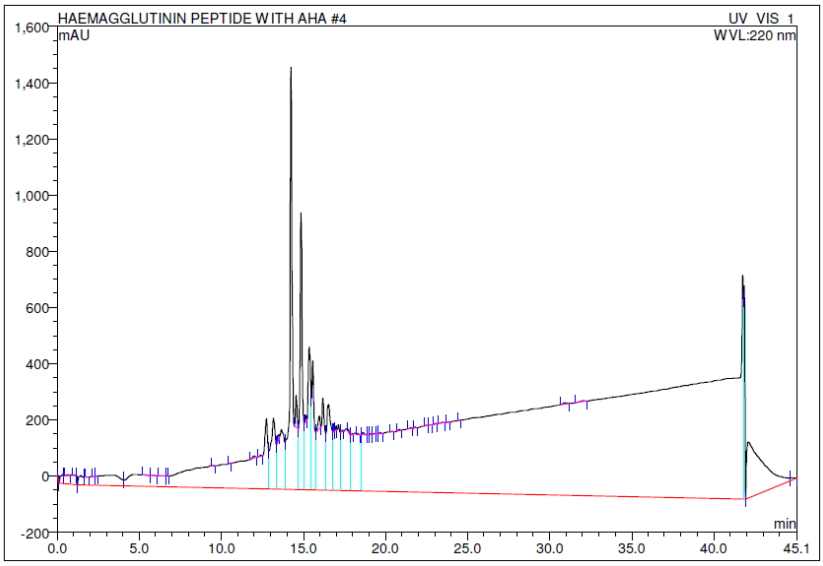


Fig. 3.33: HPLC trace for the RP purification of Aha-incorporated haemagglutinin peptide. The mass of the compound in the tallest peak corresponded to the expected peptide.

The peptide was subjected to CuAAC with EU, and the purification repeated (Fig. 3.34).

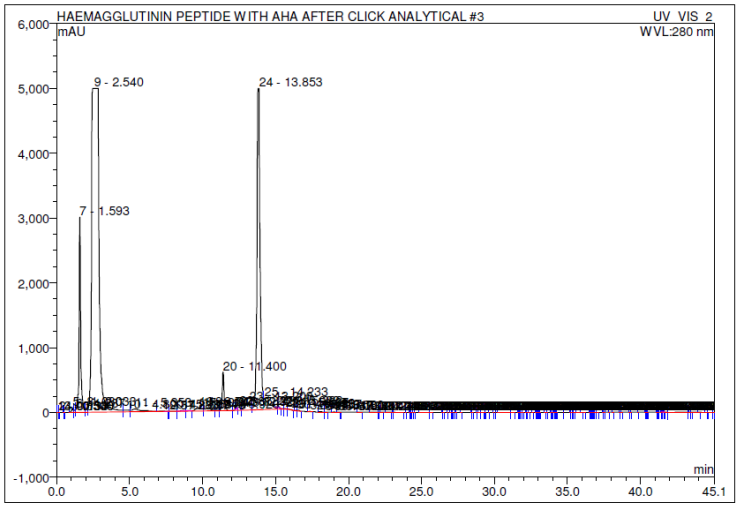


Fig. 3.34: HPLC trace of haemagglutinin peptide after CuAAC with EU. The tallest peak (at 15 min) corresponds to the desired product.

Mass spectrometric detection of COS-7 cell lysates were carried out by Benjamin Thomas at the Central Proteomics Facility at the Dunn School of Pathology. However, a poor signal to background ratio prevented detection of the Aha-EU functionalised haemagglutinin peptide at short labelling times.

3.11: Conclusion and Discussion

In this chapter, evidence in favour of a mechanism of nuclear translation and rapid turnover has been reported. Three independent labelling techniques were utilised in this regard. In the first, cells (transformed HeLa or diploid HUVECs) were pulse-labelled with Aha (a Met analogue), fixed, and a fluorophore covalently bonded via CuAAC to the nascent Aha-labeled peptides. Significant fluorescence was observed in nuclear regions of both cell lines. Unfortunately, the incorporation of Aha could not be independently verified by Raman microscopy, as the azide stretching signal at 2100 cm^{-1} was found not to be strong enough in cells labelled for short times. Simultaneous labelling with EU followed by CuAAC was found to generate a fluorescent triazole species, confirming co-localisation of labelled nascent peptides and nascent RNA. However, ascorbate (a reagent used for CuAAC) itself generated low-intensity fluorescence in unlabelled cells. Tryptophan fluorescence lifetime microscopy (FLIM) could not be used to verify the formation of the fluorescent triazole.[1]

In the second labelling method, cells were incubated in 'heavy' stable-isotope labelled amino acids, and both incorporation into nuclear regions and turnover detected by D-SIMS. Though ^{15}N-Aha and $^{13}C_2$-EU did not generate a strong enough $^{13}C^{15}N^-$ ion signal after CuAAC, the labelling with ^{13}C and ^{15}N-labelled amino acids (^{13}CAAs and ^{15}NAAs) did. To ensure most signal originated from nascent peptides, short pulses (5 s) were used, when only ~25 residues would be incorporated into growing peptide chains.

In the third labelling method, cells were pre-treated with cycloheximide to slow down translating ribosomes, pulse-labeled with puromycin (a structural mimic of tyrosyl-tRNA), fixed, and the puromycylated peptides immuno-localized using an anti-puromycin antibody.

All three different methods give signal in both nucleus and cytoplasm, with the incorporation being found to be metabolic, as confirmed by pulse-chase experiments, as well as sensitive to translation inhibitors. Surprisingly, the total amount of nuclear signal was only slightly less than the total amount seen in the cytoplasm for the short labelling times used. Furthermore, most of that signal in both compartments was seen to turn over in less than a minute, and not over many hours as expected.

The possibility that the observed signal might arise from non-specific binding of the three different labels within cells, or the incorporation of the labels into non-peptide bonded species was ruled out. Results from ^{13}CAA and ^{15}NAA incorporation make both possibilities unlikely, as the statistical likelihood of a $^{13}C^{15}N^-$ ion being produced from species other than a ^{13}CAA-^{15}NAA peptide bond was very low.[39] After 2-min pulses, signal from $^{13}C^{15}N^-$ was observed in both nucleus and cytoplasm of cells, consistent with the incorporation of these heavy labels into peptide bonds. As the relative levels of the $^{13}C^{15}N^-$ ions, as well as the effects of ribosomal inhibitors and chases corresponded to those seen in the two other independent labelling techniques (i.e., the heavy Met/Lys/Arg, as well as Aha and puromycin), this was considered to be strong evidence that all labels were incorporated by the ribosome into peptides.

Incorporation of as a result from some non-ribosomal activity (e.g., onto tRNA through the action of an aminoacyl tRNA synthetase) were also ruled out as all signals observed were sensitive to translational inhibitors, suggesting that incorporation was ribosomal.[58] Additionally, the signal was not reduced by the action of RNAse, suggesting that they did not arise from aminoacyl-tRNA complexes. An alternative explanation of these observations involves the binding of all three labels to some unknown biomolecules in a manner which biases their distributions towards those observed here. This seems highly unlikely, not only due to the observed formation of $^{13}C^{15}N^-$ ions, but also that this unknown binding must be affected by translation inhibitors, must be affected by proteases (and is therefore affected by MG132), and the concentration of such a species must change greatly during pulse-chase experiments.

About 95% newly-made nuclear RNA is degraded soon after it is transcribed due to prematurely-terminating (abortive) transcripts and intronic and non-coding RNA. However, when this turnover was only discovered in the 1950s after labelling periods shorter than the half-life of the newly-made RNA were introduced, no credible reason for why it might occur could be proposed at the time.[59] Though such synthesis and degradation of RNA appears wasteful, it is an integral part of cell metabolism, involving intronic and non-coding RNA. By analogy, the short half-lives of most newly-made peptides may have gone undetected because pulses longer than the half-life were used. If many ribosomes abort soon after initiating translation, then it is likely that the resulting short peptides will be missed using conventional approaches.

Translation has been traditionally monitored using radio-labelled amino acids, followed by acid precipitation of the resulting radio-labelled proteins. However, short peptides are not precipitated using standard conditions.[60] Also, current biochemical approaches are insufficiently sensitive to permit detection of analogues incorporated during the short pulses used here (see results from haemagglutinin mass spectrometry and Section **3.12.5** for another approach).

A question to consider is whether the cell utilises this apparently-wasteful turnover. In the case of the nuclear turnover, it seems it does. Results from RNA-FISH experiments are consistent with an active translating nuclear ribosome proof-reading nascent RNA for PTCs, and if found, degradation of the RNA by NMD. The truncated peptides formed as by-products would then need to be degraded. However, the reason for cytoplasmic turnover is not as clear. It may be that nuclear proof-reading is still error-prone enough that the cell must carry out a second round of with a cytoplasmic ribosome at degrading PTC^{+} transcripts. Alternatively, a fraction of nascent mRNA transcripts might be exported to the cytoplasm prior to having been read by nuclear ribosomes, resulting in cytoplasmic NMD and turnover of the peptides so generated. Additionally, more proteins than hitherto expected may terminate prematurely and/or misfold, and these are probably degraded quickly.[61-63] Finally, the apparently-wasteful turnover could play an important role in regulating translation.

During RNA synthesis, for every ~100 RNA polymerases that initiate at a promoter, ~99 abort prematurely to release transcripts of 2–15 nucleotides.[64] These transcripts are not detected using sequencing, as they are too short to be mapped. Subsequently, polymerases which continue transcription then terminate within 20–500 nucleotides to give the sense and anti-sense transcripts that can be detected by sequencing.[65] As a result, the production of each completed transcript is regulated by the synthesis of hundreds of shorter sense and anti-sense products which are degraded after synthesis. Although this may seem energetically wasteful, cells then go on to further degrade nine-tenths of the resulting full-length transcript when introns are removed to give the mature mRNA. Similar principles might apply to translation.

Approximately 50% of human mRNAs encode upstream open-reading frames (uORFs), which regulate the expression of downstream genes.[66,67] Ribosome profiling studies on mammalian cell systems indicate that uORFs are translated,[68,69] probably resulting in short peptides which may be quickly degraded.[61,63,69,70] If translation is like transcription, then

during the short pulses used in this chapter, many ribosomes will terminate soon after initiation (both at the end of uORFs, and close to the beginning of uORFs and ORFs). Most peptides generated in this way will be degraded quickly, and only a minority of ribosomes will go on to generate full-length proteins that would be detected conventionally using long pulses.

If so, then pulse-labels of different times would reveal proteins belonging to different metabolic pools, most newly-made peptides would be degraded as soon as they are made, and the apparently-wasteful turnover would be a by-product of an inefficient process. Some of this turnover may be exploited by the system to regulate production. Then, a previously unseen – and so 'dark-matter' – population of short peptides would coexist with its RNA counterpart.[71]

3.12: Methods

3.12.1: Cell culture and fixation

HeLa and Cos 7 cells were cultured in DMEM (PAA) supplemented with 1% penicillin/streptomycin (PAA) and FBS (Bioline; 5% for HeLa, 10% for Cos 7); HUVECs (human umbilical vein endothelial cells; Lonza; note that the supplier were responsible for obtaining the necessary ethical provision) were cultured in EGM-2 (Lonza). Prior to imaging, cells were cultured on clean glass coverslips (22×22 mm; no. 1.5) etched with 1% hydrofluoric acid (Merck). 2 mM Aha, 100 µg/ml anisomycin (Sigma), 100 µg/ml MG132 (Enzo), 91 µM puromycin (Sigma), 100 µM DRB (Sigma), 10 µM doxycycline (Calbiochem), 100 µg/ml cycloheximide (Sigma), and 100 nM spliceostatin A (a gift of M Yoshida) were added to live cells at the times indicated. Prior to pulsing with Aha, HeLa cells were grown in the absence of Met (i.e., in DMEM^{-}, which is DMEM without Met, cysteine, and L-glutamine; Sigma, then supplemented with Cys and Gln) for 15–30 min. at 37°C (to deplete endogenous pools of Met). Prior to pulsing with puromycin, cells were grown in the presence of cycloheximide (in DMEM supplemented with 5% FBS and 1% penicillin/streptomycin for HeLa, and EGM-2 for HUVECs) for 15 min. at 37°C (to slow translating ribosomes). For pulses of 5–120 s in Aha or puromycin (and for pulse-chase experiments), cells on a coverslip were dipped in medium (DMEM^{-} for HeLa and EGM-2 for HUVECs) containing the labelling species (and then into medium lacking the labelling species during a chase) in a 37°C room for the required time before plunging the coverslip immediately into ice-cold fixative. When a pulse was followed by chase, DMEM^{-} +2 mM Aha was replaced by DMEM (which contains ~200 µM Met) supplemented with 5% FBS and 1% penicillin/streptomycin, or EGM-2 in the case of HUVECs. Cells were fixed with 4% paraformaldehyde (Electron Microscopy Sciences) in 250 mM HEPES (pH 7.6; PAA) for 20 min, usually at 20°C. For transfection, Cos 7 were grown in DMEM supplemented with 10% FBS (tetracycline free; Clontech) to 80–90% occupancy in 6-well plates and transfected using 8 µl FuGENE HD (Roche/Promega) with 2 µg DNA that included 0.2 µg of Vector 1−3+0.2 µg pTetOn Advanced (Clontech) +1.6 µg sheared salmon sperm DNA (Invitrogen). After 12 h, cells were trypsinized, washed, and re-plated on coverslips in 6-well plates (to reduce background caused by input DNA). Cells were then grown for 12 h, induced with doxycycline in the presence or absence of inhibitors, and fixed.

3.12.2: Fluorescent detection of Aha by CuAAC

After fixation, cells were washed in PBS (PAA; 5 min; 20°C), permeabilized with 0.5% Triton X-100 (Sigma), washed in 3% BSA in PBS (5 min; 20°C), an alkyne-conjugated fluorophore (i.e., 100 nM Alexa555; Invitrogen) attached (using the copper-catalyzed azide-alkyne cycloaddition reaction, CuAAC) to incorporated Aha using the 'Click-iT Cell Reaction Buffer kit' (Invitrogen) as instructed by the supplier (30 min; 20°C; vigorous shaking), washed in 3% BSA in PBS (5 min; 20°C), and rinsed in PBS. After counterstaining nuclei with DAPI (Sigma) and mounting in Vectashield (Vector Laboratories), images were acquired images using a wide-field or confocal microscope.

3.12.3: Immuno-labeling

Fixed cells were washed with PBS (5 min; 20°C), permeabilized with 0.5% Triton X-100 and 0.5% saponin (Sigma), washed with 0.05% Tween 20 (Sigma) in PBS (10 min; 20°C), and blocked with 3% BSA and 0.2% cold water fish skin gelatin (Sigma). Puromycylated peptides were indirectly immuno-labeled by incubation (1 h; 20°C) with a mouse monoclonal antibody, 12D10 (1/100 dilution), cells washed 4x with 0.05% Tween 20 in PBS (1 min. each; 20°C), incubated (30 min; 20°C) with a Cy3-conjugated donkey anti-mouse (Jackson ImmunoResearch, 715-165-150; 1/200 dilution), and washed 3x with 0.05% Tween 20 in PBS (10 min. each; 20°C) followed by a wash with PBS (10 min; 20°C). After staining with DAPI and mounting in Vectashield, images were acquired with a confocal or wide-field microscope.

3.12.4: Fluorescence microscopy

Microscopes used were: (a) Wide-field fluorescence microscope (Zeiss; upright – Axioplan 2e microscope), equipped with a 175 W Xenon arc lamp (Perkin Elmer); images were acquired using a 63x Zeiss Plan-APOCHROMAT oil-immersion objective (numerical aperture 1.4), optical filters (Chroma), and a CoolSNAP$_{HQ}$ camera (Photometrics) running under MetaMorph software (Molecular Devices). (b) Confocal microscope (Olympus, FV1000, IX81); images were acquired using a 100x Olympus UPlanSApo oil-immersion objective (numerical aperture 1.4), optical filters (Olympus), a confocal aperture of 175 µm and scanning at 10 µs/pixel, 405 and 559 nm 'diode-pumped solid-state' or argon (488 nm) lasers, and FLUOVIEW v2.1b software (Olympus). Images were analyzed using ImageJ.

The fluorescence intensity over nucleus and cytoplasm (images obtained using a wide-field microscope) was determined as follows. (i) The fluorescent intensity over the whole cell and nucleus (area defined by DAPI staining) was measured, and that over the cytoplasm calculated by subtraction. (ii) Background in each compartment was subtracted; backgrounds determined separately for nucleus and cytoplasm (i.e., whole cell minus nucleus) over non-pulsed, 0-s pulsed, or untransfected cells as appropriate. (iii) Values from each experiment were exported to Excel, and mean and standard deviation (SD) calculated. *P* values (two-tailed) from unpaired Student's *t*-test were calculated using GraphPad.

3.12.5: Detecting Aha-tagged peptides biochemically

It is generally accepted that fluorescence microscopy can allow detection of fewer tagged molecules than conventional biochemical approaches. An attempt was made to detect nascent Aha-tagged after running them on gels, but failed to do so successfully. For example, one approach involved combining two analogues: after 60-s growth in Aha (with puromycin added during the last 5 s), CuAAC with alkyne-functionalised biotin with a cleavable disulfide bridged linker,[72] pull-down on Neutravidin beads, release of biotinylated peptides using 0.5 M DTT, and Western blotting using the anti-puromycin antibody, we failed to distinguish between signal and background.

3.12.6: D-SIMS

HeLa cells were grown on coverslips. For labeling with heavy Lys and Arg, cells were grown in custom-made DMEM without Lys + Arg (Thermo Scientific) for 30 min, regrown in DMEM (without Lys + Arg) supplemented with 8 mM [$^{13}C_6{}^{15}N_2$]Lys + 4 mM [$^{13}C_6{}^{15}N_4$]Arg (Cambridge Isotopes Laboratories; ^{15}N 99% enriched) for 10 or 120 s at 37˚C, immediately plunged into ice-cold fixative, and incubated in fixative for 20 min. at 4˚C. The ^{13}C tag was not utilised for D-SIMS here. For labeling with heavy Met, cells were grown in DMEM without Met, Gln and Cys (Sigma, then supplemented with Gln and Cys) for 30 min, regrown in the same medium supplemented with 2 mM [^{15}N]Met (Cambridge Isotopes Ltd; ^{15}N 97-99%) for 2 min, and fixed. For labelling with [^{13}C] amino acids plus [^{15}N] amino acids, cells were starved for 15 min. in custom-made DMEM lacking all amino acids (Thermo Scientific), regrown in the same medium supplemented with [^{13}C] algal amino acid mixture (Isotec; uniformly ^{13}C-labeled, 98 atom % ^{13}C) plus [^{15}N] cell-free amino acid mixture

(Cambridge Isotope Labs Inc; uniformly ^{15}N-labeled, 96.98%), and fixed. Here, each of the 20 amino acids was assumed to have an average molecular weight of 136, in order to have a final concentration of 1 mM during the pulse. In some cases, 100 μg/ml anisomycin or 100 μg/ml cycloheximide was present during both the starvation and the pulse, and the pulse was followed by chases in DMEM plus 5% fetal bovine serum at 37˚C or 4˚C.

After the pulse, or pulse-chase, sections of cells were prepared using a procedure commonly used for electron microscopy. Cells were fixed in 4% paraformaldehyde (Electron Microscopy Sciences) and 1% glutaraldehyde (Agar) in PBS (PAA). Samples were further fixed (1 h; room temperature) with 2.5% glutaraldehyde in 0.1M sodium cacodylate buffer (pH 7), washed in the buffer 3x (10 min. each), incubated (1 h) in 1% osmium tetroxide in the same buffer, and washed with water for 20 min. Samples were dehydrated using a graded ethanol series (50%, 70%, 90%, 95% and 100%), and embedded in Agar100 epoxy resin; after removing the coverslip, semi-thin sections (0.5 μm) were cut using an ultramicrotome (Leica UC7) with a diamond knife (Diatome), mounted on platinium-coated coverslips, and coated with 5-nm platinum. SIMS images were acquired on a CAMECA (Gennevilliers) NanoSIMS 50 to measure the ^{15}N and ^{13}C signals in the nucleus and cytoplasm. The NanoSIMS uses a 16 keV Cs^+ ion beam to bombard the sample surface, and was tuned to detect $^{12}C^-$, $^{13}C^-$, $^{12}C^{14}N^-$, $^{12}C^{15}N^-$, $^{13}C^{15}N^-$, or $^{31}P^-$ ions using a Mauttach-Herzog mass analyser with electrostatic sector and asymmetric magnet configuration. Images were analysed using the Image J plugin OpenMIMS (MIMS, Harvard University, www.nrims.harvard.edu). Regions of interest (ROIs) were selected in the nucleus and cytoplasm that avoided nucleoli and obvious cytoplasmic vesicles or oil droplets, and ratios calculated as indicated (i.e., $^{12}C^{15}N^-/[^{12}C^{15}N^- + ^{12}C^{14}N^-]$. The control samples had a $^{12}C^{15}N^-/[^{12}C^{15}N^- + ^{12}C^{14}N^-]$ ratio of 0.003838. This compares with the expected value of 0.003667 (i.e., the natural abundance of ^{15}N). Therefore other ratios were normalized relative to this expected value.

3.12.7: RNA fluorescent *in situ* hybridization (RNA FISH)

A set of forty-eight 20-mer probes (designed using Probe Designer; http://singlemoleculefish.com/) were synthesized (Biosearch Technologies, USA) targeting the intronic region (~1.6 kbp) of rat *Cd2* DNA present in Vectors 1-3. At the 3′-end of each 20-mer, an mdC(TEG-amino) modification was added. The amino group was subsequently

labeled[73] using the ARES Alexa Fluor 488 DNA labeling kit (Invitrogen) according to the manufacturers' instructions. Labeling efficiency was found to be ~5 fluorophores per 100 nucleotides. Fixed cells were processed for RNA FISH (Sequences listed in Table 3.2), and imaged as described[73] with the following exceptions: cells were immediately permeabilized as above for immuno-fluorescence, and the hybridization buffer contained 10% formamide. When RNA FISH was performed on EGFP-expressing cells, Alexa488 fluor was further detected by immuno-fluorescence as follows. After performing RNA FISH as described above, coverslips were washed thrice in 2xSSC (10 min. each; 37°C), washed with 0.05% Tween 20 in PBS (3 min; 20˚C), and blocked as above. The *Cd2* intronic RNA was detected by incubating the cells with primary antibody against Alexa 488 Fluor (rabbit polyclonal; Invitrogen, A11094; 1µg/ml; 1 h; 20°C), washed 4x with 0.05% Tween 20 in PBS (1 min. each; 20°C), then labeled further by incubating these cells with Cy3-conjugated donkey anti-rabbit (Jackson ImmunoResearch, 711-165-152; 1/2000 dilution; 30 min; 20°C), washed 4x with 0.05% Tween 20 in PBS (5 min. each; 20°C) followed by a rinse with PBS at (5 min; 20°C). After DAPI/Vectashield counter-staining and mounting, images were collected using a wide-field microscope.

#	**Sequence (5′-3′)**	#	**Sequence (5′-3′)**
1	GCACCCGTCCTAAAATGAAA	25	CTCCAAGGTTCTGAGTTCAT
2	TGTTACAGAAACGCTACTCC	26	TTACCAGAGTAAAGGCACCA
3	AACAAGAAACGACAGACAGG	27	AATGACTACATGAGAGGCCT
4	ACTGGAGTCTTCATTGTGAG	28	AGTGTAGGTGAAGTCAGTCA
5	AAGGAAAGGCAGACAGACTT	29	CACAAAGGATTCCCAAGTAC
6	AACCTGGGAGTCTTTACACA	30	ATGTCAGGCCAGAAGAAGAA
7	GCCGACCTGTTTTCTATCTT	31	CACCCGGGAATACAATTGTT
8	CAAGGATGTCCACCTTTATC	32	TACACACACACACACACACA
9	TGTCTACAACTTCATCAGCC	33	ATGTTGTAGATGCATGCGTG
10	TTAGCTCTCCAGACAAGAGA	34	TTTGCTGTTCTTCCAGAGGA
11	ATCTTGCCCTCTAACTCCTT	35	GTTGGCTGAACAGTTAAGAG
12	AGAGACTGCCTGTCTTTCTA	36	TAAGTACGACAAGTCAGAGC
13	ACTTGCCATCATCCAACAGA	37	AATGCCTAGCTTTGGGGTAA
14	GTCCTGAGTTTCTGTGTATC	38	CGTCCAAAGCTAATTGACCT
15	GCATGCAGAAATGCATTTCC	39	GTCTTCCCTTTTGCAGATGA
16	CATTTCTGCCTCCAACATAC	40	TCTAACAGCAAGCCTTCTGT

17	ATAGGCCTCCTTGGTGTAAT	41	TAAGCTGTGTCTCTACTGTG
18	TCAGATCACAGTGTCTTTCC	42	CAGGATGCGAGTAATATAGG
19	TTTTCCGATTTCCCCTCTCT	43	GACCTAAAGCATCTTGAGCT
20	TCAGTGGTCTCCATTCATCA	44	CCATCGAAGCTCTTTTGAAG
21	TTATTGAGATCAGGTCCAGG	45	GGACAAGAATCCTACCAACT
22	TCACACTGCAAATTCCACAC	46	AAACCCAGGTTTTCCTGCAT
23	ATTCCCTACCTGTCTCAAAC	47	TTTGCCAAGCTTCATGTGGT
24	GACTTGAAGGACCTCAACTT	48	TATTCTGACTCTCCCTCTAC

Table 3.2. RNA FISH probes against rat *Cd2* intron-2 against the +ve strand.

3.12.8: Half-lives

Figures 3.4 and 3.20 include a simple exponential decay curve fitted to the data (using Prism) although this may be an averaged value for however many different kinetic populations there might coexist. The half-life ($t_{1/2}$) of tagged peptides can then be calculated using $t_{1/2} = 0.693/\lambda$ where λ is the first order rate constant of exponential decay. Half-lives were 33, 10, and 48 s (and 34, 10, and 32 s) respectively. R^2 values for all exponential decay curves shown were >0.98, except for those given by heavy Lys + Arg (which gave values of 0.92 and 0.96 for cytoplasm and nucleus, respectively).

3.12.9: Molecular cloning

Three vectors were used (Fig. 3.30), all encoding an SV40 *ori* (to permit replication in Cos7 cells)[74] and a modified P^{tight} 'Tet' promoter inducible with doxycycline. In the base vector, the P^{tight} promoter drives expression of a protein containing an N-terminal mitochondrial signal sequence (mt), the rat CD2 epitope recognized by the OX34 antibody,[51,52] and encoded by exons 2 and 3 of rat *Cd2*, a shortened intron 2 of *Cd2* (with intact splice donor and acceptor sites), a ribosome pause sequence (rps), a C-terminal EGFP, and a poly-adenylation signal. Vector 1 is as the base vector except that the EGFP sequence contains a deleted segment encoding the EGFP fluorochrome (to free the green channel for other use). Vector 2 is as the base vector except that a haemagglutinin (HA) tag is inserted within the intron (in-frame with exon 2 and upstream of the first intronic stop codon). Vector 3 is as Vector 2 but also bears a PTC 67 nucleotides upstream of the exon2-intron2 junction (a position that should induce NMD).[53]

The base vector was created as follows. pLdPDH[75] encodes the mitochondrial import sequence of human *PDH1α* inserted between the *Sac*I and *Pst*I sites in pEGFP-N1 (Clontech). The P^{tight} promoter from pTRE-Tight (Clontech) was inserted between the *Xho*I and *Eco*RI sites in pLdPDH. Two fragments of rat *Cd2* were amplified from total rat DNA (Bioline) using oligo pair 1 (which target chromosome 2, between positions 196,332,778-196,333,452)[76] and oligo pair 3 nested with oligo pair 2 (which target chromosome 2 between positions 196,336,082-196,338,119)[76]. These two *Cd2* fragments of 676 and 2,057 bp (with *Nhe*I- and *Xba*I-compatible cohesive ends, respectively) were simultaneously inserted between the *Pst*I and *Bam*HI sites in pLdPDH (which now encodes P^{tight}). The base vector was formed by removing from the doubly-modified pLdPDH the fragments containing P^{tight}, the mitochondrial import sequence, and *Cd2*, and inserting the three between the *Xho*I and *Bam*HI sites in plasmid 0,*EGFP*,pA (a promoter-less form of pEGFP-N1). Vector 1 was constructed by deleting the EGFP chromophore (5′-CTGACCTACGGCGTGCAG-3′; 18 bp) from the base vector by site-directed mutagenesis (QuikChange Lightning kit; Stratagene) using oligo pair 4. Vector 2 was constructed by inserting a sequence encoding the HA epitope (5′-TACCCTTACGACGTTCCTGATTACGCT-3′; 27 bp) in the base vector by site-directed mutagenesis using oligo pair 5. Unexpectedly, the insert proved to be 81 bp (i.e., 5′-TACCCTTACGACGTTCCTGATTACGCT-TAGCAAGTTGGGGGGACGGGTGCTATG-TACCCTTACGACGTTCCTGATTACGCT-3′) in which two HA encoding sequences were interrupted by a 27-bp sequence (note that the HA tag was not used for FISH experiments, so that this modification had no impact on the experiments discussed). Vector 3 was constructed (by site-directed mutagenesis using oligo pair 6) by introducing a point mutation (5′-TAT-3′ → 5′-TAA-3′) that creates an in-frame premature termination codon (PTC) in the *Cd2* exon in Vector 3. Correct construction of the vectors was confirmed by DNA sequencing after PCR with the following oligonucleotides: Presence of P^{tight} and *Cd2* fragments in the base vector was verified using oligos 7 and 8. Absence of the EGFP chromophore in Vector 1 was verified using oligo 9. Oligo 7 was also used to verify the presence of the HA-coding sequence in Vector 2 and the point mutation in Vector 3. Bacterial transformations were performed using One Shot TOP10 chemically-competent *Escherichia coli*. All PCRs (except for site-directed mutagenesis) were performed using PfuUltra II Fusion HS DNA Polymerase (Stratagene) at a re-annealing temperature of 58.6˚C (64˚C in case of oligo pair 3). All oligonucleotide DNA primers used (supplied by Sigma) are listed in **Table 3.3**:

#	Name	Sequence (5′-3′)
1	Oligo pair 1	Fwd: GCAGA(CTGCAG)[b]AGAGACAGTGG Rev: CAGCATCTCCAGCCTTCTTC
2	Oligo pair 2	Fwd: ATCCACAGAAAGGTGCTCGT Rev: TA(GGATCC)[c]AT(TGTTCGTCGTCG)[d]CTTTGAGACCATCTCTAAAAG
3	Oligo pair 3	Fwd: AC(TCTAGA)[e]AAGACAGGCAGTCTCTG Rev: TA(GGATCC)[c]AT(TGTTCGTCGTCG)[d]CTTTGAGACCATCTCTAAAAG
4	Oligo pair 4[a]	Fwd: CACCCTCGTGACCACCTGCTTCAGCCGCTAC Rev: GTAGCGGCTGAAGCAGGTGGTCACGAGGGTG
5	Oligo pair 5[a]	Fwd: GGACGGGTGCTATG(TACCCTTACGACGTTCCTGATTACGCT)[f]TAGCAAGTTGGG Rev: CCCCAACTTGCTA(AGCGTAATCAGGAACGTCGTAAGGGTA)[f]CATAGCACCCGTCC
6	Oligo pair 6[a]	Fwd: GAGATGACAGTGGCACC(TAA)[g]AATGTAACGGTATACAGCAC Rev: GTGCTGTATACCGTTACATT(TTA)[g]GGTGCCACTGTCATCTC
7	Oligo 7	GCCATCATCCAACAGAGCAG
8	Oligo 8	CGTCGCCGTCCAGCTCGACCAG
9	Oligo 9	TGCCCTTCAGCTCGATGCGG

Table 3.3: Oligonucleotide DNA primers.

[a]: complimentary oligonucleotide primers for site-directed mutagenesis, HPLC purified

[b]: *PstI* recognition sequence

[c]: *BamHI* recognition sequence

[d]: ribosome pause sequence (rps)

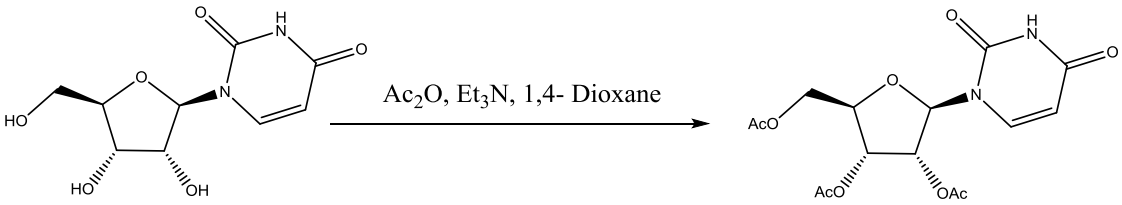
[e]: *XbaI* recognition sequence

[f]: HA-epitope encoding sequence

[g]: premature termination codon (PTC)

3.12.10: Synthesis of 5-Ethynyl Uridine (EU)

O-Acetylation of Uridine: The reaction was carried out as previously reported.[77]

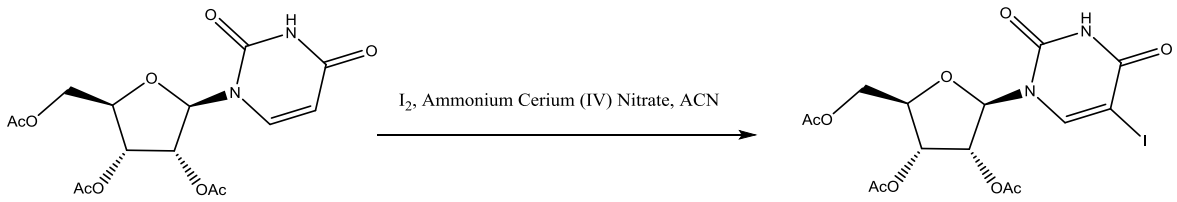


3.2

19.2 mL of Acetic Anhydride and 19.2 mL of Triethylamine were added to 6 g of Uridine in 123 mL of dry 1,4-Dioxane. The reaction was continued at room temperature for 36 h. The

solvent was evaporated in vacuo, and the residue purified by column chromatography (Hexane: EtOAc:: 1: 2). Yield was quantitative, giving 9.08 g of product. IR (ν_{max}, film): 3184, 3048, 2961, 2883, 2573, 2308, 1742, 1717, 1679, 1467, 1438, 1422, 1382, 1371, 1278, 1247, 1216, 1153, 1103, 1080, 1047, 1009, 990, 960, 941, 902, 887, 857, 838. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.01, 2.07, 2.11 (3s, 3H 2'-OAc, 3H 3'-OAc and 3H 5'-OAc), 5.29-5.34 (m, 2H, 2'-H and 3'-H), 5.77 (d, *J* = 8.1, 1H, 1'-H), 6.02 (d, J = 4.8 Hz, 1H, 5-H), 7.39 (d, J = 8.1 Hz, 1H, 6-H), 9.57 (br s, 3-NH). ^{13}C NMR (400 MHz, $CDCl_3$) δ: 20.4, 20.6, 20.7 (2'-$OCOCH_3$, 3'-$OCOCH_3$, and 5'-$OCOCH_3$), 63.1 (5'-C), 72.7, 77.0 (3'-C, 2'-C), 79.8 (4'-C), 87.4 (1'-C), 103.7 (5-C), 139.4 (6-C), 150.3 (2-C), 163.1 (4-C), 169.6, 170.1, 171.1 (2'-$OCOCH_3$, 3'-$OCOCH_3$, and 5'-$OCOCH_3$). LRMS (ESI) Calculated: [M+H] $C_{15}H_{18}N_2O_9$ (m/z): 370.31, Obtained: $[M-H]^-$: 369.11, $[2M-H]^-$: 739.21

Iodination of 2', 3', 5'-*O*-Acetyl Uridine :The reaction was carried out as previously reported.[77]



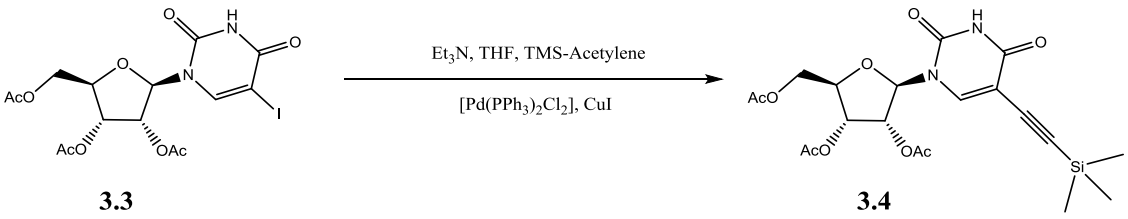
3.2 **3.3**

A mixture of 9g of the nucleoside, 3.7 g Iodine and 6.54 g Ammonium Cerium (IV) Nitrate in 386 mL MeCN was stirred at 80 ºC for 1 h. Reaction progress was monitored by TLC (Hexane: EtOAc :: 1:1). The solvent was evaporated in vacuo, and the residue was partitioned between a mixture of cold EtOAc, brine and 5% aq. $NaHCO_3$ (240, 200 and 250 mL respectively, half of which was processed at a time). The residue was extracted twice with EtOAc, and the combined organic layer was washed with cold 5% aq. $NaHCO_3$ (125 mL), then with brine (150 mL), then with water (150 mL). The organic fraction was then dried, filtered, and evaporated. The residue was purified by column chromatography (Hexane: EtOAc :: 1:1) to give the product in 86% yield.

^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.11, 2.13, 2.25 (3s, 3H 2'-OAc, 3H 3'-OAc and 3H 5'-OAc), 4.32-4.42 (m, 2H, 5'-H), 5.30-5.35 (m, 2H, 2'-H and 3'-H), 6.07 (d, J = 4.9 Hz, 1H, 1'-H), 7.9 (s, 1H, 6-H), 8.89 (br s, 3-NH). ^{13}C NMR (400 MHz, $CDCl_3$) δ: 20.5, 21.0, 21.1 (2'-$OCOCH_3$, 3'-$OCOCH_3$, and 5'-$OCOCH_3$), 63.0 (5'-C), 69.5 (5-C) 76.6, 77.0 (3'-C, 2'-C), 80.3 (4'-C), 87.1 (1'-C), 143.6 (6-C), 149.7 (2-C), 159.3 (4-C), 169.5, 169.6, 170.0 (2'-$OCOCH_3$, 3'-

$OCOCH_3$, and 5'-$OCOCH_3$). LRMS (ESI) Calculated: [M-H] $C_{15}H_{17}IN_2O_9$ (m/z): 495.00, Obtained: $[M-H]^-$: 495.01, $[2M-H]^-$: 991.01

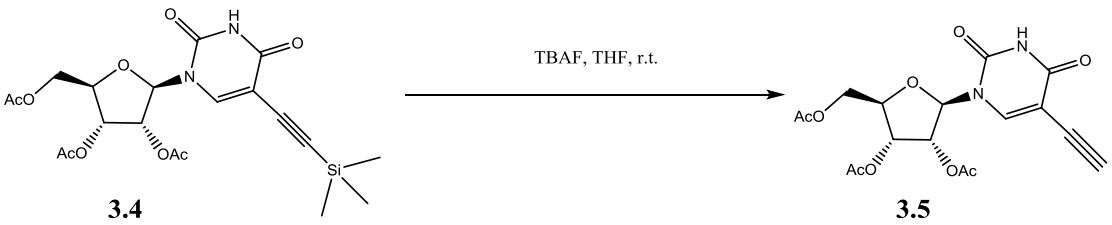
Sonogashira Coupling of TMS-Acetylene onto Iodouridine ring: The reaction was carried out as previously reported.[77]



3.3 **3.4**

4.4 mL of TMS-Acetylene was added to 10g of 5-Iodouridine in 210 mL anhydrous Et_3N and 70 mL THF together with 1.4 g of $[Pd(PPh_3)_2Cl_2]$ and 390 mg CuI. The mixture was stirred for at 45-50 ºC under Argon until TLC (Chloroform/MeOH : 30/1) showed complete conversion (3-4 h). The product mixture was filtered and washed with Et_3N, and the combined filtrate evaporated. The residue was purified by column chromatography (200:1 DCM: MeOH, then 100:1 DCM:MeOH). IR (ν_{max}, film): 3060, 2959, 2899, 2828, 2329, 2161, 1984, 1747, 1718, 1698, 1625, 1592, 1534, 1484, 1453, 1438, 1370, 1223, 1179, 1118, 1094, 1049, 996, 918, 889, 843. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 0.22 (s, 9H, TMS), 2.11, 2.13, 2.22 (3s, 3H 2'-OAc, 3H 3'-OAc and 3H 5'-OAc), 4.34-4.39 (m, 2H, 5'-H), 5.28-5.48 (m, 2H, 2'-H and 3'-H), 6.11 (d, J = 4.8 Hz, 1H, 1'-H), 7.77 (s, 1H, 6-H), 8.35 (br s, 3-NH), LRMS (ESI) Calculated: [M-H] $C_{20}H_{26}SiN_2O_9$ (m/z): 465.14, Obtained: $[M-H]^-$: 465.15, $[2M-H]^-$: 931.30

TBAF deprotection of TMS group: The reaction was carried out as previously reported.[77]

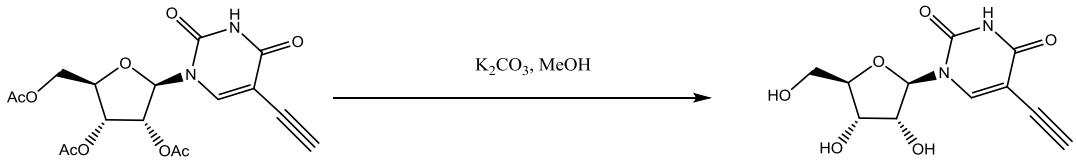


3.4 **3.5**

To a stirred solution of 2.02 g of the previous product was added 1.9 mL TBAF (1M in THF) at room temperature. TLC was regularly performed to monitor the reaction (DCM:MeOH 200:1), which was found to be incomplete after 5 hours. A fresh lot of TBAF was added and the reaction was continued overnight. The contents were filtered through a pad of silica to remove TBAF and the reaction mixture was concentrated in vacuo. The crude product was obtained in 90% yield and was used immediately for the next step.

IR (ν_{max}, film): 2963, 2878, 2360, 2341, 2112, 1746, 1717, 1646, 1486, 1460, 1439, 1379, 1344, 1235, 1182, 1119, 1056, 1033, 989, 919, 888, 844. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.01, 2.02, 2.09 (3s, 3H 2'-OAc, 3H 3'-OAc and 3H 5'-OAc), 3.19 (s, 1H, C≡CH), 4.22-4.28 (m, 2H, 5'-H), 5.22-5.31 (m, 2H, 2'-H and 3'-H), 5.93 (d, J = 4.5 Hz, 1H, 1'-H), 7.7 (s, 1H, 6-H), 8.55 (br s, 3-NH). ^{13}C NMR (400 MHz, $CDCl_3$) δ: 19.5, 20.2, 20.3 (2'-$OCOCH_3$, 3'-$OCOCH_3$, and 5'-$OCOCH_3$), 61.1 (5'-C), 77.1, 77.3, 77.4 (3'-C, 2'-C, 4'-C), 98.0 (1'-C). LRMS (ESI) Calculated: [M-H] $C_{17}H_{28}N_2O_9$ (m/z): 394.10, Obtained: $[M-H]^-$: 393.12

Acetyl deprotection: The reaction was adapted from an existing protocol.[78]



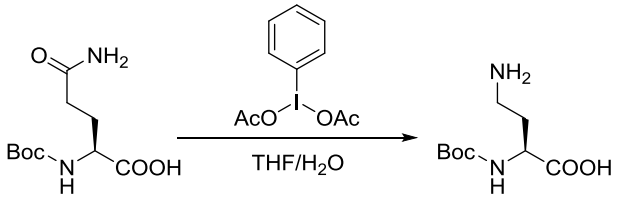
3.5 **3.1**

A solution of 2', 3', 5'O-Acetyl-5-ethynyluridine in Methanol was stirred at room temperature with 400 mg K2CO3 for 20 h. The mixture was neutralised with DOWEX-50 resin (H^+ form), and the filterate was evaporated. The residue was purified by column chromatography (DCM: MeOH 6:1). The product was obtained in 80% yield.

IR (ν_{max}, film): 3060, 2963, 2936, 2876, 2360, 2342, 2114, 1683, 1625, 1456, 1386, 1351, 1276, 1104, 1079, 1053, 988, 944, 912, 868. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 3.23 (s, 1H, C≡CH), 3.41-3.61 (m, 4H, 2'-OH, 3'-OH, 4'OH, 5'-OH), 3.88-4.28 (m, 3H, 2'-H and 3'-H and 5'-H), 5.81 (d, J = 4.9 Hz, 1H, 1'-H), 8.58 (s, 1H, 6-H). LRMS (ESI) Calculated: [M-H] $C_{11}H_{12}N_2O_6$ (m/z): 268.07, Obtained: $[M-H]^-$: 267.07, $[2M+Na]^+$: 559.60

3.12.11: Synthesis of Azidohomoalanine

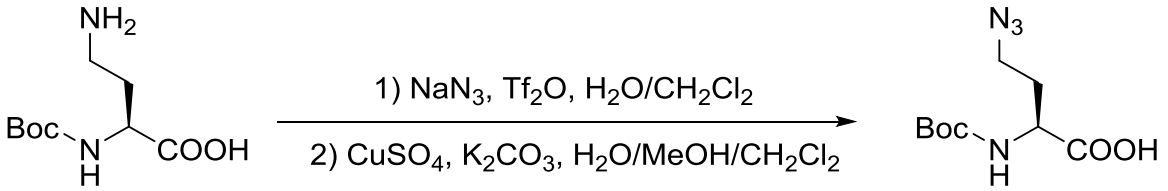
Hofmann degradation of *N*-tert-Butoxycarbonyl-(*S*)-glutamine was performed as previously reported.[79]



3.6

N-tert-butoxycarbonyl-(*S*)-glutamine (5.0 g, 20.3 mmol) was dissolved in THF (48 mL) and water (12 mL). The solution was cooled in an ice bath, and (diacetoxyiodo)benzene (7.8 g, 24.2 mmol) was added. The reaction was monitored by TLC (4:1:1 n-butanol: acetic acid: water). After 6 hours, the organic portion was evaporated *in vacuo*. The aqueous portion was diluted with water, and extracted with ethyl acetate (3 x 30 mL). The aqueous portion was then partially evaporated, frozen, and lyophilised. The product *N*-tert-butoxycarbonyl-(S)-2,4-diaminobutanoic acid was obtained as an orange-yellow solid (3.1 g, 71%). IR (ν_{max}, film): 3416, 2977, 2931, 1698, 1589, 1561, 1529, 1499, 1439, 1392, 1366, 1252, 1161, 1052, 1029, 949, 907, 868, 802. ^{1}H NMR (400 MHz, D_2O) δ: 3.86 (m, 1H, C*H*), 2.94 (t, *J* = 7.8 Hz, 2H, NH_2CH_2), 1.9 (m, 1H, CH_2), 1.85 (m, 1H, CH_2), 1.31 (s, 9H, $C(CH_3)_3$). ^{13}C (100 MHz, D_2O) δ: 28.3 ($(\underline{C}H_3)_3$), 29.9 (Cβ), 39.1 ($\underline{C}H_2N_3$), 57.1 (Cα), 80.9 ($(CH_3)_3\underline{C}CONH$), 157.9 ($(CH_3)_3C\underline{C}ONH$), 175.6 (COOH), LRMS (ESI) Calculated: [M-H] $C_9H_{18}N_2O_4$ (m/z): 218, Obtained: $[M-H]^-$: 217.12, $[2M-H]^-$: 435.26, $[M+H]^+$: 219.16, Melting pt. = 196 ºC, Lit. [79]190-191 ºC, $[\alpha]^{20}_D$ = -8.6º (c = 1.0, H_2O), Lit. [51] $[\alpha]^{20}_D$ = -9.6 (c = 1.0, H_2O).

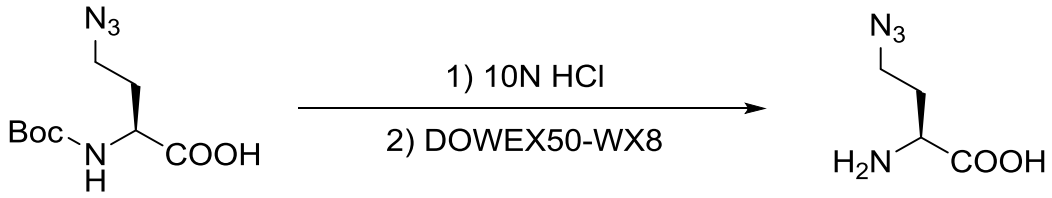
Triflyl azide was prepared in situ, and the diazo transfer reaction was adapted from literature.[80]



3.6 **3.7**

Sodium azide (9.8 g, 150 mmol) was dissolved in a mixture of water (22 mL) and dichloromethane (39 mL). The contents were cooled in an ice bath, and triflic anhydride (5.1 mL) was added via syringe over 5 min. The contents were allowed to stir for 30 min, after which the layers were separated. The aqueous layer was extracted using dichloromethane (2 x 50 mL). The organic fractions were pooled, and washed twice with 5% aq. sodium carbonate (30 mL). The triflyl azide containing organic layer was then added to a solution of *N*-tert-butoxycarbonyl-(S)-2,4-diaminobutanoic acid (3.1 g, 14.2 mmol) in water (39 mL) and methanol (90 mL). To this, potassium carbonate (3.2 g) and copper sulphate pentahydrate (38.4 mg) were added. Stirring was continued for 18 h, after which the organic fraction was evaporated *in vacuo*. The aqueous portion was extracted with dichloromethane (2 x 50 mL), diluted with water, then acidified to pH 2.0 using 1.0 N hydrochloric acid. The acidified aqueous phase was then extracted with dichloromethane (3 x 50 mL). The organic phase was then washed with brine (2 x 80 mL), dried with magnesium sulphate, filtered, and evaporated. The product was obtained as orange oil (3.2 g, 92%). IR (ν_{max}, film): 3319, 2980, 2935, 2100, 1690, 1513, 1455, 1399, 1369, 1236, 1189, 1154, 1093, 1057, 1029, 940, 853. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 5.26 (br d, *J* = 8.8 Hz, 1H, N<u>H</u>), 4.41 (m, 1H, C<u>H</u>), 3.46 (t, *J* = 6.7 Hz, 2H, $N_3C\underline{H}_2$), 2.11 (m, 1H, C<u>H</u>$_2$), 1.98 (m, 1H, C<u>H</u>$_2$), 1.46 (s, 9H, $C(C\underline{H}_3)_3$). ^{13}C (100 MHz, D_2O) δ: 28.2 ($(\underline{C}H_3)_3$), 31.4 (<u>C</u>β), 47.8 ($\underline{C}H_2N_3$), 51.3 (<u>C</u>α), 80.7 ($(CH_3)_3\underline{C}CONH$), 155.7 ($(CH_3)_3C\underline{C}ONH$), 175.8 (<u>C</u>OOH), LRMS (ESI) Calculated: [M-H] $C_9H_{16}N_4O_4$ (m/z) : 243, Obtained: $[M-H]^-$: 243.08, $[2M-H]^-$: 487.16, $[\alpha]^{20}_D$ = +16.1º (c = 1.0, $CHCl_3$), Lit.[81] $[\alpha]^{20}_D$ = +19.0º (c = 1.0, $CHCl_3$).

The Boc deprotection reaction was performed as previously reported.[82]



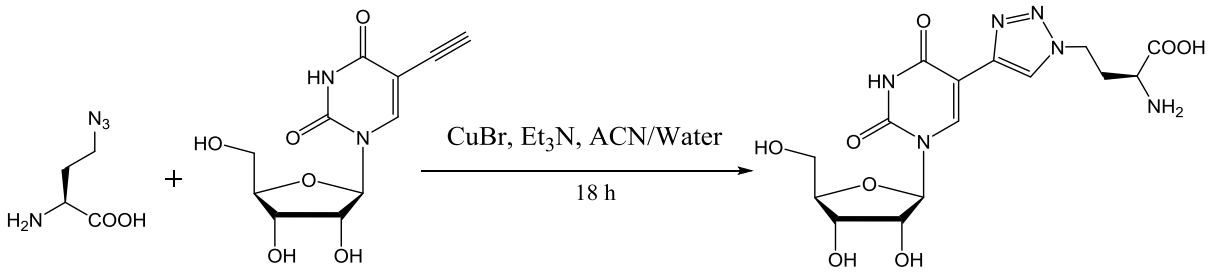
3.7 **1.3**

N-tert-butoxycarbonyl-(*S*)-azidohomoalanine (2.9 g, 11.88 mmol) was dissolved in 10 N HCl (30 mL), and the contents stirred at room temperature for 1 hour. The solution was then diluted with 170 mL water. 80 g of DOWEX50-WX8 resin was packed into a column, and conditioned by washing with 1N aq. ammonia (3 x 80 mL), water (600 mL), 1N HCl (250 mL), and water again until the eluate was at pH 7.0. The peptide solution was run through the column; the flow-through was collected and passed over the resin again. Aha was eluted from the column by washing with 500 mL of 1N aq. ammonia, and recovered as an off-white solid on evaporation. The compound was recrystallised from acetone/water to yield white crystals in 91% yield (1.54 g). IR (ν_{max}, film): 3368, 3036, 2978, 2926, 2103, 1677, 1589, 1489, 1351, 1333, 1265, 1204, 1101, 854, 771. ^{1}H NMR (400 MHz, D_2O) δ: 3.99 (t, *J* = 6.32, 1H, Cα<u>H</u>), 3.48 (T, *J* = 6.4, 2H, C<u>H</u>$_2$N$_3$), 2.07 (m, 2H, Cβ<u>H</u>$_2$). ^{13}C (100 MHz, D_2O) δ: 29.6 (Cβ), 47.6 (<u>C</u>H$_2$N$_3$), 57.8 (Cα), 171.9 (<u>C</u>OOH). LRMS (ESI) Calculated: [M+H] $C_4H_8N_4O_2$ (m/z): 144, Obtained: $[M+H]^+$: 145.08, $[M+HCl-H]^-$: 179.06, $[2M-H]^-$: 287.11, Melting pt. 180.6 ˚C, $[\alpha]^{20}_D$ = +6.60º (c = 0.50, H_2O), Lit.[81] $[\alpha]^{20}_D$ = +6.60º (c = 0.50, H_2O).

The tris-triazole ligand, triethyl 2,2',2''-(4,4',4''-nitrilotris(methylene)tris(1H-1,2,3-triazole-4,1-diyl))triacetate, was prepared following minor modification of reported literature.[83] See Chapter 4 for full details.

3.12.12: Aha - EU CuAAC

The reaction was adapted from an existing protocol.[81]

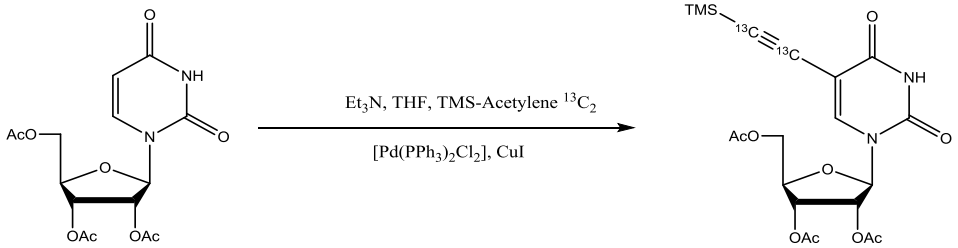


1.3 Aha **3.1** EU **3.8** 52%

27 mg Aha and 50 mg EU were dissolved in 6.2 mL of 50% acetonitrile-water. Separately, 8 mg of Cu(I)Br was dissolved in 800 µL of acetonitrile. Both solutions were placed in a stream of Argon for 20 min. The solutions were then mixed under argon, and 45 µL of Et_3N was added. The reaction was stirred for 18 hours. The product was isolated via column chromatography (4:1 EtOAc: MeOH) in 52% yield.

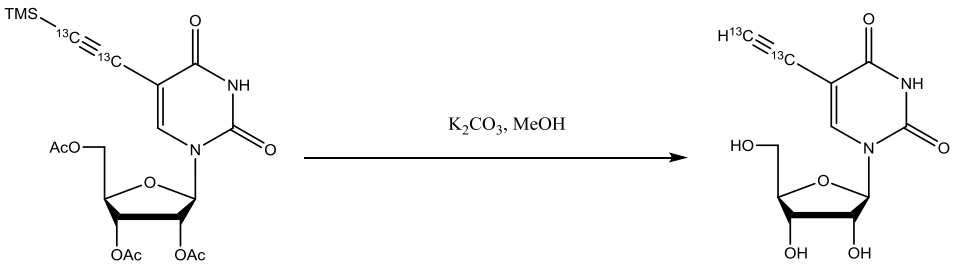
IR (ν_{max}, film): 3247, 3066, 2938, 2578, 2116, 1964, 1695, 1592, 1458, 1383, 1261, 1234, 1191, 1154, 1074, 984, 950, 853. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.57 (m, 2H, Cβ<u>H</u>), 3.52-3.57 (m, 4H, Cα<u>H</u>, 2'-OH, 3'-OH, 5'-H**)**, 3.65-3.69 (m, 4H, Cγ<u>H</u>$_2$-triazole, 5'-OH, 5'-H**)**, 4.22 (m, 1H, 2'-H), 4.50-4.53 (m, 2H, 3'-H, 4'-H), 4.96 (m, 1H, 1'-H), 6.56 (m, 1H, 1'-H), 7.92 (s, 1H, Triazole-<u>H</u>), 11.12 (s, 1H, ring-N<u>H</u>), 11.35 (s, 1H, COO<u>H</u>). ^{13}C NMR (125 MHz, D_2O) δ: 30.20 (Cβ), 33.71 (Cγ), 58.57 (Cα), 60.26 (5'C), 69.35 (3'C), 73.96 (2'C), 84.14 (4'C), 98.41 (1'C), 102.22 (C5, Uridine), 141.79 (C5, Triazole), 145.4 (C4, Triazole), 150.65 (C4, Uridine) 151. 65 (C2, Uridine), 166.23 (C4 Uridine), 177.37 (COOH). LRMS (ESI) Calculated: [M] $C_{15}H_{20}N_6O_8$ (m/z): 412.35, Obtained: $[M+1]^+$: 413.29

3.12.13: Synthesis of $^{13}C_2$-EU



3.2 **3.9**

TMS-Acetylene-$^{13}C_2$ (100 µL, 0.756 mmol) was added to O-acetyl-5-iodouridine (225 mg, 0.452 mmol) in 4.5 mL dry Et_3N and 2 mL dry THF together with $[Pd(PPh_3)_2Cl_2]$ (31.6 mg, 0.044 mmol) and CuI (8.8 mg, 0.044 mmol). The mixture was stirred at 45-50 ºC under Argon for 4 h, and monitored by TLC (30:1 :: DCM:MeOH) The product mixture was filtered and the residue washed thrice with Et_3N, and the combined filtrate evaporated. The residue was passed through a silica column (200:1 DCM: MeOH, then 100:1 DCM:MeOH). The product was obtained as a thick brown oil (210 mg, 98%). LRMS (ESI) Calculated [M] $C_{18}{}^{13}C_2H_{26}N_2O_9Si$ (m/z): 468.14, Obtained: $[M+Na]^+$: 491.13, $[2M+Na]^+$: 959.29, $[M-H]^-$: 467.14, $[2M-H]^-$: 935.27. HRMS (ESI) Calculated [M+Na] $C_{18}{}^{13}C_2H_{26}N_2NaO_9Si$ (m/z): 491.1367, Obtained $[M+Na]^+$: 491.1365. ^1H NMR (400 MHz, $CDCl_3$) δ: δ: 0.22 (s, 9H, TMS), 2.11, 2.13, 2.22 (3s, 3H 2'-OAc, 3H 3'- OAc and 3H 5'-OAc), 4.34-4.39 (m, 2H, 5'-H), 5.28-5.48 (m, 2H, 2'-H and 3'-H), 6.11 (d, J = 4.8 Hz, 1H, 1'-H), 7.77 (s, 1H, 6-H), 8.35 (br s, 3-NH), ^{13}C NMR (400 MHz, $CDCl_3$) δ: -0.20 ($(CH_3)_3$-Si), 20.38, 20.49, 20.93 (2'-OCO<u>C</u>H_3, 3'-OCO<u>C</u>H_3, and 5'-OCO<u>C</u>H_3), 70.06 (3'-C), 73.17 (2'-C), 87.11 (4'-C), 90.59 (1'-C), 94.74 (^{13}C-Uridine, J_{C-C} = 140.7 Hz), 100.21 (TMS-^{13}C, J_{C-C} = 140.7 Hz), 138.07 (C6), 149.05 (C2), 160.58 (C4), 169.63 (2'-O<u>C</u>OCH_3, 3'-O<u>C</u>OCH_3, and 5'-O<u>C</u>OCH_3), IR (νmax, film): 3084.45, 2963.12, 1745.21, 1723.13, 1544.87, 1496.98, 1453.36, 1369.87, 1219.58, 1158.79, 1046.82, 912.86, 884.73, 844.83, 757.05



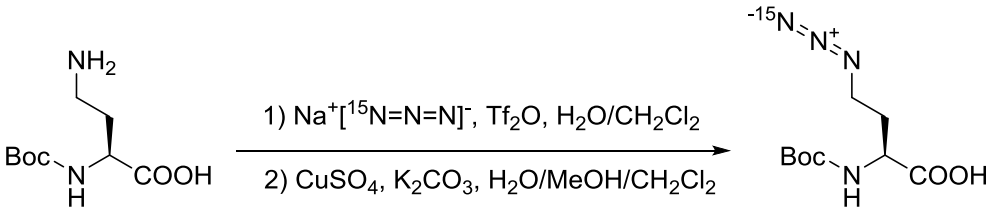
3.9 **3.10**

A solution of 2', 3', 5'O-Acetyl-5-TMS-ethynyluridine (100mg, 0.21 mmol) in Methanol (9 mL) was stirred at room temperature with K_2CO_3 (120 mg, 0.8 mmol) for 20 h. The mixture was neutralised with DOWEX-50 resin (H^+ form), and the filterate was evaporated. The residue was purified by column chromatography (DCM: MeOH 6:1). The product was obtained as a pale yellow solid (49 mg, 85%).^1H NMR (400 MHz, CDCl3) δ: 3.23 (s, 1H, C≡CH), 3.41-3.61 (m, 4H, 2'-OH, 3'-OH, 4'OH, 5'-OH), 3.88-4.28 (m, 3H, 2'-H and 3'-H and 5'-H**),** 5.81 (d, J = 4.9 Hz, 1H, 1'-H), 8.58 (s, 1H, 6-H), ^{13}C NMR (400 MHz, $CDCl_3$) δ: 68.19

(3'-C), 69.26 (2'-C), 60.60 (5'-C), 83.40 (4'-C), 90.14 (1'-C), 74.48 (^{13}C-H, J_{C-C} = 179.3 Hz), 83.46 (^{13}C-Uridine, J_{C-C} = 179.3 Hz), 130.11 (C6), LRMS (ESI) Calculated: [M] $C_9{}^{13}C_2H_{12}N_2O_6$ (m/z): 270.07, Obtained: $[M+Na]^+$: 293.06, $[2M+Na]^+$: 563.14, $[M-H]^-$: 269.06, $[2M-H]^-$: 539.13. HRMS (ESI) Calculated: $[M-H]^-$ $C_9{}^{13}C_2H_{11}N_2O_6$ (m/z): 269.0690, Obtained $[M-H]^-$: 269.0695, IR (νmax, film): 3004.01, 2970.31, 2510.09, 1738.39, 1552.70, 1426.47, 1366.07, 1228.89, 1216.77, 1110.11, 701.90, 655.46

3.12.14: Synthesis of ^{15}N-Aha

The 15N labelled Aha synthesis was carried out as in section **3.12.11**, with the diazo transfer step being carried out as below:

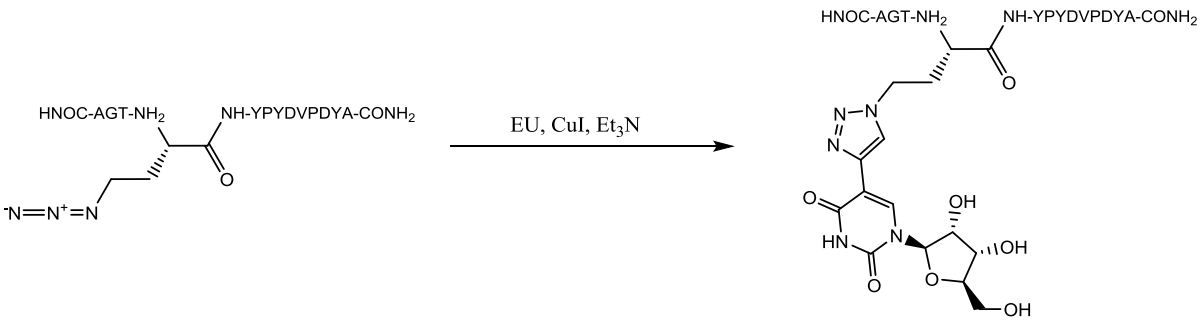


3.6 **3.11**

^{15}N-labelled Sodium Azide (500 mg, 7.57 mmol) was dissolved in a mixture of water (1.1 mL) and dichloromethane (2 mL). The contents were cooled in an ice bath, and Triflic anhydride (260 μL) was added via syringe over 5 minutes. The contents were allowed to stir for 30 min, after which the layers were separated. The aqueous layer was extracted using dichloromethane. The organic fractions were pooled, and washed twice with 5% aq. sodium carbonate. The triflyl azide containing organic layer was then added to a solution of N-tert-Butoxycarbonyl-(S)-2,4-diaminobutanoic acid (218 mg, 1 mmol) in water (2 mL) and Methanol (5 mL). To this, potassium carbonate (260 mg) and copper sulphate pentahydrate (3 mg) were added. Stirring was continued for 18 h, after which the organic fraction was evaporated in vacuo. The aqueous portion was extracted with dichloromethane (2 x 5 mL), diluted with water, and then acidified to pH 2.0 using 1.0 N hydrochloric acid. The acidified aqueous phase was then extracted with dichloromethane (3 x 20 mL). The organic phase was then washed with brine, dried with Magnesium sulphate, filtered, and evaporated. The product was obtained as orange oil, which was found to contain the desired product and the unlabelled product in equal amounts (190 mg, 78%). 1H NMR (400 MHz, CDCl3) δ: 5.26 (br d, J = 8.8 Hz, 1H, NH), 4.41 (m, 1H, CH), 3.46 (t, J = 6.7 Hz, 2H, N_3CH_2), 2.11 (m, 1H,

CH_2), 1.98 (m, 1H, CH_2), 1.46 (s, 9H, $C(CH_3)_3$).Calculated: [M-H] $C_9H_{16}N_4O_4$ (m/z) : 244 labelled, 243 unlabelled, Obtained: HRMS [M-H]- : 244.1071 and 243.1093

3.12.15: CuAAC on Aha-labelled Haemagglutinin peptide



3.12 **3.13**

5 mg of Haemagglutinin peptide and 1 mg EU were dissolved in 160 µL of 50% acetonitrile-water. Separately, 1 mg of Cu(I)Br was dissolved in 100 µL of acetonitrile. Both solutions were placed in a stream of Argon for 20 min. 12.3 µL of the copper solution was added to the former and mixed under argon. 2 µL of Et_3N was added, and the reaction was stirred for 18 hours at room temperature. Progress of the reaction was monitored by ESI-mass spectrometry. The final product was obtained by HPLC purification of the reaction mixture. LRMS (ESI) Calculated: $C_{77}H_{101}N_{19}O_{27}$ (m/z): 1723.71, Obtained: $[M+H]^+$:1724.65, $[M-H]^-$: 1722.59, $[M-2H]^{2-}$: 860.79, HRMS (ESI) Calculated (m/z): $C_{77}H_{101}N_{19}O_{27}Na$: 1746.7007, Obtained: 1746.7026

SPPS Synthesiser parameters:

Sequence of incorporated haemagglutinin peptide: H_2N-TGA-Aha-YPYDVPDYA-$CONH_2$

The peptide was synthesised on a Rink amide MBHA resin with 0.2 mmol/g loading, on a 0.25 mmol scale. The coupling reaction was carried out with HBTU/HOBt (15 min). All deprotection steps were carried out with 20% piperidine in DMF.

Protocol for manual cleavage:

Cleavage mixture used: 95:2.5:2.5 :: TFA : TIS : Water

Volume of cleavage mixture (in mL) = 10x mass of resin (in g).

1. The resin was filtered and washed with DMF (3 x 10 mL), followed by DCM (3 x 10 mL), then methanol (3 x 10 mL).

2. The resin was dried for 1 hour in a stream of Argon, then stirred with 12.5 mL of the cleavage mixture (11.8 mL TFA, 313 µL TIS. 313 µL water) for 3 hours.

3. The resin was filtered and washed with TFA. The filterate was evaporated in vacuo to 1/3 original volume (~ 4 mL).

4. A 10-fold excess of ice-cold diethyl ether was added to this solution, causing the peptide to precipitate. The suspension was left at 4 ˚C overnight to allow for complete precipitation.

5. The peptide was spun down at 3750 RPM at 4 ˚C for 5 min, and the ethereal layer was decanted off. The pellet was washed twice more with ice cold diethyl ether, followed by spinning down.

6. The pellet was redissolved in a minimum amount of water, snap-frozen in liq. Nitrogen, and lyophilised to yield the peptide as a white fluffy powder.

Protocol for reverse-phase purification of Aha-haemagglutinin-peptide:

Columns used: Hydro R.P. Synergi 4µ 80A (analytical) flow rate 1mL/min

Hydro R.P. Synergi 4µ 80A (prep) flow rate 10 mL/min

Buffers used: Buffer A: 0.1% HCOOH in MilliQ water

Buffer B: 0.1% HCOOH in CH_3CN

Gradient used: Start at 5% B from 0 to 5 min

5% to 100%B from 5 to 45 min

Equilibration at 5% B from 45 to 50 min

All buffers were filtered and purged of air before use. The crude peptide was dissolved in water and injected into the column (20 µL for analytical, 200 µL for prep). The column eluate was analysed using a UV detector at 220 and 280 nm. Peaks were collected manually, and

analysed by ESI mass spectrometry between runs. Fractions containing the peptide of interest were pooled, frozen in liq. Nitrogen and lyophilised to yield the pure peptide as a white fluffy powder.

3.13: References

(1) Baboo, S.; Bhushan, B.; Jiang, H.; Grovenor, C. R. M.; Pierre, P.; Davis, B. G.; Cook, P. R. *PLoS ONE* **2014**, *9*, e99346.
(2) Kressler, D.; Bange, G.; Ogawa, Y.; Stjepanovic, G.; Bradatsch, B.; Pratte, D.; Amlacher, S.; Strauß, D.; Yoneda, Y.; Katahira, J.; Sinning, I.; Hurt, E. *Science* **2012**, *338*, 666.
(3) Dostie, J.; Lejbkowicz, F.; Sonenberg, N. *The Journal of Cell Biology* **2000**, *148*, 239.
(4) Sarkar, S.; Azad, A. K.; Hopper, A. K. *Proceedings of the National Academy of Sciences* **1999**, *96*, 14366.
(5) Lund, E.; Dahlberg, J. E. *Science* **1998**, *282*, 2082.
(6) Dahlberg, J. E.; Lund, E.; Goodwin, E. B. *RNA* **2003**, *9*, 1.
(7) Iborra, F. J.; Jackson, D. A.; Cook, P. R. *Science* **2001**, *293*, 1139.
(8) Nathanson, L.; Xia, T.; Deutscher, M. P. *RNA* **2003**, *9*, 9.
(9) Fasken, M. B.; Corbett, A. H. **2005**, *12*, 482.
(10) Behm-Ansmant, I.; Izaurralde, E. *Genes & Development* **2006**, *20*, 391.
(11) Cheng, J.; Maquat, L. E. *Molecular and Cellular Biology* **1993**, *13*, 1892.
(12) Zhang, J.; Sun, X.; Qian, Y.; LaDuca, J. P.; Maquat, L. E. *Molecular and Cellular Biology* **1998**, *18*, 5272.
(13) Carter, M. S.; Li, S.; Wilkinson, M. F. *EMBO J* **1996**, *15*, 5965.
(14) Maquat, L. E. *Journal of Cell Science* **2005**, *118*, 1773.
(15) Dieterich, D. C.; Hodas, J. J.; Gouzer, G.; Shadrin, I. Y.; Ngo, J. T.; Triller, A.; Tirrell, D. A.; Schuman, E. M. *Nat Neurosci* **2010**, *13*, 897.
(16) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 19.
(17) Best, M. D. *Biochemistry* **2009**, *48*, 6571.
(18) Hansen, J. L.; Moore, P. B.; Steitz, T. A. *J Mol Biol* **2003**, *330*, 1061.
(19) Nielsen, P. J.; McConkey, E. H. *J Cell Physiol* **1980**, *104*, 269.
(20) Han, Y. H.; Moon, H. J.; You, B. R.; Park, W. H. *Oncol Rep* **2009**, *22*, 215.
(21) Boisvert, F. M.; Ahmad, Y.; Gierliński, M.; Charrière, F.; Lamont, D.; Scott, M.; Barton, G.; Lamond, A. I. *Mol Cell Proteomics* **2012**, *11*, M111.011429.
(22) Doherty, M. K.; Hammond, D. E.; Clague, M. J.; Gaskell, S. J.; Beynon, R. J. *J Proteome Res* **2009**, *8*, 104.
(23) Jao, C. Y.; Salic, A. *Proc Natl Acad Sci U S A* **2008**, *105*, 15779.
(24) Stryer, L. B., J. M.; Tymoczko, J. L. *Biochemistry (5th ed.), New York: W. H. Freeman, ISBN 0716746840* **2002**.
(25) Bono, V. H.; Weissman, S. M.; Frei, E. *The Journal of Clinical Investigation* **1964**, *43*, 1486.
(26) Srivatsan, S. G.; Tor, Y. *Tetrahedron* **2007**, *63*, 3601.
(27) Zhang, X.-C.; Xu, L.; Liu, W.-G.; Liu, B. *Bull. Korean Chem. Soc.*, *32*, 1692.
(28) Kapuscinski, J. *Biotech Histochem* **1995**, *70*, 220.
(29) Broess, K.; Borst, J. W.; van Amerongen, H. *Photosynth Res* **2009**, *100*, 89.
(30) Botchway, S. W.; Barba, I.; Jordan, R.; Harmston, R.; Haggie, P. M.; Williams, S.; Fulton, A. M.; Parker, A. W.; Brindle, K. M. *Biochemical Journal* **2005**.
(31) He, L.; Olson, D. P.; Wu, X.; Karpova, T. S.; McNally, J. G.; Lipsky, P. E. *Cytometry A* **2003**, *55*, 71.

(32) Dulkin, S. I.; Friedemann, T. E. *Journal of Food Science* **1956**, *21*, 519.
(33) Grandhee, S. K.; Monnier, V. M. *J Biol Chem* **1991**, *266*, 11649.
(34) Kraft, M. L.; Weber, P. K.; Longo, M. L.; Hutcheon, I. D.; Boxer, S. G. *Science* **2006**, *313*, 1948.
(35) Legent, G.; Delaune, A.; Norris, V.; Delcorte, A.; Gibouin, D.; Lefebvre, F.; Misevic, G.; Thellier, M.; Ripoll, C. *The Journal of Physical Chemistry B* **2008**, *112*, 5534.
(36) Boxer, S. G.; Kraft, M. L.; Weber, P. K. *Annu Rev Biophys* **2009**, *38*, 53.
(37) Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Molecular & Cellular Proteomics* **2002**, *1*, 376.
(38) Paul, B. D.; Holler, J. M.; Lyons, T. P.; Past, M. R. *J Anal Toxicol* **2007**, *31*, 447.
(39) McMahon, G.; Saint-Cyr, H. F.; Lechene, C.; Unkefer, C. J. *J Am Soc Mass Spectrom* **2006**, *17*, 1181.
(40) Schneider-Poetsch, T.; Ju, J.; Eyler, D. E.; Dang, Y.; Bhat, S.; Merrick, W. C.; Green, R.; Shen, B.; Liu, J. O. *Nat Chem Biol* **2010**, *6*, 209.
(41) NATHANS, D. *Proc Natl Acad Sci U S A* **1964**, *51*, 585.
(42) Pestka, S. *Annu Rev Microbiol* **1971**, *25*, 487.
(43) Eggers, D. K.; Welch, W. J.; Hansen, W. J. *Mol Biol Cell* **1997**, *8*, 1559.
(44) Starck, S. R.; Green, H. M.; Alberola-Ila, J.; Roberts, R. W. *Chem Biol* **2004**, *11*, 999.
(45) Schmidt, E. K.; Clavarino, G.; Ceppi, M.; Pierre, P. *Nat Meth* **2009**, *6*, 275.
(46) David, A.; Dolan, B. P.; Hickman, H. D.; Knowlton, J. J.; Clavarino, G.; Pierre, P.; Bennink, J. R.; Yewdell, J. W. *J Cell Biol* **2012**, *197*, 45.
(47) Redman, C. M.; Sabatini, D. D. *Proc Natl Acad Sci U S A* **1966**, *56*, 608.
(48) Stevens, A.; Maupin, M. K. *Biochem Biophys Res Commun* **1989**, *159*, 508.
(49) Maquat, L. E. *Genes Dev* **2002**, *16*, 1743.
(50) Xu, M.; Cook, P. R. *J Cell Biol* **2008**, *181*, 615.
(51) Williams, A. F.; Barclay, A. N.; Clark, S. J.; Paterson, D. J.; Willis, A. C. *J Exp Med* **1987**, *165*, 368.
(52) Davis, S. J.; Davies, E. A.; van der Merwe, P. A. *Biochem Soc Trans* **1995**, *23*, 188.
(53) Brogna, S.; Wen, J. *Nat Struct Mol Biol* **2009**, *16*, 107.
(54) Kaida, D.; Motoyoshi, H.; Tashiro, E.; Nojima, T.; Hagiwara, M.; Ishigami, K.; Watanabe, H.; Kitahara, T.; Yoshida, T.; Nakajima, H.; Tani, T.; Horinouchi, S.; Yoshida, M. *Nat Chem Biol* **2007**, *3*, 576.
(55) Carter, M. S.; Doskow, J.; Morris, P.; Li, S.; Nhim, R. P.; Sandstedt, S.; Wilkinson, M. F. *J Biol Chem* **1995**, *270*, 28995.
(56) Catic, A.; Suh, C. Y.; Hill, C. T.; Daheron, L.; Henkel, T.; Orford, K. W.; Dombkowski, D. M.; Liu, T.; Liu, X. S.; Scadden, D. T. *Cell* **2013**, *155*, 1380.
(57) Lang, G.; Wotton, D.; Owen, M. J.; Sewell, W. A.; Brown, M. H.; Mason, D. Y.; Crumpton, M. J.; Kioussis, D. *EMBO J* **1988**, *7*, 1675.
(58) Finking, R.; Marahiel, M. A. *Annu Rev Microbiol* **2004**, *58*, 453.
(59) Harris, H. *Trends Biochem Sci* **1994**, *19*, 303.
(60) Greenberg, N. A.; Shipe, W. F. *Journal of Food Science* **1979**, *44*, 735.
(61) Bachmair, A.; Finley, D.; Varshavsky, A. *Science* **1986**, *234*, 179.
(62) Yewdell, J. W.; Nicchitta, C. V. *Trends Immunol* **2006**, *27*, 368.
(63) Turner, G. C.; Varshavsky, A. *Science* **2000**, *289*, 2117.
(64) Goldman, S. R.; Ebright, R. H.; Nickels, B. E. *Science* **2009**, *324*, 927.
(65) Seila, A. C.; Calabrese, J. M.; Levine, S. S.; Yeo, G. W.; Rahl, P. B.; Flynn, R. A.; Young, R. A.; Sharp, P. A. *Science* **2008**, *322*, 1849.
(66) Jackson, R. J.; Hellen, C. U.; Pestova, T. V. *Nat Rev Mol Cell Biol* **2010**, *11*, 113.
(67) Vilela, C.; McCarthy, J. E. *Mol Microbiol* **2003**, *49*, 859.
(68) Ingolia, N. T.; Lareau, L. F.; Weissman, J. S. *Cell* **2011**, *147*, 789.

(69) Fritsch, C.; Herrmann, A.; Nothnagel, M.; Szafranski, K.; Huse, K.; Schumann, F.; Schreiber, S.; Platzer, M.; Krawczak, M.; Hampe, J.; Brosch, M. *Genome Res* **2012**, *22*, 2208.
(70) Calvo, S. E.; Pagliarini, D. J.; Mootha, V. K. *Proc Natl Acad Sci U S A* **2009**, *106*, 7507.
(71) Kapranov, P.; St Laurent, G. *Front Genet* **2012**, *3*, 60.
(72) Szychowski, J.; Mahdavi, A.; Hodas, J. J.; Bagert, J. D.; Ngo, J. T.; Landgraf, P.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *J Am Chem Soc* **2010**, *132*, 18351.
(73) Papantonis, A.; Larkin, J. D.; Wada, Y.; Ohta, Y.; Ihara, S.; Kodama, T.; Cook, P. R. *PLoS Biol* **2010**, *8*, e1000419.
(74) Mellon, P.; Parker, V.; Gluzman, Y.; Maniatis, T. *Cell*, *27*, 279.
(75) Margineantu, D. H.; Brown, R. M.; Brown, G. K.; Marcus, A. H.; Capaldi, R. A. *Mitochondrion* **2002**, *1*, 327.
(76) Gibbs, R. A.; Weinstock, G. M.; Metzker, M. L.; Muzny, D. M.; Sodergren, E. J.; Scherer, S.; Scott, G.; Steffen, D.; Worley, K. C.; Burch, P. E.; Okwuonu, G.; Hines, S.; Lewis, L.; DeRamo, C.; Delgado, O.; Dugan-Rocha, S.; Miner, G.; Morgan, M.; Hawes, A.; Gill, R.; Celera; Holt, R. A.; Adams, M. D.; Amanatides, P. G.; Baden-Tillson, H.; Barnstead, M.; Chin, S.; Evans, C. A.; Ferriera, S.; Fosler, C.; Glodek, A.; Gu, Z.; Jennings, D.; Kraft, C. L.; Nguyen, T.; Pfannkoch, C. M.; Sitter, C.; Sutton, G. G.; Venter, J. C.; Woodage, T.; Smith, D.; Lee, H. M.; Gustafson, E.; Cahill, P.; Kana, A.; Doucette-Stamm, L.; Weinstock, K.; Fechtel, K.; Weiss, R. B.; Dunn, D. M.; Green, E. D.; Blakesley, R. W.; Bouffard, G. G.; De Jong, P. J.; Osoegawa, K.; Zhu, B.; Marra, M.; Schein, J.; Bosdet, I.; Fjell, C.; Jones, S.; Krzywinski, M.; Mathewson, C.; Siddiqui, A.; Wye, N.; McPherson, J.; Zhao, S.; Fraser, C. M.; Shetty, J.; Shatsman, S.; Geer, K.; Chen, Y.; Abramzon, S.; Nierman, W. C.; Havlak, P. H.; Chen, R.; Durbin, K. J.; Egan, A.; Ren, Y.; Song, X. Z.; Li, B.; Liu, Y.; Qin, X.; Cawley, S.; Cooney, A. J.; D'Souza, L. M.; Martin, K.; Wu, J. Q.; Gonzalez-Garay, M. L.; Jackson, A. R.; Kalafus, K. J.; McLeod, M. P.; Milosavljevic, A.; Virk, D.; Volkov, A.; Wheeler, D. A.; Zhang, Z.; Bailey, J. A.; Eichler, E. E.; Tuzun, E. *Nature* **2004**, *428*, 493.
(77) Park, S. M.; Shen, Y.; Kim, B. H. *Organic & Biomolecular Chemistry* **2007**, *5*, 610.
(78) Sharma, R. A.; Kavai, I.; Hughes, R. G.; Bobek, M. *Journal of Medicinal Chemistry* **1984**, *27*, 410.
(79) Andruszkiewicz, R.; Rozkiewicz, D. *Synthetic Communications* **2004**, *34*, 1049.
(80) Lundquist, J. T.; Pelletier, J. C. *Org. Lett.* **2001**, *3*, 781.
(81) Fernandez-Gonzalez, M.; Boutureira, O.; Bernardes, G. J. L.; Chalker, J. M.; Young, M. A.; Errey, J. C.; Davis, B. G. *Chemical Science* **2010**, *1*, 709.
(82) Link, A. J.; Vink, M. K. S.; Tirrell, D. A. *Nat. Protocols* **2007**, *2*, 1879.
(83) Zhou, Z.; Fahrni, C. J. *J. Am. Chem. Soc.* **2004**, *126*, 8862.

Chapter 4

Quantitative Non-Canonical Amino acid Tagging

4.1: Introduction

The understanding of cell behaviour at a systems level is restricted by limited knowledge of proteome dynamics. Changes in the expression of proteins shows only modest correlation with transcription of their corresponding mRNAs,[1,2] with the a high correlation observed only for components of ribosomes, cell adhesion complexes and chaperonins involved in protein folding.[3] This suggests that the abundance of cellular proteins may be predominantly controlled only at the translation level,[4] and that knowledge of transcriptome dynamics alone is insufficient to understand cell behaviour. In this chapter, the development of a UAA-based proteomics method will be described, which may be applied to elucidate and quantify stimulus-induced changes in the global population of cell proteins (i.e., the cell proteome). The work was carried out in collaboration with Dr. Andrew J. M. Howden, Dr. Vincent Geoghegan, Dr. Kristin Katsch, Dr. Georgius Efstathiou, Prof. Oreste Acuto and Dr. Omar Boutureira. All UAA-based labelling and method development was carried out by BB, AJMH, VG, KK, GE and OB. Radioactive labelling experiments were carried out by KK and GE with Dr. Benedikt Kessler. MS analysis was done by AJMH, VG, KK and GE with Dr. Ben Thomas. Flow cytometry analysis was done by KK.

4.2: Stable Isotope Labelling by Amino acids in Cell culture

Stable Isotope Labelling by Amino acids in Cell culture (SILAC) based mass spectrometry allows for the quantification of steady-state proteome changes.[5] The technique involves the complete replacement of all endogenous amino acids in cells with heavier stable-isotope (i.e., non-radioactive) ^{2}H, ^{13}C and/or ^{15}N labelled amino acids, supplemented in cell culture media. This may be done over ~5 population doublings of cells grown in 'heavy' media.[6] A control cell population is also used, where the same cell line is grown for the same amount of time in normal media. The cells are then mixed in a 1:1 ratio by number, lysed, digested to form shoter peptides, and analysed by MS. For every peptide arising from the control cell sample that is detected by MS, a corresponding partner arising from the heavy-AA labelled extract,

shifted in mass corresponding to the amount of stable isotope labelled AAs it contained, thereby forming corresponding peak pairs in the MS. As isotope labelling does not alter the chemical properties of the labelled peptides so formed, the ratio of the peak pair areas then directly correspond to the relative level of expression of the same protein in the two cell samples. If the populations were in identical cell states, then the ratio for all peak pairs would be ~ 1:1. However, if the two populations consisted of different cell lines, or cells in different states, then the ratio of peak pair areas would then be indicative of the ratio of the expression of those proteins in the two populations (Fig. 4.1).

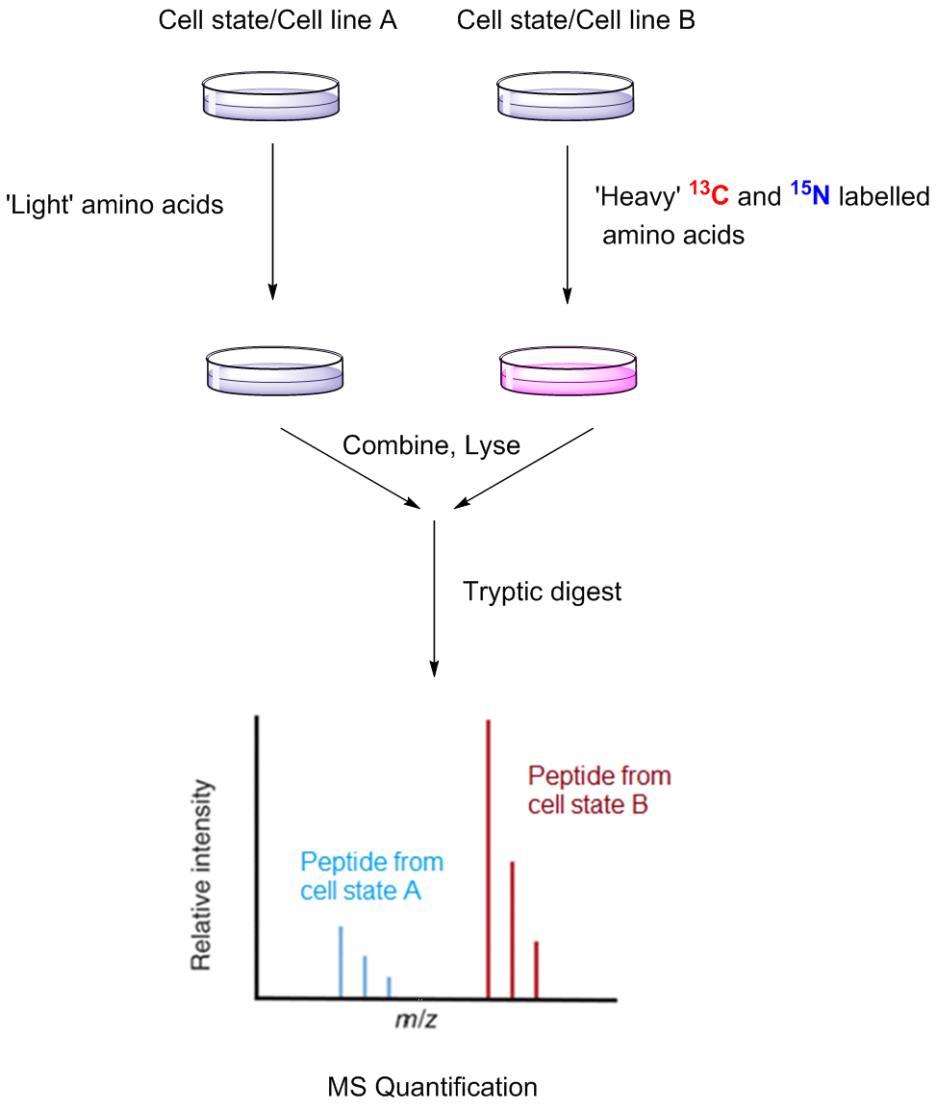


Fig. 4.1: SILAC work-flow

SILAC has been successfully applied to several cell systems such as prostate cancer cell lines of different metastatic potentials,[7,8] muscle cells in various states of differentiation,[9] and pancreatic-cancer-derived cells v/s non-tumorous pancreatic cells.[10] However, for monitoring non-steady state changes in cell proteomes (e.g. in response to external stimuli), SILAC cannot be easily applied, especially for *ex vivo* primary cells.[11] The five population doublings required to uniformly label cells with stable-isotope labelled AAs abolishes the possibility of monitoring proteome changes at lower time scales. Alternatively, shorter SILAC pulses of a few hours label only a small fraction of a protein pool,[12] thereby 'contaminating' the signal from incompletely labelled cells with that from pre-existing unlabelled proteins. The selective isolation of the fraction of stable-isotope labelled proteins would help to overcome the above limitations.

4.3: Bioorthogonal Non-Canonical Amino acid Tagging

Bioorthogonal Non-Canonical Amino acid Tagging (BONCAT) is a technique for the identification of newly synthesised proteins. The technique utilises the promiscuity of the amino acid incorporation machinery of cells to label newly synthesised proteins with unnatural amino acids bearing biologically stable but chemically reactive functional groups.[13] These labels may then be used for the site-selective modification of newly synthesized proteins, such as via the Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC), for L-azidohomoalanine (Aha, **1.3**) labelled cells.[13-15] After cell lysis, the Aha-azide may be functionalised with alkyne-linked biotinylated tag by CuAAC, allowing for the selective enrichment of the tagged protein pool using a biotin-affinity pull-down, thus purifying it from the bulk of the pre-existing non-labelled cellular proteome (Fig. 4.2). BONCAT has been used for mass spectrometry–based nonquantitative monitoring of newly synthesized proteins in cultured mammalian cells.[15]

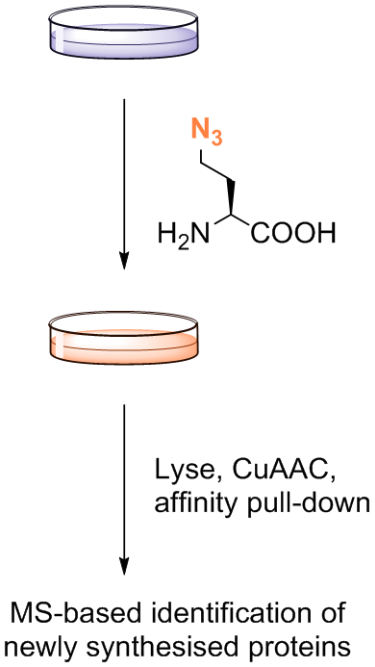


Fig. 4.2: BONCAT work-flow

However, the efficiencies and rates of bioconjugation reactions like CuAAC are site-dependent and protein-dependent, and may vary from protein to protein.[16,17] Moreover, monitoring comparative protein changes by BONCAT alone using stable isotope label-free MS would have low accuracy, as protein signals in the MS cannot be assigned to one or the other cell population being analysed.

4.4: Quantitative Non-Canonical Amino Acid Tagging Work-flow

The strengths of these two techniques were combined in order to both label and purify newly expressed protein in response to a stimulus (BONCAT) as well as quantitatively compare it with unstimulated cells using stable isotope MS contrast forming agents (SILAC), in a technique named Quantitative Non-Canonical Amino acid Tagging (QuaNCAT). This technique was applied to both Jurkat T-cells, an immortalised leukemic human T-cell line,[18] as well as freshly isolated human CD4 positive ($CD4^+$) primary T-cells, to determine proteome changes in response to T-cell stimulation. Cells were co-labelled with L-azidohomoalanine (Aha), and either 'heavy' [$^{13}C_6$,$^{15}N_4$] L-Arg and [$^{13}C_6$,$^{15}N_2$] L-Lys for stimulated cells or with 'medium' [$^{13}C_6$] L-Arg and [$^{2}H_4$] L-Lys for non-stimulated cells (Fig. 4.3). This labelling strategy allowed for distinction by MS of stimulation-induced (heavy isotope–labelled) from steady-state (medium isotope–labelled) synthesized proteins, as well

as allowing for the enrichment and distinction of the newly synthesised protein pool from the pre-existing 'light' amino acid labelled proteome. Equal cell numbers from stimulated and unstimulated populations were mixed, lysed, and the newly expressed proteins labelled by CuAAC with a cleavable alkyne-functionalised biotin. This was followed by biotin-affinity pull-down, stringent washes, and analysis of the digested eluate by MS.

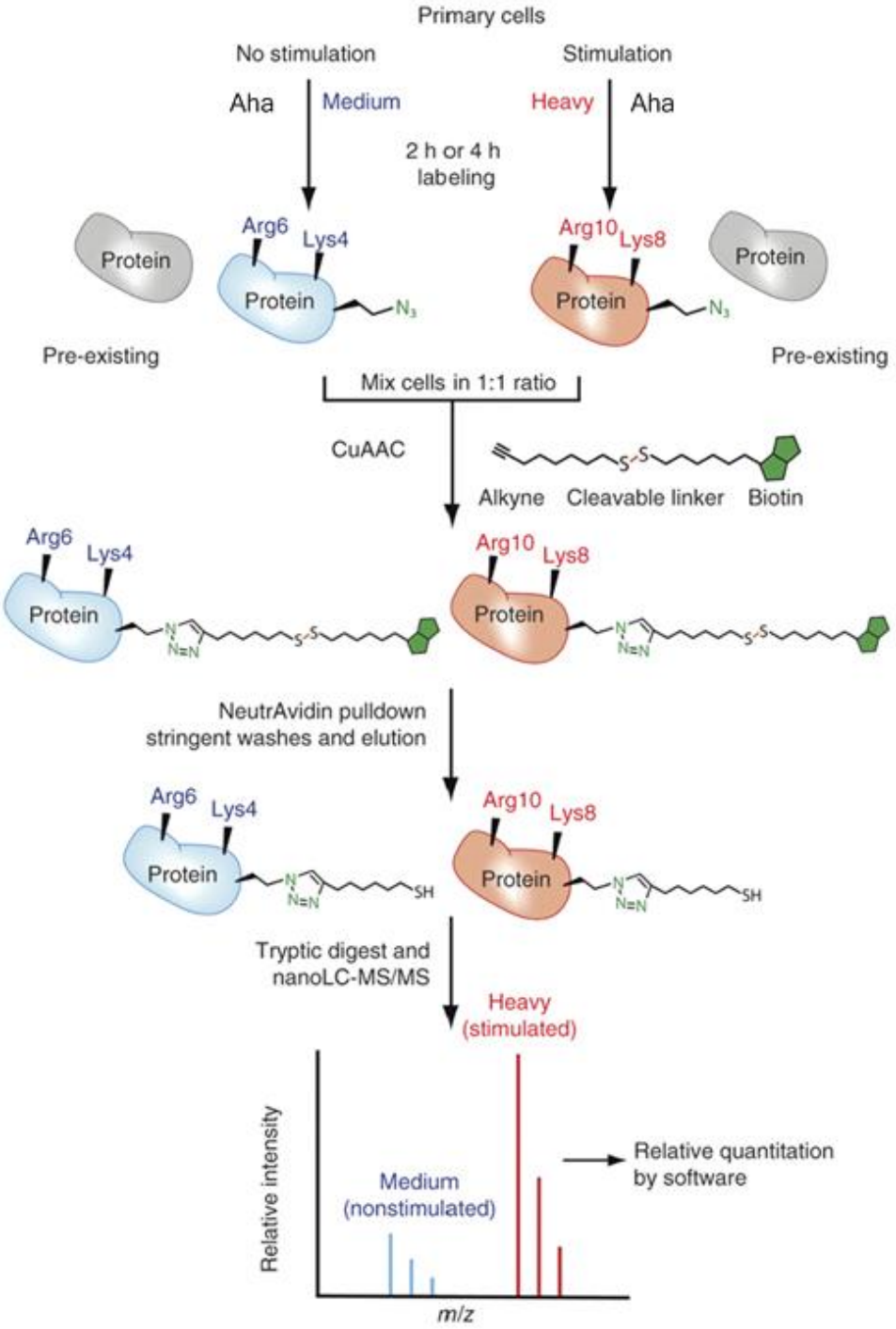


Fig. 4.3: QuaNCAT work-flow

To simulate antigenic stimulation, cells were treated with phorbol 12-myristate 13-acetate (PMA)[19,20] and ionomycin (iono).[21] These drugs closely mimic cellular activation by antigen and CD28 'co-stimuli', thereby eliciting proliferation and differentiation.

4.5: Metabolic impact of Azidohomoalanine

Aha has been reported to be non-toxic to mammalian cells for up to 2 h at a concentration of 4 mM.[13] Furthermore, Aha incorporation has been demonstrated to not alter protein structure and function.[22-25] Nevertheless, effects on T-cells incubated in media containing up to 4 mM Aha, and for up to 4h, were investigated.

Jurkat T-cells were incubated in RPMI media depleted of L-methionine (Met) and L-cysteine (Cys) for 1 h, then incubated in Met- and Cys-free RPMI supplemented with either 0.1 mM Met, 1 mM or 4 mM Aha. 10 nM (10 µCi/mL) ^{35}S-Cys + 5 µM L-Cys were then added to both cultures. After a pulse of 2 and 4 h, samples were harvested and lysed. Soluble proteins were precipitated with 5% trichloroacetic acid, spotted on filter paper and air-dried. Spots were excised and radioactivity measured with a scintillation counter (Fig. 4.4).

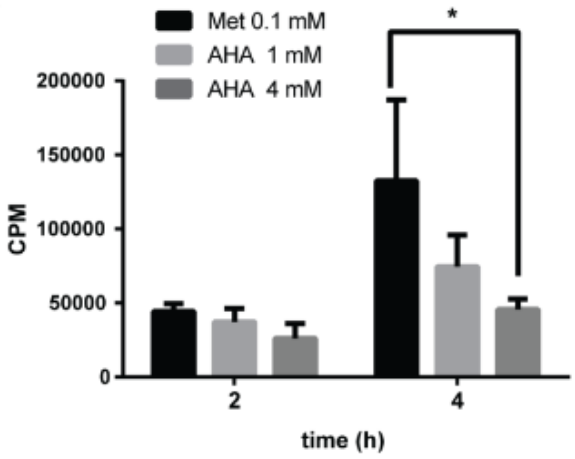


Fig. 4.4: Mean counts per minute from JurkaT-cell lysates incubated with ^{35}S and either 0.1 mM Met, 1 mM Aha or 4 mM Aha for 2 or 4 hours. (*) $P < 0.05$, $n = 3$

Statistical analysis (one-way analysis of variance (ANOVA) followed by Bonferroni's multiple comparison test) revealed a significant decrease in global protein synthesis in Jurkat cells compared to Met-incubated cells, after 4 h of incubation with 4 mM Aha. This suggests that Aha, at a concentration of 4 mM, has a time-dependent effect on protein biosynthesis

compared to Aha at 1 mM. Therefore a concentration of 1 mM was used to investigate the effect of Aha labelling on global protein degradation in a ^{35} S-Cysteine pulse-chase experiment in Jurkat cells. Jurkat T-cells were incubated in RPMI media depleted of L-methionine (Met) and L-cysteine (Cys) for 1 h, then incubated in Met- and Cys-free RPMI supplemented with either 0.1 mM Met or 1 mM Aha. 32.4 nM (30 μCi/mL) ^{35}S-Cys + 16.2 μM L-Cys were then added to both cultures. After a 15 minute pulse, media was replaced with fresh RPMI media supplemented with either Met (0.1 mM) or Aha (1 mM) without radioactive Cys. After 0, 1, 2, and 4 h, cells were harvested and lysed. Soluble protein extracts were prepared as described above, and radioactivity measured with a scintillation counter (Fig. 4.5).

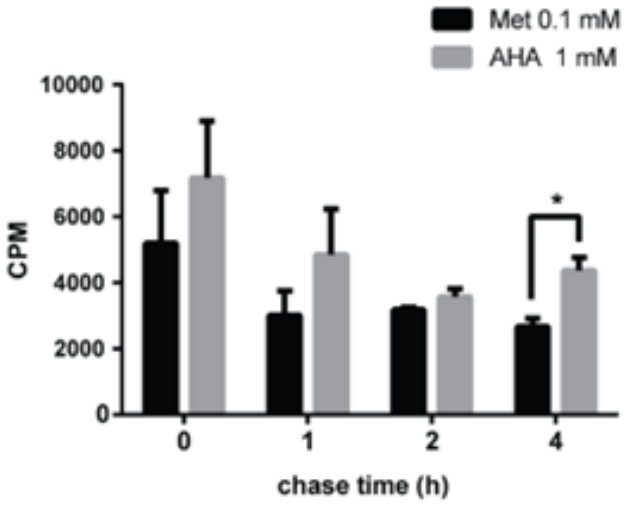


Fig. 4.5: Mean counts per minute from Jurkat T-cell extracts for a ^{35}S-Cys pulse-chase experiment co-labelled with either 0.1 mM Met or 1 mM Aha, after 0, 1, 2 and 4 h. (*) $P < 0.05$, $n = 3$

Statistical analysis (unpaired t-test) revealed significant changes in protein degradation for Aha-treated relative to Met-treated cells only after 4 h.

Next, the effect of Aha on cell morphology and viability in non-stimulated and stimulated primary human T-lymphocytes was studied. Human primary CD4+ T-lymphocytes were incubated in Met-, Arg- and Lys-free RPMI for 1 h, followed by incubation for 2 h or 4 h in RPMI supplemented with 'heavy' [$^{13}C_6$,$^{15}N_4$] L-Arg and [$^{13}C_6$,$^{15}N_2$] L-Lys and 1 mM of Aha or Met. Cells pulsed with ‘heavy’ Arg and Lys but without (w/o) Met were used as

controls. In addition, cells were either left non-stimulated or stimulated with 62.5 ng/mL PMA and 1 µg/ml iono. Cell morphology was assessed by flow cytometry, and FSC/SSC dot plots (cell size/cell granularity) were used to determine cell viability (Fig. 4.6).

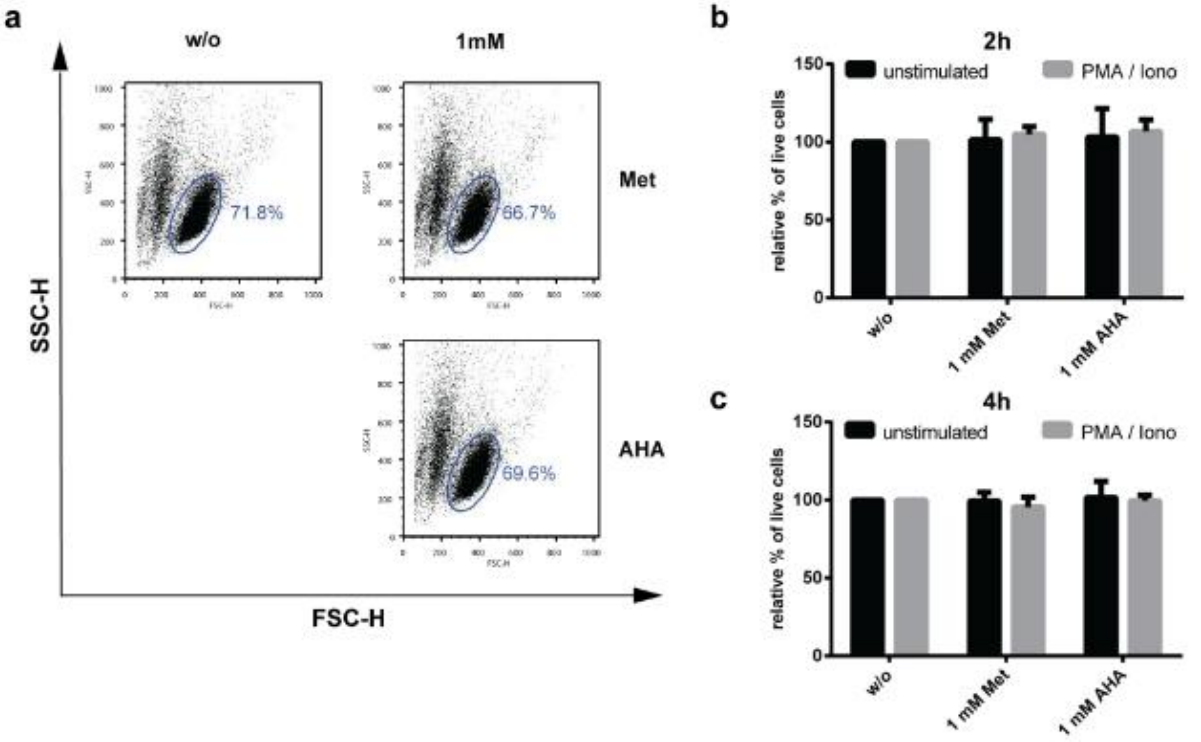
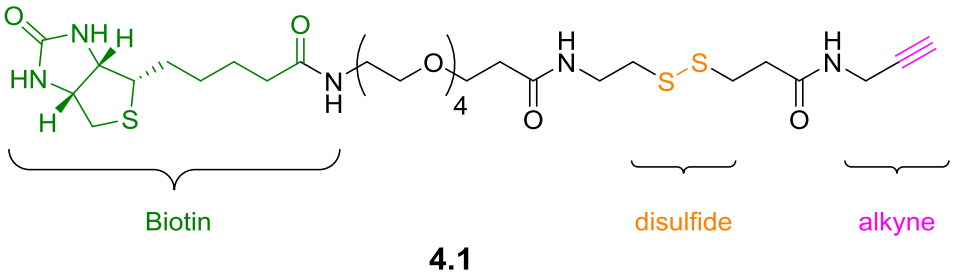


Fig. 4.6: a) Representative FSC/SSC dot plots after a 4 h pulse with 1 mM Met or Aha plus PMA/Ionomycin stimulation. b) Mean percentage of live cells from six independent blood donors after 2 h and c) 4 h, normalized to cells untreated with Met. (n=6)

Quantification of apoptotic cells was also carried out by flow cytometry by monitoring cell surface expression of phosphatidylserine by Annexin V (data not shown). Statistical analysis revealed no significant changes between all conditions tested. Labelling human primary T-cells with 1 mM Aha up to 4 h did not affect T-cell viability.

4.6: Application of QuaNCAT to primary T-cells

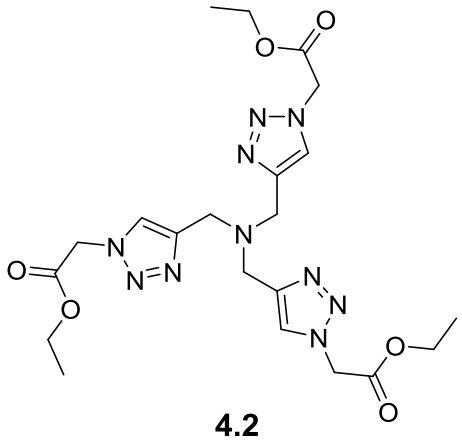
Human $CD4^+$ T-cells were incubated for 60 min in Met-, Arg-, and Lys- free RPMI, followed by incubation for either 2 or 4 h in the same media supplemented with 1 mM Aha and either 'heavy' [$^{13}C_6$,$^{15}N_4$] L-Arg and [$^{13}C_6$,$^{15}N_2$] L-Lys or with 'medium' [$^{13}C_6$]L-Arg and [$^{2}H_4$] L-Lys for nonstimulated cells. Cells incubated in heavy-AA containing media were also supplemented with 62.5 ng/mL PMA and 1 µg/ml iono to induce stimulation. Non-stimulated and stimulated cells were mixed in a 1:1 ratio, and lysed in SDS, followed by reduction and alkylation of the soluble proteins. CuAAC was then carried out with alkyne-functionalised biotin with a cleavable disulfide linker (Fig. 4.7).[26]



4.1

Fig. 4.7: Alkyne-functionalised biotin with a cleavable disulfide linker

A non-reductive CuAAC protocol was used,[16] with Cu(I)Br and the tris-triazolyl compound triethyl 2,2',2''-(4,4',4''-nitrilotris(methylene)tris(1H-1,2,3-triazole-4,1-diyl))triacetate (**4.2**) used as a ligand to stabilise Cu(I) in aqueous media (Fig.4.8).



4.2

Fig. 4.8: Tris-triazolyl ligand for stabilisation of Cu(I) during CuAAC

Excess alkyne tag was removed using a size exclusion spin column, and immunoblotting against biotin confirmed effective biotin labelling of Aha-containing proteins (Fig. 4.9).

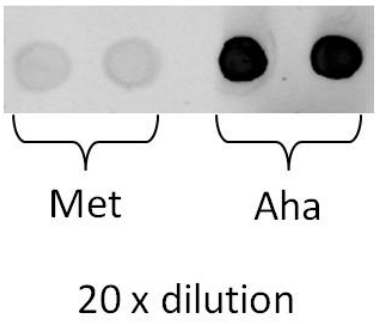


Fig. 4.9: Anti-biotin dot blots for spin-desalted cell lysates from Met-incorporated cells (control) and Aha-labelled cells

Biotin-tagged proteins were bound to NeutrAvidin[27]-immobilized beads, and non-specifically bound proteins thoroughly washed out. Biotin-tagged proteins were eluted from NeutrAvidin with DTT (Fig. 4.3), thereby substantially reducing the mass spectrometry signal from contaminating unlabelled ('light') proteins. Eluted proteins were digested by trypsin, and analysed by nanoLC-MS/MS. Simultaneous Aha and stable isotope labeling with arginine and lysine, followed by biotin tagging, allowed for strong enrichment and quantification of newly synthesized proteins to determine stimuli-induced changes in their abundance. Peak intensities of 'medium' and 'heavy' peptide pairs were quantified using both Proteome Discoverer 1.3 (Thermo Scientific) and MaxQuant Ver 1.3.0.5 software[28].

4.6.1: Statistical analysis of MS data

Each experiment was carried out in triplicate, and each replicate was analysed by nanoLC-MS/MS thrice in order to determine the reproducibility of the QuaNCAT technique. The ratios of peak intensites for corresponding peptide pairs (i.e., ratio of the peak intensity of a 'heavy' peptide from stimulated cells to the peak intensity of the corresponding 'medium' peptide from non-stimulated cells) were converted to $\log_2$(Heavy/Medium) values, and each of the nine resultant data sets (3 replicates x 3 nanoLC-MS/MS runs) were compared with each other, and the Pearson linear correlation coefficients calculated.

Statistical data is shown in Fig. 4.10, with $\log_2$(H/M) ratios for all identified proteins from one data set plotted against $\log_2$(H/M) ratios from the remaining 8 data sets, with linear

regression lines shown in red. Histograms of protein $\log_2$(H/M) ratios for each run are plotted across the matrix diagonal. Global average Pearson correlation coefficients were determined to be 0.71 (0.64-0.79) and 0.75 (0.65-0.86) for 2 h and 4 h stimulation, respectively, indicating good reproducibility of the QuaNCAT technique.

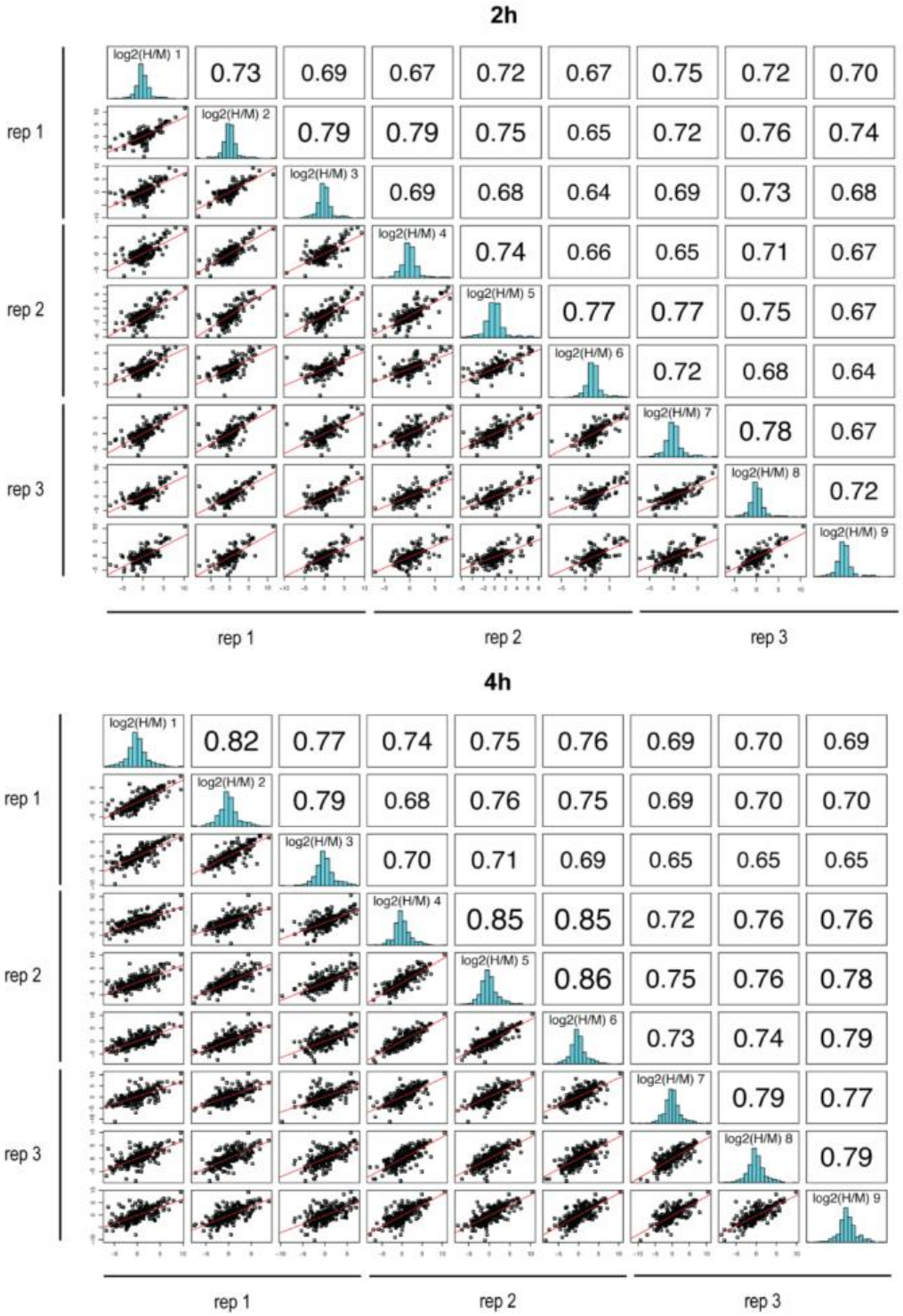


Figure 4.10. Reproducibility of protein $\log_2$(H/M) ratios between nanoLC-MS/MS runs for all replicates. Scatter-plots of protein $\log_2$(H/M) ratios for nanoLC-MS/MS run pairs and corresponding Pearson correlation coefficients for 2 and 4 h treatments.

Of the 410 quantified proteins after a 2h stimulation, consistently ~95% were observed in each replicate. Of the 550 quantified proteins after a 4h stimulation, ~89% were seen in each replicate, further indicating good reproducibility of QuaNCAT (Fig. 4.11).

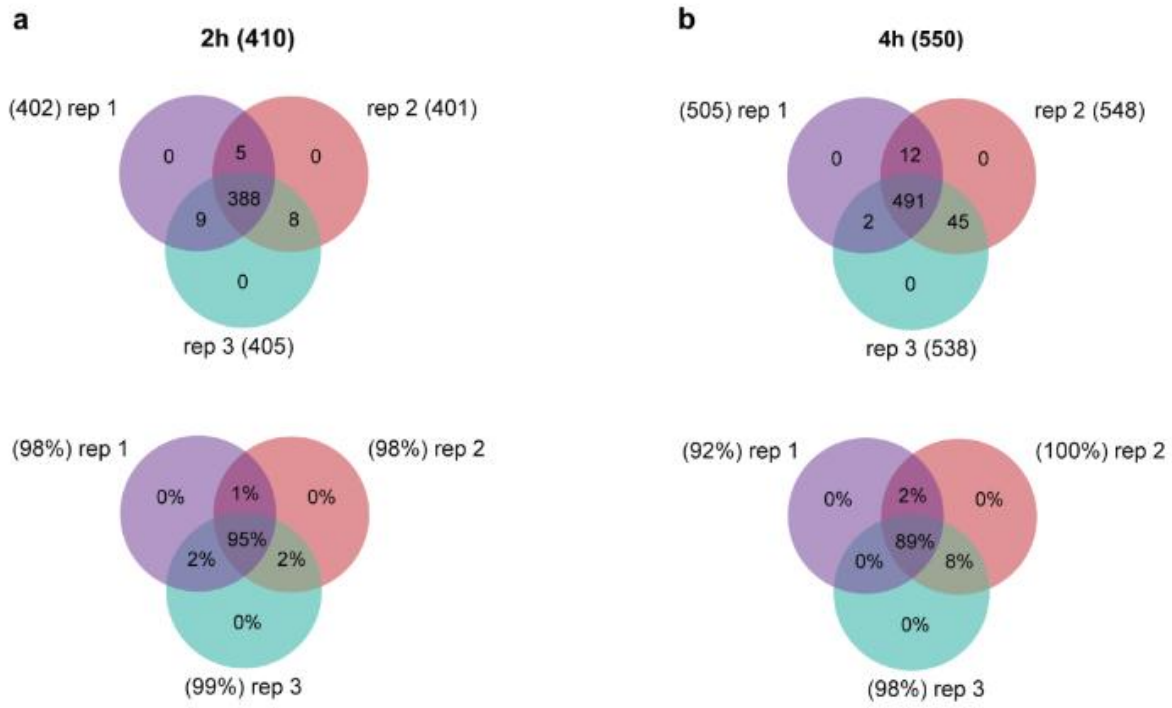


Fig. 4.11: Venn diagrams showing the overlap in number (top) and percentage (bottom) of quantified proteins,, between replicates from 2h-stimulated T-cells and 4h-stimulated T-cells.

The 410 and 550 quantified proteins after 2h and 4h of stimulation were then ordered by their respective $\log_2$(H/M) values, which were analyzed using the Linear Models for Microarray data (LIMMA) Ver. 3.14.1 statistical package. This served to provide fold changes in expression of proteins between stimulated and non-stimulated cells, with associated *P* values for differential expression of each protein (Fig. 4.12).

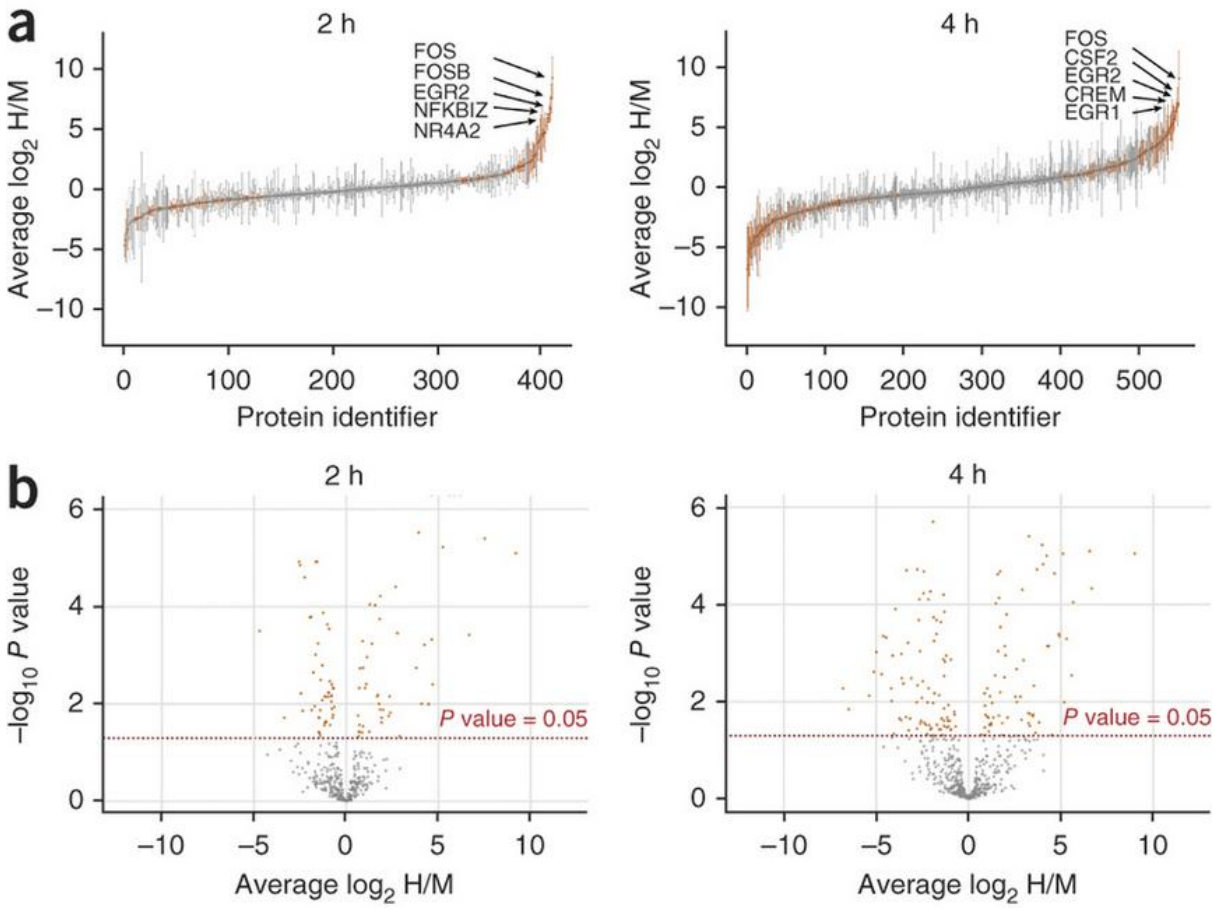


Fig. 4.12: (**a**) $\log_2$ (H/M) value-ordered plots for each quantified protein at 2 h and 4 h ± 1 standard deviation. Data for differentially expressed proteins (P < 0.05) is coloured orange. The top five up-regulated proteins after 2h and 4h of stimulation are indicated. (**b**) Volcano plots of $\log_2$ H/M versus $-\log_{10}$ P value for each quantified protein (threshold for P < 0.05 is indicated). Data for differentially expressed proteins (P < 0.05) is coloured orange.

To validate the quantification of differentially expressed proteins ($P < 0.05$), mass spectra were manually examined. After stimulation, 107 differentially expressed proteins were identified, 93 of which exhibited an increase after stimulation of T-cells (Appendix I and Fig. 4.12). 56 proteins were differentially expressed after 2h of stimulation, 33 of which (~60%) were still differentially expressed after 4h of stimulation (Appendix I). Of the 107 proteins, 50 were seen to be differentially expressed only after 4 h of stimulation. Fold increases ranged from 1.6-fold to 589-fold, with 72 of 141 measured changes being ≥ 5-fold. Some of the largest fold changes likely represent examples of *de novo* protein expression.

4.7: Results

Gene ontology analysis of function of differentially expressed proteins was carried out using Protein Center (Fig. 4.13).[29] Analysis revealed that many of these proteins are involved in regulation of transcription, including pre- and post-transcriptional processes. This is consistent with transcriptome studies of activated human T-cells,[30] which have revealed that T-cells, which are weakly metabolically active in G0 phase of the cell cycle, proliferate and differentiate into immune response effectors.

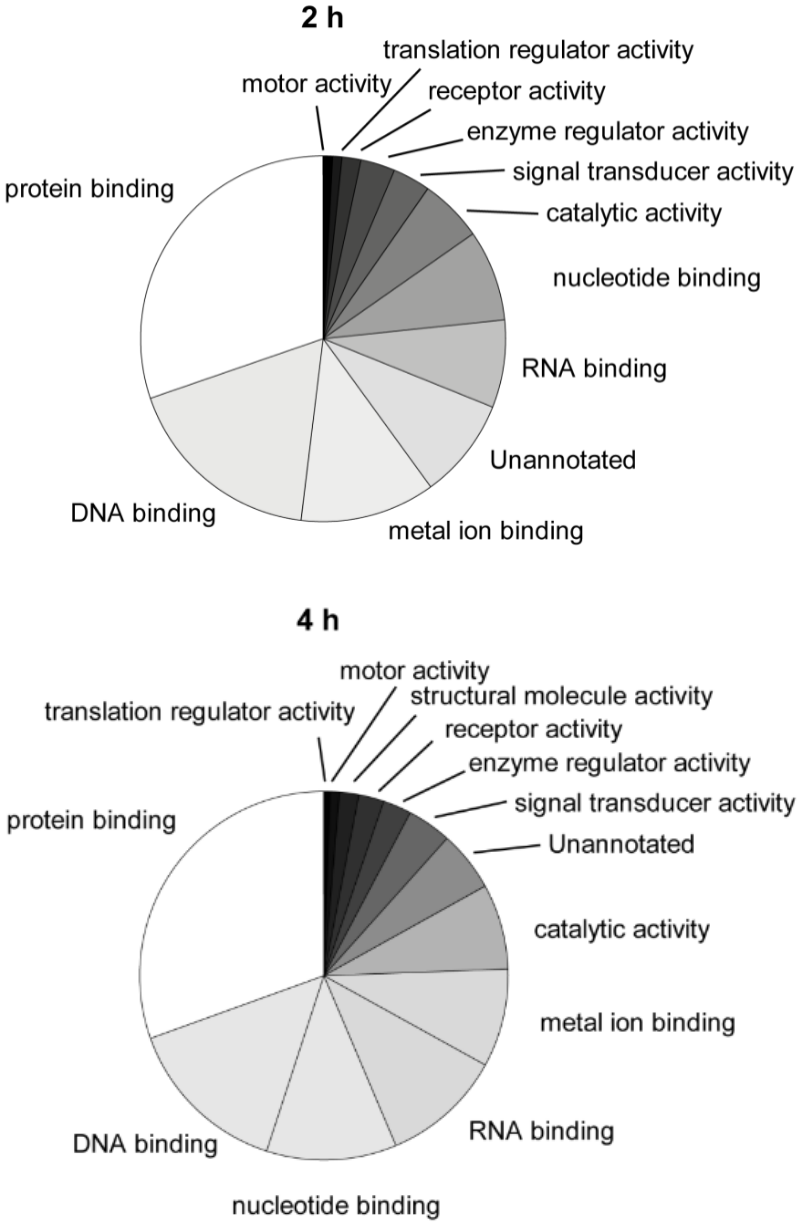


Fig. 4.13: Function of differentially expressed proteins. Differentially expressed proteins at 2 and 4 h of stimulation were uploaded to Protein Center and annotated according to Gene Ontology terms.

QuaNCAT revealed substantial increases (5-fold to 500-fold) in expression of several genes, many of which encode transcription factors and transcriptional regulators (such as Fos, Jun, NR4A, IRF and EGR family members, c-Rel, FOXp1 and a lysine-specific demethylase). Other identified proteins included cytokines (INFγ and CSF2) and activation surface markers (CD69, CD97 and HLA), protein chaperones, and proteins involved in cytoskeleton dynamics and vesicle transport. Of note was the detection of increased expression (2- to 20-fold) of several proteins that regulate mRNA metabolism, translation and the cell cycle. Some proteins were not annotated and/or not previously detected during T-cell activation, which serves to stimulate research into unravelling their role in T-cell proliferation and differentiation.

QuaNCAT data also revealed decreased expression of several proteins, perhaps caused by decreased transcription, translation, or accelerated protein degradation in response to T-cell stimulation. Important to note is that both alterations in protein translation as well as protein degradation would escape transcriptome analysis, but their resultant impact on final protein expression levels can be detected by QuaNCAT. An interaction network of proteins differentially expressed at 2 h and 4 h was constructed in MetaCore and visualized using Cytoscape. This network illustrates that more than half of the detected proteins are interlinked by functional and biochemical networks (Fig. 4.14).

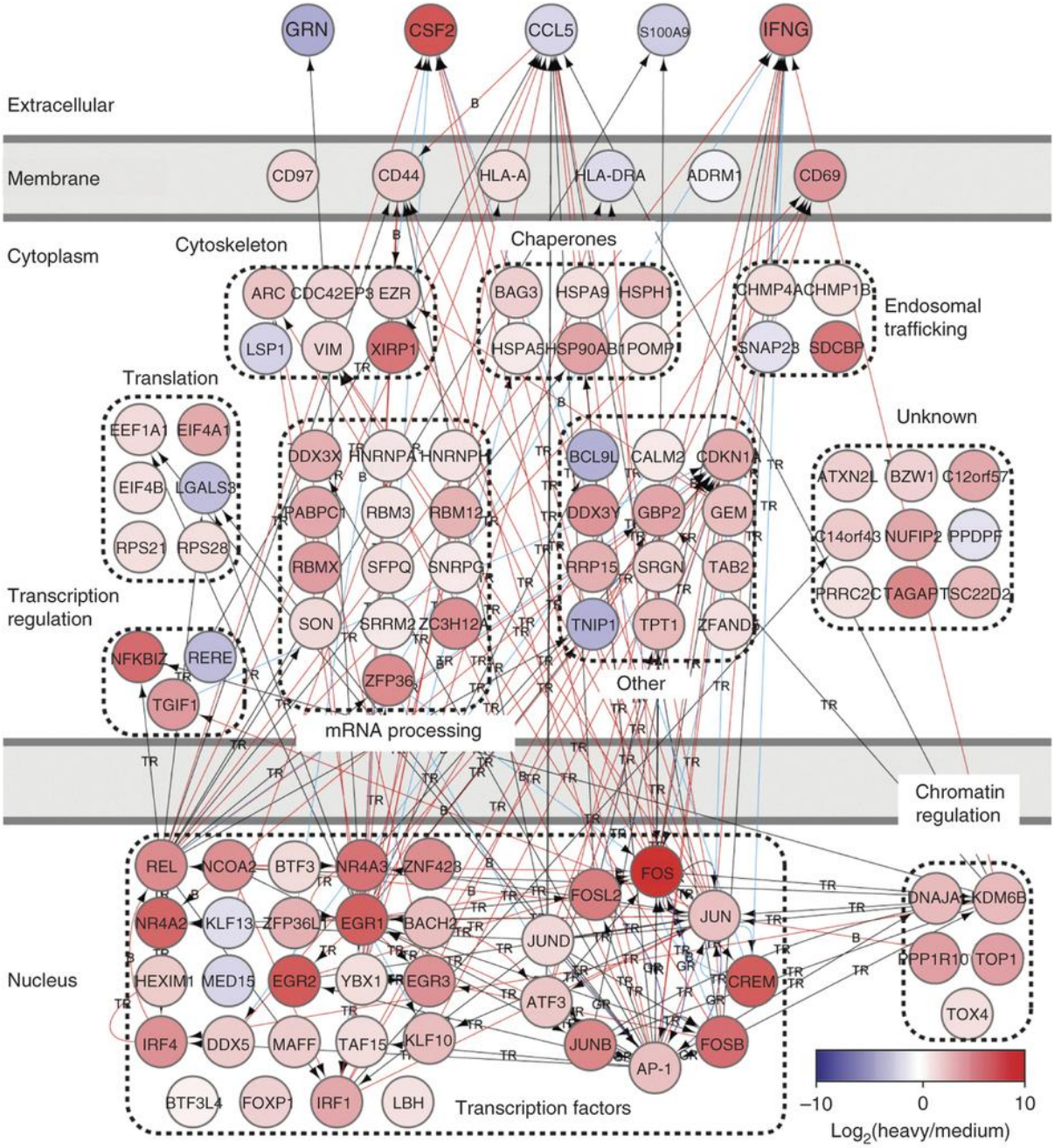


Fig. 4.14: Interaction network of differentially expressed proteins in stimulated T-cells. TR, transcriptional regulation; B, binding; and GR, group relation. Each coloured circle represents a differentially expressed protein detected by QuaNCAT. Each protein is coloured red or blue, indicating up-regulation or down-regulation, respectively (see colour bar). Red arrows indicate that the origin protein positively regulates the destination protein, blue lines indicate that the origin protein negatively regulates the destination protein, and black lines indicate unspecified interaction.

Changes in expression for each quantified protein after 2 h and 4 h stimulation and standard deviation, obtained from three independent experimental replicates (and three mass spectrometry analyses for each one) is provided in Appendix I.

4.8: Conclusion and Discussion

In this chapter, the development of a new proteomics method named Quantitative Non-Canonical Amino acid Tagging (QuaNCAT) was described. This technique may be applied to quantify proteomic changes in cells in response to external stimuli. A simultaneous labelling strategy was used to overcome problems in two existing proteomics methods, SILAC and BONCAT. Labelling with stable isotope - Lys and Arg allowed for the comparison of two different cell populations, while labelling with Aha allowed for the selective enrichment of newly expressed proteins through CuAAC-mediated biotinylation. The use of both 'medium' and 'heavy' Lys and Arg further overcame potential confounding effects in quantification owing to persistent signals from 'light' peptides, likely derived from contaminating untagged pre-existing proteins.

After affinity purification, 'heavy'- and 'medium'-labelled proteins typically comprised 10–20% of the identified proteins (data not shown), which is a substantial enrichment as < 1% of the total cellular proteome was labelled after a 2-h pulse. The ability to quantify relative differences in abundance of several cellular proteins demonstrates the utility of QuaNCAT, and this approach may generally be applicable to both shorter and longer labelling times. Furthermore, QuaNCAT may be applied to cells using identical labelling times as reported here, but only commencing several hours after stimulation. This would enable the monitoring of late stage proteome changes without necessarily increasing Aha labelling times, as depletion of cellular methionine would eventually adversely affect cellular methylation of proteins and nucleic acids.[31,32] The QuaNCAT method is also generally applicable to several cell lines, and also allows for the use of alternative metabolic labels and corresponding functionalisation chemistries. A combination of SILAC and BONCAT was also reported to quantitatively analyze the cellular secretome.[33] QuaNCAT allows for the analysis of protein changes in the entire cellular proteome, and provides a complement to the analysis of transcriptome dynamics.

4.9: Methods

4.9.1: General Consideration

Proton nuclear magnetic resonance (^{1}H NMR) spectra were recorded on a Bruker AV400 (400 MHz) or a Bruker AVII500 (500 MHz) spectrometer, as indicated. Carbon nuclear magnetic resonance (^{13}C NMR) spectra were recorded on a Bruker AV400 (100 MHz) spectrometer or on a Bruker AVII500 (125 MHz) spectrometer, as indicated. All chemical shifts are quoted on the δ scale in ppm using residual solvent as the internal standard. Coupling constants (*J*) are reported in Hertz (Hz) with the following splitting abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, quin = quintet, and a = apparent.

Infrared (IR) spectra were recorded on a Bruker Tensor 27 Fourier Transform spectrophotometer using thin films. Absorption maxima (υ_{max}) are reported in wavenumbers (cm^{-1}). Low resolution mass spectra (LRMS) were recorded on a Waters Micromass LCT Premier TOF spectrometer using electrospray ionization (ESI) and high resolution mass spectra (HRMS) were recorded on a Bruker MicroTOF ESI mass spectrometer. Nominal and exact m/z values are reported in Daltons.

Thin layer chromatography (TLC) was carried out using Merck aluminium backed sheets coated with $60F_{254}$ silica gel. Visualization of the silica plates was achieved using a UV lamp (λ_{max} = 254 nm), and/or potassium permanganate (5% $KMnO_4$ in 1M NaOH with 5% potassium carbonate). Flash column chromatography was carried out using BDH PROLAB® 40-63 mm silica gel (VWR). Mobile phases are reported in % volume of more polar solvent in less polar solvent for binary systems (e.g. 20% EtOAc in petrol = 1:4 ethyl acetate:petrol) and in relative composition for ternary systems (e.g. 1:2:4 H_2O:^{i}PrOH:EtOAc)

The anhydrous methanol used was purchased from Sigma. All other solvents were used as supplied (Analytical or HPLC grade), without prior purification. Distilled water was used for chemical reactions and Milli-Q® purified water for protein manipulations. 'Petrol' refers to the fraction of light petroleum ether boiling in the range 40-60 ºC. All reactions using anhydrous conditions were performed using flame-dried apparatus under an atmosphere of argon or nitrogen. Brine refers to a saturated solution of sodium chloride. Anhydrous magnesium sulfate ($MgSO_4$) or sodium sulfate (Na_2SO_4) was used as drying agents after reaction workup, as indicated. DOWEX 50WX8 (H^+ form) was conditioned as follows:

100 g of the commercial resin was placed in a 500 mL sintered filter funnel and allowed to swell with 200 mL of acetone for 5 minutes. The solvent was removed by suction and the resin was washed successively with 800 mL of acetone, 500 mL methanol, 500 mL 5 M HCl,

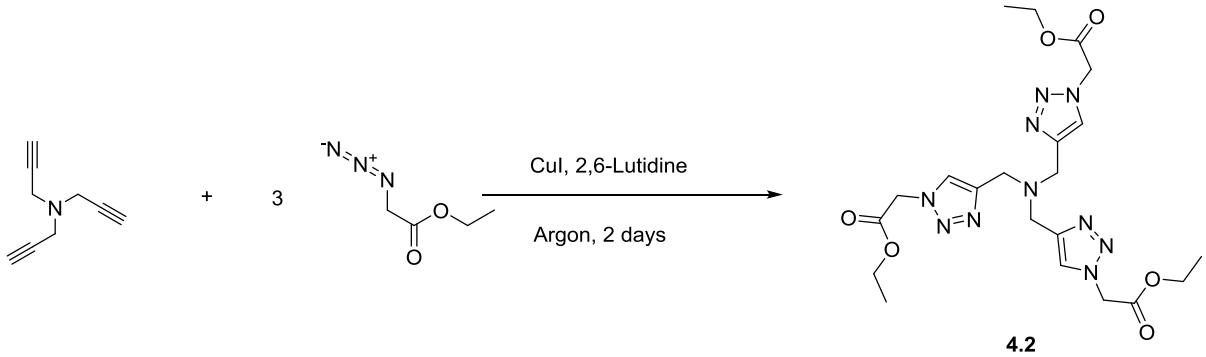
and then 1 L of water or until the pH of filtrate was ~ 7, as indicated by pH paper. The resin was partially dried on the filter and then stored and used as needed

4.9.2: Synthesis of Azidohomoalanine

See Chapter 3, Section **3.12.11**

4.9.3: Synthesis of tris-triazole ligand 4.2

The tris-triazole ligand, triethyl 2,2',2''-(4,4',4''-nitrilotris(methylene)tris(1H-1,2,3-triazole-4,1-diyl))triacetate, was prepared following minor modification of reported literature.[34]



4.2

Sodium azide (2.31 g, 35.5 mmol) was suspended in 5 mL dimethylformamide, to which ethylchloroacetate (3 mL, 28.0 mmol) was added dropwise at room temperature, and stirred for 24 hours. The solution was then diluted with diethyl ether (150 mL) and washed with water (3x 30 mL). The organic layer was then dried with sodium sulfate, filtered, and concentrated to 10 mL. This was then added to a solution of containing tripropargylamine (0.65 g, 3.8 mmol) in acetonitrile (5 mL), flushed with argon. The solution was cooled in an ice bath, and 2, 6-Lutidine (0.6 mL) and copper (I) iodide (20 mg) were added under a stream of argon. After 1 hour, the ice bath was removed, and stirring was continued for 2 days under argon. The solvent was then evaporated, and the residue purified by column chromatography (1:10 methanol: ethyl acetate) to give the product as a white solid (1.95 g, 98.7 %). IR (νmax, film): 3141, 2985, 2947, 2906, 2830, 1745, 1559, 1465, 1450, 1413, 1398, 1379, 1351, 1329, 1309, 1261, 1222, 1181, 1143, 1128, 1107, 1080, 1047, 1024, 971, 947, 931, 879, 839, 805. ^{1}H NMR (D_2O, 400 MHz) δ: 7.94 (s, 3H, triazole), 5.17 (s, 6H, CH_2COOEt), 4.26 (q, 6H, J = 7.17 Hz, $COOC\underline{H}_2CH_3$), 3.87 (s, 6H, $NC\underline{H}_2$-triazole), 1.29 (t, 9H, J = 7.17 Hz, $COOCH_2C\underline{H}_3$).

LRMS (ESI) Calculated: [M+H] $C_{21}H_{32}N_{10}O_6$ (m/z) : 519.2, Obtained: $[M+Na]^+$: 541.2, Melting pt. 106.2℃

4.9.4: SILAC and azidohomoalanine labeling

Primary human $CD4^+$ T lymphocytes were isolated from blood by negative selection (Dynal isolation kit, Invitrogen). Primary T lymphocytes were rested overnight in RPMI 1640 medium (PAA) supplemented with 10% FBS (Gibco). Purity of isolated cells was assessed by analyzing CD3 and CD4 surface expression using flow cytometry (data not shown). For metabolic labelling, cells were incubated for 1 h at 37 °C in Met-, Arg- and Lys- free RPMI medium (Dundee Cell Products) supplemented with 10% dialyzed FBS at a cell density of 2×10^6 mL^{-1}. After centrifugation, cells cultures were divided into two sets by equal numbers. One set of cells was incubated in fresh Met-, Arg- and Lys- free RPMI medium supplemented with 1.14 mM 'medium' arginine ($[^{13}C_6]$L-Arg) and 0.219 mM 'medium' lysine ($[^2H_4]$L-Lys), while the other set of cells was incubated in Met-, Arg- and Lys- free RPMI supplemented with 1.14 mM 'heavy' arginine ($[^{13}C_6,^{15}N_4]$L-Arg) and 0.219 mM 'heavy' lysine ($[^{13}C_6,^{15}N_2]$L-Lys) (CK Gas). To minimize arginine-to-proline conversion, all media was supplemented with 1.7 mM L-proline. Both cell populations were supplemented with 1 mM Aha. Cells in the heavy medium were stimulated with 62.5 ng/mL phorbol 12-myristate 13-acetate (PMA, 100 nM final concentration) and 1 μg/mL ionomycin (iono, 1.4 μM final concentration). Efficient stimulation was verified by CD69 surface staining and flow cytometry (data not shown). Cells were incubated for 2 h and 4 h at 37 °C before harvesting by centrifugation.

4.9.5: Protein extraction, reduction and alkylation.

For each replicate, 90×10^6 'heavy'-RPMI incubated stimulated and 90×10^6 'medium'-RPMI incubated unstimulated labeled primary cells were collected by centrifugation, washed with PBS, and combined. Cells were lysed in 500 μL PBS with 1% SDS containing protease and phosphatase inhibitors (Complete EDTA-free, Roche, 2 mM sodium pervanadate) by rigorous vortexing. The lysate was boiled at 100 °C for 10 min, after which 1,000 units of benzonase (Sigma) was added, and the sample incubated at room temperature for 15 min. The lysate was clarified by centrifugation at 20,000*g* for 10 min at 4 °C. 10 mM tris(2-carboxyethyl) phosphine (TCEP) (Thermo) was added, and the sample incubated at 55 °C for 45 min to reduce proteins. Samples were then cooled to room temperature, and pH 8.0 with sodium hydroxide before addition of iodoacetamide to a final concentration of 18 mM. Free cysteine

side chains were alkylated by incubation at room temperature for 30 min in the dark. Each lysate was desalted before biotinylation using spin desalting columns (Zeba 7-kDa MWCO, Thermo).

4.9.6: Copper-catalyzed azide-alkyne cycloaddition of protein samples with alkynyl-biotin

The protocol was used as previously reported[17] and was applied to all Aha-labelled samples. A freshly prepared solution of copper(I) bromide (99.999% purity) in acetonitrile (163 μL, 10 mg/mL) was premixed with an acetonitrile solution of tris-triazolyl amine ligand (63 μL, 127 mg/mL). For CuAAC, 7 μL of an alkyne-functionalised biotin with a cleavable disulfide linker[26] (25 mM stock solution in PBS at pH 7.8, 87.5 nmol) was added to 540 μL of reduced, alkylated and desalted protein lysate from each replicate and mixed. The preformed Cu(I)-ligand complex solution (30 μL) was then added to the mixture and mixed thoroughly. The reaction was incubated for 90 min at room temperature on a rotator. Tubes were then centrifuged at 2,000*g* for 5 min, and the supernatant purified by desalting columns (Zeba 7 kDa MWCO).

4.9.7: Pull-down and enzymatic digest of biotinylated proteins.

For each sample, 40 μL of high-capacity Neutravidin resin (Thermo) was washed three times in 1 mL of PBS containing 0.05% SDS. After the final wash, the supernatant was discarded, and 577 μL of the desalted lysate was added to the eppendorfs containing the beads. The pull-down was carried out by incubation at room temperature for 16 h. Samples were then loaded onto Pierce Spin Columns (Thermo), and beads collected by centrifugation. Beads were washed three times with 600 μL of each of the following buffers: 4 M urea, 6 M urea, 10% acetonitrile and 20% acetonitrile. For each washing step, beads were incubated for 5 min before they were centrifuged and collected. Proteins were then eluted from the Neutravidin beads through reductive cleavage of the biotin linker as follows: 50 μL of elution buffer (20 mM DTT, 1 M urea, 50 mM NH_4HCO_3) was added to the column and incubated with the beads for 10 min before spinning down and collecting the eluted proteins. This step was repeated twice. $CaCl_2$ was added to a final concentration of 1.25 mM in the eluate before overnight in-solution digestion of proteins with trypsin at 37 °C. Peptides were desalted using C18 desalting tips. Desalted, dried peptides were then resuspended in 0.1% TFA with 2% acetonitrile and sonicated before analysis by mass spectrometry.

4.9.8: Mass spectrometry and data analysis.

Samples were analyzed on an Ultimate 3000 RSLCnano (Dionex) system run in direct injection coupled to a QExactive mass spectrometer (Thermo Scientific). Samples were resolved on a NanoEASY (Thermo Scientific) C_{18} reverse-phase column (50 cm long, 75 μm internal diameter, 2 μm beads) at a flow rate of 300 nl min^{-1} using a linear gradient of 5–44% buffer B (80% acetonitrile and 0.1% formic acid) over 190 min. The mass spectrometer was operated in a 'top 10' data-dependent acquisition mode with dynamic exclusion enabled (40 s). Survey scans (mass range 300–1,650 Th) were acquired at a resolution of 70,000 at 200 Th with the ten most abundant multiply charged ($z \geq 2$) ions selected with a 3-Th isolation window for fragmentation by higher-energy collisional dissociation. MS/MS scans were acquired at a resolution of 17,500 at 200 Th.

Raw data was processed using MaxQuant[28] Ver. 1.3.0.5 and Proteome Discoverer 1.3 (Thermo Scientific). Searches against the UniProt[35] human database were performed using Andromeda[36] and Mascot,[37] respectively. Search parameters allowed for two missed trypsin cleavage sites, with carbamidomethylation as a fixed modification on cysteine, and methionine oxidation and N-terminal protein acetylation as variable modifications. Peptide results were filtered to 1% false discovery rate by MaxQuant and Proteome Discoverer.

Protein and peptide quantification information were extracted from MaxQuant and Proteome Discoverer and imported into the program R Ver 2.15.1.[38] Protein extracted ion chromatograms (XICs) were used from MaxQuant without additional modification. For Proteome Discoverer, 'medium' and 'heavy' protein XICs were calculated from the sum of the individual unique peptides XICs in the program R.[39] Peptides with zero medium and heavy XICs were not considered for the calculation. Proteins that were not reproducibly quantified across replicates were removed from downstream analysis. Proteins were required to have a minimum of three quantifiable unique peptides in at least two of the three biological replicates for consideration.

Differential protein expression analysis was performed with the Bioconductor package LIMMA Ver 3.14.1. After protein area distributions were quantile-normalized,[40] a linear model was fitted and a moderated *t*-test used to assess the statistical significance of heavy-to-medium protein fold changes. Proteins with a corresponding fold-change *P* value (adjusted for multiple hypothesis testing with the Benjamini-Hochberg method[41]) lower than 0.05 were

accepted as differentially expressed. MetaCore (GeneGo Inc.)[42] was used to generate an interaction network of differentially expressed proteins, which was visualized in Cytoscape.[43]

4.9.9: Biosynthetic labelling by ^{35}S-Cysteine pulse and pulse-chase

Jurkat T-cells were centrifuged and washed once in warm RPMI without L-methionine and L-cystine (R7513 SIGMA), supplemented with 10 % dialyzed FBS (Gibco) and L-glutamine (PAA). After incubation in Met- and Cys-free RPMI for 1 hour at 37 °C in the same media at a cell density of 1×10^6 /ml, equal numbers of cells were centrifuged in four sets. Two sets of cells (2×10^6 /ml) were placed in fresh, warm Met- and Cys-free RPMI media containing L-Azidohomoalanine (Aha) at a final concentration of 1 mM or 4 mM, while the other sets of cells were placed in media containing 0.1 mM L-methionine or without methionine as a control. The media for all sets was simultaneously supplemented with 10 nM (10 μCi/ mL) L-[^{35}S]-cysteine (PerkinElmer) and 5 μM L-cystine (from RPMI SILAC media, Dundee Cell Products). A sample of 2×10^6 cells of each replicate per condition was taken immediately as a reference for 0 hours of "pulse". The pulse was performed at 37 °C and samples for each replicate per condition were harvested after 2 and 4 hours. Cell lysates for protein precipitation were prepared as follows: cells were centrifuged and washed twice with 1 mL cold RPMI-1640 (PAA) supplemented with 10 % FBS (HyClone Thermo), then twice with 1.2 mL cold 50 mM iodoacetamide in PBS. Cell pellets were lysed for 10 minutes in 25 μL 1% SDS in PBS containing protease and phosphatase inhibitors (Complete EDTA-free, Roche, 2mM sodium pervanadate) and benzonase (100 Units). The samples were then centrifuged at 20,000g for 10 minutes and the supernatant was stored on ice.

All replicates for each time point were spotted together on a Watman filterpaper, which was then incubated three times for 5 minutes with 5 % trichloroacetic acid for protein precipitation. After rinsing the filter paper with 70 % ethanol and 100 % acetone, each sample was placed in 4 mL scintillation fluid (Optiphase Supermix, Perkin Elmer) and counted in a scintillation counter for 3 minutes (Packard TriCarb 2900TRLiquid Scintilliation Analyser, GMI). For the analysis of protein degradation, Jurkat T-cell were washed and starved as described for the pulse experiment. After starvation, two sets of cells (2×10^6 /ml) were placed in fresh, warm RPMI starvation media (see above) containing Aha at a final concentration of 1 mM while the other set was placed in media containing 0.1 mM L-methionine. The media for both sets was simultaneously supplemented with 32.4 nM (30 μCi/ml) L-[^{35}S]-cysteine (PerkinElmer) and 16.2 nM L-cystine (RPMI SILAC media, DundeeCell). The pulse was

carried out for 15 minutes at 37 °C. Cells were then centrifuged for 5 minutes and washed once with RPMI-1610 media (PAA) lacking L-[^{35}S]-cysteine and supplemented with 10 % FBS (HyClone Thermo). The "chase" of L-[^{35}S]-cysteine was carried out by resuspending the cells in fresh warm RPMI-1640 media at a density of 2×10^6 /ml. A sample of 2×10^6 cells for each replicate per condition was taken immediately as a reference before the chase took place. For each replicate, 2×10^6 cells were harvested after 1, 2 and 4 hours chase. Preparation of cell lysates, protein precipitation and measurement of radioactivity was performed as described previously.

4.10: References

(1) Ghazalpour, A.; Bennett, B.; Petyuk, V. A.; Orozco, L.; Hagopian, R.; Mungrue, I. N.; Farber, C. R.; Sinsheimer, J.; Kang, H. M.; Furlotte, N.; Park, C. C.; Wen, P. Z.; Brewer, H.; Weitz, K.; Camp, D. G.; Pan, C.; Yordanova, R.; Neuhaus, I.; Tilford, C.; Siemers, N.; Gargalovic, P.; Eskin, E.; Kirchgessner, T.; Smith, D. J.; Smith, R. D.; Lusis, A. J. *PLoS Genet* **2011**, *7*, e1001393.
(2) Foss, E. J.; Radulovic, D.; Shaffer, S. A.; Ruderfer, D. M.; Bedalov, A.; Goodlett, D. R.; Kruglyak, L. *Nat Genet* **2007**, *39*, 1369.
(3) Rogers, S.; Girolami, M.; Kolch, W.; Waters, K. M.; Liu, T.; Thrall, B.; Wiley, H. S. *Bioinformatics* **2008**, *24*, 2894.
(4) Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. *Nature* **2011**, *473*, 337.
(5) Mann, M. *Nat Rev Mol Cell Biol* **2006**, *7*, 952.
(6) Ong, S. E.; Foster, L. J.; Mann, M. *Methods* **2003**, *29*, 124.
(7) Everley, P. A.; Krijgsveld, J.; Zetter, B. R.; Gygi, S. P. *Mol Cell Proteomics* **2004**, *3*, 729.
(8) Everley, P. A.; Bakalarski, C. E.; Elias, J. E.; Waghorne, C. G.; Beausoleil, S. A.; Gerber, S. A.; Faherty, B. K.; Zetter, B. R.; Gygi, S. P. *J Proteome Res* **2006**, *5*, 1224.
(9) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol Cell Proteomics* **2002**, *1*, 376.
(10) Grønborg, M.; Kristiansen, T. Z.; Iwahori, A.; Chang, R.; Reddy, R.; Sato, N.; Molina, H.; Jensen, O. N.; Hruban, R. H.; Goggins, M. G.; Maitra, A.; Pandey, A. *Mol Cell Proteomics* **2006**, *5*, 157.
(11) Geiger, T.; Wisniewski, J. R.; Cox, J.; Zanivan, S.; Kruger, M.; Ishihama, Y.; Mann, M. *Nat Protoc* **2011**, *6*, 147.
(12) Schwanhäusser, B.; Gossen, M.; Dittmar, G.; Selbach, M. *Proteomics* **2009**, *9*, 205.
(13) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proceedings of the National Academy of Sciences* **2006**, *103*, 9482.
(14) Dieterich, D. C.; Hodas, J. J.; Gouzer, G.; Shadrin, I. Y.; Ngo, J. T.; Triller, A.; Tirrell, D. A.; Schuman, E. M. *Nat Neurosci* **2010**, *13*, 897.
(15) Dieterich, D. C.; Lee, J. J.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Nat Protoc* **2007**, *2*, 532.

(16) van Kasteren, S. I.; Kramer, H. B.; Jensen, H. H.; Campbell, S. J.; Kirkpatrick, J.; Oldham, N. J.; Anthony, D. C.; Davis, B. G. *Nature* **2007**, *446*, 1105.
(17) van Kasteren, S. I.; Kramer, H. B.; Gamblin, D. P.; Davis, B. G. *Nat Protoc* **2007**, *2*, 3185.
(18) Abraham, R. T.; Weiss, A. *Nat Rev Immunol* **2004**, *4*, 301.
(19) Simmons, R.; Sharp, C.; Sims, S.; Kloverpris, H.; Goulder, P.; Simmonds, P.; Bowness, P.; Klenerman, P. *J Infect Dis* **2011**, *203*, 1378.
(20) Pasquevich, K. A.; Estein, S. M.; García Samartino, C.; Samartino, C. G.; Zwerdling, A.; Coria, L. M.; Barrionuevo, P.; Fossati, C. A.; Giambartolomei, G. H.; Cassataro, J. *Infect Immun* **2009**, *77*, 436.
(21) Chatila, T.; Silverman, L.; Miller, R.; Geha, R. *J Immunol* **1989**, *143*, 1283.
(22) Banerjee, P. S.; Ostapchuk, P.; Hearing, P.; Carrico, I. S. *J Virol* **2011**, *85*, 7546.
(23) Taskent-Sezgin, H.; Chung, J.; Banerjee, P. S.; Nagarajan, S.; Dyer, R. B.; Carrico, I.; Raleigh, D. P. *Angew Chem Int Ed Engl* **2010**, *49*, 7473.
(24) Merkel, L.; Beckmann, H. S.; Wittmann, V.; Budisa, N. *Chembiochem* **2008**, *9*, 1220.
(25) Ma, Y.; Biava, H.; Contestabile, R.; Budisa, N.; di Salvo, M. L. *Molecules* **2014**, *19*, 1004.
(26) Szychowski, J.; Mahdavi, A.; Hodas, J. J.; Bagert, J. D.; Ngo, J. T.; Landgraf, P.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *J Am Chem Soc* **2010**, *132*, 18351.
(27) Oda, Y.; Nagasu, T.; Chait, B. T. *Nat Biotechnol* **2001**, *19*, 379.
(28) Cox, J.; Mann, M. *Nat Biotechnol* **2008**, *26*, 1367.
(29) Ruppen, I.; Grau, L.; Orenes-Piñero, E.; Ashman, K.; Gil, M.; Algaba, F.; Bellmunt, J.; Sánchez-Carbayo, M. *Molecular & Cellular Proteomics* **2010**, *9*, 2276.
(30) Diehn, M.; Alizadeh, A. A.; Rando, O. J.; Liu, C. L.; Stankunas, K.; Botstein, D.; Crabtree, G. R.; Brown, P. O. *Proceedings of the National Academy of Sciences* **2002**, *99*, 11796.
(31) Ulrey, C. L.; Liu, L.; Andrews, L. G.; Tollefsbol, T. O. *Hum Mol Genet* **2005**, *14 Spec No 1*, R139.
(32) Waterland, R. A. *J Nutr* **2006**, *136*, 1706S.
(33) Eichelbaum, K.; Winter, M.; Berriel Diaz, M.; Herzig, S.; Krijgsveld, J. *Nat Biotechnol* **2012**, *30*, 984.
(34) Zhou, Z.; Fahrni, C. J. *J. Am. Chem. Soc.* **2004**, *126*, 8862.
(35) Consortium, U. *Nucleic Acids Res* **2012**, *40*, D71.
(36) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. *J Proteome Res* **2011**, *10*, 1794.
(37) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551.
(38) Ihaka, R.; Gentleman, R. **1996**, *5*, 299.
(39) Carrillo, B.; Yanofsky, C.; Laboissiere, S.; Nadon, R.; Kearney, R. E. *Bioinformatics* **2010**, *26*, 98.
(40) Smyth, G. K.; Speed, T. *Candidate Genes from DNA Array Screens: application to neuroscience* **2003**, *31*, 265.
(41) Benjamini, Y.; Hochberg, Y. **1995**, *57*, 289.
(42) Nikolsky, Y.; Ekins, S.; Nikolskaya, T.; Bugrim, A. *Toxicol Lett* **2005**, *158*, 20.
(43) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res* **2003**, *13*, 2498.

Chapter 5

Outlook

5.1: Introduction

Despite intensive research being carried out in the field of UAA incorporation into biological systems, we may have yet only scratched the surface of what information this powerful technique yields. In this chapter, initial results in the application of UAA incorporation in a variety of biological systems will be discussed. Carried forward, this research has the potential to make significant impacts diverse fields such as bio-imaging and immunology.

5.2: Unnatural amino acids as contrast-forming agents in Confocal Raman Microscopy

When light is incident on a chemical sample, a small fraction of it is inelastically scattered to produce photons which are shifted in energy, a phenomenon known as the Raman effect. Raman spectra so obtained are dependent upon the rotational and vibrational modes of the molecules, and the changes which occur in their electric polarisabilities upon excitation. Raman spectroscopy thus provides structural information about the molecules being probed, as molecular vibrational modes carry information about molecular structure and symmetry. Confocal Raman microscopy is a technique which combines vibrational spectral information of molecules within the sample to be imaged, with optical information from confocal microscopy. In biological systems, the technique has been greatly utilised as confocal microscopy setups impart the Raman microscope with spatial resolutions of less than 1 μm,[1] and had been used to distinguish between closely-related bacterial species,[2] and image fibrotic heart tissue without the use of stains.[3]

Raman microscopy may also be used for the imaging of molecular distributions in cells. This can be done by raster-scanning the cells for characteristic Raman signals of the molecules of interest. The spectroscopic data is then combined with the spatial information of the spectral source to produce a false-colour image, known as a chemical map or univariate map. The chemical map thus allows the visualisation of the molecule's location.[4] To improve the contrast of the chemical map, the vibrational spectrum of the molecule of interest should be significantly different from the cellular background signal consisting of the vibrations of the

various nucleic acids, proteins, carbohydrates and lipids.[5] For this purpose, they may be tagged with specific groups that confer them with a unique vibrational fingerprint, such as heavy isotopes.[6,7,8] An alternative approach to improve contrast involves tagging the molecule of interest with a non-natural functional group, possessing a characteristic vibrational signal in a region of the spectrum where biologically relevant groups are spectroscopically silent. With the use of Raman microscopy, these systems may then be imaged non-invasively.

Via the installation of L-Azidohomoalanine (Aha), the azide group can be installed onto proteins in Met-auxotrophic *E. Coli* B834(DE3) cells,[9] and a range of mammalian cells.[10,11] The characteristic vibrational signal for its side-chain azide group lies at 2100 cm^{-1}, which is an otherwise biologically Raman-inactive region.[5] The azide group has also been utilised previously as a vibrational tag in proteins,[12] and its vibrational signal is both conformation and environment sensitive. This makes Aha an ideal vibrational Raman tag to investigate protein localisation and folding inside cells. The Nostoc punctiforme protein fusion of Np275 and Np276 (hereafter referred to as 'Np276'), a pentapeptide repeat protein[13] was chosen as the azide carrier due to its regular repeating structure. Three different mutants of Np276 were used. Of these, two mutants possessed single sites for Aha incorporation at different locations, while the third possessed two sites for Aha incorporation.

Np276 mutants were expressed containing Aha at either position 61, 101, or both. Incorporation was carried out by Met-codon reassignment using the Selenomet system (section **2.11.12**, Fig. 5.1 to 5.3).

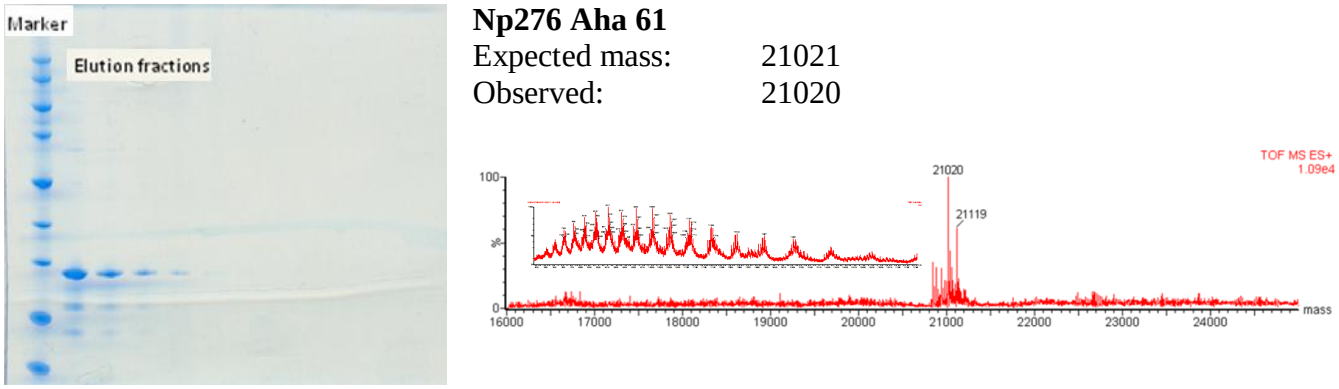
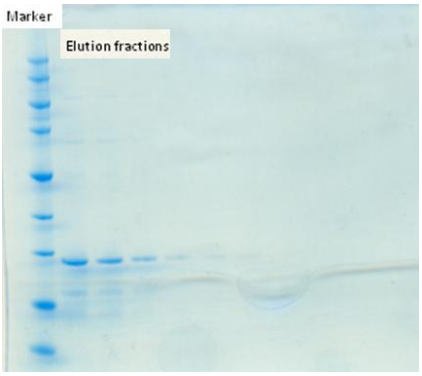


Figure 5.1: SDS PAGE and LC-MS for Np276 Aha 61



Np276 Aha 101
Expected mass: 21021
Observed: 21020

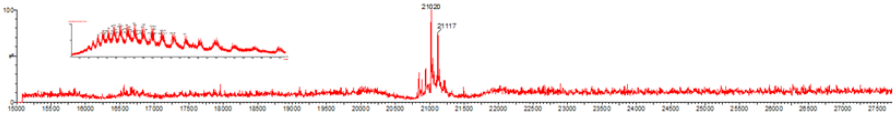
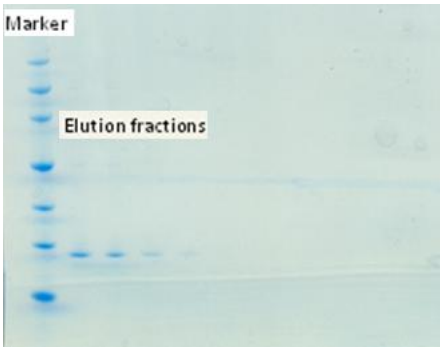


Figure 5.2: SDS PAGE and LC-MS for Np276 Aha 101



Np276 Aha61, Aha101
Expected mass: 20998
Observed: 20999

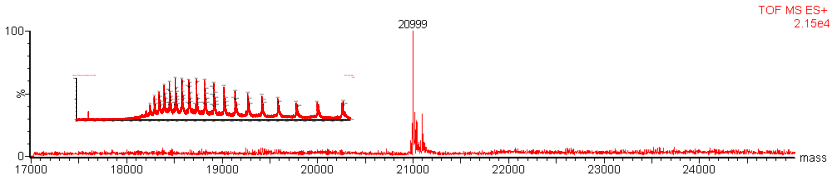


Figure 5.3: SDS PAGE and LC-MS for Np276 Aha61, Aha101

5.3: Raman data acquisition and imaging of *E. coli* B834(DE3) cells

All acquisition of Raman spectra was done at Begbroke Science Park in the lab of Prof. Alison Crossley and Prof. Colin Johnston. Protein and cell samples were stored in dry ice for transfer from the CRL to Begbroke Science Park. On arrival, the samples were collected, and allowed to thaw at room temperature. One drop of each sample was fixed onto a cleaned glass slide by drying with compressed air.

The slide was then mounted on an X-Y optical stage (75 x 50 mm, Accuracy ± 3µm, reproducibility < 1 µm, 0.1 µm step size) of a JY Horiba Labram Aramis Imaging Confocal microscope, which was coupled to a 460 mm focal length spectrograph. The sample was irradiated with either a 532 nm Nd:YAG 50mW laser, or a HeNe 633 nm 17mW laser. The objective lenses used were 50x (0.45 NA) and 100x (0.9 NA) objectives. Back-scattered photons were passed through an 1800 groove/mm diffraction grating, and the Raman-shifted photons were detected by a -70ºC air cooled multichannel CCD detector (1024 x 256 pixel).

Confocal pinhole diameter and spectrographic slit width used were 300 µm and 150 µm respectively. All spectra were recorded and analysed using LabSpec Ver. 5.28.25.

Both solid Aha as well as a dried film of Np276 Aha61 showed the presence of a band at 2100 cm-1, which was assigned to the stretching frequency of the side-chain azide (Fig. 5.4).

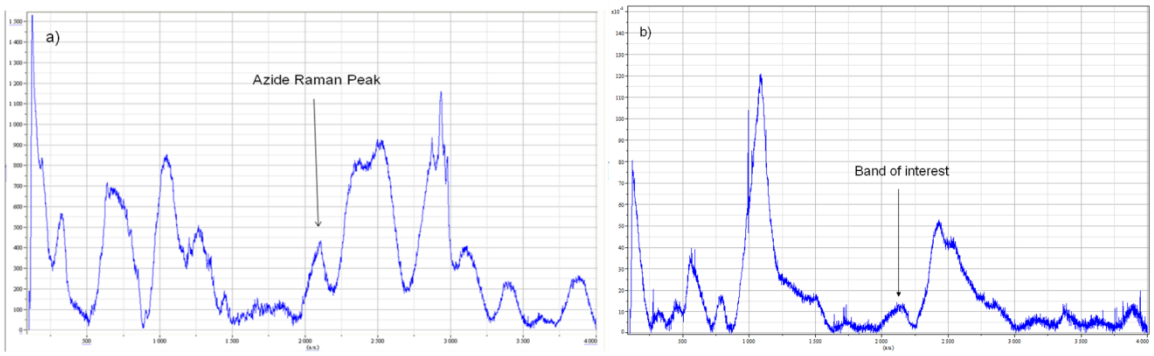


Figure 5.4: a) Raman spectrum of Aha, with location of Raman signal corresponding to azide indicated, b) Raman spectrum of Np276 Aha61, with the same band highlighted. All spectra have been baseline corrected.

E. coli cells expressing Aha-labelled Np276 were imaged at focal planes where their cross-section was largest, to ensure signal acquisition from the cellular interior. The broad cellular fluorescence signal was subtracted by fitting the baseline to an 8th order polynomial, followed by subtraction of this polynomial from the recorded spectrum. Comparison of spectra was done by normalisation of the spectral area after baseline subtraction. Approximately 35 minutes were needed to acquire a 3x3 pixel spectral map of a single *E. coli* cell at a laser power of 25 mW. However, previous reports suggest that it is possible to acquire whole-spectrum cell signals (100 to 4000 cm^{-1}) cell signals with acceptable signal-to-noise ratios with acquisition times as low as 0.2 to 0.5 ms per pixel at 10 mW laser power.[8,14]

The Raman spectrum of Np276 Aha61 Aha101 was recorded at varying acquisition times and laser exposure times per spectral window (Fig. 5.5, 490 cm^{-1} spectral window width, 8 windows needed to record whole-spectra from 100 to 4000 cm^{-1}).

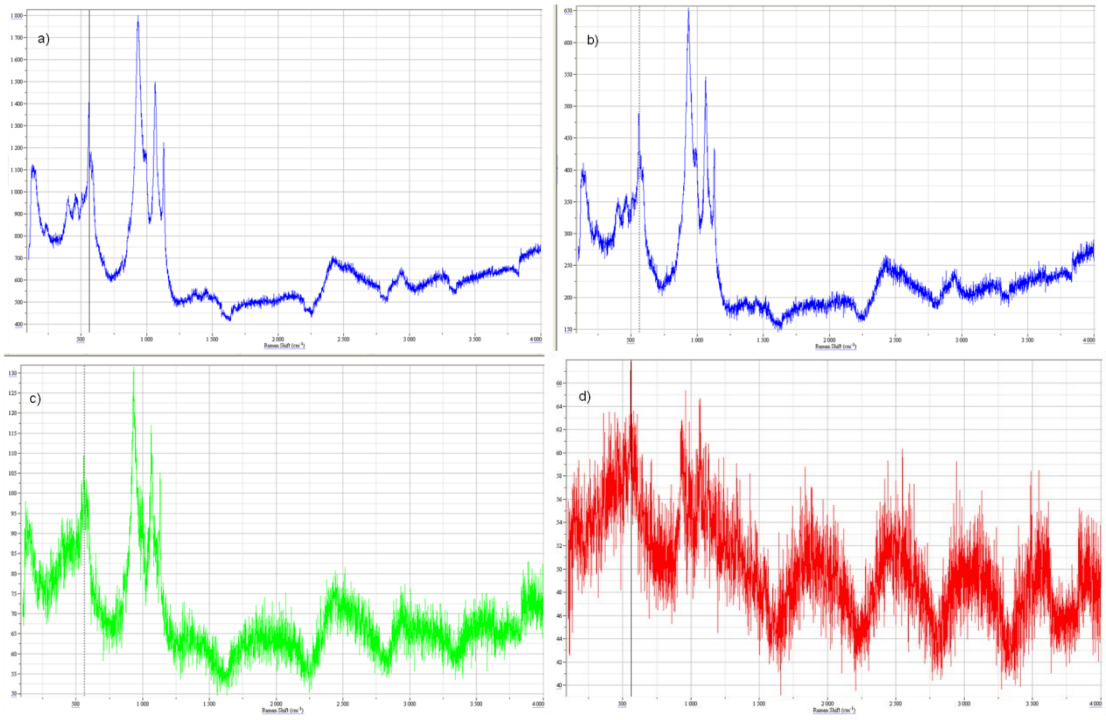


Figure 5.5: Raman spectra of Np276 Aha61 Aha 101, a) 4 min acquisition, 10 s exposure per spectral window, 2 accumulations; b) 1 min acquisition, 3.5 s exposure per spectral window, 2 accumulations; c) 15s acquisition, 0.5 s exposure per spectral window, 2 accumulations and d) 10s acquisition, 0.1 s exposure per spectral window, 2 accumulations. All acquisitions were performed at full power of 532 nm laser.

These results indicated that the detector of the Labram was not sensitive enough to acquire whole-spectrum (100 to 4000 cm^{-1}) Raman signals with acceptable signal-to-noise ratios if the acquisition time was less than 1 minute for each data point, at 50% laser power (~25 mW). For a 3x3 pixel grid per *E. coli* cell, the theoretical minimum accumulation time would be 1 min x 2 accumulations x 9 pixels = 18 minutes. A representative 3x3 pixel map, superimposed on an optical image of E. coli is shown below (Fig. 5.6).

Figure 5.6**: S**patial distribution of Raman signal from 2000 to 2200 cm^{-1} in a 3x3 pixel map of *E. coli*

Imaging time scales linearly with the area of the sample to be imaged. Applying this protocol to a mammalian cell with specialised internal structures where protein maps would yield more significant information, would therefore require a concomitant increase in laser exposure time. For example, imaging a CHO cell with 14 μM diameters would require ~88 times longer exposure times as compared to an *E. coli* cell with a 1.5 μM diameter.

When imaged for long exposure times or higher laser power (in order to achieve better signal to noise ratio), *E. coli* cells were seen to be damaged, as evidenced by discolouration of the fixed sample and drastic alterations in raman signals. Reducing the 532 nm laser power, or switching to the less energetic 633 nm laser, lead to reduced signal-to-noise and even longer spectral acquisition time.

5.4: Raman microscopy outlook

These results indicate that the Raman signal produced by Aha-labelled Np276 is not strong enough to be detected at rates faster than of the order of a few minutes per spectrum. However, the technique used here, utilising the spontaneous Raman scattering from the sample, has since been widely replaced with stimulated Raman scattering,[15] or coherent anti-stokes Raman scattering techniques,[16-18] to improve signal strengths for Aha-based contrast formation. Furthermore, alkyne groups also possess a characteristic Raman vibrational signal at ~2100 cm^{-1}, and have been successfully utilised to label proteins via the unnatural amino acid and Met-analogue homopropargylglycine (Hpg).[19] Alkyne-containing molecules have also been used to label nascent DNA via 5-ethynyl-2'-deoxyuridine (EdU),[20] and nascent RNA via 5-ethynyluridine (EU).[21] The use of these alternative labels and imaging modes may prove useful in the development of minimally-labelled Raman microscopy of biological systems.

5.5 Towards a new HIV vaccine by chemical cross-linking of viral envelope glycoprotein gp140:

This work was carried out in collaboration with Torben Schiffner and Prof. Quentin Sattentau at the Dunn School of Pathology and Guillaume Stewart-Jones at the Weatherall Institue of Molecular Medicine.

Envelope glycoprotein gp120 is expressed on the viral envelope of HIV, and is involved in the infection mechanism of the virus via binding to host cell CD4 receptors.[22] gp120 is tethered to the viral surface as a trimer via gp41, both of which are cleavage products formed by the processing of the precursor glycoprotein gp160 by furin. gp41 mediates viral and host cell membrane fusion.[23] However, recombinant gp120 expresses as a monomer, and itself does not elicit neutralising antibodies against primary isolates of HIV.[24] The corresponding construct gp140 consists of gp120 covalently bonded to gp41 as well as truncation of the transmembrane domain rendering it soluble. This conjugation of gp120 and gp41 may be done either through the deletion or mutation of the cleavage site residues resulting in a fusion protein,[25,26] or by introducing an additional disulfide bridge between the cleaved products.[27] Both approaches generate proteins which express as trimers, and are better at inducing neutralising antibodies than their respective monomeric gp120 fragments.

This project aimed to improve the antigenicity of HIV envelope protein gp140 via chemical cross-linking of amino acids on the protein surface. This was attempted via the directed incorporation of Aha into gp140 at selected incorporation sites, followed by the copper-catalysed azide-alkyne cycloaddition (CuAAC) reaction to cross-link these species to produce a more rigid trimeric protein framework for recognition by the immune system. Though mutation-based work on conformational stabilisation of gp120s have been attempted previously,[28-30] this is to our knowledge the first work in this area which seeks to utilise CuAAC to impart structural rigidity.

In order to generate single sites for Aha incorporation, sites containing Methionine and Isoleucine on regions corresponding to gp120 and gp41 on gp140 based on cryo-ET models by Liu et al were identified (Fig. 5.7).[31] Methionines or Isoleucines which were both solvent accessible as well as proximal to the trimerisation contact region were considered, while those buried in regions of high glycosylation were excluded.[32] Of these, a conserved, solvent-

accessible methionine (Met 434, Fig. 5.8) on each monomer at the apex of the trimer spike was chosen for Aha incorporation.

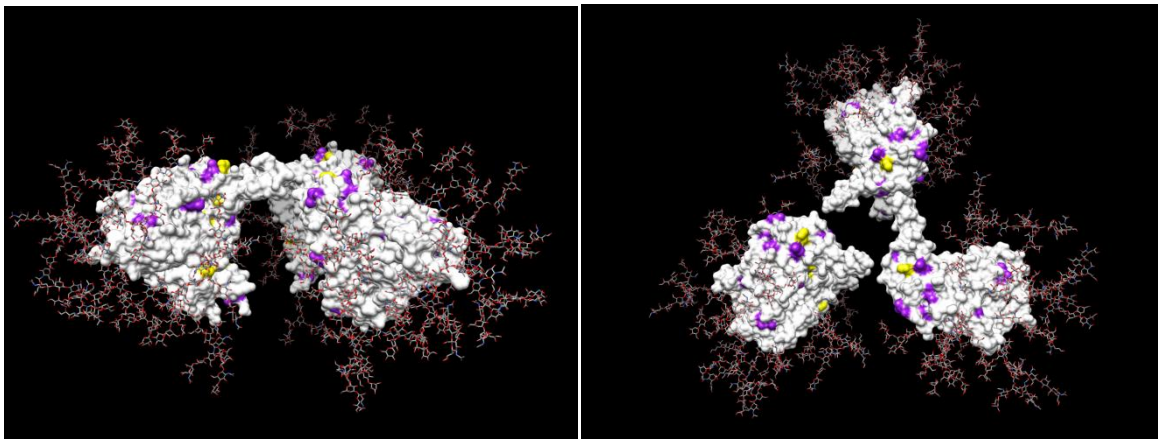


Figure 5.7: Isoleucines (Purple) and Methionines (Yellow) on gp120, 'side' and 'top' view

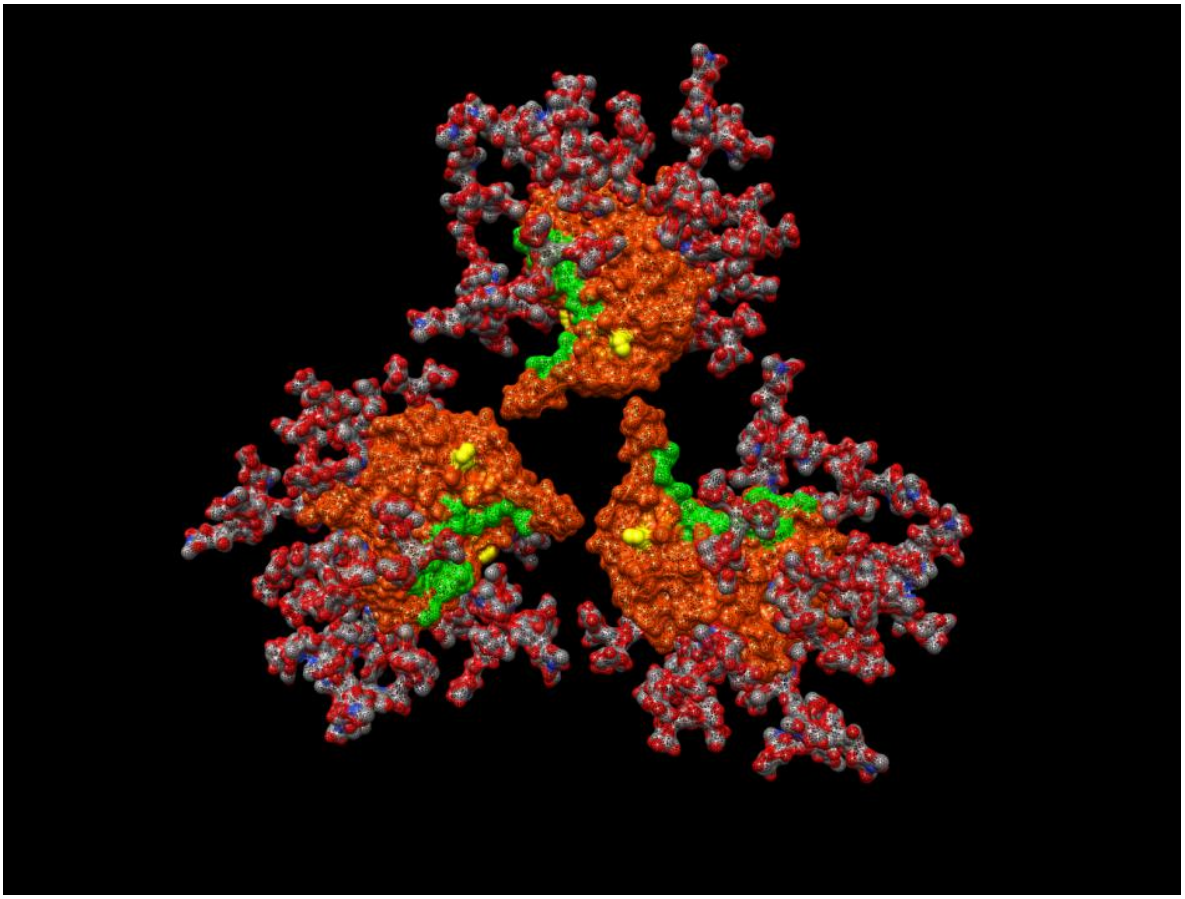
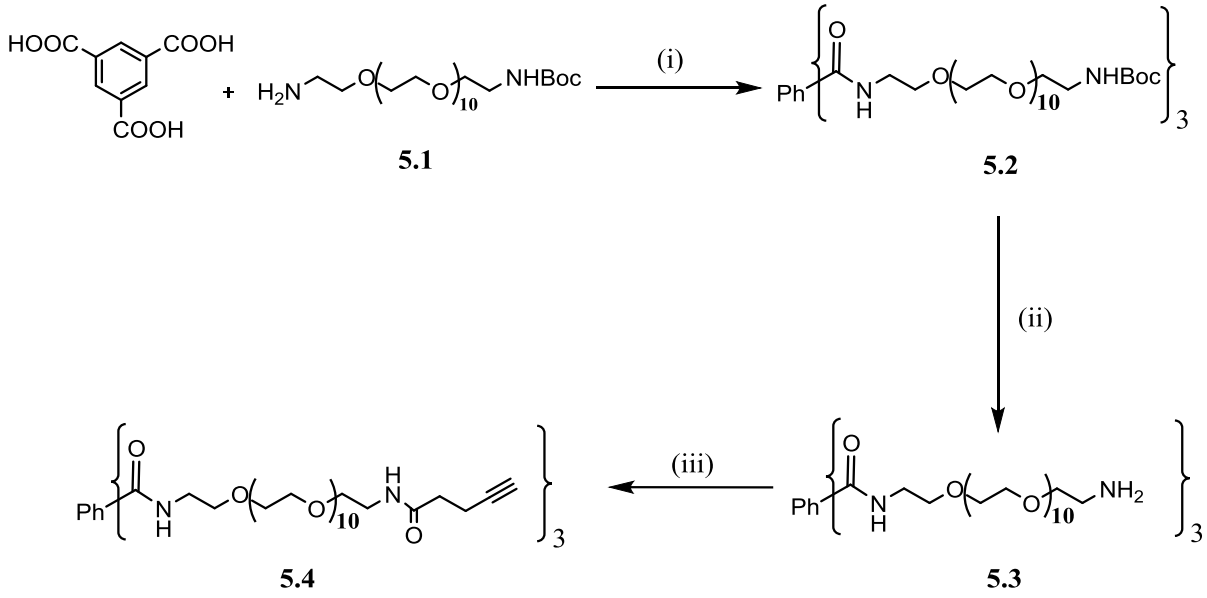


Figure 5.8: Met434 highlighted on trimer apex (Yellow) on gp140

A gene for UG37 gp140[33] was ordered from Genscript retaining Met434, with the remaining Met replaced with Ile. The tolerance of proteins toward the mutation of Met to Ile, has previously been demonstrated.[34] A complementary trimeric water-soluble linker was then synthesised, triply functionalised with an alkyne (Scheme 5.1). The distance between the the

side chains of Met434, and therefore incorporated Aha434, between monomers in the gp140 trimer was 46.6Å. Thus, linker arms of 48Å were synthesised. A PEG-based linker system was chosen to in order to be non-toxic and non-immunogenic by itself.[35]



Scheme 5.1: Synthesis of trimeric alkyne linker for stabilisation of gp140 (i): HATU, DIPEA, DMF, 70˚C 18h, (ii) a) TFA, DCM, b) 2N HCl (iii) Pent-4-ynoic acid, HATU, DIPEA, DMF, 60˚C, 8h. (R = Phenyl)

The triplex alkyne was then used for on-protein CuAAC, to form a covalently bonded 'cap' on the gp140, making the resultant protein less conformationally flexible.

An alternative approach was also outlined for this project, utilising two incorporation sites on each monomer. These sites were chosen such that each was proximal to corresponding sites on the two adjacent monomers. In this case, Met95 and Ile213 on each monomer were spatially proximal to Ile213 and Met95 of adjacent monomers. After mutating Ile213 to Met, the cells would be fed with a growth medium containing both azide-containing Aha as well as alkyne-containing Hpg. Statistically, an incorporation pattern would be expected biased by the incorporation efficiency of Aha and Hpg.. Nevertheless, some homo-trimers formed would contain one Aha and one Hpg, and would be able to be covalently bonded to each other via CuAAC. Also, hetero-trimers formed from a doubly Aha-substituted monomer, a doubly Hpg-substituted monomer and a single Aha and Hpg substituted monomer would also form a similar complex upon reaction, utilising a much shorter dimeric linker.

Once cross-linked, the timers would be tested for binding to known antibodies, that are known to bind to the CD4bs region,[36-38] which is distant from our proposed sites of incorporation and reaction.

However, since the models used herein contain truncated forms of the variable V1/V2 loop, one concern was that in the actual gp140, the full-length loops might sterically inhibit the approach of the reagents for CuAAC.[39] However, previous work has shown that the V1/V2 loop serves to shield the V3 loop of a neighbouring gp120 monomer,[40] and that major conformational changes upon binding to CD4 may cause the loops to migrate away from the apex,[41] potentially leaving the residues free for reaction.

The protocol for the purification of HIV env glycoproteins was provided by Dr. Guillaume Stewart-Jones and has been outlined in brief in section **5.11.2**. This procedure involved the growth of HEK 293T cells,[42] which were transiently transfected with env gp-coding plasmids. The glycoproteins were expressed in the media, and are purified by His-tag affinity chromatography (using Talon resin), glycan-affinity chromatography (using lectin beads),[43] followed by size-exclusion chromatography. gp140 constructs remained uncleaved in the absence of furin during expression in HEK 293T cells.

An aliquot of the expressed gp140 protein (~140 KDa) was also treated with PNGase-F (NEB) to remove all N-linked glycans to yield a ~70 KDa deglycosylated protein as confirmed by SDS-PAGE (Fig. 5.9).

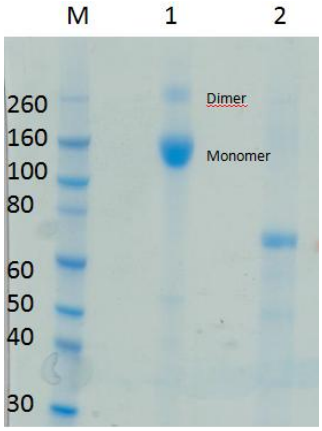


Fig. 5.9: SDS-PAGE: M: Protein marker, 1: gp140 Aha 434 in non-reducing gel-loading buffer, 2: After treatment with PNGase-F

5.6: CuAAC on gp140

To confirm the labelling of gp140 with Aha, the copper-catalysed cycloaddition of propargyl biotinamide onto gp140 was carried out, followed by visualisation via anti-biotin Western blot (Fig. 5.10). The protein stocks were concentrated to 0.2 mg/mL using a Vivaspin 500 centrifugal concentrator (MWCO 10,000 Da). CuAAC was carried out using a protocol as previously reported.[44]

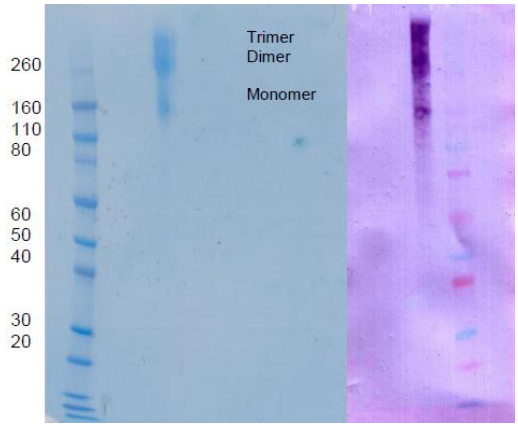


Fig. 5.10: SDS-PAGE and anti biotin western-blot of UG37 gp140 single Aha434 after 18h CuAAC reaction with propargyl biotinamide

Following the confirmation of functionalisation of gp140, attempts were made to chemically trimerise gp140 monomers via CuAAC using **5.4** as a linker (Fig. 5.11).

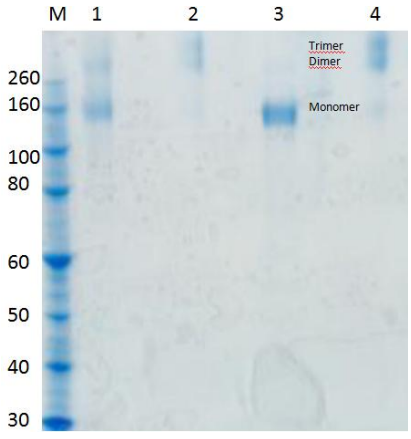


Fig. 5.11: gp140 Aha434 1) After CuAAC with trimeric linker (1.1 eq protein monomer per alkyne on linker) **reducing** gel loading buffer**,** 2) After CuAAC with trimeric linker, **non-reducing** gel loading buffer, 3) starting material, **reducing** gel loading buffer, 4) starting material, **non-reducing** gel loading buffer.

A comparison of lanes 1 and 3 shows that there is an increase in the amount of dimeric gp140 species in the reaction mixture as compared to the starting material. The reducing gel loading buffer in these lanes suggests that this dimerisation is not through mixing of disulfides on the protein, but may be through covalent conjugation via triazole formation.

5.7: Mass spectrometric analysis of cross-linked gp140

For further confirmation of functionalisation, the protein was analysed by MALDI-TOF, using a Waters MALDI Micro MX operating in linear mode, both in positive and negative mode. Sinapinic acid (3 mg in 300 µL 2:3 Acetonitrile: 0.1% TFA in H_2O) was used as the matrix. 2 µL of this solution was pipetted onto 9 spots on a MALDI plate, and together with either 1.5, 1.0, or 0.5 µL of 0.2 mg/mL solutions of:

1) gp140 Aha 434
2) gp140 Aha 434 incubated with pngase-F (deglycosylated)
3) gp140 Aha 434 after reaction with trimeric alkyne linker

The spots were quickly mixed with the matrix solution and allowed to dry. Separately, 1 µL each of the protein solutions was mixed with 4 µL of 0.1% TFA in water. Subsequently, 0.5 µL of each solution was mixed with 2 µL of the matrix solution on the MALDI plate. A BSA standard (1 mg/mL) was also used as a control. However, apart from BSA, none of the protein samples could be detected (Fig. 5.12 to 5.15):

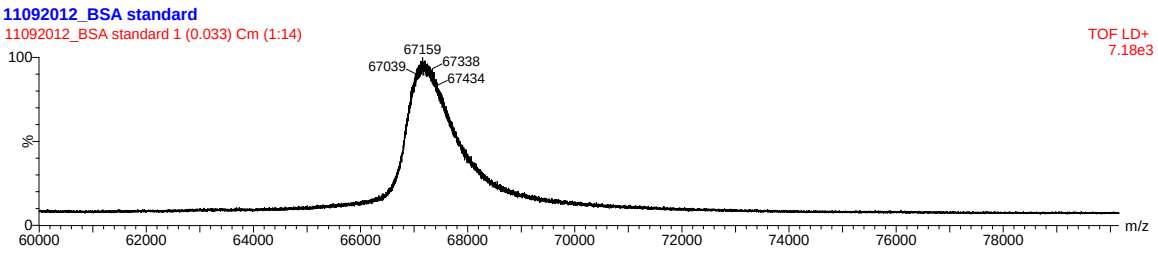


Fig. 5.12: MALDI spectrum of BSA Standard

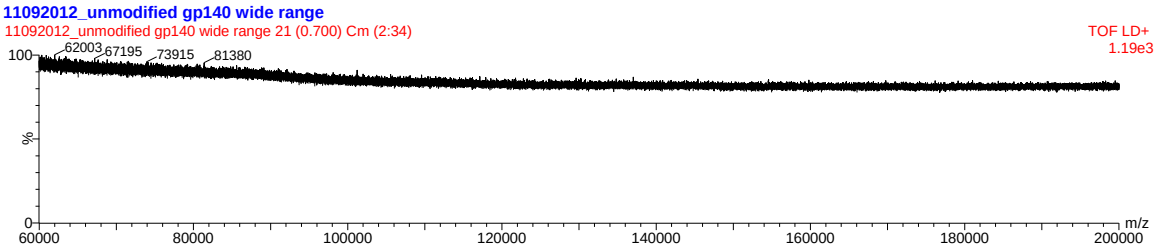


Fig. 5.13: gp140 Aha 434 starting material

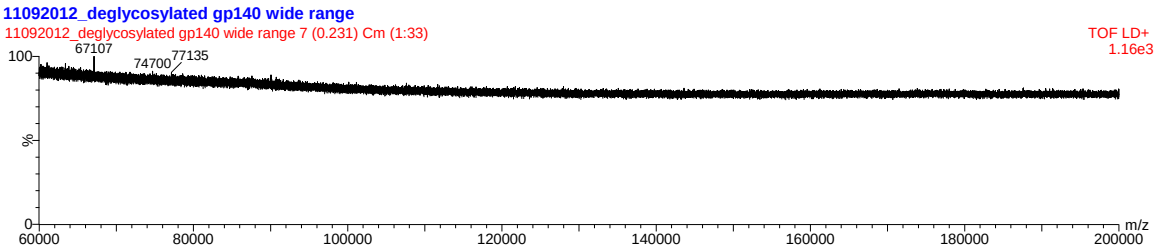


Fig. 5.14: deglycosylated gp140 Aha 434

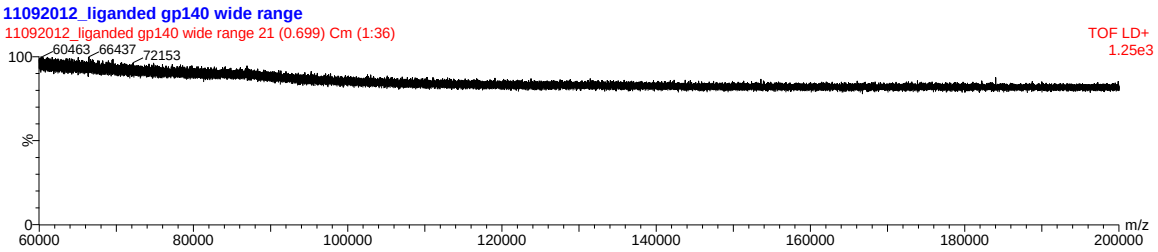


Fig. 5.15: gp140 Aha 434 after 18h reaction with trimeric alkyne linker

Next, optimisation of the MALDI conditions was attempted. Factors were:

1. Protein solutions:
    a. gp140 Aha 434 starting material
    b. Deglycosylated gp140
    c. gp140 Aha 434 after reaction with trimeric alkyne linker
    d. gp140 Aha 434 after reaction with alkynyl-biotin amide

2. Matrix used:
    a. sinapinic acid
    b. α-Cyano-4-hydroxycinnamic acid

3. Volume of 0.2 mg/mL protein solution used:
    a. 0.5 µL
    b. 1.0 µL

4. Volume of matrix used
    a. 1.0 µL
    b. 1.5 µL
    c. 2.0 µL
    d. 2.5 µL

=> Total number of conditions = 4 x 2 x 2 x 4 = 64

However, no protein signal was detected in any of these experiments. Following this, the Ultra Thin Layer MALDI preparative technique was employed.[45] Once again, no protein signal could be obseerved.

With the aid of Georg Hochberg in the Justin Benesch group, a Quadrupole-TOF was used for the analysis of gp140. The machine had been successfully used in the past by the Benesch and Robinson groups for the analysis of large protein complexes.[46,47] gp140 Aha 434 was buffer exchanged into 200 mM ammonium acetate for injection in native conditions, and denatured with 1:1 formic acid: methanol for non-native injections. Though gp140 was observed in the mass spec (Fig. 5.16), the protein was observed to ionise poorly, as seen by low ion counts and low 'z' values leading to high m/z values (m/z values > 2000)

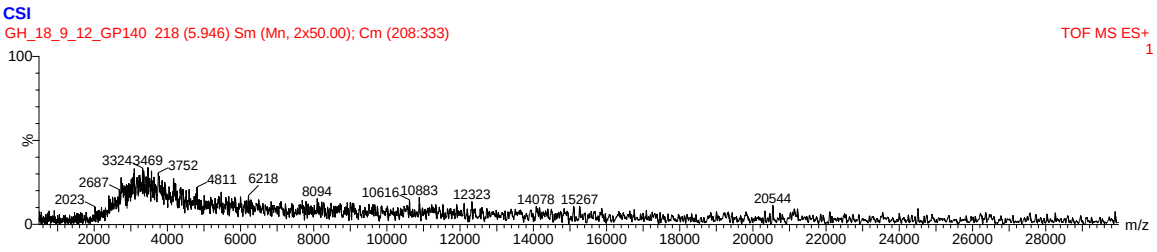


Fig. 5.16: Q-TOF MS of gp140 Aha 434 0.2 mg/mL

The protein stock was concentrated in order to obtain a greater signal from the mass spec, from 0.2 mg/ mL to 1.0 mg/ mL using a Vivaspin centrifugal concentrator. The ion counts were seen to increase (Fig. 5.17.), but the low charge per molecule, together with the heterogeneous glycosylation on gp140 caused the formation of low-intensity overlapping ion series, thus preventing the software from deconvoluting. Furthermore, the m/z was seen to increase, indicating poorer ionisation of the protein upon concentration (Fig. 5.17).

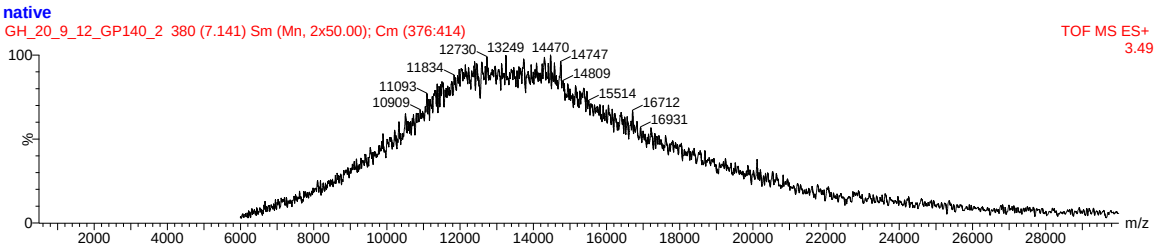


Fig. 5.17: Q-TOF MS of gp140 Aha 434 1.0 mg/mL

In order to circumvent the problem caused by heterogeneous glycosylation, a 1 mg/mL deglycosylated sample of gp140 was then analysed after buffer-exchanging into 200 mM ammonium acetate. A protein signal was observed at a lower m/z than for glycosylated gp140, as was expected. However, once again, an ion series was not seen to form, suggesting that the lack of ionisation may be due to the primary sequence of the protein itself (Fig. 5.18)

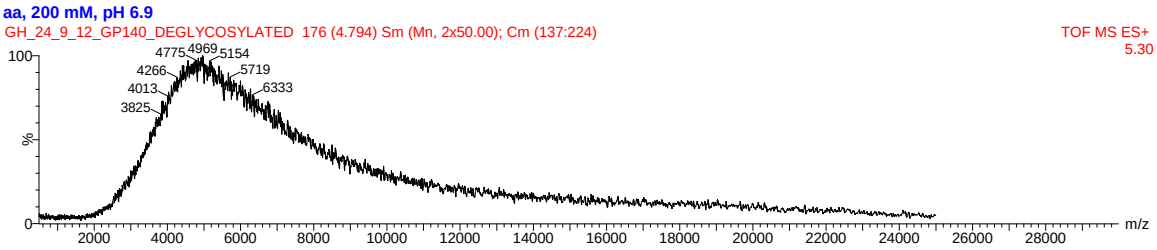


Fig. 5.18: Q-TOF MS of deglycosylated gp140 Aha 434 1.0 mg/mL

5.8: gp140 Chemical Cross-Linking Outlook

BG505 SOSIP.664, is a trimeric cleavable gp140 construct containing a disulfide bond between gp120 and gp41 (termed SOS), an Ile to Pro mutation at 559 which strengthens the gp120-gp41 association (termed IP),[48] as well as a truncated membrane proximal external region (MPER) on gp41.[49] This trimer has been crystallised,[50] and has been found to bind effectively to broadly neutralising antibodies raised against neutralising epitopes on the HIV viral envelope. In particular, antibodies such as PG9, PG16 and PGT145 which recognise quarternary epitopes on the viral protein bind effectively to BG505 SOSIP.664, but not to uncleaved gp140 trimers.[27,51,52] Furthermore, unlike non-cleaved gp140 constructs, BG505 SOSIP.664 does not display epitopes for non-neutralising antibodies.[53,54] With this information in hand, it was decided to use BG505 SOSIP.664 as a new model to generate conformationally restricted trimers via CuAAC. Residues in the crystal structure of the trimer

were identified as possible targets for Aha incorporation. Specifically, K117 and L122 were chosen as two such possible sites (Fig. 5.19).

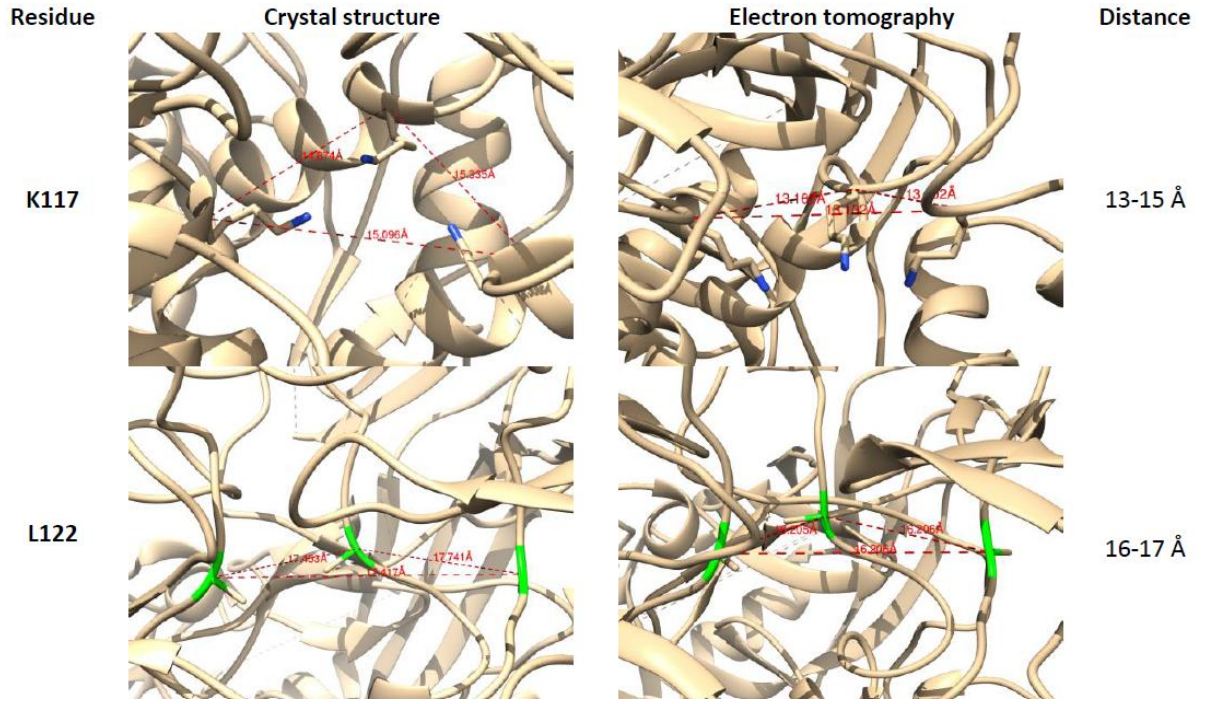
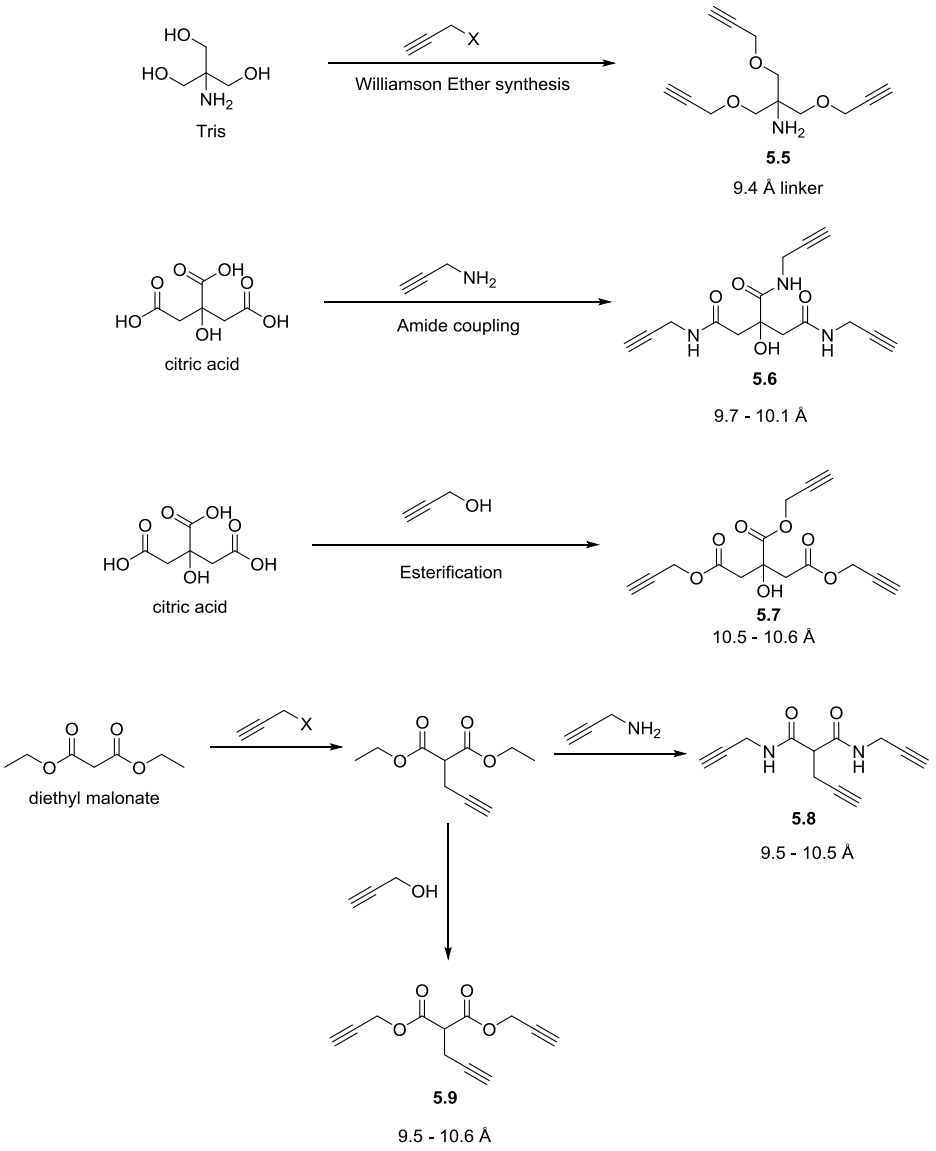


Fig. 5.19: Crystal Stucture and ET images of BG505 gp140, with proximal residues highlighted. Image used with permission of Torben Schiffner.

A construct was ordered from GeneArt, where all native methionines were mutated to isoleucines. Next, two constructs are to be made, a K117M mutant and an L122M mutant, followed by the subsequent incorporation of Aha. These two constructs will be cross-linked with a range of trimeric alkyne-functionalised linkers with different lengths and chain flexibilities via CuAAC. A few potential linkers may be synthesised as follows (Scheme 5.2):



Scheme 5.2: Synthesis plan for various tri-alkyne linkers for BG505 SOSIP.664 single Aha cross-linking. Linker lengths are displayed below each structure.

5.9: T-cell receptor cross-linking

The aim of this project is to investigate the possibility of activation of the T-cell receptor (TCR) complex by chemically cross-linking surface Azidohomoalanine and Homopropargyl glycine residues, as a surrogate for the cis-interaction of the TCR with antigens presented by major histocompatibility complex molecules.

The T-cell receptor complex is anchored in the cell membrane of T-cells, and consists of an α-chain and a β-chain. The interaction of the T-cell receptor complex with an antigen bound

to Major Histocompatibility Complex molecules results in T-cell activation through a relay of biochemical events, including the dynamic formation of TCR microclusters on the surface of the T-cell.[55,56] Another downstream event of TCR activation and clustering is the phosphorylation of tyrosine on 18-19 amino acid long peptide sequences found in the flanking CD3 and ζ-chains near the TCR complex,[57] known as immunoreceptor tyrosine-based activation motifs (ITAMs).[58] The phosphorylated tyrosines provide docking sites for other molecules in the signalling pathway.[59,60] This phosphorylation is therefore a significant component in the further propagation of the antigen signal into the cell. It was decided to use this signal as an indicator for activation of the TCR. Covalent cross-linking of azide and alkyne residues on the TCR would then be used to induce association, thereby acting as a chemical surrogate for peptide-MHC based activation.

Jurkat Cells were grown in RPMI medium, then placed in Met-free SILAC medium for 30 minutes to deplete the intracellular pools of Methionine. Test samples were pulse-labelled with 100 mg/L Aha and 100 mg/L Hpg for 4 hours, enough time for the internalisation and recycling of all T-cell receptors.[61] This would generate non-specifically Aha- and Hpg-labelled TCRs on the cell surface, some of which might be close enough for cross-linking via CuAAC. CuAAC was carried out on the cell surface for 5 minutes. The first attempt was carried out by pre-dissolving the Copper(I) bromide and triazole ligand in acetonitrile, such that the final concentration of acetonitrile in water was 5%.[34]

Cells were then spun down, washed with Met-free medium, lysed, followed by SDS-PAGE and Western blot with anti-phosphate 4g10 mouse antibody to test for increase in phosphorylation, as well as an anti-actin rabbit antibody as a loading control. The blot was visualised by secondary labelling with fluorescent anti-mouse (green) and fluorescent anti-rabbit (red) antibodies.

On performing the blot, it was seen that the anti-actin loading control showed that the cellular matter in all samples where the CuAAC was performed was considerably less than in the control well (Fig. 5.20).

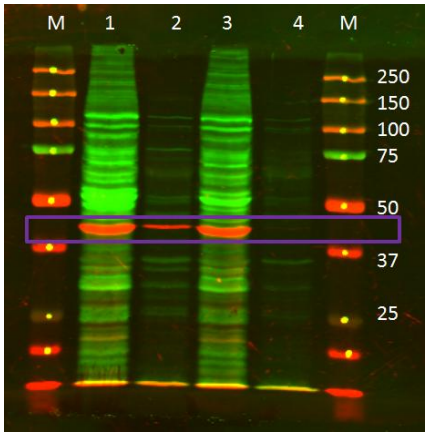


Figure 5.20. Western blot for TCR cross-linking experiment with acetonitrile as co-solvent. M: Marker, 1: lysate from unlabelled Jurkat cells, 2: lysate from unlabelled Jurkat cells after CuAAC, 3: lysate from Aha- and Hpg-labelled Jurkat cells, 4: lysate from Aha- and Hpg-labelled Jurkat cells after CuAAC. The loading control, G-actin has been highlighted at 42 kDa.

This result may be explained by the lysis of Jurkat cells during CuAAC in 5% acetonitrile. If a significant amount were lysed during the reaction, the subsequent cell debris would have been lost when the intact cells were spun down. Cross-linking was then attempted using DMSO to pre-dissolve the Cu(I)-*tris*-triazole complex, as well as reducing the amount of DMSO in the final volume of the cell suspension 10-fold, i.e., down to 0.5% from 5% v/v. This also necessitated the use of a 10-fold more concentrated copper catalyst stock. Once again however, very little cellular matter was seen in the lanes where Jurkats were subjected to CuAAC (Fig. 5.21):

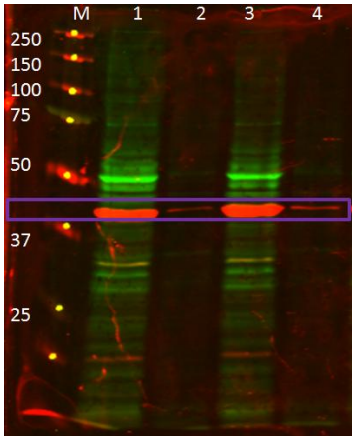


Figure 5.21. Western blot for TCR cross-linking experiment with DMSO as co-solvent. M: Marker, 1: lysate from unlabelled Jurkat cells, 2: lysate from unlabelled Jurkat cells after CuAAC, 3: lysate from Aha- and Hpg-labelled Jurkat cells, 4: lysate from Aha- and Hpg-labelled Jurkat cells after CuAAC. The loading control, G-actin has been highlighted at 42 kDa.

Referring to a previously published protocol involving the copper-catalysed Huisgen cycloaddition on live cells,[62] the amount of copper and ligand was reduced 10-fold. The Jurkat cells were seen to remain viable after subjecting them to the reaction conditions, which may be inferred from the unchanged amount of actin in the western-blot. However, though protocol referred to suggests that the reaction is complete within 3 minutes, there was no observable alteration in the intracellular phosphorylation signal even after 5 minutes (Fig. 5.22), implying that the T-cell receptors were not significantly activated.

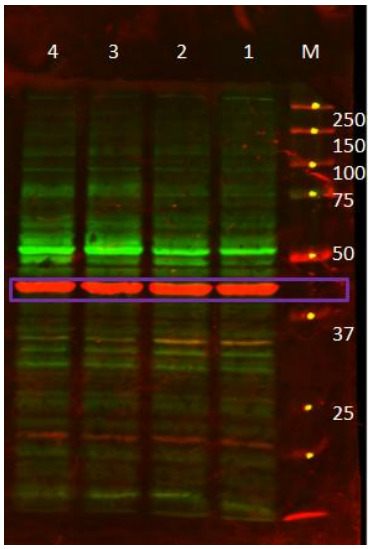


Figure 5.22. Western blot for TCR cross-linking experiment using 10-fold less Cu. M: Marker, 1: lysate from unlabelled Jurkat cells, 2: lysate from unlabelled Jurkat cells after CuAAC, 3: lysate from Aha- and Hpg-labelled Jurkat cells, 4: lysate from Aha- and Hpg-labelled Jurkat cells after CuAAC. The loading control, G-actin has been highlighted at 42 kDa.

5.10: T-Cell Receptor Chemical Cross-Linking Outlook

The crystal structure of a human αβ T-cell receptor liganded with an influenza haemagglutinin peptide and an MHC molecule HLA-DR1 displays several surface-accessible methionine side-chains,[63] which may be specifically targeted for TCR cross-linking. To better induce TCR microclustering, an alternate strategy may be to use a single Met surrogate, such as only Aha, rather than both Aha and Hpg. Subsequently, a designed dendrimeric alkyne ligand may then be used for further attempts at cross-linking by CuAAC.

5.11: Methods

5.11.1: General Procedure for Expression of NP276 proteins

An E. Coli B834(DE3) glycerol stock transformed with either Np276 (M21I) or Np276 M21I, M61I, I101M or Np276 (G2F, M21I, I101Aha) in pET28d was used to inoculate L.B. media (20 mL) containing kanamycin (50 µg/mL) and incubated for 16 hours at 37 °C. The starter cultures were used to inoculate Selenomet media (1 L), supplemented with Selenomet nutrient mix, L-methionine (20 mg/L) and kanamycin (50 µg/mL), the cells incubated at 37 °C, and the optical density (OD_{600}) monitored until a value of 0.7 was obtained (~3-4 hours). The cells were then pelleted by centrifugation (8 minutes, JLA-9.1000 rotor, 8000 rpm). The pellets were washed three times with SelenoMet base media (3 × 300 mL) containing kanamycin (50 µg/mL) and then transferred to prewarmed SelenoMet media (1 L) containing kanamycin (50 µg/mL) and supplemented with Selenomet nutrient mix as well as the desired unnatural amino acid (DL-Ahc, 500 mg/L). The culture was incubated for 30 minutes at 37 °C, followed by 30 min at 30 °C. Expression was induced by addition of IPTG (1 mM) and was continued for 16 hours at 30 °C (220 rpm). The cells were harvested by centrifugation (8 minutes, 8000 rpm) to give pellets, which were collected in a 50 mL falcon tube and resuspended in binding buffer (20 mM Tris-HCl, 15 mM imidazole, 0.5 M NaCl, 1% glycerol, pH 7.8) (20 mL). The cell suspension was treated with DNase (10 µL), and proteinase inhibitor tablet (1/4 tablet), and incubated at 4 °C for 2 hours with occasional votexing. The cell was then lysed by sonication (Soniprep at 10 micron amplitude, 30 seconds on with 30 s rest, 5 cycles, on ice). The cell debris was removed from the lysis mixture by centrifugation (40 minutes, JA 25.5 rotor, 20000 rpm). The protein was purified by incubation of the cleared cell lysate with HisBind Ni-NTA resin (3 mL per L of culture, Novagen) overnight at 4 °C. The resin was packed into a glass econo-column and unbound proteins washed out with binding buffer (**A**) (20 mM Tris-HCl, 15 mM imidazole, 0.5 M NaCl, 1% glycerol, pH 7.8) (15-20 CV, ~50 mL). The bound protein was eluted from the resin using elution buffer (**B**) (20 mM Tris-HCl, 500 mM imidazole, 0.5 M NaCl, 5% glycerol, pH 7.8) using a step gradient (20% B (5 mL), 40% B (5 mL), 60% B (5 mL), 80% B (5 mL), 100% B (5 mL × 4)). All eluted proteins were collected in 5 mL fractions, which were analysed by SDS-PAGE (NuPAGE 4-12% Bis-Tris gel, Invitrogen, MES buffer, 200 V, 35 min).

Fractions containing pure Np276 protein were combined subjected to ammonium sulfate precipitation (85% saturation) for 16 h at 4 °C. The precipitated proteins were pelleted by centrifugation (45 minutes, JA 25.5 rotor, 20000 RPM 4°C), desalted using a PD10 column, with 50 mM Sodium phosphate buffer, pH 8.0 as a running buffer, then analyzed by LC-MS. Protein solutions were flash frozen, and stored at -20 ˚C.

5.11.2: General procedure for gp120 / gp140 expression

To test the expression of the single Met434 gp140 mutant, DH5α were transformed either with plasmid containing UG37 gp140 single Met434 mutant or WT Bal gp120 (used as an expression control),[64] and used to inoculate 25 mL of L.B. medium supplemented with 100 µg/L carbenicillin. This starter culture was incubated overnight at 37 °C, and then used to inoculate 2 L each of L.B. supplemented with 100 µg/L carbenicillin.

Cells were grown until the OD at 600 nm was seen to be ~1.5 for both flasks. The contents were spun down at 8000 rpm, the cell pellets combined, and the DNA extracted using a QiaGEN Plasmid Giga Kit.

Total amount of gp140 DNA = 17.79 mg
Total amount of gp120 DNA = 17.44 mg

1. 1 vial of HEK 293T cell were thawed from Liq. N_2 in a 37 °C bath. The contents were transferred to a sterile 50 mL falcon tube containing pre-warmed DMEM media (supplemented with 10% FBS, 500 µg/mL penicillin, 100 µg/mL normocin, and 50 µg/mL streptomycin).
2. The contents were centrifuged at 1200 rpm for 5 minutes, and the supernatant (containing DMSO as a cryoprotectant) was poured off.
3. The cells were re-suspended in 1 mL DMEM, and transferred to a T-25 (25 cm^2 horizontal cross-sectional area) cell culture flask containing 5 mL medium. The flask was incubated at 37 °C with 5% CO_2 for 3-4 days.
4. As the cells reached confluence, they were split into two T-75 flasks (containing 15 mL DMEM), and stored in the incubator at 37 °C with 5% CO_2 for 3-4 days.

5. Upon reaching confluence, the cells were washed with 5 mL PBS, re-suspended in a further 5 mL PBS by manual agitation, then diluted with 15 mL DMEM (+ 10% FBS + 1% Pen/Strep), or "**D10**" media. Four T-175 flasks (containing 50 mL DMEM) were seeded with 10 mL of the cell suspension each, and stored in the incubator at 37 °C with 5% CO_2 for 3-4 days.
6. Upon reaching confluence, two of these T-175 flasks were used to make new frozen cell-stocks for the BGD group (see Methods Section). Cells in the two remaining flasks were washed with 10 mL PBS, and then re-suspended in a further 30 mL PBS. 6 x 10 mL of this cell suspension was used to inoculate six T-175 flasks containing 50 mL D10 media. Flasks were stored in the incubator at 37 °C with 5% CO_2 for 3-4 days.
7. Upon reaching confluence, the cells in each T-175 flask were washed with 10 mL PBS, the re-suspended in a further 10 mL of PBS per flask. Two Corning hyperflasks were filled with 470 mL of D10 each, and 30 mL of the cell suspension was added to each flask. These were stored in the incubator at 37 °C with 5% CO_2 for 3-4 days.
8. Upon reaching confluence, cells were suspended separately in 40 mL PBS per hyperflask, and the cell mass decanted into two fresh T-25 flasks. The cell suspensions were directly added to nine 550 mL bottles of freshly prepared D10 media, and the media used to fill ten hyperflasks (the two flasks from the previous step were re-used without washing). The ten flasks were stored in the incubator at 37 °C with 5% CO_2 for 4 days.
9. The D10 media was seen to change colour after 4 days of incubation from red to yellow, indicating depletion of nutrients and acidification of the media. As the cells were only ~70 % confluent, a media swap was performed. Depleted D10 media was removed by decantation, and the flasks were supplemented with fresh D10 media. The flasks were stored in the incubator at 37 °C with 5% CO_2 for 2 days.
10. HEK cells were transiently transfected with gp120 or gp140 plasmid DNA using PEI, the contents thoroughly mixed, and the media allowed to stand at room temperature for 10 min.
11. This was repeated in another 500 mL of DMEM with 9.1 mg DNA containing gp120 DNA, with 163 µL of the PEI solution. All flasks were placed in the incubator at 37 °C with 5% CO_2 for 3 days.
12. 0.5 mL aliquots from each hyperflask were removed for analysis of excreted gp-proteins by western blot with Anti-His6 antibody (performed by Guillaume Stewart-Jones). Media from gp120 and gp140-expressing flasks was pooled separately, and stored at 4 °C with 0.01% NaN_3.

13. Fresh D2 media was made up, and each flask supplemented with it. The hyperflasks were stored in the incubator at 37 °C with 5% CO_2 for 3 days.
14. After 3 days, the media was harvested, stored at 4 °C with 0.01% NaN_3, and the hyperflasks filled with autoclaved water. The flasks were incubated for 1h at r.t., and the contents discarded.
15. Media was centrifuged to remove cell pellets, and extracellular gp120 / gp140 purified by His-tag affinity chromatography, GNL lectin bead based affinity chromatography, and finally by size-exclusion chromatography.
16. Fractions were analysed by gel electrophoresis (Fig. 5.23) and anti-his tag western blot.

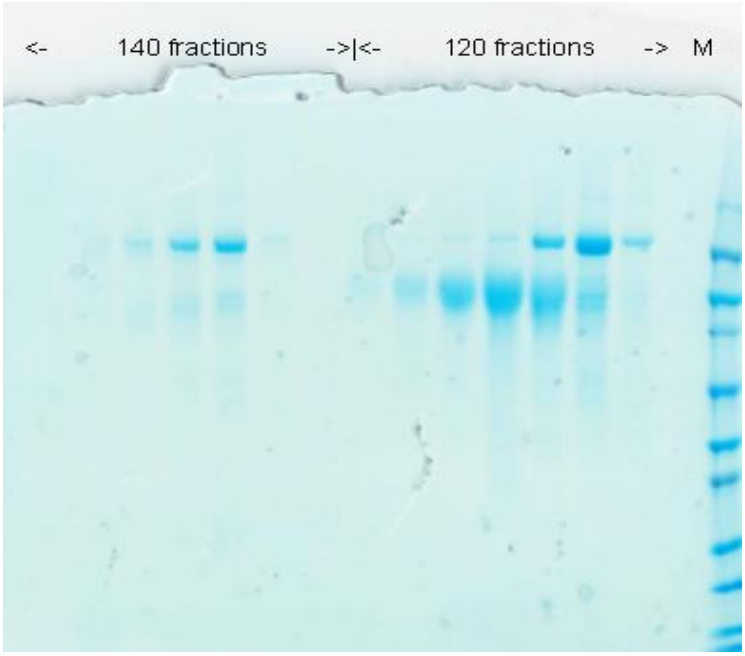


Fig. 5.23: SDS-Page of purified WT Bal gp120 and UG37 single Met434 gp140

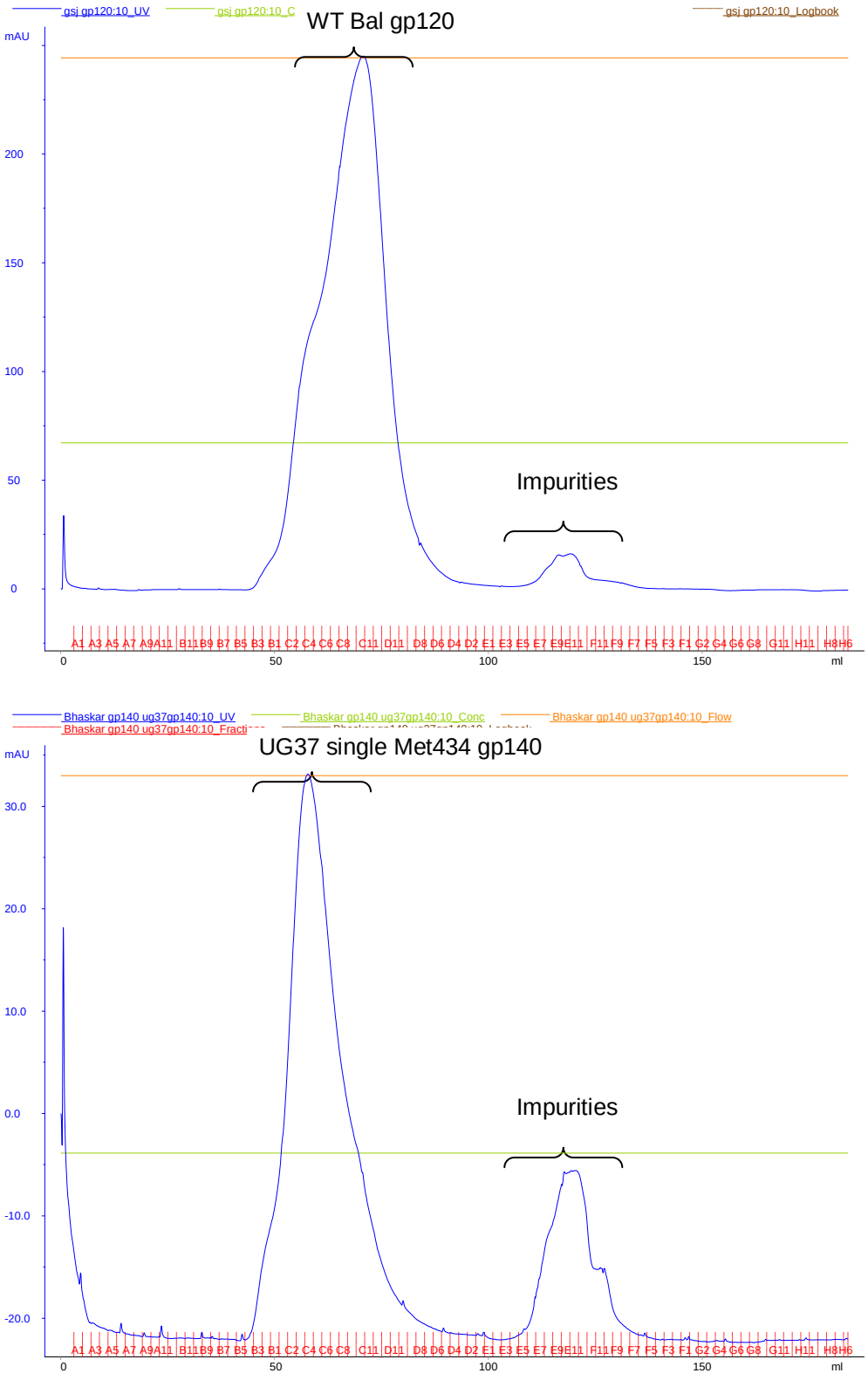


Fig. 5.24: FPLC chromatograms for the purification of WT Bal gp120 and UG37 Met434 gp140

5.11.3: Procedure for Cross-linking Jurkat T-cell receptor

1. Jurkat Cells were grown in RPMI medium, then transferred to met-free SILAC (20 mL) for 30 min.
2. The cells were spun down at 1200 RPM for 5 min at 21 ˚C, and resuspended in a minimal amount of medium.
3. Half the pellet was transferred to met-supplemented SILAC (20 mL) and split into 2 flasks of 10 mL each, and used as the 'unlabelled' and 'unlabelled + click' samples.
4. The remaining half of the pellet was resuspended in met-free SILAC (20 mL total), supplemented with L-Aha (5.7 mg, 0.04 mmol) and rac-Hpg (20.1 mg, 0.16 mmol) This was split into 2 flasks of 10 mL each, and used as the 'labelled' and 'labelled + click' samples.
5. All samples were placed in an incubator for 4 hours, then spun down at 1200 RPM for 5 min at 21 ˚C, and suspended in 5 mL each of PBS.
6. Solutions of Copper (I) Bromide and tris-((1-Ethoxycarbonylmethyl) triazolylmethyl)amine were prepared either in Acetonitrile or DMSO, premixed, and added to each sample.
7. The CuAAC reaction was allowed to run for 5 minutes, then cells were spun down at 1200 RPM for 5 min at 21 ˚C, and cell pellets were snap-frozen in dry ice for storage.
8. The cell pellets were lysed in RIPA buffer, and cell debris was spun down.
9. 200 µL of the supernatant was added to 60 µL of Laemmli buffer for loading onto a 10% SDS-PAGE gel. 10 µL of each sample was added, and run at 120-140 V.
10. The protein bands were transferred onto a nitrocellulose membrane for Western-blotting

5.11.4: Procedure for Western Blot

1. The nitrocellulose membrane as well as both filter-paper supports were dipped in semi-dry transfer buffer, and then placed in the blotting system with the PAGE gel.
2. The system was run at 18 V for 1 hour, after which time the proteins were transferred from the gel to the membrane. The membrane was then washed with PBS.
3. Blocking Buffer (25 mL, Li-Cor) was used to wash the membrane for 15 minutes to block non- proteinaceous regions.
4. The membrane was then transferred to a Falcon tube containing both anti-phosphate 4g10 mouse and anti-actin rabbit primary antibody solutions. The contents were

placed on a roller for 90 min. Excess antibody was then washed off with PBS and 0.1% Tween (3 x 10 min).

5. A solution containing fluorophores-linked anti-mouse and anti-rabbit secondary antibodies was then added to the membrane, and allowed to bind for 30 min, then washed with PBS (3 x 10 min).
6. The membrane was then visualised under a fluorescent dock.

Buffers used for Western Blot:

Primary Antibody Solution:	1000 x diluted stock of 4g10
	5000 x diluted stock of anti-actin
	1.5 g BSA
	1 mL of 10% Tween
	500 μL of 10% NaN_3
	Made up to 50 mL in water.
Secondary Antibody Solution:	1.5 g BSA
	1 mL of 10% Tween
	500 μL of 10% NaN_3
	50 μL of 10% SDS
	Made up to 50 mL in water.
Semi-dry Transfer Buffer:	14.4 g Glycine
	15.0 g Tris
	0.5 g SDS
	200 mL MeOH
	800 mL water

5.11.5: Creating Liq. N2 stocks for HEK 293T cells

1. Cells in 2 confluent T175 flasks were washed with 10 mL PBS, then re-suspended in 10 mL PBS.
2. Cells were counted using an optical microscope, and a total cell count of ~ 7.44×10^7 was determined.

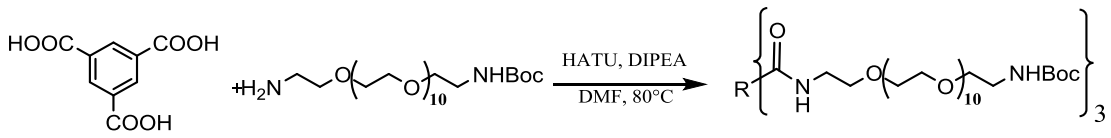
3. Cells were spun down at 1200 RPM for 5 min, then re-suspended in 12.4 mL of 10% DMSO in DMEM (+ 10% FBS + 1% Pen/Strep), to make a final concentration of 6 million cells per mL.
4. Cells were transferred to twelve 1 mL cryo-vials, and slowly frozen to - 80 ℃ in an isopropanol bath.
5. Cells were then transferred to a Liq. N_2 Dewar for long-term storage.

5.11.6: QIAGEN Gigaprep Kit Protocol

1. Bacterial cultures containing UG37 gp140 single Met434, and gp120 each in 2 L of L.B. media with 100 µg/L carbenicillin were grown to an O.D. at 600 nm of 1.2, then growth continued for a further 4 hours.
2. The cells were then harvested by centrifugation, corresponding cell pellets combined, and each sample re-suspended in 125 mL buffer P1 in a fresh 2 L bottle.
3. 125 mL buffer P2 was added to each bottle, the contents thoroughly mixed by inverting 4-6 times, and then incubated at r.t. for 5 min. The cell suspensions were seen to turn blue due to the presence of LyseBlue reagent.
4. 125 mL of chilled buffer P3 was added to each bottle, and the contents mixed by vigorous inversion 4 to 6 times. The blue colour was seen to disappear, the viscosity of the suspension was seen to reduce, and a white fluffy material was seen to form. The contents were incubated on ice for 30 min.
5. The contents were centrifuged at 20,000 g for 30 min at 4 ℃, and the supernatant removed.
6. Two ‘QIAGEN Tip 10000’ columns were equilibrated with 75 mL buffer QBT. The columns were allowed to empty by gravity flow.
7. The gp140-DNA-containing supernatant was loaded onto one column, while the gp120-containing supernatant was loaded onto the second. Both were allowed to enter the column resin by gravity flow.
8. Each column was washed with 600 mL buffer QC by gravity flow. The eluate was discarded.
9. DNA was eluted from the columns with 100 mL buffer QF, by gravity flow.
10. DNA was precipitated from the eluates by the addition of 70 mL isopropanol. The suspensions were centrifuged at 15,000 g for 30 min at 4 C, and the supernatant carefully poured off.

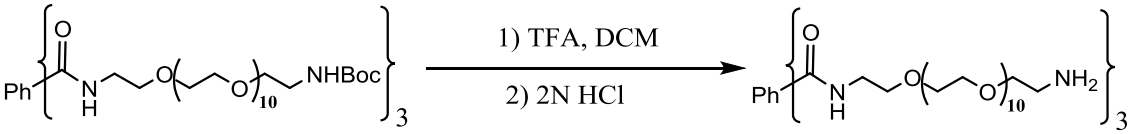
11. The translucent DNA pellet was allowed to air-dry for 1 hour, then re-dissolved in 4 mL MilliQ water. The concentration of the solutions was determined by Nanodrop.

5.11.7: Synthesis of Trimeric PEG linker



5.1 **5.2**

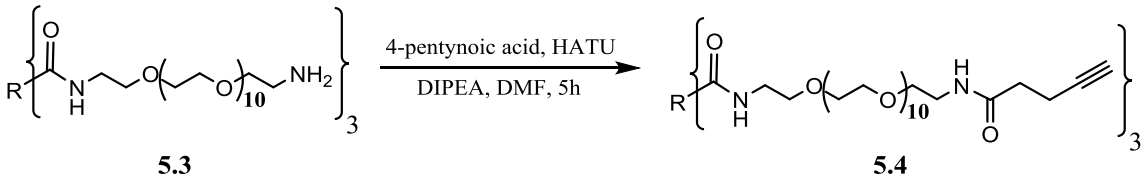
Trimesic acid (21 mg, 0.1 mmol) was dissolved in 6 mL DMF, to which was added Boc-Amino-Peg_{11}-Amine (260 mg, 0.4 mmol). The mixture was stirred at 70 ˚C, and HATU (196 mg, 0.5 mmol) and DIPEA (212 µL, 1.2 mmol) were added. The product was seen to form over 18h via TLC (10: 1 DCM: MeOH). The contents were evaporated, and the residue purified by column chromatography (10:1 DCM: MeOH), as a thick yellow oil (194 mg, 92%). ^{1}H NMR (500 MHz, $CDCl_3$) δ: 1.39 (s, 27H, Boc), 2.85 (t, 108H, J = 40.6 Hz, PEG), 3.59 (t, 24H, J = 4.1 Hz, PEG), 3.61 (m, 18H, PEG), 8.45 (s, 3H, C_6H_3). ^{13}C NMR (500 MHz, $CDCl_3$) δ: 12.52, 17.09, 18.50 (CH_3), 31.39, 36.47, 38.54 (PEG), 69.96, 70.15 (PEG), 135.00 (phenyl C-CONH), 128.82 (Phenyl C-H), 162.57 (PhCONH). LRMS: $C_{96}H_{180}N_6O_{42}$ (m/z) Calculated [M] = 2089.21 Observed: $[M+Na]^+$ = 2112.85, HRMS: $C_{96}H_{180}N_6Na_2O_{42}$ (m/z) Calculated = 1068.0976, Observed = 1068.0973



5.2 **5.3**

The trimeric boc-protected linker (90 mg, 0.043 mmol) was dissolved in 2 mL DCM. To this, 350 µL TFA was added, and the contents stirred at room temperature for 3h. The solvent was then evaporated, and the residue dissolved in 10 mL of 2N HCl. The contents were stirred for 2 minutes, and then the solvent was evaporated. This process was repeated twice more. Some contamination from dirty glassware was observed, so the product was purified by column chromatography (10:1 DCM: MeOH, then 6:1 DCM: MeOH). The product was obtained as a thick yellow liquid (61.4 mg, 79.6 %). ^{1}H NMR (400 MHz, D_2O) δ: 3.07 (m, 12H, PEG) 3.54, (m, 132 H, O-CH_2CH_2-O), 8.23 (s, 3H, C_6H_3). ^{13}C NMR (125 MHz, $CDCl_3$) δ: 39.49 ($CONH_2$-$\underline{C}H_2$), 42.96 ($CH_2\underline{C}H_2$-NH_2), 69.87 (-O-$\underline{C}H_2\underline{C}H_2$-O-), 129.47 (Phenyl $\underline{C}$-H), 135.21 (phenyl C-CONH) LRMS: $C_{81}H_{156}N_6O_{36}$ (m/z) Calculated $[M+3H]^{3+}$ = 597.35 Observed:

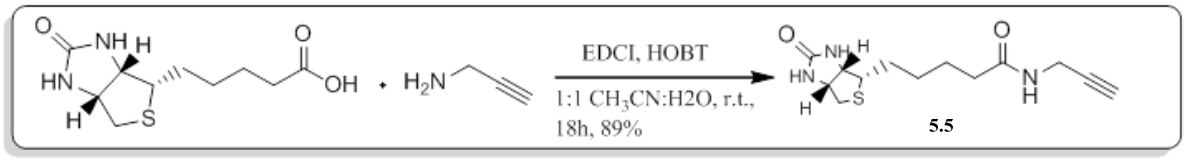
$[M+3H]^{3+}$ =597.34, HRMS: $C_{81}H_{159}N_6O_{36}$ (m/z) Calculated $[M+3H]^{3+}$ = 597.3593 Observed: $[M+3H]^{3+}$ =597.3583



5.3 **5.4**

The previous compound (30 mg, 0.016 mmol) was dissolved in DMF (6 mL) along with 4-pentynoic acid (8mg, 0.080 mg), HATU (30.4 mg, 0.08 mmol) and DIPEA (34 µL, 0.192 mmol). The contents were stirred at 60 ˚C for 5 h. The reaction was monitored by TLC (10:1 DCM: MeOH). The solvent was then evaporated, and the residue purified by column chromatography (10:1 DCM: MeOH) to yield the product as a clear oil (31 mg, 90%)
^{1}H NMR (400 MHz, D_2O) δ: 2.55 (m, 12H, C<u>H</u>$_2$C<u>H</u>$_2$-C≡CH), 2.89 (s, 3H, C≡C<u>H</u>), 3.15 (m, 12H, PEG), 3.68, (m, 138 H, O-CH_2CH_2-O), 8.59 (s, 3H, C_6H_3).^{13}C NMR (125 MHz, $CDCl_3$) δ: 18.81 (<u>C</u>H$_2$- C≡CH), 33.29(<u>C</u>H$_2$-CH_2-C≡CH), 42.54 (Phenyl C-$CONH_2$-<u>C</u>H$_2$), 54.37 (CH_2<u>C</u>H$_2$-NHCO), 70.03 (-O-<u>C</u>H$_2$<u>C</u>H$_2$-O-), 68.94**(**C≡<u>C</u>H), 82.59**(**<u>C</u>≡CH), 126.82 (Phenyl <u>C</u>-H), 134.89 (phenyl C-<u>C</u>ONH), 168.75 (PEG-NH-<u>C</u>O-CH_2-) LRMS (ESI): $C_{96}H_{168}N_6O_{39}$ (m/z) Calculated $[M+Na]^+$: 2052.12, Observed: $[M+Na]^+$: 2052.84, HRMS (ESI) $C_{96}H_{168}N_6Na_2O_{39}$ (m/z) Calculated $[M+2Na]^{2+}$ = 1037.5566, Observed: $[M+2Na]^{2+}$ = 1037.5535. IR (νmax, film): 3414.26, 3015.84, 2970.46, 2946.32, 2476.98, 2124.85, 1738.48, 1455.29, 1372.36, 1227.09, 1215.89, 1205.09, 1126.71, 1092.52, 854.53.

5.11.8: Synthesis of Alkynyl biotin



5.5

2-Propyn-1-amine (0.044 mL, 0.648 mmol) was added to an ice bath cooled solution of biotin (100 mg, .41 mmol) in CH_3CN / H_2O 1:1 (18 mL) under stirring. HOBt (84 mg, 0.62 mmol) and EDCI (123 mg, 0.64 mmol) were added at room temperature and the mixture was stirred overnight. The reaction mixture was concentrated under reduced pressure and the residue was purified by chromatography (DCM: MeOH, 15:1, 12:1 then 10:1) to give the product in 89% yield (102 mg) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$)**:** *δ*= 1.43 (m, 2H, C<u>H</u>$_2$CH$_2$CH$_2$CO), 1.66 (m, 4H, SCHC<u>H</u>$_2$- and CH$_2$C<u>H</u>$_2$CH$_2$CO), 2.18 (t, 1H, *J* = 7.4 Hz, CH-

$_2$C$\underline{H}_2$CO), 2.83 (m, 2H, C$\underline{H}_2$SCHCH$_2$-), 3.15 (m, 1H, S-C$\underline{H}$-CH2-), 3.66 (s, 1H, alkyne), 4.03 (s, 2H, NHC$\underline{H}_2$-alkyne) 4.35 (dd, 1H, J = 7.4, 4.5 Hz, C$\underline{H}$-CH(S)-CH$_2$CH$_2$), 4.55 (dd, 1H, J = 12.3, 7.4 Hz ,C$\underline{H}$-CH$_2$SCH-CH$_2$). HRMS (ESI): *m*/*z*: calcd for $C_{13}H_{19}N_3O_2SNa$: 304.1090, Found $[M+Na]^+$: 304.1084

5.12: References

(1) Schlucker, S.; Schaeberle, M. D.; Huffman, S. W.; Levin, I. W. *Analytical Chemistry* **2003**, *75*, 4312.
(2) Huang, W. E.; Griffiths, R. I.; Thompson, I. P.; Bailey, M. J.; Whiteley, A. S. *Analytical Chemistry* **2004**, *76*, 4452.
(3) Ogawa, M.; Harada, Y.; Yamaoka, Y.; Fujita, K.; Yaku, H.; Takamatsu, T. *Biochemical and Biophysical Research Communications* **2009**, *382*, 370.
(4) Walter, A.; Erdmann, S.; Bocklitz, T.; Jung, E.-M.; Vogler, N.; Akimov, D.; Dietzek, B.; Rosch, P.; Kothe, E.; Popp, J. *Analyst* **2010**, *135*, 908.
(5) Naumann, D. *Applied Spectroscopy Reviews* **2001**, *36*, 239
(6) Ye, J.; Fox, S. A.; Cudic, M.; Rezler, E. M.; Lauer, J. L.; Fields, G. B.; Terentis, A. C. *J. Am. Chem. Soc.* **2009**, *132*, 980.
(7) Liu, Z.; Li, X.; Tabakman, S. M.; Jiang, K.; Fan, S.; Dai, H. *J. Am. Chem. Soc.* **2008**, *130*, 13540.
(8) Miljkovic, M.; Chernenko, T.; Romeo, M. J.; Bird, B.; Matthaus, C.; Diem, M. *Analyst* **2010**, *135*, 2002.
(9) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 19.
(10) Baboo, S.; Bhushan, B.; Jiang, H.; Grovenor, C. R. M.; Pierre, P.; Davis, B. G.; Cook, P. R. *PLoS ONE* **2014**, *9*, e99346.
(11) Howden, A. J.; Geoghegan, V.; Katsch, K.; Efstathiou, G.; Bhushan, B.; Boutureira, O.; Thomas, B.; Trudgian, D. C.; Kessler, B. M.; Dieterich, D. C.; Davis, B. G.; Acuto, O. *Nature Methods* **2013**, *10*, 343.
(12) Oh, K.-I.; Lee, J.-H.; Joo, C.; Han, H.; Cho, M. *The Journal of Physical Chemistry B* **2008**, *112*, 10352.
(13) Vetting, M. W.; Hegde, S. S.; Hazleton, K. Z.; Blanchard, J. S. *Protein Science* **2007**, *16*, 755.
(14) Zuser, E.; Chernenko, T.; Newmark, J.; Miljkovic, M.; Diem, M. *Analyst* **2010**, *135*, 3030.
(15) Saar, B. G.; Freudiger, C. W.; Reichman, J.; Stanley, C. M.; Holtom, G. R.; Xie, X. S. *Science* **2010**, *330*, 1368.
(16) Downes, A.; Mouras, R.; Elfick, A. *Journal of Raman Spectroscopy* **2009**, *40*, 757.
(17) Evans, C. L.; Potma, E. O.; Puoris'haag, M.; CÃ´tÃ©, D.; Lin, C. P.; Xie, X. S. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 16807.
(18) Pezacki, J. P.; Blake, J. A.; Danielson, D. C.; Kennedy, D. C.; Lyn, R. K.; Singaravelu, R. *Nat Chem Biol* **2011**, *7*, 137.

(19) Ribeiro-Viana, R.; Sánchez-Navarro, M.; Luczkowiak, J.; Koeppe, J. R.; Delgado, R.; Rojo, J.; Davis, B. G. *Nat Commun* **2012**, *3*, 1303.
(20) Zeng, C.; Pan, F.; Jones, L. A.; Lim, M. M.; Griffin, E. A.; Sheline, Y. I.; Mintun, M. A.; Holtzman, D. M.; Mach, R. H. *Brain Research* **2010**, *1319*, 21.
(21) Best, M. D. *Biochemistry* **2009**, *48*, 6571.
(22) Sattentau, Q. J.; Weiss, R. A. *Cell* **1988**, *52*, 631.
(23) Chan, D. C.; Fass, D.; Berger, J. M.; Kim, P. S. *Cell* **1997**, *89*, 263.
(24) Pantophlet, R.; Burton, D. R. *Annual Review of Immunology* **2006**, *24*, 739.
(25) Earl, P. L.; Sugiura, W.; Montefiori, D. C.; Broder, C. C.; Lee, S. A.; Wild, C.; Lifson, J.; Moss, B. *Journal of Virology* **2001**, *75*, 645.
(26) Yasmeen, A.; Ringe, R.; Derking, R.; Cupo, A.; Julien, J.-P.; Burton, D.; Ward, A.; Wilson, I.; Sanders, R.; Moore, J.; Klasse, P. *Retrovirology* **2014**, *11*, 41.
(27) Julien, J.-P.; Lee, J. H.; Cupo, A.; Murin, C. D.; Derking, R.; Hoffenberg, S.; Caulfield, M. J.; King, C. R.; Marozsan, A. J.; Klasse, P. J.; Sanders, R. W.; Moore, J. P.; Wilson, I. A.; Ward, A. B. *Proceedings of the National Academy of Sciences* **2013**, *110*, 4351.
(28) Dey, B.; Pancera, M.; Svehla, K.; Shu, Y.; Xiang, S.-H.; Vainshtein, J.; Li, Y.; Sodroski, J.; Kwong, P. D.; Mascola, J. R.; Wyatt, R. *J. Virol.* **2007**, *81*, 5579.
(29) Xiang, S.-H.; Kwong, P. D.; Gupta, R.; Rizzuto, C. D.; Casper, D. J.; Wyatt, R.; Wang, L.; Hendrickson, W. A.; Doyle, M. L.; Sodroski, J. *J. Virol.* **2002**, *76*, 9888.
(30) Dey, B.; Svehla, K.; Xu, L.; Wycuff, D.; Zhou, T.; Voss, G.; Phogat, A.; Chakrabarti, B. K.; Li, Y.; Shaw, G.; Kwong, P. D.; Nabel, G. J.; Mascola, J. R.; Wyatt, R. T. *PLoS Pathog* **2009**, *5*, e1000445.
(31) Liu, J.; Bartesaghi, A.; Borgnia, M. J.; Sapiro, G.; Subramaniam, S. *Nature* **2008**, *455*, 109.
(32) Kong, L.; Sheppard, N. C.; Stewart-Jones, G. B. E.; Robson, C. L.; Chen, H.; Xu, X.; Krashias, G.; Bonomelli, C.; Scanlan, C. N.; Kwong, P. D.; Jeffs, S. A.; Jones, I. M.; Sattentau, Q. J. *Journal of Molecular Biology*, *403*, 131.
(33) Pruzina, S.; Williams, G. T.; Kaneva, G.; Davies, S. L.; Martín-López, A.; Brüggemann, M.; Vieira, S. M.; Jeffs, S. A.; Sattentau, Q. J.; Neuberger, M. S. *Protein Engineering Design and Selection* **2011**, *24*, 791.
(34) Fernandez-Gonzalez, M.; Boutureira, O.; Bernardes, G. J. L.; Chalker, J. M.; Young, M. A.; Errey, J. C.; Davis, B. G. *Chemical Science*, *1*, 709.
(35) Veronese, F. M.; Pasut, G. *Drug Discovery Today* **2005**, *10*, 1451.
(36) Roben, P.; Moore, J. P.; Thali, M.; Sodroski, J.; Barbas, C. F., 3rd; Burton, D. R. *J. Virol.* **1994**, *68*, 4821.
(37) Wu, X.; Yang, Z.-Y.; Li, Y.; Hogerkorp, C.-M.; Schief, W. R.; Seaman, M. S.; Zhou, T.; Schmidt, S. D.; Wu, L.; Xu, L.; Longo, N. S.; McKee, K.; O'Dell, S.; Louder, M. K.; Wycuff, D. L.; Feng, Y.; Nason, M.; Doria-Rose, N.; Connors, M.; Kwong, P. D.; Roederer, M.; Wyatt, R. T.; Nabel, G. J.; Mascola, J. R. *Science* **2010**, *329*, 856.
(38) Zhou, T.; Georgiev, I.; Wu, X.; Yang, Z. Y.; Dai, K.; Finzi, A.; Kwon, Y. D.; Scheid, J. F.; Shi, W.; Xu, L.; Yang, Y.; Zhu, J.; Nussenzweig, M. C.; Sodroski, J.; Shapiro, L.; Nabel, G. J.; Mascola, J. R.; Kwong, P. D. *Science* **2010**, *329*, 811.
(39) Hu, G.; Liu, J.; Taylor, K. A.; Roux, K. H. *J. Virol.* **2011**, *85*, 2741.
(40) Rusert, P.; Krarup, A.; Magnus, C.; Brandenberg, O. F.; Weber, J.; Ehlert, A.-K.; Regoes, R. R.; GÃ¼nthard, H. F.; Trkola, A. *The Journal of Experimental Medicine* **2011**.
(41) White, T. A.; Bartesaghi, A.; Borgnia, M. J.; Meyerson, J. R.; de la Cruz, M. J. V.; Bess, J. W.; Nandwani, R.; Hoxie, J. A.; Lifson, J. D.; Milne, J. L. S.; Subramaniam, S. *PLoS Pathog* **2010**

6, e1001249.

(42) Bonomelli, C.; Doores, K. J.; Dunlop, D. C.; Thaney, V.; Dwek, R. A.; Burton, D. R.; Crispin, M.; Scanlan, C. N. *PLoS ONE* **2011**, *6*, e23521.
(43) Swanson, M. D.; Winter, H. C.; Goldstein, I. J.; Markovitz, D. M. *Journal of Biological Chemistry* **2010**, *285*, 8646.
(44) van Kasteren, S. I.; Kramer, H. B.; Gamblin, D. P.; Davis, B. G. *Nat Protoc* **2007**, *2*, 3185.
(45) Fenyo, D.; Wang, Q.; DeGrasse, J. A.; Padovan, J. C.; Cadene, M.; Chait, B. T. *J Vis Exp* **2007**, 192.
(46) Barrera, N. P.; Di Bartolo, N.; Booth, P. J.; Robinson, C. V. *Science* **2008**, *321*, 243.
(47) Zhou, M.; Sandercock, A. M.; Fraser, C. S.; Ridlova, G.; Stephens, E.; Schenauer, M. R.; Yokoi-Fong, T.; Barsky, D.; Leary, J. A.; Hershey, J. W.; Doudna, J. A.; Robinson, C. V. *Proceedings of the National Academy of Sciences* **2008**, *105*, 18139.
(48) Sanders, R. W.; Vesanen, M.; Schuelke, N.; Master, A.; Schiffner, L.; Kalyanaraman, R.; Paluch, M.; Berkhout, B.; Maddon, P. J.; Olson, W. C.; Lu, M.; Moore, J. P. *Journal of Virology* **2002**, *76*, 8875.
(49) Khayat, R.; Lee, J. H.; Julien, J.-P.; Cupo, A.; Klasse, P. J.; Sanders, R. W.; Moore, J. P.; Wilson, I. A.; Ward, A. B. *Journal of Virology* **2013**, *87*, 9865.
(50) Julien, J.-P.; Cupo, A.; Sok, D.; Stanfield, R. L.; Lyumkis, D.; Deller, M. C.; Klasse, P.-J.; Burton, D. R.; Sanders, R. W.; Moore, J. P.; Ward, A. B.; Wilson, I. A. *Science* **2013**, *342*, 1477.
(51) Davenport, T. M.; Friend, D.; Ellingson, K.; Xu, H.; Caldwell, Z.; Sellhorn, G.; Kraft, Z.; Strong, R. K.; Stamatatos, L. *Journal of Virology* **2011**, *85*, 7095.
(52) Hoffenberg, S.; Powell, R.; Carpov, A.; Wagner, D.; Wilson, A.; Kosakovsky Pond, S.; Lindsay, R.; Arendt, H.; DeStefano, J.; Phogat, S.; Poignard, P.; Fling, S. P.; Simek, M.; LaBranche, C.; Montefiori, D.; Wrin, T.; Phung, P.; Burton, D.; Koff, W.; King, C. R.; Parks, C. L.; Caulfield, M. J. *Journal of Virology* **2013**, *87*, 5372.
(53) Dey, A. K.; David, K. B.; Lu, M.; Moore, J. P. **2009**, *385*, 275.
(54) Pancera, M.; Wyatt, R. **2005**, *332*, 145.
(55) Zhong, L.; Zeng, G.; Lu, X.; Wang, R. C.; Gong, G.; Yan, L.; Huang, D.; Chen, Z. W. *PLoS ONE* **2009**, *4*, e5945.
(56) Schamel, W. W. A.; Alarcón, B. *Immunological Reviews* **2013**, *251*, 13.
(57) Irving, B. A.; Weiss, A. *Cell* **1991**, *64*, 891.
(58) Reth, M. *Nature* **1989**, *338*, 383.
(59) Iwashima, M.; Irving, B. A.; van Oers, N. S.; Chan, A. C.; Weiss, A. *Science* **1994**, *263*, 1136.
(60) Vély, F.; Nunès, J. A.; Malissen, B.; Hedgecock, C. J. R. *European Journal of Immunology* **1997**, *27*, 3010.
(61) Liu, H.; Rhodes, M.; Wiest, D. L.; Vignali, D. A. A. **2000**, *13*, 665.
(62) Besanceney-Webler, C.; Jiang, H.; Zheng, T.; Feng, L.; Soriano del Amo, D.; Wang, W.; Klivansky, L. M.; Marlow, F. L.; Liu, Y.; Wu, P. *Angewandte Chemie International Edition*, *50*, 8051.
(63) Hennecke, J. C. A. W. D. C. *The EMBO Journal* **2000**, *19*, 5611.
(64) Liao, H.-X.; Alam, S. M.; Mascola, J. R.; Robinson, J.; Ma, B.; Montefiori, D. C.; Rhein, M.; Sutherland, L. L.; Scearce, R.; Haynes, B. F. *Journal of Virology* **2004**, *78*, 5270.